

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Approximation Two Independent Sets Based Connected Dominating Set Construction Algorithm for Wireless Sensor Networks

Z. Liu, B. Wang and Q. Tang

Department of Control Science and Engineering,
Huazhong University of Science and Technology, Wuhan Hubei, 430074, China

Abstract: In WSNs (Wireless Sensor Networks), an optimized way of prolonging the networks lifetime and flooding packets is to find the minimum CDS (Connected Dominating Set). In this study, a new method called ATISA (Approximation Two Independent Sets based Algorithm) for constructing CDS is proposed. The ATISA has three stages. The first stage is constructing a connected set CS (connected set) and the second stage is constructing a connected dominating set CDS and the third stage is pruning the redundant dominators of CDS. The performance ratio of ATISA is approximately $(9.67+19n_k)$ when the number of nodes is bigger enough and the message complexity is $O(n)$. Compared with some famous CDS construction algorithms, ATISA constructs the CDS with the smallest size.

Key words: WSN, algorithm, CDS, connected set, virtual backbone

INTRODUCTION

In wireless sensor network, all nodes are energy constrained and designing an energy efficient routing is very important (Dai *et al.*, 2009). Clustering routing protocol is a kind of energy efficient routing (Zhou and Meng, 2008), while using a virtual backbone to organize the nodes and perform an efficient routing is a better way.

A Connected Dominating Set (CDS) is a kind of virtual backbone and design an algorithm for constructing the CDS with the minimum size is a hot issue in recent years. The CDS problem can be defined as follows: find a minimum size subset S of nodes, such that the sub-graph induced by S is connected and S forms a dominating set. This problem is shown to be NP-hard (Jeremy *et al.*, 2004).

In WSN, the routing process can be restricted to the reduced graph formed by the CDS, every node form the CDS will be considered as a gateway and only the gateway needs to keep routing information. However, an important aspect is how to select an optimal CDS.

The CDS construction algorithms are mainly classified into two classes, which are Unit Disk Graph (UDG) based algorithms and Disk Graphs with Bidirectional links (DGB) based algorithms (Thai *et al.*, 2007). In UDG and DGB, the link between any pair of nodes is bidirectional.

In this study, a new CDS construction algorithm ATISA is proposed. This algorithm can be implemented in both UDG and DGB. ATISA has three stages. The first

stage is constructing a connected set, which are composed of two independent sets. The second stage is an assistant process, which makes the Connected Set (CS) to dominate all the nodes in the network and thus a Connected Dominating Set (CDS) is constructed. The last stage is a pruning process, which reduces the size of the CDS. The performance ratio of ATISA is given and the properties of the CDS are analyzed. In the performance evaluation, ATISA is compared with some classic algorithms (Wan *et al.*, 2002; Thai *et al.*, 2007; Yingchang *et al.*, 2009; Wu and Li, 1999).

Some CDS construction algorithms are introduced. Because the link between any two neighboring nodes in UDG and DGB is bidirectional and most of the CDS construction algorithms are based on the bidirectional links, thus the algorithms designed in UDG can be easily extended to DGB.

In the graph $G(V, E)$, $C(C \subseteq V)$, is a subset of V . If any node $u(u \in V)$ is either in C or adjacent to one node in C , then C is one of the dominating set of $G(V, E)$. Besides if there is at least one path between any pair of nodes in C , then C is one of the connected dominating set (CDS) of $G(V, E)$. Independent Set (IS ($IS \subseteq V$)) is a subset of V . In IS, any pair of nodes is disconnected. Given an Independent Set (IS), if any one node $v(v \in V-IS)$ is added into the IS, there will be a pair of nodes connected, then the IS is a Maximum Independent Set (MIS). In MIS, any pair of nodes is disconnected and any node not belonging to MIS is adjacent to at least one node in MIS.

In Jeremy *et al.* (2004), Blum *et al.* introduced a lot of approximation MCDS algorithms in sensor networks and MANETs. Blum point out that the two algorithms, Guha and Khuller (1998) have motivated many designs of CDS construction algorithms. One of the algorithms (Guha and Khuller, 1998) is growing a spanning tree in a single phase and the other is growing separate components in the first phase and connecting the components together in the second phase.

Some distributed approximation MCDS algorithms are proposed by Alzoubi *et al.* (2002a), Wan *et al.* (2002) and Alzoubi *et al.* (2002b), the distance between any pair of complementary subsets in the constructed MIS is exactly two hops. A rooted spanning tree is constructed at first and all of the nodes in the network have their tree levels, respectively. The level of the root is 0. If the parent node's level is i , then the child's level is $i+1$. All nodes have to get their neighbors information. The algorithm has two phases and the first phase is constructing a MIS and the lowest level is the dominators election criterion. The second phase is CDS construction phase, where the connectors are selected. If a black node receives an INVITE message from a gray neighbor, it immediately chooses the gray node as its parent and the gray node becomes a connector. This algorithm has the performance ratio $8\text{opt}+1$. The message complexity is $O(n \log n)$ and the time complexity is $O(n)$. The algorithm proposed by Wan *et al.* (2002) has two phases and in the first phase the dominator selection is based on the lower level neighbors' states. If all of a white node's lower level neighbors are marked gray, then the white node becomes a dominator immediately. In the second phase, the gray node with the maximum dominator nodes degree is selected as the connector. The performance ratio is $8\text{opt}-2$. The time complexity is $O(n)$ and the message complexity is $O(n \log n)$. Alzoubi *et al.* (2002a) proposed a more complex algorithm containing two phases. The first phase is constructing a MIS and the second phase is finding some connectors with more complex steps. The performance ratio is $192\text{opt}+48$. The time complexity is $O(n)$ and the message complexity is $O(n)$.

