

Advanced Graph Theory
Prof. Rajiv Misra
Department of Computer Science and Engineering
Indian Institute of Technology, Patna

Lecture – 23
Dominating Set and Distributed Algorithm

Dominating Set and Distributed Algorithms.

(Refer Slide Time: 00:17)

Preface

Recap of Previous Lecture:

In previous lecture, we have discussed the Hamiltonian Graph Traveling Salesman Problem and NP-Completeness.

Content of this Lecture:

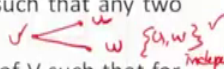
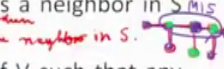
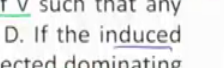
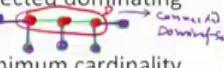
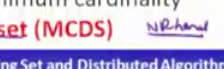
In this lecture, we will discuss the Dominating Set, Connected Dominating Set and Distributed Algorithm on Graphs.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Recap of previous lecture we have discussed Hamiltonian Graph Travelling Salesman Problem and NP-Completeness. Content of this lecture we will discuss dominating sets connecting dominating sets and distributed algorithms on graph.

(Refer Slide Time: 00:36)

Preliminaries

- Given **graph** $G=(V,E)$, two vertices are independent if they are not neighbors. For any vertex v , the set of independent neighbors of v is a subset of v 's neighbors such that any two vertices in this subset are independent. 
- **An independent set (IS)** S of G is a subset of V such that for all $u, v \in S$, $(u, v) \notin E$. 
- **MIS:** S is maximal if any vertex not in S has a neighbor in S (denoted by MIS). 
- **A dominating set (DS)** D of G is a subset of V such that any node not in D has at least one neighbor in D . If the induced subgraph of D is connected, then D is a **connected dominating set (CDS)**. $D \subseteq V(G)$ D is Dominating Set nodes in $V-D$'s neighbor in D . 
- Among all CDSs of graph G , the one with minimum cardinality is called a **minimum connected dominating set (MCDS)**. 

Advanced Graph Theory
Connected Dominating Set and Distributed Algorithm

Preliminaries given a graph G 2 vertices are independent if they are not neighbors. For any vertex b the set of independent neighbors of v is a subset of v 's neighbors such that any 2 vertices in this subset are independent.

For example, let us say this is the vertex b it has the neighbors let us say u and w . So, u and w they are independent neighbors, why because there is no edge connection between them and they are independent. So, hence for any vertex v the set of independent neighbors is a subset of the v 's neighbor. Such that any 2 vertices in these subsets are independent u and w they belong to the neighbor of v .

Now, let us see another definition independent set S of G is a subset of set of vertices such that for all vertices u and v in s they do not have any edge between them; that means, such that let us say u and v they are in S and also u and v they do not have an edge in G .

Then it is then that set of vertices is called independent set. The maximal independent set if the vertex not in S has a neighbor in S for example, so; that means, the maximal independent set means S is an independent set is maximal independent set when the nodes, which are in V minus S must have neighbor in S . For example, let us take a simple graph if this is the graph then let us say take the independent set which are shown by the a green color vertices.

So, together these 2 vertices are maximal independent set why because the nodes which are not green or which are not in the maximal independent set are having the neighbor in this particular set. For example, this node is a neighbor, this node which is not in s is also a neighbor, this node is the neighbor of both the elements in the independent set. So, this also should be this also should be in the independent set. So, this also node is a neighbor and this particular node is also a neighbor and this node is also a neighbor. So, these 3 set of vertices is in MIS. Now this particular node, if it is present, then they are not independent let us not take this particular node let us take some other node in the independent set.

Let us take this particular node in the independence. So, all the green color nodes are in the independent set, because they do not have any edge between them and all other nodes all other nodes are basically the neighbor of the other set. Hence this is an independent set example. So, we have shown the independent set example.

Now the dominating set D of G is a subset of V , such that any node not in D is at least one neighbor in D and let us consider the independent set same example, if we if we construct, then these particular 3 different set of nodes is the dominating set why because all other nodes which are not in D that is this is a neighbor of at least one node in D similarly this node also is a neighbor of at least one node in D this is also a neighbor of at least one node in D so, also this node and this node. Therefore, D is the dominating set and this D is a subset of the vertices of G such that such that the nodes in V minus D is the neighbor in D .

Now, if the induced graph of D if the induced sub graph of D is connected then D is called the connected dominating set. For example, here the induced graph of D is also connected therefore; it is a connected dominating set.

Now among all CTSS connecting dominating set of a graph G the one with the minimal cardinality is called minimum connected dominating set. Now find out a minimum connected dominating set in a given graph requires to explore the large amount of solution space hence the problem of finding the maximum or a minimum connected dominating set is basically NP hard. If let us say an algorithm is developed to find out a minimum connected dominant acing that becomes an NP hard problem.

