

Location-based Clustering for Skewed-topology Wireless Sensor Networks

Hae-Won Choi, Myung-Chun Ryu, Sang-Jin Kim

Dept. of Computer Engineering Kyungwoon University

편향된 토폴로지를 가진 무선센서네트워크를 위한 위치기반 클러스터링

최해원, 류명춘, 김상진
경운대학교 컴퓨터공학과

Abstract The energy consumption problem in wireless sensor networks is investigated. The problem is to expend as little energy as possible receiving and transmitting data, because of constrained battery. In this paper, in order to extend the lifetime of the network, we proposed a location-based clustering algorithm for wireless sensor network with skewed-topology. The proposed algorithm is to deploy multiple child nodes at the sink to avoid bottleneck near the sink and to save energy. Proposed algorithm can reduce control traffic overhead by creating a dynamic cluster. We have evaluated the performance of our clustering algorithm through an analysis and a simulation. We compare our algorithm's performance to the best known centralized algorithm, and demonstrate that it achieves a good performance in terms of the life time.

Key Words : Skewd Topology, Wireless Sensor Network, Energy Consumption, Centralized Algorithms

요 약 제한된 배터리를 사용하는 무선 센서 네트워크에서 에너지 소비 문제는 중요한 이슈이다. 이 논문에서는 네트워크의 수명을 연장하기 위해 위치기반 클러스터링 알고리즘을 제안한다. 제안된 알고리즘은 싱크노드 주변의 병목 현상을 방지하고 에너지를 절약하기 위해 싱크노드에 여러 자식 노드를 분산 배치한다. 알고리즘은 subscription query, query relay, position aware and cluster head selection의 네 단계를 거치면서, 제어 트래픽 오버 헤드를 줄일 수 있는 동적 클러스터를 생성한다. 이를 통해 수명을 연장하고 네트워크의 효율을 증대시킬 수 있다. 알고리즘 검증을 위해 가장 잘 알려진 집중화 알고리즘과 비교 분석 및 시뮬레이션을 통해 성능을 평가하고, 수명의 관점에서 양호한 성능을 달성하는 것을 보여준다.

주제어 : 편향 토폴로지, 무선 센서네트워크, 에너지 소비 문제, 집중화 알고리즘

Received 3 November 2015, Revised 4 December 2015

Accepted 20 January 2016

Corresponding Author: Chu-Seock Chang

(The Society of Digital Policy)

Email: happychw@ikw.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

A Wireless sensor networks have been identified as one of the most important technologies for the future century. A sensor network is composed of a large number of battery-operated sensor nodes, which are most closely deployed in the interest area [1,2,3,4,5,6]. If sensors detect an event, such sensors transmit to neighbor nodes using multi-hop manner the collected information. The cost of transmitting an event is higher than a processing cost and hence it may be advantageous to organize the sensors into clusters. Clustering is a fundamental mechanism in the design of scalable sensor network protocols. Clustering splits the networks into disjoint sets of nodes, each centering on a chosen cluster header. The main function of the clustering is to minimize the exchange of flooding messages. It is pointless to waste valuable resources to pro-actively maintain such an elaborate structure between floods, when there is no traffic that can utilize it [7,8,9,10,13,15].

Although there is already a huge number of routing protocols in wireless sensor networks, we extend an existing cluster-based protocol, defined as the low-energy adaptive clustering hierarchy [11], probably the best-known cluster-based protocol. The position models described in this paper form a relatively skew-deployment. Cluster headers are concentrated on the right-hand or left-hand side in the sensor network topology. In this case, if the cluster headers rapidly deplete the remaining energy, the inter-cluster predefined path may fail. This can jeopardize the entire mission in some cases. The unbalanced energy depletion is caused by the variation in the distance from the sink.

To the best of our knowledge, clustering algorithms to minimizing the energy consumption of a sensor node have been addressed. However, these mechanisms are mostly heuristic in nature and aim at generating the minimum numbers of clusters such that nodes in any

cluster are at most a certain number of hops away from the cluster header. In favorable environments, the deployment of the nodes and the network topology need to be predefined and carefully engineered. However, for application employing the sensor network in hostile environments, e.g., skewed, concentrated, linear environments, to support effective monitoring, a large number of sensor nodes are densely deployed either inside the phenomenon or very close to it. In some application, replacement of power resources might be impossible. Node's life time has been depended on battery lifetime. Generally, information collected from sensors is aggregated to sink or base station. Nodes closer to the sink node have heavier traffic load, since they not only collect data within their sensing range but relay data for nodes further away as well. In such a scenario, the nodes that are further away from the base station will have their energy drained much faster than those nodes that are closer to the base station. The unbalance of energy depletion is caused by different distance from the sink. Because of the malfunction of few nodes, a significant topology change can be caused, and this might require re-routing of packets and re-organization of the networks.