Li *et al.* (2004, 2005), Thai *et al.* (2007), proposed some distributed connected dominating set construction algorithms. A greedy algorithm called S-MIS is proposed. S-MIS has two phases. In the first phase, the MIS construction algorithm is described by Wan *et al.* (2002). The second phase is the main contribution of Li *et al.* (2005) and in this phase, black-blue component is proposed, which is a connected component of the sub-graph induced only by black and blue nodes and by ignoring connections between blue nodes. The black node is the dominator of the first phase and the blue node

is the dominator of the second phase. In the second phase, a greedy algorithm is used to find some gray nodes adjacent to at least 2 black nodes in different black-blue components. The performance of S-MIS is $(5.8+\ln 4)\text{opt}+1.2$. Thai *et al.* (2007) proposed two algorithms called TFA and TSA, respectively. TFA is an extension of S-MIS in the DGB and the algorithm codes are not changed, which explain that the algorithms implemented in UDG can be easily extended in DGB. The second algorithm TSA is an improved algorithm for TFA. TSA has two phases and in the first phase, an MIS is constructed by choosing the nodes with the biggest radius as the dominators. The size of MIS constructed by TSA is smaller than that constructed by TFA. The second phases of TFA and TSA are the same. The performance ratio of TFA is $(K+2+\ln K)\text{opt}$ and TSA's performance ratio is not introduced by Thai *et al.* (2007). A one-phase based CDS construction algorithm is proposed by Li *et al.* (2004). The algorithm is better at maintenance since there is no need to build a tree or select a leader. In this algorithm, each node needs to know the connectivity information within its 2-hop neighborhood. If a node has the smallest 2-hop number of neighbors, it becomes a dominator1 and broadcasts a BLACK message. If a node u receives a BLACK message from a node w and if u has already received a notification that there is a 2-hop black neighbor v sent by a neighbor m and w has not been connected to v yet, then u and m become dominator 2 nodes and connect node v and w . The dominator1 nodes form an MIS and the performance ratio of the algorithm is $172\text{opt}+43$. The time complexity is $O(\Delta)$ and message complexity is $O(n\Delta^2)$, where Δ is the maximum degree.

Many MIS based algorithms were reported by Das and Bharghavan (1997), Raei *et al.* (2008, 2009), Rai *et al.* (2009), Zeng *et al.* (2006), Funke *et al.* (2006), Yingchang *et al.* (2009), Wu and Li (1999) and Dai and Wu (2004) are mainly have two stages. The first stage is constructing a MIS and the second stage is finding some connectors to connect the nodes in MIS and finally get a CDS. The CDS construction algorithms can have different criterions. The criterion is the node degree as reported by Das and Bharghavan (1997), Raei *et al.* (2008) and Rai *et al.* (2009). For example, Das and Bharghavan (1997) mentioned the node degree is the number of the unmarked neighbors in the first stage and in the second stage, the node degree is the number of unmarked nodes adjacent to the non-fragment nodes. The algorithm reported by Das and Bharghavan (1997) has the performance ratio $2H(\Delta)+1$. The time and message complexities are $O(n+|C|\Delta)$ and $O(n|C|+m+n \log n)$ respectively. The criterion for constructing a CDS is a timer, which is based on the transmission ranges of nodes (Raei *et al.*, 2009).

The criterion is the weight of node and the weight is composed of the node degree and the battery power. In (Funke *et al.*, 2006; Yingchang *et al.*, 2009), the criterion is the node id. The node with the largest id will become a dominator (Funke *et al.*, 2006), while the node with the smallest id will become a dominator (Yingchang *et al.*, 2009).

The algorithms described above are all MIS based. There are also some non-MIS based algorithms reported by Jie *et al.* (2008), Dai and Wu (2004) and Rong *et al.* (2009).

Jie *et al.* (2008) and Dai and Wu (2004) proposed some distributed algorithms in UDG and DGB. A very simple marking process is proposed. In the marking process every node v exchanges its open neighbor set $N(v)$ with its neighbors and if there exist two unconnected neighbors, v will mark itself. All the marked nodes form a CDS. The CDS constructed in the first phase contains many redundant nodes. Jie *et al.* (2008) proposed two prune algorithms called Rule-1 and Rule-2 to delete some redundant marked nodes. The performance ratio is not given by Wu and Li (1999), but in (Wan *et al.*, 2002) the performance ratio of this algorithm is proved $O(n)$, where n is the number of nodes. The message complexity is $\Theta(m)$, where m is the number of edges in the unit-disk graph. The time complexity is $O(\Delta)^3$, where Δ is the maximum degree. Dai and Wu (2004) proposed an extended algorithm for the algorithm by Wu and Li (1999). The marking process is the same as Wu and Li (1999), but the pruning process is improved and denoted by Rule-k. Based on the Rule-k, the size of the final constructed CDS is reduced. A hierarchical graph is constructed at first and then the essential nodes are determinate in the graph according to the Rule-1 by Rong *et al.* (2009). After the essential nodes determination phase is ended, there are some redundant dominators, which are deleted according to the Rule-2 by Rong *et al.* (2009). The performance ratio of the algorithm by Rong *et al.* (2009) is $O(\log n)$.

The objectives of the algorithms described above are minimizing the size of the CDS. But in some researches such as (Donghyn *et al.*, 2009; Thai *et al.*, 2008; Feng *et al.*, 2009; Shuhui *et al.*, 2008; Jie *et al.*, 2006, 2008; Adjih *et al.*, 2005; Yu *et al.*, 2006; Ivan *et al.*, 2002; Jie and Fai, 2006), the objectives of CDS construction are not or not only minimizing the size of the CDS. For example, in Donghyn *et al.* (2009), Kim *et al.* proposed two centralized algorithms. The two algorithms have the constant performance ratios of the size and diameter of the constructed CDS. The diameter of the CDS is a new term considered for the constructed CDS. Thai *et al.* (2008) proposed a constant approximation algorithm for the Strongly Connected Dominating Set

(SCDS) problem. Feng *et al.* (2009), proposed a new algorithm CDSA to construct a 2-connected virtual backbone. CDSA can resist the failure of one node. Shuhui *et al.* (2008) and Jie *et al.* (2008) developed a notation of directional network backbone by using the directional antenna model and then formed the problem of constructing a directional connected dominating set (DCDS). Jie *et al.* (2008) also proposed a localized heuristic algorithm as well as two extensions of the algorithm for constructing a DCDS by Shuhui *et al.* (2008).

NETWORK MODEL AND PRELIMINARY WORK

This study assumes that N sensor nodes are randomly scattered in a two-dimensional square field:

- This network is a static deployed network. It means sensor nodes do not move any more after deployment
- Sensor nodes have different transmission ranges.
- The nodes can get their position coordinates through GPS or location algorithms
- After the establishment phase of the topology, the network is duty-cycle based
- The link between any pair of nodes is bidirectional and the neighbor set of a node is defined by Rule 1

Rule 1: If and only if $d(u, v) < \min\{r_u, r_v\}$, node v is a neighbor of u and node u is a neighbor of v .

In the CDS construction process, there are three states, which are white state, gray state and black state. The white state is the initial state and in the beginning, all the nodes are marked white.