(Refer Slide Time: 08:08)

Introduction

- The idea of **virtual backbone routing for ad hoc wireless networks** is to operate routing protocols over a virtual backbone. *no infrastructure*
- One purpose of virtual backbone routing is to alleviate the serious **broadcast storm problem** suffered by many existing on-demand routing protocols for route detection. Thus constructing a virtual backbone is very important. *no infrastructure*
- In this lecture we study, the virtual backbone is approximated by a **minimum connected dominating set (MCDS)** in a unit-disk graph. This is a **NP-hard problem**.
- A distributed approximation algorithm with **performance ratio at most 8** will be covered.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, let us take an application of dominating set and that application we can find in the wireless communication and also the wireless ad hoc networks, let us see how this particular dominating set in a graphs are used in constructing the algorithms for wireless networks.

So, the idea of virtual backbone routing for ad hoc wireless network is to operate routing protocol over a virtual backbone, because there is no infrastructure in an ad hoc network. So, the routing is to be done if it is in a systematic manner without too much of communication, then it requires to create a virtual backbone and operate the routing protocol over it that is the method.

Now, here one purpose of virtual backbone based routing is to alleviate the problem, which is called a broadcast storm problem that will arise if the destination is not known. So, the source who wants to send the information has to depend upon the flooding and the flooding will create the serious problem, which is called a broadcast storm problem. Thus the construction of a virtual backbone is very important in the wireless networks to operate the routing protocol without suffering from a broadcast storm problem.

So, in this lecture we will focus on the virtual backbone creation, which is nothing, but is approximated by a minimum connected dominating set and the kind of graph, which we will assume for ad hoc wireless networks is called a unit disk graph that we will explain. Now we will take up an algorithm to solve the minimum connected dominating set

which will be used for construction of a virtual backbone. Now this particular problem is through the algorithm becomes NP hard problems. So, we will look upon the different algorithms, which are the approximation algorithms we will look into that particular algorithm to see how the algorithms are designed? And these algorithms are distributed algorithms why because the nature of the ad hoc network is large number of nodes which are deployed and it cannot be done in a centralized manner. So, the algorithm has to be a distributed algorithm.

So, the distributed approximation algorithm we will discuss and the performance ratio is 8 and that performance ratio means, that it takes the guarantees of the algorithm how much it can deviate from the optimum size of the minimum connected dominating set.

(Refer Slide Time: 11:32)

Sensor Network as Adhoc network

- **Ad hoc wireless and Sensor network** has applications in emergency search-and-rescue operations, decision making in the battlefield, data acquisition operations in inhospitable terrain, etc.
- It is featured by **dynamic topology (infrastructureless), multihop communication, limited resources** (bandwidth, CPU, battery, etc) and **limited security**.
- These characteristics put special challenges in **routing protocol design** inspired by the **physical backbone** in a wired network, many researchers proposed the concept of virtual backbone for unicast, multicast/broadcast in ad hoc wireless networks .

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

So, let us see that sensor network is also a network sensor network is also a wireless network and we can view it as an ad hoc network. So, ad hoc wireless and sensor network has an application in emergency search and rescue operation decision making in a battlefield, data acquisition operation in inhospitable terrains etcetera.

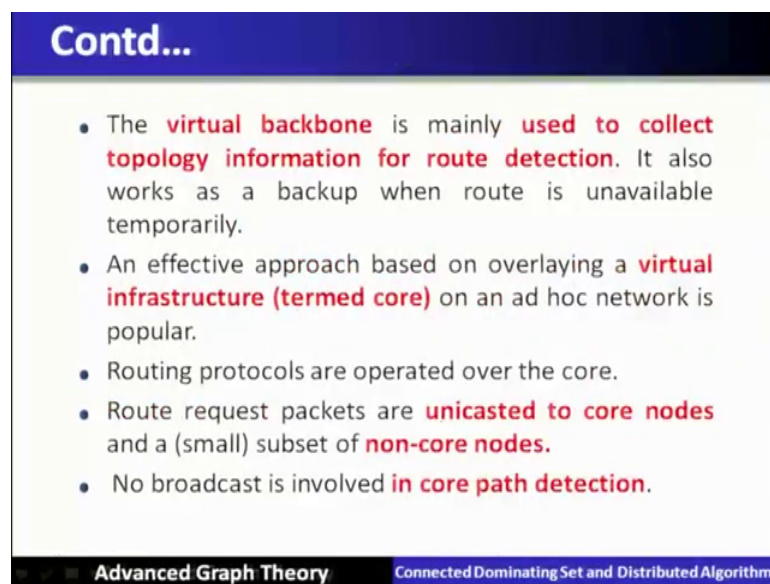
It is featured by a dynamic topology here in such applications there is no possibility of establishing the infrastructure and operate the wireless communications, hence it is a dynamic topology where in the infrastructures are dynamically created on the fly on demand. Now it is also featured by the multi hoc communication and it also having a

limited resource that is the bandwidth CPU and battery are all limited in nature and also such networks are having a limited security.

So, these characteristics puts my special challenges in routing protocol design and hence they are inspired to create a virtual backbone why because there is no possibility to create a physical backbone, but only the inspiration of a physical like backbone as if we have seen in the wired network is ve very much required to be created.

