The First Constant Factor Approximation for Minimum Partial Connected Dominating Set Problem in Growth-bounded Graphs

**Xianliang Liu · Wei Wang · Donghyun Kim · Zishen Yang ·
Alade O. Tokuta · Yaolin Jiang**

Abstract The idea of virtual backbone has emerged to improve the efficiency of flooding based routing algorithms for wireless networks. The effectiveness of virtual backbone can be improved as its size decreases. The minimum connected dominating set (CDS) problem was used to compute minimum size virtual backbone. However, as this formulation requires the virtual backbone nodes to connect all other nodes, even the size of minimum virtual backbone can be large. This observation leads to consider the minimum partial CDS problem, whose goal is to compute a CDS serving only more than a certain portion of the nodes in a given network. So far, the performance ratio of the best approximation algorithm for the problem is $O(\ln \Delta)$, where Δ is the maximum degree of the input general graph. In this paper, we first assume the input graph is a growth-bounded graph and introduce the first constant factor approximation for the problem. Later, we show that our algorithm is an approximation for the problem in unit disk graph with a much smaller performance ratio, which is of practical interest since unit disk graph is popular to abstract homogeneous wireless networks.

Xianliang Liu, Wei Wang ✉, Zishen Yang, Yaolin Jiang
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.
E-mail: liuxianliang1@126.com, wang_weiw@163.com, 1164926717@qq.com, yljiang@mail.xjtu.edu.cn

Donghyun Kim, Alade O. Tokuta
Department of Mathematics and Physics, North Carolina Central University, Durham, NC 27707, USA.
E-mail: {donghyun.kim, atokuta}@nccu.edu

Finally, we conduct simulations to evaluate the average performance of our algorithm.

1 Introduction

Recently, several kinds of self-organizing wireless networks such as ad-hoc networks and wireless sensor networks have received lots of attentions [1,2]. One well-known merit of those networks is that they can be instantly deployed without any existing infrastructure. Therefore, the networks can be employed for various applications, for which traditional wireless network technologies can be hardly adopted. This kind of wireless networks usually consist of a number of wireless nodes which use a limited power supply such as a battery as their primary energy source. Consequently, energy efficiency is a significant issue of those wireless networks.

Due to the lack of an efficient supporting infrastructure, many existing routing protocols in wireless networks have to exploit a flooding-like strategy to discover a new routing path or maintaining a routing table. Under the circumstance, the battery of each node tends to drain quickly due to significant wireless signal interference and collision while perforating the routing-related tasks. This problem is known as the broadcast storm problem and is a significant issue of wireless networks [3]. Recently, a decent idea to address this issue, in which only a small subset of nodes are in charge of routing related tasks, has been introduced [4]. In the literature, the subset of nodes used for this purpose is widely referred as *virtual backbone* since the subset is required to form a connected subgraph and any pair
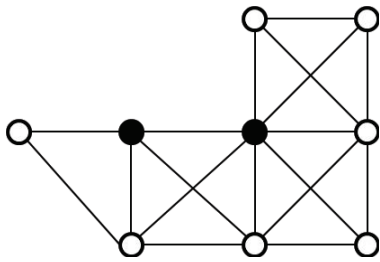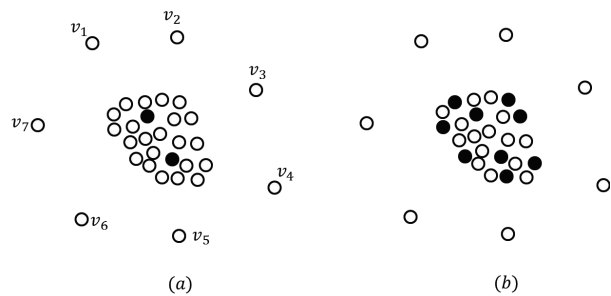
Fig. 1 In this figure, the set of black nodes can serve as the virtual backbone of the whole network to route messages.

Fig. 2 In Fig.(a), by ignoring a few nodes (e.g. $v_1, \cdots, v_7$), the size of the CDS (the set of black nodes) is significantly reduced (compared to Fig.(b)). Note that a wireless network may have a number of sub network area with such a shape, the benefit of partial connected dominating set can potentially scale up as the size of the network grows.

of nodes can communicate with each other through a routing path which consists of the nodes in the subset (see Fig. 1). Clearly, this strategy can suppress the overhead caused by the flooding-based routing algorithms such as managing routing tables, forwarding messages, etc. This means that the amount of signal collision and interference in the networks can be greatly lowered and the whole networks become more energy-efficient [5].

It is quite straightforward to understand the efficiency of a virtual backbone can be enhanced by decreasing its size. Given the unmeasurable benefits of the virtual backbone, the problem of generating a smaller size virtual backbone is a problem of great importance. Formally speaking, given a graph $G = (V, E)$, a subset D of nodes in V is called a *dominating set (DS)* if for each node $u \in V$, if either $u \in D$ or there exists another node $v \in D$ such that $(u, v) \in E$. The subset is also called as a CDS if the subgraph of G induced by D, notated by $G[D]$, is a connected graph. In [6], Guha and Kuller modeled the problem of computing a minimum size virtual backbone as the minimum *connected dominating set (CDS)* problem. Unfortunately, this problem is NP-hard [7], which means that it is unlikely to find a polynomial time exact algorithm for the problem unless $P = NP$. As a result, a significant amount efforts have been made to find a polynomial time heuristic algorithm for the problem with a theoretical worst case performance guarantee, which is also known as an *approximation algorithm* [8–16].

One significant drawback of using a CDS as a virtual backbone of a wireless network is that the size of CDS can be significantly magnified by a few outliers which are far from the majority of the nodes in the network. In this case, due to the requirement that a CDS has to connect "all" nodes in the network, even the size of a minimum CDS can be very huge. Based on this observation, Liu and Liang [17] introduced the minimum *partial connected dominating set (PCDS)* problem, whose goal is to find a minimum cardinality subset of nodes whose induced graph is still required to be connected, but it only needs to connect a certain portion of the nodes

in the network (see Fig. 2). It is easy to understand that the minimum PCDS problem is NP-hard. Unfortunately, they only introduced a heuristic algorithm for the problem, which does not have any theoretical worst case performance guarantee.