Before describing the centralized and distributed algorithms, we introduce some symbols in Table 1. The

Table 1: Symbols definitions

Symbol	Meaning	Contents
BMSG	black message	id, level
GMSG	gray message	id
BCMSG	black children message	id, level, black children set
IMSG	invite parent message	id, id
ISMSG	invite schedule message	id, ids
SMSG	schedule message	id, ids
IBSMMSG	invite black set message	id
BSMMSG	black set message	id, ids
N	open neighbor set	ids, positions
BN	black neighbor set	ids, levels
WN	white neighbor set	ids
GN	gray neighbor set	ids, levels
BP	black parent set	ids, levels
BC	black children set	ids, levels
B	black-node set	ids, levels
G	gray-node set	ids, levels
W	white-node set	ids, levels
state	node state	white/black/gray
level	node level	an integer

Centralized Algorithm 2: The second stage algorithm

```

1: while  $W \neq \Phi$ 
2:   foreach  $i$  in  $W$  do
3:     if  $\exists j \in B \cap N_i$ 
4:        $G \leftarrow G \cup \{i\}$ ,  $W \leftarrow W \setminus \{i\}$ 
5:     elseif  $\exists j \in G \cap N_i$ 
6:        $B \leftarrow B \cup \{i\}$ ,  $W \leftarrow W \setminus \{i\}$ ,  $G \leftarrow G \setminus \{i\}$ 
7:     end if
8:   end for
9: end while
10: return  $B, G, W$ 

```

The second stage constructs a connected dominating set and all the nodes are either black or gray and there is no white node left in the network. But there are some redundant black nodes in the network and in the third stage the redundant black nodes are deleted.

In the third stage, if a black node with no children and if the neighbors of the black node are all adjacent to at least two black nodes (include the black node itself), then the black node is put into G and deleted from B . The algorithm is shown as follows:

Centralized Algorithm 3: The third stage algorithm

```

1: foreach  $i$  in  $B$  do
2:   if  $C_{bi} = \Phi$ 
3:     if  $\forall j \in G \cap N_i$  has two or more black neighbors
4:        $G \leftarrow G \cup \{i\}$ ,  $B \leftarrow B \setminus \{i\}$ 
5:     end if
6:   end for
7: end for
8: return  $B, G$ 

```

Distributed implementation: In the distributed implementation, all the nodes exchange their positions information with their neighbors and put their one-hop neighbors information into set N_i , respectively at first and then all of the nodes have to exchange their neighbors information with their neighbors and get their two-hop neighbors information.

Algorithms in the first stage: In the first stage, all nodes are initialized white. Before the first stage ended, there are white nodes, gray nodes and black nodes.

In the first stage, if a node is a black node, it will calculate its farthest independent neighbors and put the neighbors into the corresponding set and then broadcast a BCMSG containing the independent neighbors in the network. If a black node has no white neighbor and has obtained its children black nodes, the black node will do nothing. Before calculate the farthest independent neighbors, a black node should wait for a period of time and get all of the latest neighbors information. If a node is a gray node, it will update its black neighbor set according to the BMSGs it

received. If a node is a white node, it will do something according to the message it received. If a white node receives a BCMSG and it is included in the message, it will mark itself black and put the black node where the BCMSG comes from into the corresponding set and then broadcast a BMSG. If a white node receives a BMSG, it will update its corresponding sets and check whether it has received two or more black messages. If it has received two or more black messages, it will change its stage into gray and broadcast a GMSG. If a white node receives a GMSG, it only updates its corresponding sets. The algorithm is shown as follows:

Distributed Algorithm 1: The first stage algorithm

```

1: if  $state_i = \text{black}$ 
2:   if  $WN_i \neq \Phi$  &  $CB_i = \Phi$ 
3:     sets a timer
4:     While timer > 0
5:       If receives GMSG from  $x$ 
6:          $GN_i \leftarrow WN_i \cup \{x\}$ ,  $WN_i \leftarrow WN_i \setminus \{x\}$ 
7:       end if
8:     end while
9:   while  $WN_i \neq \Phi$ 
10:     $u \leftarrow$  a farthest node in  $WN_i$ 
11:     $WN_i \leftarrow WN_i \setminus \{WN_i \cap WN_u\}$ 
12:     $WN_i \leftarrow WN_i \cup \{WN_u \cap WN_i\}$ 
13:     $BC_i \leftarrow BC_i \cup \{u\}$ ,  $BN_i \leftarrow BN_i \cup \{u\}$ 
14:     $WC_i \leftarrow WN_i \setminus \{u\}$ 
15:  end while
16:  if  $WN_i = \Phi$ 
17:    broadcasts a BCMSG
18:  end if
19:elseif  $state_i = \text{white}$ 
20:  if receives a BCMSG from  $x$ 
21:    if  $i$  is included by  $CB_x$ 
22:       $state_i \leftarrow \text{black}$ 
23:       $level_i \leftarrow level_x + 1$ 
24:       $BP_i \leftarrow BP_x \cup \{x\}$ ,  $BN_i \leftarrow BN_x \cup \{x\}$ 
25:      broadcast BMSG
26:    end if
27:  elseif receives a BMSG from  $x$ 
28:     $BN_i \leftarrow BN_i \cup \{x\}$ ,  $WN_i \leftarrow WN_i \setminus \{x\}$ 
29:    if  $size(BN_i) \geq 2$ 
30:       $state_i \leftarrow \text{gray}$ , broadcasts a GMSG from  $x$ 
31:    end if
32:  elseif receives a GMSG from  $x$ 
33:     $GN_i \leftarrow GN_i \cup \{x\}$ ,  $WN_i \leftarrow WN_i \setminus \{x\}$ 
34:  end if
35:elseif  $state_i = \text{gray}$ 
36:  if receives a BMSG from  $x$ 
37:     $BN_i \leftarrow BN_i \cup \{x\}$ ,  $WN_i \leftarrow WN_i \setminus \{x\}$ 
38:  end if
39:endif

```

If there are no new black nodes generated, the first stage is ended. When the first stage is ended, the white nodes which have received only one black message should change their states into gray.

The algorithms executed in the first stage can make most of the network nodes change their states into either black or gray. Sometimes there are little white nodes left after the first stage. The black nodes generated in the first stage construct a connected set (CS), but the CS is not a dominating set sometimes. In the second stage, the connected dominating set (CDS) is constructed.

Algorithms in the second stage: In the second stage, there are black nodes, gray nodes and sometimes white nodes in the network. In order to make the constructed set in the first stage dominating all the nodes, the white nodes should select some gray nodes as black nodes. Only white nodes can change their states into gray and only gray nodes can change their states into black in this stage.

