

<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

A Survey on Connected Dominating Set Construction Algorithm for Wireless Sensor Networks

Zhuo Liu, Bingwen Wang and Lejiang Guo

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

Abstract: In Wireless Sensor Network (WSN), all nodes are energy constrained. Clustering is a kind of energy efficient algorithm, while using a virtual backbone to organize the nodes is a better way. Although, there is no physical backbone infrastructure, a virtual backbone can be formed by constructing a Connected Dominating Set (CDS). The CDS of a graph representing a network has a significant impact on an efficient design of routing algorithms in WSN. A good CDS should first and foremost be small, additionally, it should have other characteristics such as robustness to node failures and low stretch. In this paper, we present a taxonomy and general classification of CDS construction algorithms. We survey different CDS construction algorithms for WSNs.

Key words: WSN, CDS, MIS, connected dominating set, construction algorithm

INTRODUCTION

Recent advances in miniaturization and low-power design have led to the development of small-sized battery-operated sensors that are capable of detecting ambient conditions such as temperature and sound. Sensors are generally equipped with data processing and communication capabilities. The sensing circuitry measures parameters from the environment surrounding the sensor and transforms them into an electric signal. Processing such a signal reveals some properties about objects located and events happening in the vicinity of

the sensor. Each sensor has an onboard radio that can be used to send the collected data to interested parties. Such technological development has encouraged practitioners to envision aggregating the limited capabilities of the individual sensors in a large scale network that can operate unattended (Abbasi and Younis, 2007; Akyildiz *et al.*, 2002; Chong and Kumar, 2003; Deborah *et al.*, 1999; Pottie and Kaiser, 2000; Wang *et al.*, 2003; Rabaey *et al.*, 2000; Kahn *et al.*, 1999).

Numerous civil and military applications can be leveraged by networked sensors (Fig. 1). A network of sensors can be employed to gather meteorological



Fig. 1: WSN's application examples

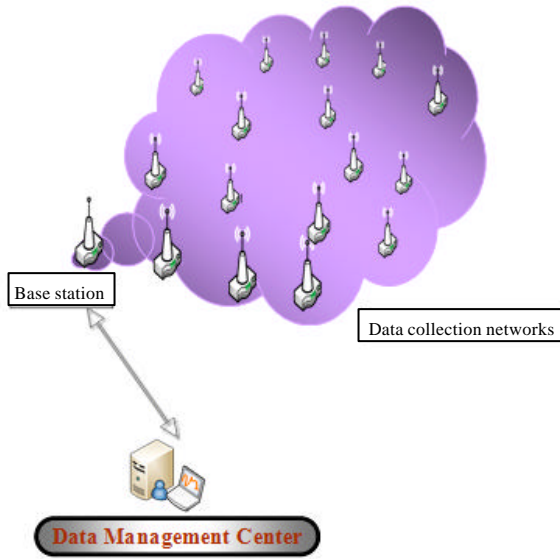


Fig. 2: Architecture of WSNs

variables such as temperature and pressure. These measurements can be then used in preparing forecasts or detecting harsh natural phenomena. In disaster management situations such as earthquakes, sensor networks can be used to selectively map the affected regions directing emergency response units to survivors.

In military situations, sensor networks can be used in surveillance missions and can be used to detect moving targets, chemical gases, or the presence of micro-agents.

Wireless Sensor Networks (WSN) consist a number of wireless nodes. The architecture of WSN is shown in Fig. 2. The WSNs can be divided into three parts: data collection networks, base-station and data management center.

In WSN, there is no fixed or pre-defined infrastructure. Nodes in wireless networks communication via a shared medium, either through a single hop or multi-hops. Such sharing reduces the network performance due to aggravated radio interference. At same time it raises energy consumption since packet retransmission is needed when interference occurs. Moreover, the energy consumption could be increased if we don't limit number of active sensor nodes for the data relays. Backbone can be utilized to address the above problems. Backbone will remove unnecessary transmission links by shutting down some of redundant nodes. Nevertheless backbone will still guarantee network connectivity in order to deliver data efficiently in a wireless sensor network. In virtual backbone based WSNs, some nodes are chosen as dominator node (backbone node) in the backbone construction process.

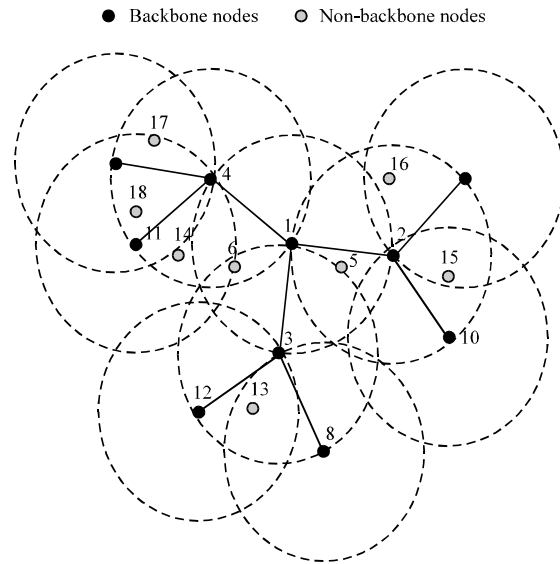


Fig. 3: CDS construction schematic diagram

All nodes then can directly or indirectly communicate with other nodes via these dominator nodes. The dominator form the backbone and the non-selected nodes can perform sleeping schedule or turn-off the radio to save the energy consumption. Although, there is no physical backbone infrastructure, a virtual backbone can be formed by constructing a Connected Dominating Set (CDS). Given an undirected graph $G = (V, E)$, a subset $V' \subseteq V$ is a CDS of G if for each node $u \in V$, u is either in V' or there exist a node $v \in V'$ such that $uv \in E$ and the sub-graph induced by V' , i.e., $G(V')$, is connected. The node in the CDS are called dominator (backbone node), other nodes are called dominatee (non-backbone node) (Fig. 3). With the help of CDS, routing is easier and can adapt quickly to network topology changes. To reduce the traffic during communication and simplify the connectivity management, it is desirable to construct a minimum CDS. Constructing a MCDS is proved a NP-hard problem (Jeremy *et al.*, 2004) and in recent years many algorithms of constructing an approximate MCDS have been proposed.