Ever since the minimum PCDS problem has been introduced in LCN 2005 [17], an approximation algorithm for this problem has not been invented for years. In SODA 2014, Khuller et al. [18] has finally introduced the first approximation algorithm for the problem in *general graphs*. In detail, they introduced an approximation algorithm for the minimum PCDS problem whose performance ratio is $O(\ln \Delta)$, more precisely $4 \ln \Delta + 2 + o(1)$, where Δ is the maximum degree of a given general graph.

Inspired by Khuller et al's work [18], in this paper, we study the minimum PCDS problem in a subclass of general graph called *growth-bounded graphs (GBG)* and a subclass of GBG, namely *unit disk graph (UDG)*. By relying on the special properties of each subgraph class, we introduce the *first constant factor approximation algorithm* for each of the subgraph classes. We claim our result is very important since

(a) most wireless network topology can be more accurately abstracted using the subgraphs of our interest rather than general graphs, and

(b) even though Khuller et al.'s algorithm for general graphs still works for the subgraph classes of our interest, GBG and UDG, each of our algorithms, which is specifically designed for GBG and UDG, respectively, has a constant performance ratio, which is independent from Δ or another other variables relying on the input of a problem instance.

The rest of this paper is organized as follows. Section 2 discuss some related works. Section 3 introduces important notations, definitions, and preliminaries. Our

main results which include our algorithm for the minimum PCDS problem and its approximation ratio analysis given growth-bounded graph and given unit disk graph are in Section 4. The simulation results and corresponding analysis are in Section 5. Finally, we conclude this paper in Section 6.

2 Related Work

Recently, the concept of virtual backbone was emerged as a promising tool to deal with the broadcasting storm problem in wireless networks. The most of the efforts related this topic have been dedicated to design an approximation algorithm to produce a CDS with smaller cardinality under various circumstances [8–16,19–22, 24–30].

In [6], Guha and Khuller proposed a $(\ln \Delta + 3)$-approximation algorithm and Ruan et al. [11] introduced a $(\ln \Delta + 2)$-approximation algorithm for the minimum CDS problem in general graphs, where Δ is the maximum node degree of an input graph. In [10], the authors introduced the first polynomial-time constant-factor approximation algorithm for the minimum CDS problem in UDG. In this work, a CDS of a given graph is computed throughout the following two phases, which becomes a very popular approach. In the first phase, a subset of nodes are selected to form a DS of a given graph. Then, in the following phase, some additional connecting nodes are merged to the DS nodes so that the union of them can form a CDS. Given a graph, an *independent set (IS)* is a subset of nodes in the graph such that no two nodes in the subset is adjacent in the graph. An IS is called an *maximal IS (MIS)* if no node u in the graph which is not in the IS can be merged with the IS to form a larger IS. Clearly, an MIS is also a DS of a graph since once an MIS I is computed, all nodes in a connected graph G is either in I or adjacent to a node in I, otherwise we can add such a node to I and make the I to be a new larger MIS, which is against the definition of MIS. Since computing a minimum DS is NP-hard, a simple coloring algorithm to compute an MIS becomes a popular heuristic algorithm to find a suboptimal DS. In [10], the authors has proved that an MIS computed by such a coloring strategy is an approximation of computing a minimum DS with performance ratio of 4. This bound has been improved for many times and currently is smaller than 3.5 [15,16].

In [9], Cheng et al. showed that there exists a full polynomial-time approximation scheme, i.e. for any ε, there exist a polynomial-time $(1 + \varepsilon)$-approximation algorithm. Several distributed algorithms are also proposed, such as in [12,31]. Thai et al. [21] studied the minimum CDS problem in disk graph, and proposed a

constant factor approximation algorithm. Li et al. [20] studied the minimum power *strongly connected dominating set (SCDS)* problem in directed graph, and the authors gave an $O(\ln n)$-approximation algorithm, where n is the number of the nodes in the graph. The main idea of their algorithm is selecting a random root node and building a broadcast tree of an input graph first, and getting another broadcast tree by reversing the edges later. Then, the union of the two broadcast trees is a SCDS of the graph. The weighted dominating set (or connected dominating set) problem has been studied, either. In [8], Guha and Khuller gave an $(1.35 + \varepsilon) \ln n$-approximation algorithm in node-weighted graphs by exploiting existing minimum node-weight Steiner tree algorithms, where n is the number of the nodes in an input graph. In addition to those mentioned so far, many efforts are made to study the minimum CDS problem under various consideration such as routing cost [24, 26–28], 3-dimensional topology [25], fault-tolerance [30], etc.

The concept of PCDS is originally introduced by Lie and Liang [17] many years ago (2005), but its first approximation algorithm in general graph is introduced very recently (2014) by Khuller et al. [18], and its performance ratio is $O(\ln \Delta)$. This paper aims to study the problem in two subclass of general graphs, namely GBG and UDG, and introduce a constant factor approximation algorithm for the problem in each subgraph class. Our research is motivated by the fact that after Guha and Khuller introduced a $O(\ln \Delta)$-approximation algorithm for the minimum CDS problem in general graph, Wan et al. has introduce the first constant factor approximation algorithm for the problem in UDG [10], which is later used as a seed result for a number of papers to design an efficient algorithm for computing virtual backbone in homogenous wireless networks.