In order to minimize the energy expended and to meet the needs of clustering in sensor networks with skewed-topology, we propose a location-based cluster header selection algorithm for establishing virtual clusters of variable length. This can make a variable size cluster. We separate the entire network into a few clusters based on the distance from between nodes and from the sink, then only a cluster head is selected at each grid each times. Consequently, cluster sizes should be as close as possible to the specified density. Therefore, once the cluster head is assigned to a cluster, the total number of messages forwarded is reduced by a fraction of the number of neighbors.

Our algorithm differs substantially from [5, 8], because we do not make any assumptions about the

specific locations of the sink, and data forwarding does not rely on direct connectivity. In our proposed scheme, we help to avoid "hot spots" in the network. Previous works have been aimed at generating the minimum number of clusters, but not at minimizing the energy expended in a sensor node. For these reasons, it is very important to design a fast algorithm to organize sensors into clusters, to minimize the energy used to communicate information from all nodes to the processing center. Proposed scheme is validated through the simulation, which shows that our algorithm increases the network lifetime by reducing the energy consumption of the sensor.

The rest of the paper is organized as follows. Section 2 describes related works. Section 3 presents preliminary works and describes our algorithm. In Section 4, our proposed scheme is described, and numerical results are evaluated in Section 5. Simulation results for evaluating performance are presented in section 6. Finally, Section 7 concludes the paper and discusses possible future research directions.

2. Related Work

Wireless sensor networks have the characteristic that the forwarding nodes are clearly engaged in a balancing act between reduced transmission energy and increased received energy. Hops that are too short lead to excessive received energy, and hops that are too long lead to excessive path loss [12].

There exists a considerable amount of previous work addressing the topology control problem of minimizing nodal transmission power, with guarantees of network connectivity. Bandyopadhyay et al. [13], proposed a clustering-based protocol that forms clusters with one node acting as the cluster head. However, because the cluster heads would deplete their energy much faster than the rest of the nodes, each node can only be a cluster head temporarily, which implies that

the clustering global synchronization would have to be done rather frequently. Potente [14] does not control the number of clusters in the current round, because the sensor nodes elect themselves to be local cluster-headers. However, this algorithm cannot maintain an optimal number of clusters so there is much energy dissipation. In order to ensure time-bounded delay, Raje *et al.* [15] has shown a routing approach which constrains the minimum transmission range. However, this might require the deployment of multiple gateways to guarantee high sensor coverage.

3. Network Model

3.1 Communication model

This chapter has been observed that the communication energy, i.e., which consists of transmitter, receiver, antennae, and amplifier, is usually a significant component of the total energy consumed in a sensor network. To determine the optimal parameters, we consider a set of sensors deployed in a field. All sensors are arbitrarily deployed in a two-dimensional plane, and have homogeneous capability, i.e., they are equipped with GPS-capable antennae. Let d_i be the location information, and d be the distance between i and $(i+1)$, $i=1, 2, \dots, l$, for all nodes in the network. We denote the Euclidean distance between node x_i and y_i as

$$d_{(x_i, y_i)} = \sqrt{(x_i - y_i)^2 + (y_i - x_i)^2} \quad (1)$$

We use the following model for the energy dissipation used for communication. A transmit and receive energy costs for the transfer of k -bit data message between two nodes separated by a distance of d meters is given by equation 1 and equation 2, respectively.

$$e_T(k, d) = e_{Tx}k + e_{amp}(d)k \quad (2)$$

$$e_R(k) = e_{Rx}k \quad (3)$$

where $e_T(k, r)$ in equation 1 denote the total energy dissipated in the transmitter of the source node, and $e_R(k)$ represents the energy cost incurred in the receiver of the neighbor node, we use e_{Tx} and e_{Rx} to denote the per bit energy dissipations for transmission and reception, as shown in equation 2, respectively. $e_{amp}(d)$ is the energy required by the transmit amplifier to maintain an acceptable signal-to-noise ratio in order to transfer data messages reliably. We use the free-space propagation model to approximate the path loss sustained due to wireless channel transmission. Given a threshold transmission distance of r_0 , the free space model is applied for case where r less than r_0 , the energy required by the transmit amplifier $e_{amp}(r)$ is given by

$$e_{amp}(r) = \epsilon_{FS}r^4, \quad d \leq r_0 \quad (4)$$

where ϵ_{FS} shows transmit amplifier parameters corresponding to the free-space. In addition the radios on every node are switched off when they do not need to participate in any communication.