A good CDS should first and foremost be small, additionally, it should have other characteristics such as robustness to node failures and low stretch (routing in the backbone should not be much longer than the shortest routes over the network topology graph).

In this study, we opt to categorize CDS construction algorithms proposed in the literature for WSNs. We report on the state of the research and summarize their feature and shortcomings. We also compare the different approaches.

CDS CONSTRUCTION ALGORITHM COMPARISON

We can compare the CDS construction algorithms in some main aspects. In particular, we have considered the following metrics:

- **Algorithm duration:** The time needed by each protocol to complete CDS construction
- **The overhead per node (in bytes):** Associated to the protocol operations. (These are physical layer measurements, which account for collisions and for the corresponding automatic packet retransmissions at the MAC layer)
- **Energy consumption per node:** This metric is important for assessing the application of CDS construction to networks whose nodes are energy constrained. If the energy and the overhead cost of maintaining a CDS is high, then CDS reorganization may impose a non-negligible burden to the network. On the other hand, CDS reorganization in terms of gateway rotation is needed to prevent nodes with more demanding roles from depleting their energy, possibly impairing network operations
- **CDS size:** The fraction of the network nodes that are in the CDS (percentage). A smaller CDS is desirable for minimizing the routing overhead. In sensor networks with nodes that can turn off their radio interface (energy saving sleep mode), the CDS size is also a measure of how many nodes need to stay awake for data transport. Usually, the nodes in the CDS stay up or have a higher duty cycle for guaranteeing routes to the sink and hence, the smaller the CDS, the more nodes can be in energy conserving mode
- **Route length:** Over the sub-graph G_b of the network topology graph induced by the CDS construction. (G_b is the graph where there are no links between ordinary nodes). This metric gives a measure of how well a routing protocol can perform over the CDS
- **CDS robustness:** Defined as the number of CDS nodes whose removal (because of failure or energy depletion), causes CDS disconnection or the uncovering of ordinary nodes. This is the metric that provides an indirect measure on how long the network will be operational before requiring a CDS re-computation

TAXONOMY OF CDS CONSTRUCTION ALGORITHM

From different aspects, CDS construction algorithms can be classified into different types. Here, some classifications are proposed below.

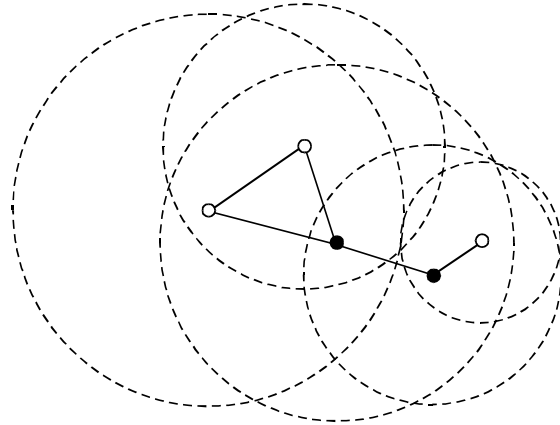


Fig. 4: DGB link

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

The CDS problem has been studied intensively in UDG, in which each node has the same transmission range. The minimum problem in UDG has been shown to be NP-hard.

However in practice, the transmission ranges of all nodes are not necessary equal. In this case, a WSN can be modeled using a directed graph ($G = V, E$) (Fig. 4). The nodes in the V are located in a Euclidean plane and each node $v_i \in V$ has a transmission range $r_i \in [r_{min}, r_{max}]$. A directed edge $(v_i, v_j) \in E$ if and only if $d(v_i, v_j) \leq r_i$ where $d(v_i, v_j)$ denotes the Euclidean distance between v_i and v_j . Such graph are called disk graphs. An edge (v_i, v_j) is bidirectional if both (v_i, v_j) and (v_j, v_i) are in E , $d(v_i, v_j) \leq \min\{r_i, r_j\}$. In other words, the edge (v_i, v_j) is bidirectional if v_i is in the disk D_j centered at v_j with radius r_j and v_j is in the disk D_i centered at v_i with radius r_i . The minimum CDS problem in DGB is NP-hard since the minimum problem in UDG is NP-hard and UDG is a special case of DGB.

MIS based and Non-MIS based: 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 both UDG and DGB, the CDS construction algorithms are mainly based on the Maximum Independent Set (MIS).

Li *et al.* (2005) proposed some distributed connected dominating set construction algorithms. Li *et al.* (2005) proposed a greedy algorithm called S-MIS. The 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 $(5.8 + \ln 4)\text{opt} + 1.2$.

The MIS based algorithms mainly have two kinds of realization. The first one is the two phase based realization, which constructs a MIS at first and then find some optimal nodes to connect the MIS nodes together. The second one is the one phase based realization, which selects the MIS nodes and the optimal intermediate nodes simultaneously. The optimal nodes selection is based on some criterions such as node degree, residual energy of node and node id, etc.