3 Notations, Problem Definition, and Preliminaries

In this paper, $G = (V, E) = (V(G), E(G))$ is an abstraction of a wireless network. Depending on the context, G can be either growth-bounded graph (GBG) (see Definition 3) which is a subclass of general graph, or Unit Disk Graph (UDG) (see Definition 4). For any pair of nodes u, v, $Euc(u, v)$ is the Euclidean distance between them. For any node subset $V' \subseteq V$, $G[V']$ means a subgraph of G induced by V'. Similarly, for any edge subset $E' \subseteq E$, $G[E']$ will imply a subgraph of G induced by E'. Also, denote by

$$\Gamma_r(v) = \{u \in V \mid hopdist(u, v) \leq r\}.$$

Now, we introduce some important definitions.

Definition 1 (Independent Set (IS)) Given $G = (V, E)$, a subset $I \subset V$ is an independent set of G if for each pair $u, v \in I$, $(u, v) \notin E$.

Definition 2 (Maximal IS (MIS)) An independent set I is referred as a maximal independent set if for any vertex $v \in V \setminus I$, $I \cup \{v\}$ is not an independent set.

Let $S \subset V$. Then we denote by $I(S)$ a maximum independent set of the induced graph $G[S]$.

Definition 3 (Growth-bounded Graph (GBG)) For any given function $f : \mathbb{N}_+ \to \mathbb{R}$, we call a graph $G = (V, E)$ is $f(\cdot)$ GBG, if it satisfies that for any vertex $v \in G$, $|I(\Gamma_r(v))| \leq f(r), \forall r \in \mathbb{N}_+$ holds, where the definition of $\Gamma_r(v)$ and $I(\cdot)$ are shown above. Particularly, the graph is called a polynomial GBG, when the function $f(\cdot)$ is a polynomial.

Definition 4 (Unit Disk Graph (UDG)) A graph $G = (V, E)$ is a unit disk graph if it can be embedded in the Euclidean plane such that for each pair of nodes $u, v \in V$, there exits an edge between them, i.e. $(u, v) \in E$, if any only if $Euc(u, v) \leq 1$.

A UDG is also a GBG, which follows from the fact that a disk with radius r contains at most $(2r + 1)^2$ independent nodes. Actually, let I be an independent set contained in a disk with radius $r + 1/2$. We draw a disk with radius $1/2$ centered at each node $v \in I$, then all small disks are pairwise disjoint and contained in the larger disk with radius $r + 1/2$. Thus the maximum number of independent nodes is at most

$$\frac{\pi(r + 1/2)^2}{\pi(1/2)^2} = (2r + 1)^2.$$

Definition 5 (Dominating Set (DS)) Given $G = (V, E)$, a subset $D \subset V$ is a dominating set of G if for each $u \in V$ either $u \in D$ or there exists another node $v \in D$ such that $(u, v) \in E$.

Definition 6 (Connected DS (CDS)) A dominating set D, whose induced graph $G[D]$ is connected, is called as a connected dominating set.

Now, we provide the formal definition of the partial connected dominating set. Suppose for any $v \in G$ and $r \in \mathbb{N}_+$, $\Gamma_r(v)$ is the vertex set in which the distance between vertices and v are no more than r.

Definition 7 (Partial Connected Dominating Set (PCDS)) Given $G = (V, E)$ and a positive integer n', where $n' \in \{n' \in \mathbb{N}_+ | n' \leq |V(G)|\}$, a subset $C \subset V$ is a partial connected dominating set, or in short $PCDS(G, n')$ if

(a) $G[C]$ is connected, and
(b) the number of vertices dominated by C (includes C itself) is at least n'.

Definition 8 (Minimum PCDS Problem) Given G, n', the minimum PCDS problem is to find a PCDS (G, n') with minimum cardinality.

Definition 9 (Quota Steiner Tree) Given a graph G with weights on both vertices and edges and an integer n', a quota Steiner tree is a tree T in G such that

$$\sum_{v \in V(T)} w(v) \geq n'.$$

Definition 10 (Minimum Quota Steiner Tree Problem) Given a graph G with weights on both vertices and edges and an integer n', the minimum quota Steiner tree problem is to find a Quota Steiner Tree of G such that

$$\sum_{e \in E(T)} w(e)$$

becomes minimum.

Johnson et al. [34] studied the QST problem and showed that an α-approximation algorithm for the k-MST problem (that is, given an edge weighted graph, find a minimum cost tree with at most k vertices) can be adapted to obtain an α-approximation algorithm for the Quota Steiner Tree problem. Using this result along with the 2-approximation for k-MST by Garg [35], gives us the following theorem.

Theorem 1 ([34, 35]) *There exists a 2-approximation algorithm for the minimum Quota Steiner tree problem.*

During the rest of this paper, a Quota Steiner tree for an input pair $\langle G, n' \rangle$ will be denoted by $\text{QST}(G, n')$.

4 Main Results

In this section, we propose a polynomial time algorithm for the minimum PCDS problem, namely the partial connected dominating set algorithm (PCDSA). Then, we prove the proposed algorithm has a constant factor given the input graph is GBG. Finally, we improve the constant approximation factor based on the assumption that the input graph is a UDG.

4.1 General Idea

There are many literatures which show a "partial" problem is much more difficult than its "complete" version. For example, it is well known that the minimum

spanning tree (MST) problem can be solved efficiently by some greedy approach such as Kruskal's algorithm. However, its one "partial version", the k-MST problem, is NP-hard and a 2-approximation algorithm is known [35]. We found that this is true to the case of the PCDS problem. There are various approximation algorithms available for the minimum CDS problem in the past decades, whereas the first approximation algorithm for the PCDS problem managed to appear very recently [18].

Now we give some general idea about PCDSA. Basically, PCDSA follows from the ideas from [18]. First we construct a maximal independent set (which is also a dominating set) $D = \{v_1, v_2, \cdots, v_k\}$ by using, say, the greedy approach (note that unlike in [18], greedy approach is not essential here; instead, we can use any existing method such as coloring strategy for MIS). During this process, each vertex $v_i \in D$ newly covers w_i number of uncovered vertices in G (i.e., the contribution of v_i is w_i in the covering process, and clearly $\sum_{i=1}^{k} w_i = |V|$). Next, applying the Quota Steiner Tree algorithm with vertex weight w_i (the weight of the nodes not in D are set to zero) and edge weight one gives the approximated solution PD. Note that by setting the vertex weight in this way, we are guaranteed to find a CDS $D' = V(T)$ (T is the resulting QST) which dominates a required n' number of nodes in G; while by setting the edge weight to one, we actually try to minimize $|D'| = |E(T)| + 1$.