In this stage, if a node is a white node, it will check whether its BN is empty. If it's BN is empty, it will check whether its GN is empty. If it GN is not empty and BN is empty, it will broadcast a GMSG and send an IMSG to the node with the minimum id in GN and change its state into gray. If a white node's BN is not empty, it will change its state into gray and broadcast a GMSG. At the same time, the white nodes update their corresponding sets according to the messages they have received such as GMSG, BMSG. If a node is a gray node, if it receives an IMSG, it will change its state into black and broadcast a BMSG and then send an IMSG to a black neighbor with the minimum id. At the same time, a gray node should update its corresponding sets according to the messages it has received such as BMSG, GMSG and IMSG. If a node is a black node, it only updates its corresponding sets according to the messages it has received. The algorithm is shown as follows:

Distributed Algorithm 2: The second stage algorithm

```

1:  if statei = white
2:    if BNi =  $\Phi$ 
3:      if GNi  $\neq \Phi$ 
4:        statei  $\leftarrow$  gray, broadcasts a GMSG
5:        Minid  $\leftarrow$  the minimum id node in GNi
6:        sends an IMSG to the minid
7:        BNi  $\leftarrow$  BNi  $\cup$  {minid}, GNi  $\leftarrow$  GNi \ {minid}
8:      end if
9:    else
11:   statei  $\leftarrow$  gray, broadcasts a GMSG
13:   end if
14:   if receives an BMSG from node x
15:     BNi  $\leftarrow$  BNi  $\cup$  {x}, GNi  $\leftarrow$  GNi \ {x}
16:   elseif receives a GMSG from node x
17:     GNi  $\leftarrow$  GNi  $\cup$  {x}, WNi  $\leftarrow$  WNi \ {x}
18:   end if
19: elseif statei = gray

```

Continued

```

20:  if receives an IMSG from node x
21:    statei  $\leftarrow$  black, broadcasts BMSG
22:    mind  $\leftarrow$  the minimum id node in BNi
23:    sends an IMSG to the minid
24:    GNi  $\leftarrow$  GNi  $\cup$  {x}, BPi  $\leftarrow$  BPi  $\cup$  {mind}
25:  elseif receives a BMSG from node x
26:    BNi  $\leftarrow$  BNi  $\cup$  {x}, GNi  $\leftarrow$  GNi \ {x}
27:  elseif receives a GMSG from node x
28:    GNi  $\leftarrow$  GNi  $\cup$  {x}, WNi  $\leftarrow$  WNi \ {x}
29:  end if
30: elseif statei = black
31:   if receives an IMSG from node x
32:     BCi  $\leftarrow$  BCi  $\cup$  {x}
33:   elseif receives a BMSG from node x
34:     BNi  $\leftarrow$  BNi  $\cup$  {x}, GNi  $\leftarrow$  GNi \ {x}
35:   elseif receives a GMSG from node x
36:     GNi  $\leftarrow$  GNi  $\cup$  {x}, WNi  $\leftarrow$  WNi \ {x}
37:   end if
38: end

```

When there are no white nodes left, the second stage is ended. In this stage, the black nodes construct a Connected Dominating Set (CDS). But the CDS has some redundant nodes, which should be deleted in the next stage.

Algorithms in the third stage: In the third stage, the redundant black nodes are deleted. This stage contains two sub-stages, which are generating schedule stage and deleting redundant black nodes stage. The first sub-stage generates a schedule for all the leaf black nodes. All the leaf black nodes are allocated their time slots in the schedule, respectively. If a node is a black node without children, it will broadcast an ISMSG to its parent and then wait the SMSG from its parent. If a node is a black node with children, it waits the ISMSGs from all the children and if it has received all the ISMSGs from its children, it will fuse all the information in the ISMSGs into a new ISMSG, which will be sent to its parent. The algorithm is shown as follow:

Distributed Algorithm 3-I: The third stage algorithm I

```

1:  if statei = black
2:    if BCi =  $\Phi$ 
3:      sends a ISMSG to j  $\in$  BPi
5:    elseif BCi  $\neq \Phi$ 
7:      if receives all the ISMSGs from j  $\in$  BPi
8:        sends a ISMSG to j  $\in$  BPi
10:       if receives a SMSG from j  $\in$  BPi
11:         sends the SMSG to j  $\in$  BCi
12:       end if
13:     end if
14:   end if
15: end if

```

In the algorithm described above, all leaf black nodes send their ids to the root by sending the ISMSGs. The

non-leaf black nodes fuse all the ISMSGs from its children and send a new ISMSG to their parent nodes, respectively. The new ISMSG contains the ids of the leaf black nodes of the branch where the black node sending the new ISMSG stays. The root black node allocates the time slots according to the leaf nodes' ids. The leaf node with the minimum id has the first time slot in the schedule and the node with the maximum id has the last time slot in the schedule. When all the leaf nodes receive the schedule, they can get their time slot from the schedule, respectively and execute the second sub-stage algorithm in the corresponding time slot. If a time slot of a black node without children arrives, the black node will broadcast an IBSMSG to request all the neighbors' black neighbor set information. If the black node receives all the reply messages (BSMSGs) from its neighbors, it will calculate the number of the black neighbors of every neighbor. If every neighbor has at least two black neighbors, the black node changes its state into gray and broadcast a GMSG. The algorithm in this sub-stage is shown as follows:

Distributed Algorithm 3-II: The third stage algorithm II

```

1: if statei = black & BCi = Φ
2:   broadcasts an IBSMSG
3:   receives BSMSGs
4:   if size(BNi) > 2, ∀ j ∈ Ni
5:     statei ← gray, broadcasts a GMSG
6:   end if
7: end if

```

In the second sub-stage, only the leaf black nodes execute the algorithm described above and all the other nodes only receive the messages and update the neighbors' information according to the messages they have received.

THEORETICAL ANALYSIS

We first give the maximum number of independent neighbors in both heterogeneous and homogeneous network and then introduce the size of the connected set constructed in the first stage. Finally the properties of the connected set as well as the approximation performance ratio and the message complexity of CDS are analyzed.