The MIS based algorithms such as 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) 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 (Das and Bharghavan, 1997; Raei *et al.*, 2008; Rai *et al.*, 2009). For example, 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 (Das and Bharghavan, 1997). The algorithm has the performance ratio $2H(\Delta)+1$ (Das and Bharghavan, 1997). 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. The criterion is the node id (Funke *et al.*, 2006; Yingchang *et al.*, 2009). 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).

Many algorithms are MIS based. There are also some non-MIS based algorithms (Dai and Wu, 2004; Rong *et al.*, 2009).

Wu and Li (1999) and Dai and Wu (2004) proposed some distributed algorithms in UDG and DGB. A very simple marking process is proposed by Wu and Li (1999). 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. Wu and Li (1999) proposed two prune algorithms called Rule-1 and 2 to delete some redundant marked nodes. The performance ratio is not given by Wu and Li (1999), but by 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. The marking process is the same as that by 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 (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 in (Rong *et al.*, 2009). The performance ratio of the algorithm is $O(\log n)$ (Rong *et al.*, 2009).

Centralized algorithm and decentralized algorithm: Algorithms that construct a CDS can be divided into two categories: centralized algorithms and decentralized algorithm.

Centralized algorithms: The centralized algorithms in general yield a smaller CDS with a better performance ratio than that of decentralized algorithm. Guha and Khuller (1998) proposed two polynomial time algorithms to construct a CDS in general graph G . These algorithms are greedy and centralized. The first one has the approximation ratio of $2(H(\Delta)+1)$ where Δ is the maximum degree of G and H is a harmonic function. The idea of this algorithm is to build a spanning tree T rooted at the nodes with maximum degree and grow T until all nodes are added to T . The non-leaf nodes in T form a CDS. In particular, all nodes in a given network are white initially. The greedy function that the algorithm uses to add nodes into T is the number of the white neighbors of each node or a pair of nodes. The one with the largest such number is marked black and its neighbors are marked grey. These nodes are then added into T . The algorithm stops when no white

node exists in G . The second algorithm is an improvement of the first one. This algorithm consists of two phases. The first phase is to construct a dominating set and the second phase is to connect the dominating set using a Steiner tree algorithm. With such improved, the second algorithm has the performance factor of $H(\Delta)+2$. These algorithms later were studied and implemented by Das *et al.* (1997), Das and Bharghavan (1997) and Sivakumar *et al.* (1998). Ruan *et al.* (2004) introduced another centralized and greedy algorithm of which the approximation ratio is $(2+\ln \Delta)$.

Decentralized algorithms: The decentralized algorithms can be further divided into two categories: distributed algorithms and localized algorithms. In distributed algorithm, the decision process is decentralized. In the localized algorithm, the decision process is not only distributed but also requires only a constant number of communication rounds. Most of the distributed algorithm find a MIS and connect this set. Note that in an undirected graph, an MIS is also a Dominating Set (DS). Wan *et al.* (2002) and Alzoubi *et al.* (2002a) proposed a distributed algorithm for a CDS problem in UDG. This algorithm consists two phases and has the constant approximation ratio of 8. The algorithm first constructs a spanning tree. Then each node in a tree is examined to find an MIS for the first phase. All nodes in an MIS are colored black. At the second phase, more nodes are added (color blue) to connect the black nodes. Later, Cadei *et al.* (2002) presented another 2-phase distributed algorithm for a CDS in UDG. This algorithm has the same performance ratio as the previous one. However, the improvement over (Wan *et al.*, 2002) is the message complexity. The root does not need to wait for the complete message from the furthest nodes. Recently, Li *et al.* (2005) proposed another distributed algorithm with a better approximation ratio, which is $(4.8+\ln 5)$. This algorithm also has two phases. At the first phase, an MIS is found. At the second phase, a Steiner tree algorithm is used to connect the MIS.

For the localized algorithms, Wu and Li (1999) proposed a simple algorithm that can quickly determine a CDS based on the connectivity information within the 2-hops neighbors. This approach uses a marking process. In particular, each node is marked true if it has two unconnected neighbors. All the marked nodes form a CDS. The researchers also introduced some dominant pruning rules to reduce the size of the CDS. Wan *et al.* (2002) showed that the performance ratio of Wu and Li (1999) is within a factor of $O(n)$ where n is the number of nodes in a network. Taxonomy of CDS construction algorithm is shown in Fig. 5.

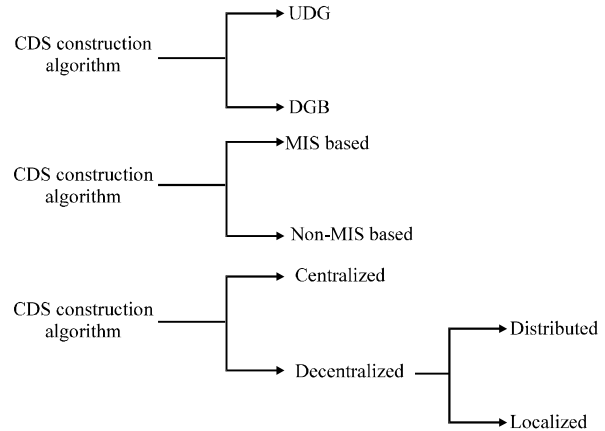


Fig. 5: Taxonomy of CDS construction algorithm

Alzoubi *et al.* (2002b) proposed another localized 2-phase algorithms with the performance ratio of 192. At the first phase, an MIS is constructed using the 2-hops neighbors information. Specifically, once a node knows that it has the smallest ID within its neighbors, it becomes a dominator. At the second phase, the dominators are responsible for identifying a path to connect the MIS. Li *et al.* (2004) proposed another localized algorithm with the performance ratio of 172. This localized algorithm has only 1 phase. A node marks itself as a dominator if it can cover the most white nodes compared to its 2-hops neighbors.