The key point why PCDSA is a constant approximation lies in the fact that above D is an independent set. If we restrict D to the 2-hop neighborhood of an optimal solution OPT of PCDS to obtain a D', then D' can dominate the required n' number of nodes, and $|D'|$ can be upper bounded by a constant factor, $f(2)$, of $|OPT|$, for any GBG G. Next by adding a few additional nodes (the number of which can be upper bounded by a constant factor of $|OPT|$), we can modify D' into a connected CDS D'' which also dominates n' number of nodes. Let T'' be a spanning tree of $G[D'']$, then T'' is a feasible solution to QST problem. By comparing T'' with the optimal solution OPT^* of QST problem, we get $|E(OPT^*)| \leq |E(T'')| = |D''| - 1$ since OPT^* is optimal. On the other hand, the QST gives a feasible solution PD of PCDS, which can be upper bounded by $2|E(OPT^*)| + 1$, by using the fact QST is a 2-approximation. Finally, combing the above together gives that $|PD|$ is upper bounded by a constant factor of $|OPT|$, which shows that PCDSA is a constant approximation algorithm.

Algorithm 1 PCDSA $(G = (V, E), n')$

1: Set $D \leftarrow \emptyset$, and $Q \leftarrow \emptyset$.
2: Set $G' = (V' \leftarrow V, E' \leftarrow E, w_N \leftarrow 0, w_E \leftarrow 1)$, where w_N and w_E are the node and edge functions.
3: **while** $V \neq D \bigcup Q$ **do**
4:     Find a node $v \in V \setminus (Q \bigcup D)$ such that $|N_v(G) \bigcap (V \setminus (D \bigcup Q))|$ is maximum, where $N_v(G)$ is the subset of nodes in G neighboring to v.
5:     Set $D \leftarrow D \bigcup \{v\}$ and $Q \leftarrow Q \bigcup \{N_v(G)\}$.
6:     Set $w_N(v) \leftarrow w_N(v) + |N_v(G) \bigcap (V \setminus (D \bigcup Q))|$.
7: **end while**
8: Apply an existing 2-approximation algorithm ([34, 35]) for the quota Steiner tree problem over a problem instance $\langle G, n' \rangle$ and obtain an output quota Steiner tree QST (G', n').
9: Output PCDS (G, n'), which is the set of the non-leaf nodes in QST (G', n').

4.2 Algorithm Description

Algorithm 1 is the brief description of the proposed algorithm for the minimum PCDS problem. This algorithm largely consists of two phases: the first phase (Lines 1-7) is mainly about the greedy algorithm to compute a DS of G. The second phase (Line 8) is to find the connector nodes so that the DS nodes in the first phase can be connected in a way that the resulting output satisfies the constraints.

In detail, Line 1 prepares two empty sets D and Q, where D will be used to record the dominating nodes and Q will be used to record the dominated nodes. Therefore, they have to be initially empty. Line 2 is to initialize the input graph G' for the quato Steiner tree problem. Initially, the weight of each node is set to 0 and the weight of each edge is set to 1.

Lines 3-7 are describing a round-based greedy strategy to compute a DS of the input graph G, and at the same time G' is modified. Specifically, in Line 4, the algorithm finds a node v which is not in $Q \bigcup D$ such that v has the most number of neighbors (say contribution) in $V \setminus (Q \bigcup D)$. In Line 5, v is added to the DS, D, and its neighbors, which are not in $(Q \bigcup D)$, are added to Q. At last, in Line 6, the weight of v in G' is set to the contribution of v in this greedy process.

After a DS D is computed, in Line 8, D is used along with n', as an input pair of an existing 2-approximation algorithm for the quota Steiner tree problem. Then, at the end, we obtain a spanning tree QST (G', n'). Finally, the algorithm outputs the non-leaf nodes of QST (G', n') as the result of the whole algorithm in Line 9.

4.3 Theoretical Analysis

First, we prove the proposed algorithm is correct and its running time is polynomial.

Theorem 2 *The output of Algorithm 1 is correct, i.e. it is a $\mathrm{PCDS}(G, n')$ of the minimum PCDS problem instance $\langle G, n' \rangle$.*

Proof Obviously, the graph induced by the output of the algorithm is connected because it is the set of non-leaf nodes in an output of quota Steiner tree algorithm. Considering the way we construct the weighted graph G from it without weight, obviously, $V(\mathrm{QST}(G, n'))$ can dominate at least n' vertices since for any vertex the increased contribution for dominating will never less than its weight. Therefore, the output of the algorithm is a connected vertex set of the graph G which dominating at least n' vertices. As a result, this theorem is true.

Theorem 3 *The running time of Algorithm 1 is polynomial.*

Proof Algorithm 1 mainly consists of two stages. The first stage is using greedy strategy to compute a maximal independent set; clearly this can be done in polynomial time. The second stage is applying 2-approximation algorithm in [34,35] for Quota Steiner tree, which can also be done in polynomial time. Therfore, Algorithm 1 is a polynomial time algorithm.

Next, we prove the algorithm is a constant factor approximation algorithm for the minimum PCDS problem in GBGs.