3.2 System model

We assume that the radius of a cluster is limited by the node density. These sensors with higher power levels should cover an area of at least two or more cluster diameters to guarantee that the resulting inter-cluster overlay will be connected.

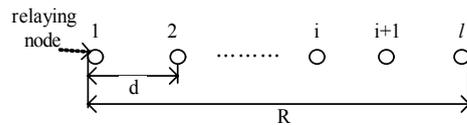
Let the clustering time taken to re-establish a grid topology, T_{grid} be the time interval taken by the sink.

Clustering terminates within a fixed number of iterations without regard to cluster range. Assuming that the process for selecting a head was completed within T_{head} , we ensure that $T_{grid} \geq T_{head}$. Clustering and the head selection process are triggered every

second, to select new cluster heads.

We consider a topology model for sensing field length (R), in which n nodes are randomly distributed in a square field.

We consider the location-based self-organization scheme, which based on grid topology a set of nodes placed along a long. Each grid have only one cluster head, which consist of relaying node or cluster head and a general set of nodes. Each relaying node also relays sensed data from each node toward sink node at the end.



[Fig. 1] Network Topology

As shown in [Fig. 1], node $i+1$ also forwards the data collected by node l to node i , each node within a cluster sends data to cluster head repeatedly. At the end, the forwarding node relays information to the sink node.

4. Proposed Scheme

4.1 Location-based Clustering

The clustering algorithm consists of four phases: subscription query, query relay, position aware and cluster head selection, as illustrated in [Fig. 2]. The position aware phase starts with the received position information from sensors. When a sink receives a reply from nodes via a back path, the sink starts deciding the path toward the sink in the backward paths, because it receives the location of route path.

At the subscription query phase, a sink directs a join query to the source. At the query relay phase, each sensor node forwards reply message to a pre-determined back path via a neighbor sensor node. At the position aware phase, our algorithm separates a

sensing field into a certain number of grid areas; the second phase, referred to as the head selection, locally selects cluster headers on this grid topology. Generally, the size of a cluster is the number of nodes belonging to it. However, in our algorithm a density (or size) of nodes d is imposed by the average number of nodes.

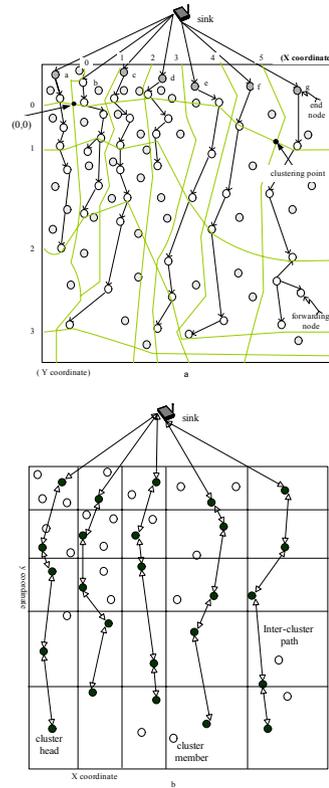
Our clustering algorithm starts with a skewed topology T_{skew} of wireless sensor networks. In this scheme, before [Fig. 2(a)], based on the information received from each sensor, the sink broadcasts an advertisement message using its maximum power range, referred to as the cluster setup (ES) message. The ES includes information of each cluster number, i.e., the cluster number C_i and cluster coordinates (x_i, y_i) , $i=1,2,\dots,z$. Upon receiving such an ES , each node a_j , where $j=1,2,\dots,m$, selects the nearest node of its adjacent nodes as a cluster header. In [Fig 2(a)], the source of the route path is $F[0]$, e.g. a, b, c, d, e, f , and g and that node denotes a relay node in order to find the route path from the sink to the gate relay nodes. Let S_i denote the accumulated sum of the hop count from the terminal node to the gate relay node along the route path.