CDS CONSTRUCTION ALGORITHMS FOR WSNs

Generally, WSNs involve a large number of sensors ranging in the hundreds or even thousands. The CDS construction is an effective mean for managing such high population of nodes. Here, we present a literature survey of published algorithms for constructing CDS in WSN. Following the increased interest in wireless sensor networks, many distributed approaches have been proposed because of no requirements for global network topology knowledge.

Distributed approximation algorithms for MCDS in WSN were first developed in a series of papers (Das and Bharghavan, 1997; Sivakumar *et al.*, 1998). These algorithms provided distributed implementations of the greedy heuristics (Guha and Khuller, 1998). The CDS is referred to in these papers as a spine and functions as a virtual backbone. The primary task of the spine is the route computation and maintenance. Nodes in the spine maintain up to date information about their domains (neighbors of the CDS nodes) and are used to exchange

such information between each other to store global information of the network. The advantage of this strategy is the storage of global information in fewer nodes than the number of nodes in the network, which reduces the access overhead for this information and reduces the update overhead. According to this strategy, it is highly desirable to reduce the size of the CDS in order to minimize the access and update overhead. The spine is not necessarily used to route packets in the network, even though it can be used to provide a temporary backup routes for fault tolerance.

We notice that these algorithms lack mechanisms to bridge two consecutive stages. Specifically, individual nodes have no way to tell when the next stage should begin. Furthermore, these approximation algorithms suffer from high performance ratio and high implementation complexities, $O(n)$ time and messages. The construction of the minimum CDS requires 2-hop neighborhood knowledge, which means larger message size, frequent updates, slower convergence speed and more memory.

Alzoubi *et al.* (2002b) proposed two famous distributed heuristics algorithms: (1) ID-based approach for CDS and (2) level-based approach for CDS.

ID-based approach: The construction of the CDS uses the MIS generated by the ID-Based approach for rank assignment. A variation of this algorithm was proposed by Alzoubi *et al.* (2002b), but the level-based approach was used for rank assignment. Thus, the performance ratio was 8 instead of 12. Here, the distributed algorithm for CDS consists of three procedures

- Leader Election
- MIS construction
- Dominating tree construction

The leader election procedure elects a node e.g., with the smallest ID, as the leader. The distributed algorithm in (Cidon and Mokryn, 1998) for leader election can be adopted. This algorithm has a message complexity of $O(n \log n)$. When the leader is found, it broadcasts its identity to all the nodes in the network.

When a node receives the leader's identity, it starts the MIS construction procedure, which is described earlier and using the ID-based approach for rank assignment. Each node will either be colored with black (as a dominator) or gray (as a dominatee). The leader will also select a black node as the root of the dominating tree as follows: If the leader is marked black, it selects itself as the root of the tree, otherwise it selects one of the black neighbors to be the root.

The dominating tree construction procedure is initiated by the root to construct a tree containing all black nodes in addition to some gray nodes. All nodes in this dominating tree form a CDS. The root joins the dominating tree first, then it sends an invitation message to all black nodes within 3-hop distance to join the tree. When a node receives an invitation message and all its neighbors have been marked either black or gray, it responds to the message. When each black node joins the dominating tree, it will also send an invitation to all black nodes within 3-hop distance to join the tree. This invitation will be relayed through the gray nodes within 2-hops distance. Each black node will join the tree when it receives the invitation for the first time together with the gray nodes which relays the invitation to itself. This process should be repeated until all black nodes are in the tree.

Result: All nodes in the dominating tree form a CDS with size at most $12\text{opt}+3$. In addition, the algorithm has $O(n \log n)$ message complexity and $O(n)$ time complexity.

Level-based approach: The construction of the CDS uses the MIS generated by the level-based approach for rank assignment. This algorithm can be divided into three phases: The leader election phase, the level calculation phase and the color marking phase. The distributed algorithm for leader election can be adopted by Cidon and Mokryn (1998). This algorithm has $O(n \log n)$ message complexity and $O(n)$ time complexity. After the construction of the spanning tree is completed and when the root receives level-complete messages from all its children, each node knows the level numbers of all its neighbors. The pair (level, ID) of a node defines the rank of this node. The ranks of all nodes are sorted in the lexicographic order. Thus the leader, which is at level 0, has the lowest rank. In the color marking phase all nodes are initially unmarked (white) and will eventually get marked either black or gray. Two types of messages are used by the nodes during this phase, the dominator message and the dominatee message. The dominator message is sent by a node after it marks itself black and the dominatee message is sent by a node after it marks itself gray. Both messages contain the sender's ID. The algorithm can be described as a color markup process. Initially, the root marks itself black and then broadcasts to its neighbors a dominator message. All other nodes act according to the following principles.

- Whenever a white node receives a dominator message from a white neighbor for the first time, it marks itself gray and broadcasts the dominatee message

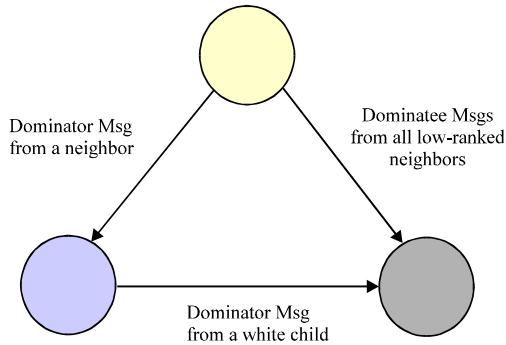


Fig. 6: State transition diagram

- When a white node has received a dominatee message from each of its neighbors of lower rank, it marks itself black and broadcasts the dominator message
- When a gray node receives a dominator message for the first time from one of its children in T, which has never sent a dominatee message, it remarks itself black and broadcasts the dominator message

Figure 6 shows the state transition diagram of this phase. Eventually each node will be either black (a dominator) or gray (dominatee). A reporting process, if necessary, can be performed to notify the root of the completion. When a leaf node has determined its status, it transmits a report complete message to its parent. Each internal node will wait till it receives this report complete message from each of its children and then forward it up the tree to the root. When the root receives the report complete message from all its children then the tree is completed.