Theorem 4 *Given any connected $f(\cdot)$ GBG G and a positive integer $n' \leq |V(G)|$, one always obtain a solution for $\mathrm{PCDS}(G, n')$ with constant performance ratio via Algorithm 1.*

Proof According to Theorem 1, we have

$$\Big| \sum_{e \in E(\mathrm{QST}(G, n'))} w(e) \Big| \leq 2 \Big| \sum_{e \in E(\mathrm{OPT}^*)} w(e) \Big|,$$

where $w(\cdot)$ denotes the weight of the edge, and OPT^* denotes the optimal tree for the Quota Steiner Tree problem. Notice that all the edges in graph G have weight 1, therefore,

$$|E(\mathrm{QST}(G, n'))| \leq 2|E(\mathrm{OPT}^*)|.$$

Furthermore, we have

$$|PD| = |V(\mathrm{QST}(G, n'))|$$
$$= |E(\mathrm{QST}(G, n'))| + 1 \leq 2|E(\mathrm{OPT}^*)| + 1,$$

where PD is the output of Algorithm 1.

Denote the optimal solution for $\mathrm{PCDS}(G, n')$ by OPT, and its i-neighborhood by

$$\mathrm{OPT}_i = \mathrm{N}^i(\mathrm{OPT}) \backslash \mathrm{OPT}_{i-1}, i = 1, 2, \ldots,$$

where $\mathrm{OPT}_0 = \mathrm{OPT}$.

Let $D' = D \cap (\mathrm{OPT} \cup \mathrm{OPT}_1 \cup \mathrm{OPT}_2)$, where D is vertex set introduced in both Algorithm 1. We claim that D' can dominate all the vertices in $\mathrm{OPT} \cup \mathrm{OPT}_1$. In fact, D is a dominating set itself and all the vertices in $\mathrm{OPT} \cup \mathrm{OPT}_1$ cannot be dominated by the vertices outside. Therefore, the number of vertices dominated by D' must not be less than $|\mathrm{OPT} \cup \mathrm{OPT}_1|$, which is exactly the number of vertices dominated by the optimal solution OPT. In other words, D' is a feasible partial dominating set but not connected (it is an independent set).

On the other hand,

$$|D'| = |D \cap (\mathrm{OPT} \cup \mathrm{OPT}_1 \cup \mathrm{OPT}_2)|$$
$$= \Big| D \cap \bigcup_{v \in \mathrm{OPT}} \Gamma_2(v) \Big|$$
$$= \Big| \bigcup_{v \in \mathrm{OPT}} \big(D \cap \Gamma_2(v) \big) \Big|$$
$$\leq \sum_{v \in \mathrm{OPT}} |D \cap \Gamma_2(v)|,$$

since $\mathrm{OPT} \cup \mathrm{OPT}_1 \cup \mathrm{OPT}_2$ and $\bigcup_{v \in \mathrm{OPT}} \Gamma_2(v)$ represent 2-neighborhood of OPT and the union of 2-neighborhood of all the vertices in OPT respectively. In addition,

$$\sum_{v \in \mathrm{OPT}} |D \cap \Gamma_2(v)| \leq \sum_{v \in \mathrm{OPT}} |I(\Gamma_2(v))|$$
$$\leq \sum_{v \in \mathrm{OPT}} f(2) = f(2)|\mathrm{OPT}|$$

holds for a $f(\cdot)$ GBG G, since D is an independent set.

To make set D' connected, we need only to add a few vertices to D'. One possible way is combining the set with all vertices in OPT, and adding at most a vertex in OPT_1 to connect the vertex in $D' \cap \mathrm{OPT}_2$. Denote the connected vertex set obtained by above method by D'', then we have
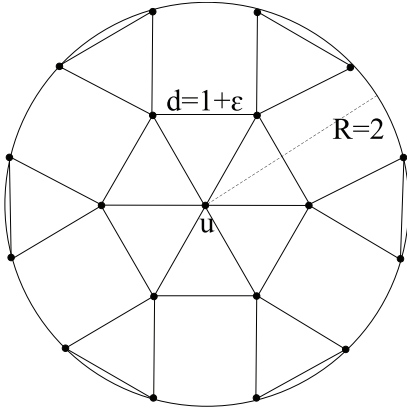
$$|D''| \leq |\mathrm{OPT}| + 2|D'|.$$

Let T'' be a spanning tree of the induce graph $G[D'']$. We claim that T'' is a feasible solution for $\mathrm{QST}(G, n')$ problem, since T'' is a tree and its subset D' has already satisfied the constraint. Thus,

$$|E(T'')| \geq |E(\mathrm{OPT}^*)|.$$

Furthermore,

$$|D''| = |V(T'')| = |E(T'')| + 1 \geq |E(\mathrm{OPT}^*)| + 1.$$

Fig. 3 There are 19 independent vertices in disk with radius 2

Fig. 4 Two disks associated with two adjacent nodes in OPT have a large overlap.

It follows from above equations that

$$|PD| \leq 2|E(\text{OPT}^*)| + 1$$
$$\leq 2|D''| - 1$$
$$\leq 2(|\text{OPT}| + 2|D'|) - 1$$
$$= 2|\text{OPT}| + 4|D'| - 1$$
$$\leq 2|\text{OPT}| + 4 \sum_{v \in \text{OPT}} |D \cap \Gamma_2(v)| - 1$$
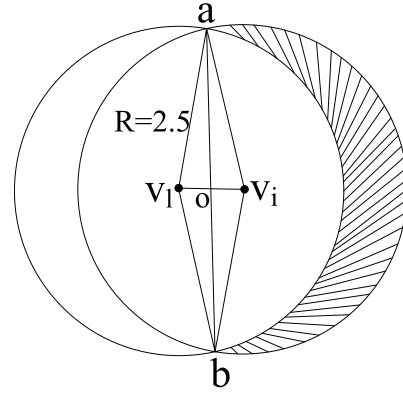$$\leq 2|\text{OPT}| + 4f(2)|\text{OPT}| - 1$$
$$= (4f(2) + 2)|\text{OPT}| - 1,$$

which shows Algorithm 1 is a $(4f(2)+2)$-approximation algorithm for a polynomial GBG. This completes the proof.