The maximum number of independent neighbors: According to Thai *et al.* (2007), the maximum number of independent neighbors is described by the following equation:

$$K = \begin{cases} 5 & \text{if } k = 1 \\ 10 \left\lfloor \frac{\ln k}{\ln(2 \cos(\pi/5))} \right\rfloor & \text{otherwise} \end{cases} \quad (1)$$

In Eq. 1, $k = r_{\max}/r_{\min}$. If $k \rightarrow 1$, the network tend to be a homogeneous network, where the transmission radii are the same. According to Eq. 1, the maximum number of independent neighbors of the homogeneous network is 5 and if $k \rightarrow 1$, the maximum number of independent neighbors of the heterogeneous network which tend to be a homogeneous network tend to be zero, which contradicts the Eq. 1. In order to unify the description of the maximum number of independent neighbors of the homogeneous and heterogeneous network, we give the following equation:

$$K = 5 + 10 \left\lfloor \frac{\ln k}{\ln(2 \cos(\pi/5))} \right\rfloor \quad (2)$$

In Eq. 2, K is bigger than the bounds in Eq. 1 and if $k \rightarrow 1$, the K in Eq. 2 is tend to 5, which is the first case in Eq. 1. Thus, the Eq. 2 is a bound for UDG and DGB.

The size of CT: According to the first stage algorithm, the constructed connected set sometimes can not dominate all the nodes in the network, thus the second stage should be adopted to make the connected set to be a connected dominating set. But the connected set can dominate most of the network nodes, when the nodes number is big enough. In other words, if the number of nodes is bigger than a critical value, the connected set can dominate all most all the nodes in the network.

Let r_{\max} and r_{\min} represent the maximum radius and the minimum radius, respectively. M represents the side length of the square area. N represents the number of nodes in the network. Assume all the nodes are uniformly distributed in the network area. The average occupied size of one node is:

$$S_{oc} = \frac{M^2}{N} \quad (3)$$

The expected transmission radius is:

$$R_{exp} = \frac{r_{\min} + r_{\max}}{2} \quad (4)$$

The size of expected coverage area is:

$$S_{ca} = \pi R_{exp}^2 \quad (5)$$

The average covered node number is:

$$C_{ave} = \frac{S_{ca}}{S_{oc}} \quad (6)$$

In order to make the network have approximately full coverage, a node in the network should have at least K neighbors. In other words, if a node covers at least K neighbor nodes and itself, the network can have big enough number of node and the connected set can dominate all most all the nodes. So, the following equation is got:

$$C_{ave} > K+1 \tag{7}$$

According to Eq. 2-6, the following relationship is deduced:

$$N > \frac{\left(10 \left[\frac{\ln \frac{r_{max}}{r_{min}}}{\ln(2 \cos \frac{\pi}{5})} + 5 + 1 \right] M^2 \right)}{\pi \left(\frac{r_{min} + r_{max}}{2} \right)^2} \tag{8}$$

If the number of nodes satisfies the condition in Eq. 8, the connected set can dominate approximately all the nodes in the network. The performance simulation is shown in Fig. 1.

Properties

Lemma 1: Given a connected graph, the set constructed by ATISA is a connected dominating tree.

Proof: Assume there are some white nodes denoted by A, B, C left in the network after the connected dominating set constructed by ATISA. According to the construction algorithm, A and B and C has no gray neighbors and black neighbors, or they will invite some nodes to be black nodes and join into the set. Thus A, B and C are isolated and the original graph is not connected, which contradicts the conditions. So, the set is a Dominating Set (DS).

According to ATISA, the black nodes generated in the first stage can communicate with the root, because they are derived from the root. Any generated black node in the second stage can communicate with the black node generated in the first stage, because the first generated black node in the second stage is the gray node. Thus, any pair of black nodes can communicate with each. Because any gray node is dominated by one or more black nodes, so any pair of nodes is connected. So, the tree constructed by ATISA is a Connected Set (CS).

Because all the black nodes have their parent or children respectively and the set constructed by ATISA is a connected dominating set, thus the set constructed

by ATISA is a Connected Dominating Tree (CDT). The set constructed by the first stage of ATISA is a Connected Tree (CT) and the set constructed by the first and second stage is a CDT.

Lemma 2: Any black node has only one parent black node.

Proof: Assume a black node A has two or more parent black nodes. A should receive two or more black messages before it becomes a black node and according to ATISA, if a white node receives two black messages, it will become a gray node, thus A is a gray node, which contradicts that A is a black node in the network.

Lemma 3: in the CT, any two black nodes which are not parent child relation are not adjacent.

Proof: Assume two black nodes denoted by A and B. They are not parent child relation but adjacent with each other. Before A and B become the black nodes, they should receive their parent black node's message. Because they are adjacent, at least one of them should receive a black message from the other. Thus, one of them should receive two black messages and become a gray node, which contradicts that A and B are black nodes.

Black nodes in the network have their levels, respectively and the level of black node is either even or odd.

Lemma 4: In the CT, any two even level or odd level black nodes are not adjacent.

Proof: According to lemma 3, in the CT, any black node can only adjacent with its parent black node and children black nodes. In other words, if a black node with even level, its parent black node and children black nodes have odd levels. Thus, any two even level or odd level black nodes are not adjacent.

Theorem 1: The CT is composed of two independent sets.

Proof: according to lemma 4, any black node in the connected tree has either even level or odd level. Thus, the connected tree is composed of two black nodes sets, which are even level black node set and odd level black node set. Because any two even level or odd level black nodes are not adjacent, thus the even level black node set and odd level black node set are both independent sets.

We use Even Independent Set (EIS) and Odd Independent Set (OIS) denote the even level black node set and odd level black node set, respectively.

Performance ratio: The performance ratio analysis is based on Eq. 8. In other words, if the number of nodes is bigger than the critical value, the constructed tree in the first stage can approximately dominate all of the network nodes and the black nodes generated in the second stage can be ignored. So, the size of CDS constructed in the first and second stages of ATISA is approximately equal to the size of the tree constructed in the first stage of ATISA.

Because the size of CDS constructed ATISA is smaller than that of the tree constructed in the first and second stage and the size of the tree constructed in the first stage is approximately equal to that of the tree constructed in the first and second stage, so we only need to analyze the performance ratio of the tree constructed in the first stage, which is bigger than the performance ratio of ATISA.

According to Yingchang *et al.* (2009), we give the following equation:

$$n_k = \left\lceil \frac{\ln \frac{r_{max}}{r_{min}}}{\ln(2 \cos \frac{\pi}{5})} \right\rceil \quad (9)$$

According to Yingchang *et al.* (2009), there are two cases:

Case 1: If a black node *i* does not have a parent the biggest number of children black nodes it can generate is $5+10n_k$.

Case 2: if a black node *i* has a parent the biggest number of children black nodes it can generate is $4+7n_k$.