Result: All nodes form a CDS with size at most $8\text{opt}+1$. In addition, the message complexity of the algorithm is $O(n \log n)$ and the time complexity is $O(n)$.

Das et al.'s algorithm: The centralized version of the distributed algorithm proposed by Das et al. (1997) consists of three stages. The first stage finds an approximation to minimum dominating set, which is essentially the well-studied set cover problem. Not surprisingly, the heuristic proposed by Das et al. (1997), Bharghavan and Das (1999), Das and Bharghavan (1997) and Sivakumar et al. (1998) is a translation of Chvatal's greedy algorithm (Chvátal, 1979) for set cover and thus guarantees an approximation factor of $H(\Delta)$, where Δ is the maximum degree and H is the harmonic function. Let U denote the dominating set output in this stage. The

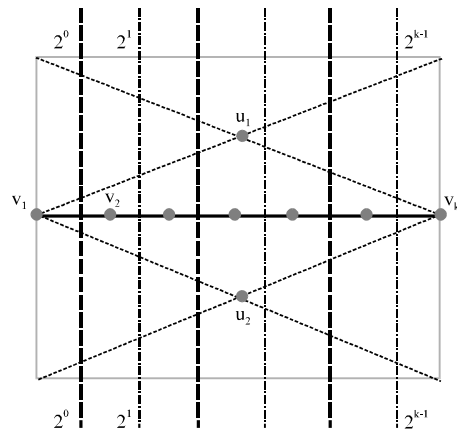


Fig. 7: Instance for which the size of solution output by Das et al.'s algorithm

second stage constructs a spanning forest F . Each tree component in F is a union of stars centered at the nodes in U . The stars are generated by letting each dominatee node pick up an arbitrary neighbor in U . The third stage expands the spanning forest F to a spanning tree T . All internal nodes in T form a CDS. It is easy to show that the CDS generated in this way contains at most $3|U|$ nodes and therefore is a $3H(\Delta)$ approximation of minimum CDS.

Figure 7 shows a family of instances for which the size of the solution computed by the above greedy algorithms is larger than the optimum solution by a logarithm factor. All points lie in a rectangle whose horizontal side has length one and whose vertical side has length $2\sqrt{1-(1/(2(k-1)))^2}$ (the selection of the value $2\sqrt{1-(1/(2(k-1)))^2}$ is based on the fact for any two points whose y-coordinates are differed by $\sqrt{1-(1/(2(k-1)))^2}$, they are connected if and only if their x-coordinates are differed by at most $1/(2(k-1))$). The two nodes v_1 and v_k are the centers of the left and right vertical sides, respectively. The $k-2$ nodes v_2, v_3, \dots, v_{k-1} are evenly distributed within the line segment between and from left to right. The two nodes u_1 and u_2 are the centers of the two sub-rectangles above and below the segment between v_1 and v_k , respectively. The rest points lie in the two horizontal sides. In each horizontal side, $2^0 = 1$ node lies to the left of (and excluding) the perpendicular bisector of v_1v_2 , 2^{k-1} nodes lie to the right of the perpendicular bisector of $v_{k-1}v_k$ and 2^{i-1} nodes lie between the perpendicular bisector of $v_{i-1}v_i$ and the perpendicular bisector of $v_i v_{i+1}$. Thus, the total number of nodes is:

$$n = k + 2 + 2 \sum_{i=1}^k 2^{i-1} = k + 2^{k+1}$$

where, u_1 is adjacent to all nodes lying in the top subrectangle, u_2 is adjacent to all nodes lying in the bottom subrectangle and they are adjacent to each other. Thus, $\{u_1, u_2\}$ forms a minimum CDS.

Result: All nodes form a CDS with size at most $3H(\Delta)$. In addition, the message complexity of the algorithm is $O(n^2)$ and the time complexity is $O(n^2)$.

Wu and Li's algorithm: While the algorithm proposed by Das *et al.* first finds a DS and then grow this DS into a CDS, the algorithm proposed by Wu and Li (1999) takes an opposite approach. The algorithm in Wu and Li (1999) first finds a CDS and then prune certain redundant nodes from the CDS. The initial CDS U consists of all nodes which have at least two non-adjacent neighbors. A node u in U is considered as locally redundant if it has either a neighbor in U with larger ID which dominates all other neighbors of u or two adjacent neighbors with larger IDs which together dominate all other neighbors of u . The algorithm then removes all locally redundant nodes from U . This algorithm applies only to wireless ad hoc networks whose unit-disk graph is not a complete graph. As indicated by Wu and Li (1999), the approximation factor of this algorithm remains unspecified. Obviously, the MCDS of any wireless ad hoc network whose unit-disk graph is not complete graph consists of at least two nodes. On the other hand, any CDS contains at most n nodes. Thus, the approximation factor of the above algorithm is at most $n/2$ where n is the number of nodes. Next, we show that the approximation factor of the above algorithm is exactly $n/2$. This means that the above algorithm does perform extremely poorly over certain instances.