Now, we assume the input graph is a UDG and improve the performance ratio of Algorithm 1. As a direct consequence of Theorem 4, we have

Theorem 5 *Suppose G is a UDG. Then Algorithm 1 is a 102-approximation for the minimum PCDS problem.*

Proof According to the above Theorem 4, for any $f(r)$ polynomial GBG, we have $4f(2) + 2$-approximation for the minimum PCDS problem. In addition, we know that for any vertex u of UDG G, the cardinality of maximum independent set of $\Gamma_2(u)$ is at most $f(2) = (2 \times 2 + 1)^2$. So, we have at most 25 independent nodes in a disk with radius 2. Hence, Algorithm 1 is a $102-$approximation for the minimum PCDS problem.

One way to improve the approximation ratio in Theorem 5 is to find a better way to compute $f(2)$, i.e., the number of independent nodes in a disk with radius 2. The problem is closely related to circles packing in circle. It is easy to show that there are 19 independent nodes in a disk with radius 2, see Fig. 3. According to a conjecture [33] of circle packing, the maximum number of independent nodes in a disk with radius 2

is exactly 19. So, if the conjecture is true, we have a $78-$approximation for the minimum PCDS problem in UDG.

Next, we will give a better analysis of the performance ratio of Algorithm 1 by employing a closely relationship between the cardinality of an independent set and that of an optimal solution of a connected dominating set. From the proof of Theorem 4, we know that the performance ratio of our algorithm is mainly dependent on the estimated numbers of dominating subset $D \cap (\text{OPT} \cup \text{OPT}_1 \cup \text{OPT}_2)$. Since the subset $D \cap (\text{OPT} \cup \text{OPT}_1 \cup \text{OPT}_2)$ is also an independent subset, we will focus on the estimation of how many independent nodes can be contained in a 2-hop neighborhood area of the optimal solution OPT. We have the following lemma.

Lemma 1 $|D \cap (OPT \cup OPT_1 \cup OPT_2)| \leq 6.25|\text{OPT}| + 19$.

Proof We prove the lemma by an area argument (see also [36]). First, for each node in $v \in$ OPT, draw a disk with radius 2.5 centered at v. Denote by the union of these disks by R. Then for each node u in an independent set, draw a disk with radius $1/2$ centered at u. Clearly, all the small disks are pairwise disjoint and contained in R. So an upper bound for the number of independent nodes is given by the ratio of the area of R and the area of a small disk. Next, we give an estimation of the area of R. A key observation is that two neighboring disks with radius 2.5 have a large overlap; see Fig. 4.

Let us compute the dashed area (see Fig. 4). Note the centers of two larger disks associated with vertices $v_l, v_i \in OPT$ have distance at most 1. Assume $\angle av_lv_i = \frac{\theta}{2}$ and $\angle av_lb = \theta$. Then, we know that $\cos\theta = -\frac{23}{25}$ (since $av_l = R = 2.5, v_lv_i = r = 1$). Thus, it is easy to

know that the area of the dashed region is at most

$$(\pi R^2 - \frac{\theta R^2}{2}) - (\frac{\theta R^2}{2} - 4 \cdot \frac{1}{2} \cdot R\cos\frac{\theta}{2} \cdot R\sin\frac{\theta}{2}) = (\pi - \theta + \sin\theta)R^2.$$

That is,

$$\left(\pi - \arccos(-\frac{23}{25}) + \sin\arccos(-\frac{23}{25})\right) \cdot (\frac{5}{2})^2$$
$$\leq (\pi - 0.87\pi + 0.12\pi) \cdot (\frac{5}{2})^2.$$

In order to estimate the area of R. Let OPT $= \{v_1, v_2, \cdots, v_s\}$. Since OPT is connected, we can construct a spanning tree T on the nodes in OPT, which is rooted, say at v_1. Let us examine the area of R by iteratively adding disks with radius 2.5 to existing disks one by one, starting from the disks associated with v_1. Note when adding a new disk with radius 2.5, the area increases by at most

$$(\pi - 0.87\pi + 0.12\pi) \cdot (\frac{5}{2})^2.$$

Thus, the total area of R is at most

$$\pi(\frac{5}{2})^2 + (\pi - 0.87\pi + 0.12\pi) \cdot (\frac{5}{2})^2(|\text{OPT}| - 1).$$

Thus, the number of independent nodes in R is at most

$$\frac{\pi(\frac{5}{2})^2 + (\pi - 0.87\pi + 0.12\pi) \cdot (\frac{5}{2})^2(|\text{OPT}| - 1)}{\pi(\frac{1}{2})^2}$$
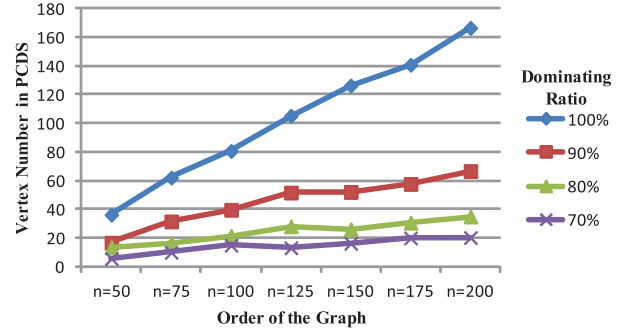$$\leq 6.25(|\text{OPT}| - 1) + 25$$
$$\leq 6.25|\text{OPT}| + 19.$$

The proof is complete.