Thus, $F[0]$ denotes the first node. In [Fig. 2(a)], let a, b, c, d, e, f , and g be a set of any relay nodes, i.e., $F[a], F[b], F[c], F[d], F[e], F[f]$, and $F[g]$, and $F[0]$'s children, i.e., nodes involved in a set of relay nodes $F[0]$. The sink finds a set of relay nodes to minimize the hop count of the route path and to maximize the power range R_m among elements of relay nodes $F[0]$, as follows:

$$\{S_i = a_1 + a_2 + \dots + a_{n-1} + a_n\} < \lfloor P_h \rfloor \quad (5)$$

Based on P_h and y -coordinate, i.e., $h=1,2,3,\dots,H$, over the route path, the sink could be identified as the relay nodes, i.e., a, b, c, d, e, f , and g .

$$E_{Rx}(l) = E_{Rx-elec}(l) = lE_{elec} \quad (6)$$



[Fig. 2] The position aware process consists of four phase: (1) query reporting phase, from the sink to sensors, (2) sink broadcasts the partition information to all node, (3) position aware: two hops distances, (4) cluster selection: After broadcasting information from sink, clusters of grid size appear.

```

For each node, j located in  $a_j \in T_{skew}$  to be included at each cluster,  $a_i \in C_i$ :
for re-establishing cluster,
if completes then
repeat
send request_message from sink to node
if sink receives information then
compute grid topology for clustering
send partition information to nodes
end if
until position_aware_time  $\leq$  threshold;
else
repeat
if one round is expired then
get backoff_time;
broadcast remaining energy;
select as a cluster head a node with  $W_i$ ;
else
while all node have transmission energy do
send its event to cluster head;
end while
end if
end if
until one_round_time;
end if
end for
    
```

[Fig. 3] Pseudo-code for clustering algorithm

Proposed algorithm is formulated, as shown in [Fig. 3] The clustering algorithm is recursively repeated along the route path. The failure of a node anywhere in the network requires that the algorithm is executed again for the entire network. It is a fully centralized algorithm, because each node determines the designated cluster based on information received from the sink. Sensor nodes do not have global knowledge of either the hop count or the geometric length of each hop. The geometric distance between nodes is approximately proportional to the hop count between them.

5. Performance Evaluation

The transmitter dissipates energy in the running of the radio electronics and the power amplifier, and the receiver dissipates energy in the running of the radio electronics, as shown in [Fig. 2] In the experiments, both the free space (d^2 power loss) and the multi-path fading (d^4 power loss) channel models were used, depending on the distance between the transmitter and receiver. Power control can be used to invert this loss by appropriately setting the power amplifier—if the distance is less than a threshold d_0 , the free space (*fs*) model is used; otherwise, the multi-path (*mp*) model is used. Thus, to transmit 1-bit message in the distance d , the radio consumes $E_{Tx}(l,d)$, and to receive this message, the radio expends:

$$E_{Tx}(l,d) = E_{Tx}(l) + E_{Tx-amp}(l,d) = \begin{cases} lE_{elec} + l_{e_{fs}} d^2, & d < d_0 \\ lE_{elec} + l_{e_{mp}} d^4, & d \geq d_0 \end{cases} \quad (7)$$

The electronics energy, E_{elec} , depends on factors such as the digital coding, modulation, filtering, and spreading of the signal, whereas the amplifier energy, $E_{fs}d^2$ or $E_{mp}d^4$, depends on the distance to the receiver and the acceptable bit-error rate, respectively. The

cluster formation algorithm was created to ensure that the expected number of clusters per round is k , a system parameter.

In the communication energy models, described in Heinzelman et al., we can determine the optimal number of cluster. Assume that there are N nodes distributed uniformly in an $M \times M$ region. If there are k clusters, there are on average N/k nodes per cluster (one cluster header and $(N/k)-1$ non-cluster header nodes). Each cluster header dissipates energy when receiving signals from the nodes, aggregating the signals, and transmitting the aggregate signal to the sink. Because the sink is far from the nodes, presumably the energy dissipation occurs according to the multi-path model (d^4 power loss). Therefore, the energy dissipated in the cluster header node during a single frame is

$$E_{CH} = lE_{elec} \left(\frac{N}{k} - 1 \right) + lE_{DA} \frac{N}{k} + lE_{elec} + lE_{mp} d_{toSink}^4 \quad (8)$$

where d_{toSink} is the distance from the cluster header node to the sink and we have assumed perfect data aggregation.