When n is even, we consider the instance illustrated in Fig. 8a. These nodes are evenly distributed over the two horizontal sides of a unit-square. Each node has exactly m neighbors, one in the opposite horizontal side and the rest in the same horizontal side. The two nodes at the left two corners of the unit-square form an minimum CDS. However, the CDS output by the algorithm in (Wu and Li, 1999) consists of all nodes. Indeed, for each node u , the unique neighbor lying in the opposite horizontal side is not adjacent to all other neighbors of u . Thus, the initial CDS U consists of all nodes. In addition, no single neighbor of a node u can dominate all other neighbors of u . Furthermore, if a pair of neighbors of u is adjacent, they must lie in the same horizontal side as u and therefore neither of them is adjacent to the unique neighbor of u lying in the opposite horizontal side. So, no node is locally redundant. Consequently the output CDS still consists of all nodes.

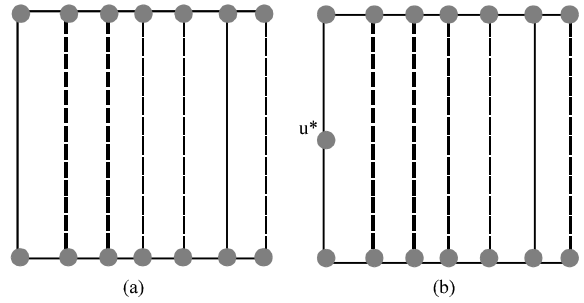


Fig. 8: Instance for which the CDS output by Wu and Li's algorithm

When n is odd, we consider the instance illustrated in Fig. 8b. The node with the largest ID, denoted by u^* , is the center of the left vertical side of a unit-square and all other $n-1$ nodes are evenly distributed over the two horizontal sides of the unit-square. The two nodes at the left two corners of the unit-square form a minimum CDS. On the other hand, the CDS output by the algorithm by Wu and Li (1999) also consists of all nodes. In fact, following the same argument as in the even case, all nodes other than u^* are in the initial CDS U . The node u^* is also in the initial CDS U as well. Since, u^* is not adjacent to the two nodes at the right corners of the unit-square, all nodes other than u^* are not locally redundant. The u^* itself is also not locally redundant as it has the maximum ID. Therefore, the output CDS still consists of all nodes.

Result: All nodes form a CDS with size at most $n/2$. In addition, the message complexity of the algorithm is $\Theta(m)$ and the time complexity is $o(\Delta^3)$.

Stojmenovic et al.'s algorithm: In the context of clustering and broadcasting, Stojmenovic *et al.* (2001) presented three synchronized distributed constructions of CDS. In each of the three constructions, the CDS consists of two types of nodes: the cluster-heads and the border-nodes. The cluster-heads form a Maximal Independent Set (MIS), i.e., a dominating set in which any pair nodes are nonadjacent. A node is a border-node if it is not a cluster-head and there are at least two cluster-heads within its 2-hop neighborhood. The set of cluster-heads is induced by a ranking of nodes which give rise to a total ordering of all nodes. Three rankings are used: the ID only (Gerla *et al.*, 1995; Lin and Gerla, 1997), an ordered pair of degree and ID (Chen and Stojmenovic, 1999) and an order pair of degree and location (Stojmenovic *et al.*, 2001). The selection of the cluster-heads is given by a synchronized distributed algorithm, which can be generalized to the following framework.

Initially all nodes are colored white. In each stage of the synchronized distributed algorithm, all white nodes which have the lowest rank among all white neighbors are colored black; then all white nodes adjacent to the these black nodes are colored gray; finally the ranks of the remaining white nodes are updated. The algorithm ends when all nodes are colored either black or gray. All black nodes then form the cluster-heads.

Regardless of the choice of the ranking, the algorithms by Stojmenovic *et al.* (2001) have a $\Theta(n)$ approximation factor. Such inefficiency stems from the non-selective inclusion of all order-nodes. In fact, if the rank is ID only, Fig. 9 shows a family of instances which would imply the approximation factor to be exactly n , the worst possible. In these instances, the node with the largest ID is located at the center of a unit-disk and all other nodes are evenly distributed in the boundary of the unit-disk. After the cluster-heads are selected, all other nodes become border-nodes. Thus the CDS would consist of all nodes. On the other hand, the node at the center dominates all other nodes. If the rank is an ordered pair of degree and ID or an order pair of degree and location, the instances shown in Fig. 8 imply that their approximation factors are at least $n/2$.

All algorithms by Stojmenovic *et al.* (2001) have $o(n^2)$ message complexity and $\Omega(n)$ time complexity. This can be illustrated in the following instance: All n nodes are evenly distributed in an interval of length $n \pm 1$ with two nodes being the endpoints of the interval. The i th node from the left endpoint of interval has ID i .

Result: All nodes form a CDS with size at most n . In addition, the message complexity of the algorithm is $O(n^2)$ and the time complexity is $\Omega(n)$.

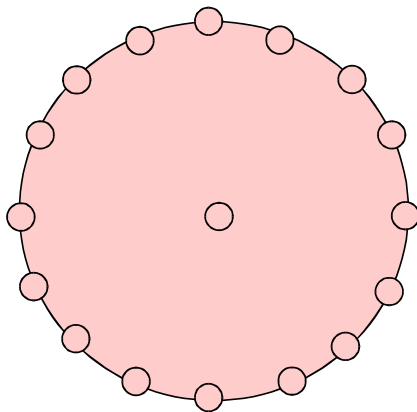


Fig. 9: Instance for which the CDS output by Stojmenovic *et al.*'s algorithm

Zhuo Liu *et al.*'s algorithm: Liu *et al.* (2010a, b) proposed a good performance CDS construction algorithm with small size for WSNs. The algorithm 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.