As a consequences of Lemma 1 and Theorem 4, we have the following theorem.
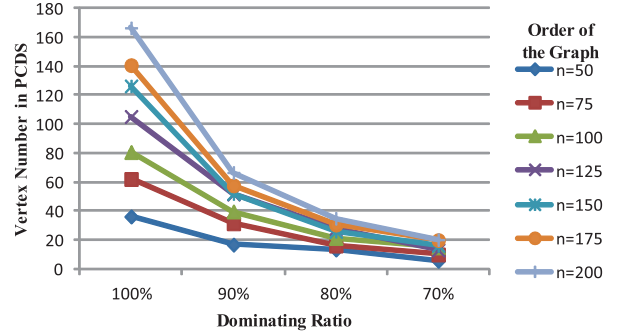
Theorem 6 *Suppose G is a UDG. Then Algorithm 1 is a $27-approximation$ for the minimum PCDS problem asymptotically.*

Proof According to the proof of Theorem 4 and Lemma 1, we have

$$|PD| \leq 2|D''| - 1$$
$$\leq 2(|\text{OPT}| + 2|D'|)$$
$$\leq 2|\text{OPT}| + 4(6.25|\text{OPT}| + 19)$$
$$\leq 27|\text{OPT}| + 76.$$

(a) Performance of Algorithm 1 while the size of input graph n is growing.

(b) Performance of Algorithm 1 while the dominating ratio, $(n'/n) \times 100\%$, is decreasing.

Fig. 5 Average performance of Algorithm 1.

5 Simulation Result and Analysis

In this section, we conduct simulations to observe the averaged behaviors of the proposed algorithm against parameter changes and analyze the result. Each result is an averaged result of 100 trials. In each trial, under the same parameter setting, we generate a random connected unit ball graph (a growth-bound graph). In detail, we first deploy a number of n nodes in 3-dimensional Euclidean space, and check if its induced unit ball graph is connected. Otherwise, we discard it and generate a new one to make sure of its connectivity. Once a connected graph G is obtained, we apply our algorithm over the minimum partial connected dominating set problem instance $\langle G, n' \rangle$ for a given n' and obtain the output.

Fig. 5 shows the result of our simulations. Fig. 5(a) and Fig. 5(b) show the average performance of the proposed algorithm while n is increasing and n' is decreasing. We set the required dominating ratio, which is

$$(n'/n) \times 100\%,$$

from 100% to 70%, which means that n' is decreasing proportionally. As we can see from the figures, the size of the (ordinary) connected dominating set (with

dominating ratio to be 100%) is much greater than its partial connected dominating set counterpart. That is, even with the dominating ratio of 90%, we can reduce the size of the connected dominating set significantly. This decreasing trend is not so obvious when the dominating ratio is reduced form 90% to 70%. This implies that our proposed partial connected dominating set algorithm is effectively reducing the size of the dominating set. Our result also shows that the effective of the algorithm on the size of the output partial connected dominating set is related to n'.

6 Concluding Remarks and Future Work

Over years, the minimum connected dominating set problem and its variations have attracted lots of attentions. Since the problem is NP-hard, many efforts are made to introduce approximation algorithms for them, for many of which, either a constant factor approximation algorithm or a polynomial time approximation scheme (PTAS) is discovered. In this paper, we have investigated a general case of the classical minimum connected dominating set problem called the minimum partial connected dominating set problem, which is originally introduced by Liu and Liang [17] in 2005. This new problem is known to be very challenge. Recently, the first approximation algorithm for this problem in general graph has been introduced by Khuller et al. at SODA 2014 [18]. Given the significance of the connected dominating set in the quality virtual backbone construction in wireless networks, it is important to consider the problem in more specific graphs which are used to abstract wireless networks. Motivated by this observation, we propose a new polynomial time algorithm for the minimum partial connected dominating set problem, and prove the algorithm has a constant factor approximation ratio in growth-bounded graph and in unit disk graph. Most importantly, compared to the performance ratio of the algorithm proposed by Khuller et al., which is $O(\ln \Delta)$, we prove the performance ratio of our algorithm is 27 (asymptotically), in unit disk graph. As our future work, we plan to investigate the PTAS of this problem.

References

1. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Communications Magazine*, vol. 40, pp. 102-114, August 2002.
2. C. R. Dow, P. J. Lin, S. C. Chen, J. H. Lin, and S. F. Hwang, "A Study of Recent Research Trends and Experimental Guidelines in Mobile Ad Hoc Networks," *the 19th International Conference on Advanced Information Network-*
ing and Applications (AINA'05), pp. 72-77, Taiwan, March 2005.
3. S. Ni, Y. Tseng, Y. Chen, and J. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," *the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM'99)*, pp. 152-162, Washington, USA, August 1999.
4. A. Ephremides, J. Wieselthier, and D. Baker, "A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling," *Proceeding of the IEEE*, vol. 75, issue 1, pp. 56-73, 1987.
5. P. Sinha, R. Sivakumar, and V. Bharghavan, "Enhancing Ad Hoc Routing with Dynamic Virtual Infrastructures," *the 20th annual joint conference of the IEEE computer and communications societies*, vol. 3, pp. 1763-1772, 2001.
6. S. Guha and S. Khuller, "Approximation Algorithms for Connected Dominating Sets," *Algorithmica*, vol. 20, pp. 374-387, April 1998.
7. M.R. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-completeness," Freeman, San Francisco, 1978.
8. S. Guha and S. Khuller, "Improved Methods for Approximating Node Weighted Steiner Trees and Connected Dominating Sets," *Information and Computation*, vol. 150, issue 1, pp. 57-74, 1999.
9. X. Cheng, X. Huang, D. Li, W. Wu, and D.-Z. Du, "A Polynomial-Time Approximation Scheme for the Minimum-Connected Dominating Set in Ad Hoc Wireless Networks," *Networks*, vol. 42, issue 4, pp. 202-208, 2003.
10. P.-J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks," *the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'02)*, New York, USA, June 2002.
11. L. Ruan, H. Du, X. Jia, W. Wu, Y. Li, and K.-I. Ko, "A Greedy Approximation for Minimum Connected Dominating Sets," *Theoretical Computer Science*, vol. 329, issue 1-3, pp. 325-330, 2004.
12. P.-J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks," *ACM Journal on Mobile Networks and Applications (MONET)*, vol. 9, issue 2, pp. 141-149, 2004.
13. W. Wu, H. Du, X. Jia, Y. Li, S.C.-H. Huang, "Minimum Connected Dominating Sets and Maximal Independent Sets in Unit Disk Graphs," *Theoretical Computer Science*, vol. 352, 2006.
14. S. Funke, A. Kesselman, U. Meyer, and M. Segal, "A Simple Improved Distributed Algorithm for Minimum CDS in Unit Disk Graphs," *ACM Transactions on Sensor Networks (TOSN)*, vol. 2, issue 3, pp. 444-453, 2006.
15. P.-J. Wan, L. Wang, and F. Yao, "Two-phased Approximation Algorithms for Minimum CDS in Wireless Ad Hoc Networks," *Proceedings of the 2008 The 28th International Conference on Distributed Computing Systems (ICDCS)*, pp. 337–344, 2008.
16. M. Li, P.-J. Wan, and F. Yao, "Tighter Approximation Bounds for Minimum CDS in Wireless Ad Hoc Networks," *The 20th International Symposium on Algorithms and Computation (ISAAC 2009)*, December 16-18, Hawaii, USA
17. Y. Liu and W. Liang, "Approximate Coverage in Wireless Sensor Networks," *Proceedings of 30th IEEE Conference on Local Computer Networks (LCN)*, 2005.
18. S. Khuller, M. Purohit, and K.K. Sarpatwar, "Analyzing the Optimal Neighborhood: Algorithms for Budgeted and Partial Connected Dominating Set Problems," *Proceedings of 2014 ACM-SIAM Symposium on Discrete Algorithms (SODA14)*, Jan. 5-7, 2014, Portland, OR, USA.