Yingchang *et al.* (2009), the size of a maximum independent set *S* satisfies the following condition:

$$|S| \leq (opt - t)(4 + 7n_k) + t(5 + 10n_k) \quad (10)$$

$$\frac{t}{opt - t} \leq \frac{5}{1} \quad (11)$$

In Eq. 11 and 10, *t* represents the number of black nodes belonging to case 1 and *opt-t* represents the number of black nodes belonging to case 2. *opt* represents an optimal solution of the CDS problem. Thus, the size of *S* satisfies:

$$|S| \leq \left(\frac{29}{6} + \frac{19}{2} n_k \right) opt \quad (12)$$

Because EIS and OIS are independent sets and *S* is a maximum independent set and the size of CDS constructed by ATISA is approximately bounded by the size of the tree constructed in the first stage, so we have:

$$|EIS| \leq |S| \quad (13)$$

$$|OIS| \leq |S| \quad (14)$$

$$|CDS| \leq |EIS| + |OIS| \quad (15)$$

Thus, the performance ratio is:

$$|CDS| \leq (9.67 + 19n_k)opt \quad (16)$$

So, the approximate ratio $< o(n)$.

Message complexity: According to the algorithm in the first stage of ATISA, every black nodes at most broadcasts 2 black messages and every gray nodes at most broadcasts 1 gray messages. The sum of the black nodes and gray nodes in this stage is not bigger than *n*, which is the total number of nodes. Thus, the total number of messages in the first stage is not bigger than 2*n*.

According to the algorithm in the second stage of ATISA, every white nodes at most broadcasts an invite parent message and a gray message and every gray nodes at most broadcasts a black message and an invite message. The sum of the gray nodes and white nodes are bounded by 2*n*, thus the total number of messages in the second stage is not bigger than 4*n*.

According to the algorithm in the third stage of ATISA, every black nodes at most broadcasts 4 messages and every gray nodes at most broadcasts $5+10n_k$ messages. The sum of black nodes and gray nodes is not bigger than *n*. Thus, the total number of messages in the third stage is not bigger than $(5+10n_k)n$.

So, the total number of messages of ATISA is not bigger than $(11+10n_k)n$. The message complexity is $o(n)$.

PERFORMANCE EVALUATION

Here, the performance of ATISA is simulated in the MATLAB7 (Hahn and Valentine, 2010). We did the project in 2010 between January and February and completed this project under the guidance of Professor Wang in control theory laboratory.

We compare the size of CDS constructed by ATISA with four algorithms: Wu's Algorithm (Wu and Li

(1999)(denoted by WUA), Wang’s Algorithm (Wan *et al.*, 2002) (denoted by WAA), the first algorithm of Xiang’s algorithms in (Yingchang *et al.*, 2009) (denoted by XA) and the first algorithm of Thai’s algorithms in (Li *et al.*, 2004) (denoted by TFA).

In the simulation, we simulate the size of the tree constructed in the first stage of ATISA at first. And then we compare the size of CDS constructed by ATISA with that of WUA, WAA, XA and TFA. According to Thai *et al.* (2007), the following network parameters may affect the algorithm performance: the number of nodes, the ratio of the largest transmission range to the smallest transmission range and the node density.

Before simulation, we define some symbols: M representing the side length of the square network area, N representing the number of nodes, r_{min} representing the minimum transmission radius, r_{max} representing the maximum transmission radius. Every point in all the simulated figures is simulated 50 times and the results of the 50 times are averaged.

The size of CT: Here, we define 5 scenarios. M is set to 200 m in the 5 scenarios. (R_{min}, r_{max}) is set to (45, 90), (50, 60), (40, 120), (50, 50), (60, 61), respectively and according to Eq. 8, the critical number of nodes are 57, 41, 57, 30, 22, respectively. N varies from 22 to 66. In order to introduce the size of the tree constructed in the first stage, the number of the nodes which are not dominated by the tree (the number of white nodes when the first stage is ended) is shown in the simulation process.

Figure 2 shows that if the number of nodes in the network is bigger than the critical value in Eq. 8, the average white nodes number is very little when the first stage is ended. In the 5 scenarios, if the number of nodes is bigger than their respective critical values, the average white nodes number is less than 1 when the first stage is ended. So, if the number of nodes in the network is bigger than the critical value, the tree constructed in the first stage can dominate almost all the nodes in the network and the number of the black nodes generated in the second stage can be ignored.

Number of nodes: Here, the network area is 200*200 m. In order to simulate the size of CDS in UDG and DGB, we design two scenarios. At the first scenario, (r_{min}, r_{max}) is set to (60, 60) and the number of nodes varies from 20 to 200, while at the second scenario, (r_{min}, r_{max}) is set to (30, 90) and the number of nodes varies from 20 to 200.

Figure 3 shows that the size of the CDS constructed by ATISA is the smallest.

Figure 4 also shows that the size of the CDS constructed by ATISA is the smallest. And the result in