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 by Yingchang *et al.* (2009) (denoted by XA) and the first algorithm of Thai's algorithms by Thai *et al.* (2007) (denoted by TFA) (Fig. 10).

Result: All nodes form a CDS with size at most $(9.67+19n_k)opt$. In addition, the message complexity of the algorithm is $O(n)$ and the time complexity is $O(n)$.

Table 1: Performance comparison

	Msg complexity	Time complexity	Approximate factor
Bharghavan and Das (1999)	$o(n^2)$	$o(n^2)$	$\Theta(\log n)$
Das and Bharghavan (1997)	$o(n^2)$	$o(n^2)$	$\Theta(\log n)$
Sivakumar <i>et al.</i> (1998)	$o(n^2)$	$o(n^2)$	$\Theta(\log n)$
Wu and Li (1999)	$\Theta(m)$	$o(\Delta^3)$	n
Stojmenovic <i>et al.</i> (2001)	$o(n^2)$	$\Omega(n)$	n
Misra and Mandal (2010)	$o(n\Delta^2)$	$o(n)$	$4.8opt+$ $\ln 5opt+1.2$ $o(n)$
Raei <i>et al.</i> (2009)	$o(n)$	$o(n)$	$o(n)$
Raei <i>et al.</i> (2008)	$o(n)$	$o(1)$	$7.6opt$
Butenko <i>et al.</i> (2004)	$o(n^2 \log^3(n))$	$o(n \log^3(n))$	Constant
Alzoubi <i>et al.</i> (2003)	$o(n \log n)$	$o(n)$	Constant
Gao <i>et al.</i> (2005)	$o(n)$	$o(n)$	Constant
Li <i>et al.</i> (2005)	$o(n\Delta^2)$	$o(\Delta)$	$o(4.8+\ln 5)$
Dai and Wu (2004)	$o(\Delta)$	$o(\Delta^2)$	Constant
Alzoubi <i>et al.</i> (2002b)	$o(n \log n)$	$o(n)$	$4opt$
Cadei <i>et al.</i> (2002)	$o(n \log n)$	$o(n)$	$8opt+1$
Cheng <i>et al.</i> (2004)	$o(n)$	$o(n)$	$147opt+33$
Yingchang <i>et al.</i> (2009)	$o(n\Delta)$	$o(n)$	$4opt$
Guha and Khuller (1998)	$o(n^2)$	$o(n^2)$	$o(n \log n)$

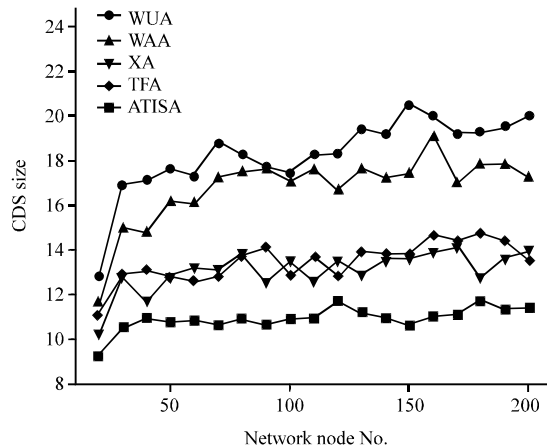


Fig. 10: The size of the CDS

We have surveyed some famous CDS construction algorithms. Performance comparison of more algorithms is shown in Table 1.

CONCLUSIONS

The CDS have proven to be an effective construct within which to solve a variety of problems that arise in wireless networks. Significant attention has been paid to CDS construction algorithms yielding a large number of publications. In this study, we surveyed and classified the CDS construction algorithms and classified.

As WSN moves from the research labs to the real world, additional requirements will certainly arise. Issues that will likely arise include the security and survivability of CDS-based applications. In infrastructure-based wireless networks, the infrastructure plays a key role for security purposes. A perimeter is established around the infrastructure, regulating access to the network. A CDS could perform similar actions through forming a virtual network infrastructure. However, significant issues must first be addressed including the real-time assurance of the trust worthiness of mobile hosts. Similarly, performance of the CDS construction algorithms in the face of security attacks must be evaluated in order to ensure robust, survivable network functionality.