Each non-cluster header node only needs to transmit its data to the cluster header once during a frame. Presumably, the distance to the cluster header is small, so the energy dissipation occurs according to the Friss free-space model (d^2 power loss). Thus, the energy used in each non-cluster header node is

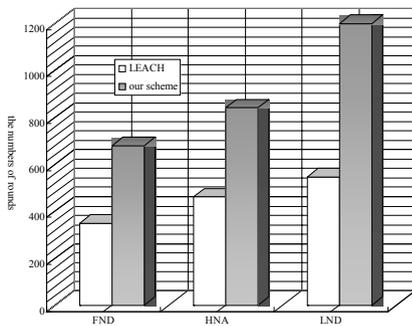
$$E_{non-CH} = lE_{elec} + lE_{fs} d_{toCH}^2 \quad (9)$$

where d_{toCH} is the distance from the node to the cluster header.

6. Simulations

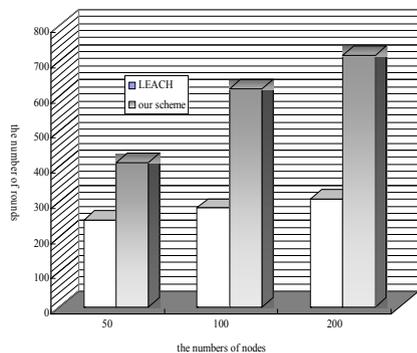
We evaluate the performance of our algorithm via a simulation study. For simulation, we used a 100-node

network, where nodes were randomly distributed in a $100 * 100$ area and the location of the sink was fixed. Each data message was 500 bytes long and the packet header for each type of packet was 25 bytes long. The communication energy parameters are set as: $E_{elec}=50nJ/bit$, $E_{fs}=10pJ/bit/m^2$, and $E_{mp}=0.0013pJ/bit/m^4$. The energy for data aggregation is set as $E_{DA}=5nJ/bit/signal$. For simulation, we use the same radio model which was discussed in [2].



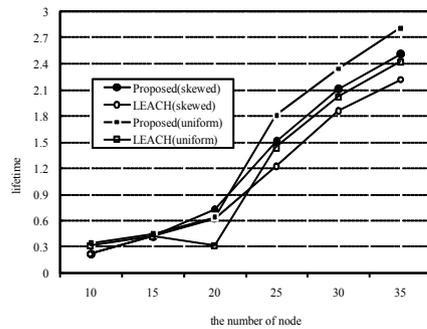
[Fig. 4] System Life Time

[Fig. 4] shows the number of rounds at which the first node dies (*FND*), half of the nodes die (*HNA*) and the last node dies (*LND*). We can see that our proposed algorithm yields approximately twice the life time than for [12], in all cases for a $100m*100m$ network. The energy consumption for forming clusters in [12], and our algorithm are similar.



[Fig. 5] Life time versus the number of nodes

[Fig. 5] shows the number of rounds completed at the FNA with respect to the number of nodes. When the node density is high, our algorithm still yields approximately twice the life time than for [12].



[Fig. 6] Compare the network lifetime of the proposed scheme to LEACH in terms of uniform and skewed

[Fig. 6] show the numerical results of proposed scheme, i.e., uniform and skewed and proposed scheme, in terms of total power consumption and network lifetime. The network lifetime increases while the total energy consumption decreases with more nodes relaying the data in the region, because the burden is evened out when the total number of nodes increase.

Our algorithm has a series of advantage for maintaining the optimal number of cluster headers, without any negotiation between the sensor nodes for the election of cluster headers. However, the proposed algorithm has scalability problems for large sensor networks and is not directly applicable to the support of variant data delivery models such as [11]. Our scheme is more efficient than [12] and there is less variation of energy consumption. Because this protocol uses cluster header nodes with more energy than the nodes along the shorter routes, the optimal routes are chosen. Therefore, the remaining energy in all nodes is evenly maintained. This is mainly due to the load balance algorithm used in our algorithm. Thus, all nodes try to evenly share their lifetime.

7. Conclusion and Future Work

To investigate the energy consumption problem in skewed-topology wireless sensor networks, a location-based clustering scheme presented. Proposed scheme is an energy efficient location-based clustering scheme in wireless sensor networks with skewed-topology.

Performance of the proposed scheme is assessed by simulation and compared to other clustering-based protocol (LEACH), because of having the characteristic of centralized topology. Simulation results show that proposed scheme yields a much longer network life time, by reducing the sensor power consumption. Our scheme outperforms its comparatives by uniformly placing cluster heads based-on location throughout the entire interesting fields, performing balanced clustering.