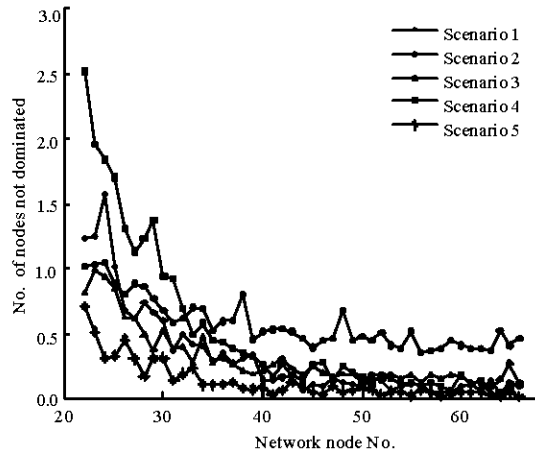


Fig. 2: The size of the CDS at the first scenario

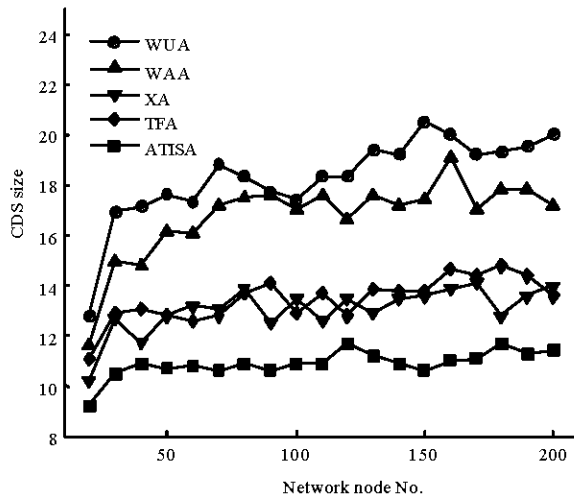


Fig. 3: The size of the CDS at the first scenario

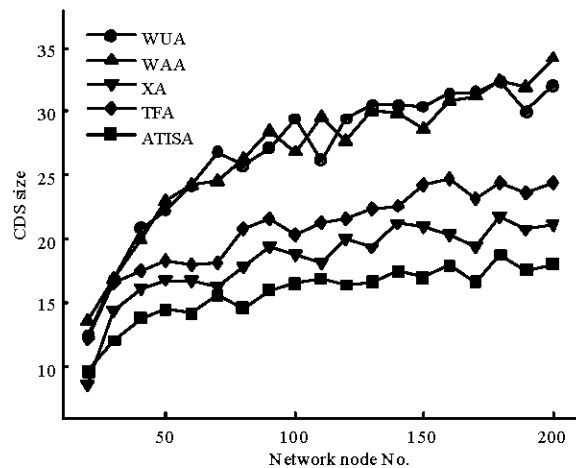


Fig. 4: The size of the CDS at the second scenario

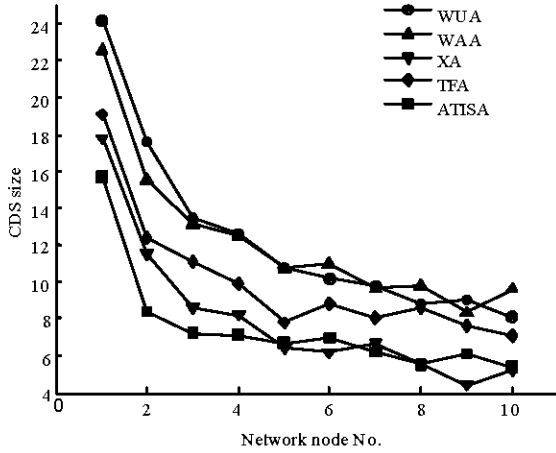


Fig. 5: The relationship between the size of the CDS and k

Fig. 4 presents the size of CDS increases as the number of nodes increases, which is because as the number of nodes increases the more dominators are needed to dominate the nodes in the network. The increment of ATISA is smaller than that of any one algorithm of WUA, WAA, XA and TFA. This is because ATISA adopts the distance as the criterion for selecting the children nodes. The number of dominators is mainly determined by the transmission radii and the side length of the network area, which is simulated in the next sections.

Transmission radii: The network area is set to 400*400 m. The number of nodes is set to 100. (r_{min}, r_{max}) is set to (100, 100), (100, 200), (100, 300), (100, 400), (100, 500), (100, 600), (100, 700), (100, 800), (100, 900), (100, 1000). Thus $k = r_{min}/r_{max}$ varies from 1 to 10. We compare the size of CDS of ATISA with that of WUA, WAA, XA and TSA.

The simulation result show that the number of dominators decreases as the k increases. This is because the k increasing makes the radii to increase and a dominator will dominate more nodes in the network. Thus the size of CDS decreases as the k increases.

Figure 5 also shows that ATISA has the smallest size of CDS at the most of the time compared with the four algorithms.

Node density: We set the number of nodes to 50. The side length of the network area varies from 400 to 1400 m and the step is 50 m. The maximum transmission radius is set to 600 m and the minimum transmission radius is set to 200 m.

Figure 6 shows that the number of dominators increases as the node density decreases. This is because

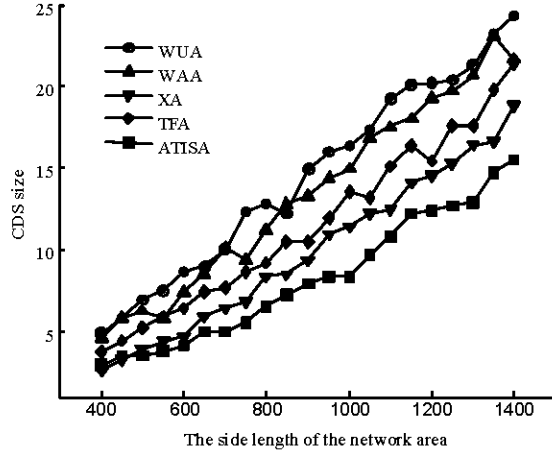


Fig. 6: The relationship between the CDS size and node density

Table 2: Performance comparison

References	MSG complexity	Approximate ratio
Das and Bharghavan (1997)	$O(n^2)$	$O(\log n)$
Wu and Li (1999)	$O(n^2)$	$O(n)$
Ivan <i>et al.</i> (2002)	$O(n)-O(n^2)$	$O(n)$
Alzoubi <i>et al.</i> (2002a,b)	$O(n)$	$O(n \log n)$
Wan <i>et al.</i> (2002)	$<O(n)$	$O(n \log n)$
ATISA (our algorithm)	$<O(n)$	$O(n)$

the node density decreasing makes a dominator to dominate less number of nodes in the network and thus more dominators are needed.

And when the side length of the network area is 1000, the CDS size of WUA, WAA, XA, TSA and ATISA are 15, 11, 14, 10 and 6. No matter the size of the density, the size of ATISA's CDS is the smallest among the five compared algorithms.

DISCUSSION

In this study, we have proposed a simple and efficient algorithm for constructing connected dominating set in wireless sensor networks. A simulation has been conducted to compare our proposed algorithm with Wu and Li (1999), Wan *et al.* (2002), Yingchang *et al.* (2009) and Li *et al.* (2004) in term of the size of connected dominating set generated. In all the cases, our algorithm has the smallest size of CDS.

The performance comparison of some reviews algorithms is shown in Table 2.

Message complexity and approximate ratio are two important factors to measure the quality of the CDS algorithm. So we can conclude from Table 2 that our approach outperforms the list approach.

ACKNOWLEDGMENTS

This study is supported by two grants from the National Natural Science Foundation of China (No. 60773190, No. 60802002).