ACKNOWLEDGMENTS

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

REFERENCES

- Abbasi, A.A. and M. Younis, 2007. A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.*, 30: 2826-2841.
- Akyildiz, I.F., W. Su, Y. Sankarasubramanian and E. Cayirci, 2002a. Wireless sensor networks: A survey. *Comput. Networks*, 38: 393-422.
- Alzoubi, K.M., P.J. Wan and O. Frieder, 2002a. Distributed heuristics for connected dominating sets in wireless ad hoc networks. *J. Commun. Networks*, 4: 22-29.
- 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.
- Alzoubi, K.M., P.J. Wan and O. Frieder, 2003. Maximal independent set, weakly connected dominating set and induced spanners for mobile ad hoc networks. *Int. J. Found. Comput. Sci.*, 14: 287-303.
- Bharghavan, V. and B. Das, 1999. Routing in ad hoc networks using minimum connected dominating sets. *Proceedings of the International Conference on Communications, (ICC'99)*, Montreal, Canada, pp: 376-380.
- Butenko, S., X. Cheng, C. Oliveira and P.M. Pardalos, 2004. A New Heuristic for the Minimum Connected Dominating Set Problem on Ad Hoc Wireless Networks. In: *Recent Developments in Cooperative Control and Optimization*, Butenko, S., R. Murphey and P. Pardalos (Eds.). Kluwer Academic Publishers, Boston, pp: 61-73.
- Cadei, M., X. Cheng and D.Z. Du, 2002. Connected domination in ad hoc wireless networks. *Proceedings of the 6th International Conference on Computer Science and Informatics, (ICCSI'02)*, Durham, NC, USA., pp: 1-5.
- Chen, G. and I. Stojmenovic, 1999. Clustering and routing in wireless ad hoc networks. *Technical Report TR-99-05*, Computer Science, SITE, University of Ottawa.
- Cheng, X., M. Ding, D.H. Du and X. Jia, 2004. On the construction of connected dominating set in ad hoc wireless networks. To Appear in *Special Issue on Ad Hoc Networks of Wireless Communications and Mobile Computing*.
- Chong, C.Y. and S.P. Kumar, 2003. Sensor networks: Evolution, opportunities and challenges. *Proc. IEEE*, 91: 1247-1256.
- Chvátal, V., 1979. A greedy heuristic for the set-covering problem. *Math. Operat. Res.*, 4: 233-235.
- Cidon, I. and O. Mokryn, 1998. Propagation and leader election in multihop broadcast environment. *Proceedings of the 12th International Symposium on Distributed Computing*, Sept. 24-26, Springer-Verlag, pp: 104-119.
- 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.
- Das, B. and V. Bharghavan, 1997. Routing in ad hoc networks using minimum connected dominating sets. *Proc. IEEE Int. Conf. Commun.*, 1: 376-380.
- Das, B., R. Sivakumar and V. Bharghavan, 1997. Routing in ad hoc networks using a spine. *Proceedings of the 6th International Conference on Computer Communications and Networks*, Sept. 22-25, Washington, DC, USA., pp: 34-34.

- Deborah, E., G. Ramesh, H. John and S. Kumar, 1999. Next century challenges: Scalable coordination in sensor networks. Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Aug. 15-19, Seattle, Washington, USA., pp: 263-270.
- Funke, S., A. Kesselman, U. Meyer and M. Segal, 2006. A simple improved distributed algorithm for minimum CDS in unit disk graphs. *ACM Trans. Sensor Networks*, 2: 444-453.
- Gao, B., Y. Yang and H. Ma, 2005. A new distributed approximation algorithm for constructing minimum connected dominating set in wireless ad hoc networks. *Int. J. Commun. Syst.*, 18: 743-762.
- Gerla, M. and J.T. Tsai, 1995. Multicluster, mobile, multimedia radio network. *ACM-Baltzer J. Wireless Networks*, 1: 255-265.
- Guha, S. and S. Khuller, 1998. Approximation algorithms for connected dominating sets. *Algorithmica*, 20: 374-387.
- Jeremy, B., D. Min, T. Andrew and C. Xiuzhen, 2004. Connected Dominating Set in Sensor Networks and MANETs. *Handbook of Combinatorial Optimization*, Springer, US..
- Kahn, J.M., R.H. Katz and K.S.J. Pister, 1999. Next century challenges: Mobile networking for smart dust. Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Aug. 15-19, Seattle, WA., pp: 271-278.
- 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 AdHoc, 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. *Wireless Commun. Mobile Comput.*, 5: 927-932.
- Lin, C.R. and M. Gerla, 1997. Adaptive clustering for mobile wireless networks. *IEEE J. Selected Areas Commun.*, 15: 1265-1275.
- Liu, Z., B. Wang and Q. Tang, 2010a. Approximation two independent sets based connected dominating set construction algorithm for wireless sensor networks. *Inform. Technol. J.*, Volume: 9.
- Liu, Z., B. Wang and W. Yang, 2010b. Design and deployment of bridge structural health monitoring system based on wireless sensor network. 6th International conference on wireless communications networking and mobile computing.
- Misra, R. and C. Mandal, 2010. Minimum connected dominating set using a collaborative cover heuristic for ad hoc sensor networks. *IEEE Trans. Parallel Distributed Syst.*, 21: 292-302.
- Pottie, G.J. and W.J. Kaiser, 2000. Embedding the internet: Wireless integrated network sensors. *ACM Commun.*, 43: 51-58.
- Rabaey, J.M., M. Josie Ammer, J.L. da Silva, D. Patel and S. Roundy, 2000. PicoRadio supports ad hoc ultra low power wireless networking. *Computer*, 33: 42-48.
- 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 8th International Conference on Networks, March 1-6, 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.
- Ruan, L., H. Du, X. Jia, W. Wu, Y. Li and K. Ker-I, 2004. A greedy approximation for minimum connected dominating sets. *Theor. Comput. Sci.*, 329: 325-330.
- Sivakumar, R., B. Das and V. Bharghavan, 1998. An improved spine-based infrastructure for routing in ad hoc networks. Proceedings of IEEE Symposium on Computer and Communication, June 1998, Athens, Greece, pp: 356-362.
- Stojmenovic, I., M. Seddigh and J. Zunic, 2001. Dominating sets and neighbor elimination based broadcasting algorithms in wireless networks. *Proc. IEEE Hawaii Int. Conf. Syst. Sci.*, 13: 14-25.
- 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.
- 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.

- Wang, H., J. Elson, L. Girod, D. Estrin and K. Yao, 2003. Target classification and localization in habitat monitoring. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 6-10, Hong Kong, China, pp: 1-4.
- 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, Aug. 20-20, Seattle, Washington, United States, 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.
- Zeng, Y., X. Jia and Y. He, 2006. Energy efficient distributed connected dominating sets construction in wireless sensor networks. Proceedings of the International Conference on Wireless Communications and Mobile Computing, July 3-6, Vancouver, British Columbia, Canada, pp: 797-802.