However, we could not remove a detour routing problems for forwarding fused data to the sink (or base station) in our scheme. Therefore, we plan to extend our simulation results by studying additional network parameters and more general topologies.

REFERENCES

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: a survey," *Computer Networks*, Vol. 38, pp. 393-422, 2002.
- [2] C. Chong, S.P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges," in *Proc. of the IEEE*, Vol. 9, No. 8, pp. 1247-1256, 2003.
- [3] http://en.wikipedia.org/wiki/wireless_sensor_network
- [4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor networks: A survey", *Computer Networks Elsevier Journal*, Vol 38, No. 4, pp.393-422, 2003.
- [5] T. Haensekmann, *Sensornetworks, GFDL Wireless Sensor Networks textbook*, 2007.
- [6] Ian F. Akyildiz, Mehmet Can Vuran, "Wireless Sensor Networks", Wiley, 2010.
- [7] H. Chan and A. Perrig, "ACE: An Emergent Algorithm for Highly Uniform Cluster Formation," in *Proc. of 1st European Workshop Wireless Sensor Networks, Lecture Note in Computer Science(LNCS)*, Vol. 2920, pp. 157-171, 2004.
- [8] S. Bandyopadhyay and E.J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks," in *Proc. of IEEE INFOCOM 2003*, Vol. 3, pp. 1713-1723, 2003.
- [9] SNico Loose, Hannens Frey, "Accuracy over Energy Metrics for Evaluating Sensor Network MAC Protocols", *IEEE 8th International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 74-80, 2010.
- [10] Pei Huang, Li Xiao, Soroosh Soltani, Matt W. Mutka, Ning Xi, "The Evolution of MAC Protocols in Wireless Sensor Networks : A Survey", *IEEE Communications Surveys & Tutorials*, Vol. 15, No 1, pp. 101-120, 2013.
- [11] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," *IEEE Trans. on Mobile Computing*, Vol. 3, No. 4, pp. 366-379, 2004.
- [12] Mahapatro. Arunanshu, Khilar. Pabitra, "Energy-efficient distributed approach for clustering-based fault detection and diagnosis in image sensor networks", *IET Wireless Sensor System*, Vol. 3, No. 1, 2013.
- [13] Bari, A., Chaturvedi, R., Jaekel, A., Bandyopadhyay, S., "Energy Aware Distributed Clustering in Two-Tiered Sensor Networks *Computer Communications and Networks*", *Proceedings of 18th International Conference on ICCCN*, pp. 1-8, 2009.
- [14] Potente. Michael, "An energy-sensor network takes center stage during endothelial aging", *Circulation research : a journal of the American Heart Association*, Vol. 106, No. 8, pp. 1316-1318, 2010.
- [15] Raje. RadhikaA., Sakhare, ApekshaV., "Routing in Wireless Sensor Network Using Fuzzy Based Trust Model", *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on 2014*, pp. 529-532, 2014.

최 해 원(Choi, Hae Won)



- 1996년 2월 : 경일대학교 컴퓨터공학과(공학사)
- 2000년 2월 : 경북대학교 컴퓨터공학과(공학석사)
- 2009년 2월 : 경북대학교 컴퓨터공학과(공학박사)
- 2006년 3월 : 경운대학교 컴퓨터공학과 교수

· 관심분야 : 알고리즘, 유비쿼터스 컴퓨팅
· E-Mail : chw@ikw.ac.kr

류 명 춘(Ryoo, Myung Chun)



- 1989년 2월 : 영남대학교 컴퓨터학과(공학사)
- 1991년 2월 : 영남대학교 컴퓨터공학과(공학석사)
- 2009년 2월 : 영남대학교 컴퓨터공학과(공학박사)
- 1997년 3월 ~ 현재 : 경운대학교 컴퓨터공학과 교수

· 관심분야 : 지능정보시스템, Bioinformatics
· E-Mail : mcryoo@ikw.ac.kr

김 상 진(Kim, Sang Jin)



- 1994년 2월 : 계명대학교 컴퓨터공학과 (공학사)
- 1996년 2월 : 경북대학교 컴퓨터공학과 (공학석사)
- 2000년 8월 : 경북대학교 컴퓨터공학과 (공학박사)
- 1999년 9월 ~ 현재 : 경운대학교 컴퓨터공학과 교수

· 관심분야 : 알고리즘, 게임이론
· E-Mail : sjkim@ikw.ac.kr