19. Y. Li, S. Zhu, My T. Thai, and D-Zhu Du, "Localized Construction of Connected Dominating Set in Wireless Networks," *NSF International Workshop on Theoretical Aspects of Wireless Ad Hoc, Sensor and Peer-to-Peer Networks (TAWN04)*, Chicago, June 2004.

20. D. Li, H. Du, P.-J. Wan, X. Gao, Z. Zhang, and W. Wu, "Minimum Power Strongly Connected Dominating Sets in Wireless Networks," in Proc. of *the 2009 International Conference on Wireless Networks (ICWN'09)*, pp. 447-451, 2008.

21. M. T. Thai, F. Wang, D. Liu, S. Zhu and D-Z. Du, "Connected Dominating Sets in Wireless Networks with Different Transmission Ranges," *IEEE Transactions on Mobile Computing*, Vol. 6, No. 7, July 2007.

22. M.T. Thai, R. Tiwari, and D.-Z. Du, "On Construction of Virtual Backbone in Wireless Ad Hoc Networks with Unidirectional Links," *IEEE Transactions on Mobile Computing (TMC)*, vol. 7, no. 9, pp. 1098-1109, 2008.

23. D. Kim, X. Li, F. Zou, Z. Zhang, and W. Wu, "Recyclable Connected Dominating Set for Large Scale Dynamic Wireless Networks," Proceedings of *the 3rd International Conference on Wireless Algorithms, Systems and Applications (WASA 2008)*, Dallas, TX, USA, October 26-28, 2008.

24. D. Kim, Y. Wu, Y. Li, F. Zou, and D.-Z. Du, "Constructing Minimum Connected Dominating Sets with Bounded Diameters in Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 20, no. 2, pp. 147 157, February, 2009.

25. D. Kim, Z. Zhang, X. Li, W. Wang, W. Wu, and D.-Z. Du, "A Better Approximation Algorithm For Computing Connected Dominating Sets in Unit Ball Graphs," *IEEE Transactions on Mobile Computing (TMC)*, published online.

26. L. Ding, X. Gao, W. Wu, W. Lee, X. Zhu, and D.-Z. Du, "Distributed Construction of Connected Dominating Sets with Minimum Routing Cost in Wireless Networks," *2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, pp. 448-457, 2010.

27. L. Ding, W. Wu, J. Willson, H. Du, W. Lee, and D.-Z. Du, "Efficient Algorithms for Topology Control Problem with Routing Cost Constraints in Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 22, no. 10, pp. 1601-1609, 2011.

28. H. Du, W. Wu, Q. Ye, D. Li, W. Lee, and X. Xu, "CDS-based Virtual Backbone Construction with Guaranteed Routing Cost in Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 24, issue 4, pp. 652-661, 2013.

29. H. Du, W. Wu, S. Shan, D. Kim, and W. Lee, "Constructing Weakly Connected Dominating Set for Secure Clustering in Distrubuted Sensor Network," *Journal Of Combinatorial Optimization (JOCO)*, vol. 23, pp. 301-307, February 2012.

30. W. Wang, D. Kim, M.K. An, W. Gao, X. Li, Z. Zhang, and W. Wu, "On Construction of Quality Fault-Tolerant Virtual Backbone in Wireless Networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, issue 5, pp. 1499-1510, October 2013.

31. J. Wu and H. Li, "On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks," in Proc. of *the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIAL-M'99)*, pp. 7-14, 1999.

32. B. N. Clark, C. J. Colbourn, and D. S. Johnson, "Unit Disk Graphs", *Discrete Mathematics*, vol. 86, pp. 165-177, Dec. 1990.

33. R.L. Graham, B.D. Lubachevsky, K.J. Nurmela, P. R.J. Ostergard, "Dense packings of congruent circles in a circle", *Discrete Math* , vol. 181, pp. 139-154, 1998.

34. D.S. Johnson, M. Minkoff, and S. Phillips, "The prize collecting Steiner tree problem: theory and practice", In *SODA*, pages 760-69, 2000.

35. N. Garg, "Saving an epsilon: a 2-approximation for the k-MST problem in graphs", In *STOC*, 2005.

36. S. Funke, A. Kesselman, U. Meyer. "A simple improved distributed algorithm for minimum CDS in unit disk graphs", *ACM Trans. Sensor Netw.* ,vol. 2, pp. 444-453, 2006.