REFERENCES

- Adjih, C., P. Jacquet and L. Viennot, 2005. Computing connected dominated sets with multipoint relays. *Ad Hoc Sen. Networks*, 1: 27-39.
- Alzoubi, K., P.J. Wan and O. Frieder, 2002a. New distributed algorithm for connected dominating set in wireless ad hoc networks. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, Jan. 7-10, IEEE Computer Society Washington, DC, USA., pp: 297-297.
- Alzoubi, K.M., P.J. Wan and O. Frieder, 2002b. Message-optimal connected dominating sets in mobile ad hoc networks. *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing*, June 9-11, Lausanne, Switzerland, pp: 157-164.
- Dai, F. and J. Wu, 2004. An extended localized algorithms for connected dominating set formation in ad hoc wireless networks. *IEEE Trans. Parallel Distributed Syst.*, 15: 908-920.
- Dai, Z., B. Wang, Q. Tang and Z. Li., 2009. An-energy-aware cluster-based routing protocol for wireless sensor and actor network. *Inform. Technol. J.*, 8: 1044-1048.
- Das, B. and V. Bharghavan, 1997. Routing in ad-hoc networks using minimum connected dominating sets. *Proceedings of the IEEE International Conference on Communications*, June 8-12, Montreal, Que, pp: 376-380.
- Donghyn, K., W. Yiwei, L. Yingshu, Z. Feng and D. Ding-Zhu, 2009. Constructing minimum connected dominating sets with bounded diameters in wireless networks. *IEEE Trans. Parallel Distributed Syst.*, 20: 147-157.
- Feng, W., M.T. Thai and D. Ding-Zhu, 2009. On the construction of 2-connected virtual backbone in wireless networks. *IEEE Trans. Wireless Commun.*, 8: 1230-1237.
- Funke, S., A. Kesselman, U. Meyer and M. Segal, 2006. A simple improved distributed algorithm for minimum CDS in unit disk graphs. *ACM Trans. Sen. Network*, 2: 444-453.
- Guha, S. and S. Khuller, 1998. Approximation algorithms for connected dominating sets. *Algorithmica*, 20: 374-387.
- Hahn, B.H. and D.T. Valentine, 2010. *Essential MATLAB for Engineers and Scientists*. 4th Edn., Elsevier, USA.
- Ivan, S., S. Mahtab and Z. Jovisa, 2002. Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks. *IEEE Trans. Parallel Distributed Syst.*, 13: 14-25.
- Jeremy, B., D. Min, T. Andrew and C. Xiuzhen, 2004. Connected Dominating Set in Sensor Networks and MANETs. In: *Handbook of Combinatorial Optimization*, Du, D.Z. and P. Pardalos (Eds.). Springer, USA., pp: 329-369.
- Jie, W. and D. Fai, 2006. Virtual backbone construction in MANETs using adjustable transmission ranges. *IEEE Trans. Mobile Comput.*, 5: 1188-1200.
- Jie, W., L. Wei and D. Fei, 2006. Extended multipoint relays to determine connected dominating sets in manets. *IEEE Trans. Comput.*, 55: 334-347.
- Jie, W., D. Fei and Y. Shuhui, 2008. Iterative local solutions for connected dominating sets in ad hoc wireless networks. *IEEE Trans. Comput.*, 57: 702-715.
- Li, Y., S. Zhu, M.T. Thai and D.Z. Du, 2004. Localized construction of connected dominating set in wireless networks. *Proceedings of US Nat'l Science Foundation Int'l Workshop Theoretical Aspects of Wireless Ad Hoc, Sensor and Peer-to-Peer Networks*.
- Li, Y., M.T. Thai, W. Feng, C.W. Yi, P.J. Wan and D.Z. Du, 2005. On greedy construction of connected dominating sets in wireless networks: Research articles. *Wireless Commun. Mobile Comput.*, 5: 927-932.
- Raei, H., M. Sarram, F. Adibniya and F. Tashtarian, 2008. Optimal distributed algorithm for minimum connected dominating sets in wireless sensor networks. *Proceedings of the 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, Sept. 29-Oct. 2, Atlanta, GA., pp: 695-700.
- Raei, H., M.A. Fathi, A. Akhlaghi and B. Ahmadipoor, 2009. A new distributed algorithm for virtual backbone in wireless sensor networks with different transmission ranges. *Proceeding of the IEEE/ACS International Conference on Computer Systems and Applications*, May 10-13, USA., pp: 983-988.
- Rai, M., S. Verma and S. Tapaswi, 2009. A heuristic for minimum connected dominating set with local repair for wireless sensor networks. *Proceedings of the 2009 Eighth International Conference on Networks*, March 1-6, IEEE Computer Society Washington, DC, USA., pp: 106-111.
- Rong, X., Q. Deyu, L. Yongjun and Z.W. James, 2009. A novel distributed MCDS approximation algorithm for wireless sensor networks. *Wireless Commun. Mobile Comput.*, 9: 427-437.

- Shuhui, Y., W. Jie and D. Fei, 2008. Efficient directional network backbone construction in mobile Ad Hoc networks. *IEEE Trans. Parallel Distributed Syst.*, 19: 1601-1613.
- Thai, M.T., W. Feng, L. Dan, Z. Shiwei and D. Ding-Zhu, 2007. Connected dominating sets in wireless networks with different transmission ranges. *IEEE Trans. Mobile Comput.*, 6: 721-730.
- Thai, M.T., T. Ravi and D. Ding-Zhu, 2008. On construction of virtual backbone in wireless ad hoc networks with unidirectional links. *IEEE Trans. Mobile Comput.*, 7: 1098-1109.
- Wan, P.J., K.M. Alzoubi and O. Frieder, 2002. Distributed construction of connected dominating set in wireless ad hoc networks. *Proceedings of IEEE 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, June 23-27, New York, pp: 1597-1604.
- Wu, J. and H. Li, 1999. On calculating connected dominating set for efficient routing in ad hoc wireless networks. *Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, August 1999, Seattle, Washington, pp: 7-14.
- Yingchang, X., X. Kai, C. Wei, E.K. Park and R. Shmuel, 2009. Distributed virtual backbone construction in sensor networks with asymmetric links. *Wireless Commun. Mobile Comput.*, 10.1002/wcm.810
- Yu, W., W. Weizhao and L. Xiang-Yang, 2006. Efficient distributed low-cost backbone formation for wireless networks. *IEEE Trans. Parallel Distributed Syst.*, 17: 681-693.
- Zeng, Y., X. Jia and Y. He, 2006. Energy efficient distributed connected dominating sets construction in wireless sensor networks. *Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing*, July 3-6, British Columbia, Canada, pp: 797-802.
- Zhou, K. and L. Meng, 2008. A dynamic clustering-based-routing algorithm for wireless sensor networks. *Inform. Technol. J.*, 7: 694-697.