So, many researchers they have proposed the concept of a virtual backbone and that particular virtual backbone will be used for different kind of communication. For example, unicast communication, multicast, broadcast kind of communication in adhoc wireless networks.

(Refer Slide Time: 13:20)



Contd...

- The **virtual backbone** is mainly **used to collect topology information for route detection**. It also works as a backup when route is unavailable temporarily.
- An effective approach based on overlaying a **virtual infrastructure (termed core)** on an ad hoc network is popular.
- Routing protocols are operated over the core.
- Route request packets are **unicasted to core nodes** and a (small) subset of **non-core nodes**.
- No broadcast is involved **in core path detection**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

The virtual backbone is mainly used to collect the topology information for the route detection. So, it also works up as the backup when the route is unavailable temporarily. An effective approach based on overlaying a virtual infrastructure, which is also called as a core on an adhoc network becomes a popular.

So, routing protocols are operated over the core; route request packets are unique hosted to the code nodes and a small subset of noncore nodes are also used up in this broadcasting or a routing of the information. Here no broadcast is involved and only the nodes in the core they are involved in the communication.

(Refer Slide Time: 14:14)

Classification of Routing Protocols

- Existing routing protocols can be classified into two categories: (i) *Proactive* and (ii) *Reactive*.

(i) Proactive routing protocols ask each host (or many hosts) to maintain global topology information, thus a route can be provided immediately when requested.

- But large amount of control messages are required to keep each host updated for the newest topology changes.

(ii) Reactive routing protocols have the feature on-demand. Each host computes route for a specific destination only when necessary.

- Topology changes which do not influence active routes do not trigger any route maintenance function, thus communication overhead is lower compared to proactive routing protocol.

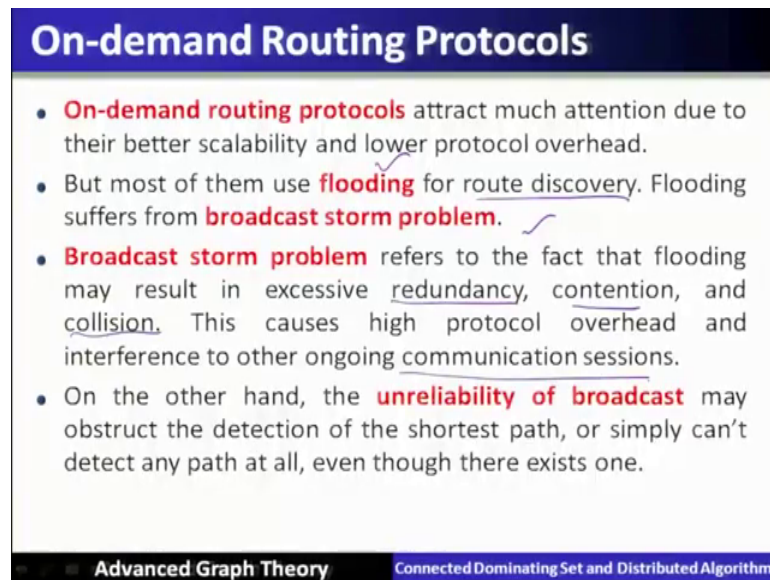
Advanced Graph Theory Connected Dominating Set and Distributed Algorithms

The classification of routing protocols we can do in 2 categories proactive and reactive routing protocols, proactive routing protocols ask each host or a many hosts to maintain a global topology information, thus a route can be provided immediately whenever it is required, but large amount of control messages are required to keep each host updated for the newest topology changes in proactive routing protocols.

Now, the scenario where there is a scarcity of the resources and also there is a change in the infrastructure there is no infrastructure. In those scenarios like wireless ad hoc networks proactive routing protocol becomes a costlier affair to maintain the information by sending all the time information, whenever there is a little change. Rather than another routing another way of handling the routing protocol is called reactive protocols reactive protocol have a feature which is called a on demand. So, our host computes a route for a specific destination only whenever it is necessary and hence it is called reactive routing protocols.

So, topology changes, which do not influence the active roads, do not trigger any route maintenance function thus the communication overhead is lower compared to the proactive routing protocols.

(Refer Slide Time: 15:46)



On-demand Routing Protocols

- **On-demand routing protocols** attract much attention due to their better scalability and lower protocol overhead.
- But most of them use **flooding** for route discovery. Flooding suffers from **broadcast storm problem**.
- **Broadcast storm problem** refers to the fact that flooding may result in excessive redundancy, contention, and collision. This causes high protocol overhead and interference to other ongoing communication sessions.
- On the other hand, the **unreliability of broadcast** may obstruct the detection of the shortest path, or simply can't detect any path at all, even though there exists one.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithms

So, on demand routing protocol attract much attention due to their scalability and lower protocol overhead, but most of them uses the flooding for the route discovery why because the destination is not known. So, they have to basically depend upon the flooding for the route discovery, flooding will have the disadvantage that it suffers from the broadcast storm problem.

So, broadcast storm problem refers to the fact that flooding may result into an excessive redundancy contention and the collision. This causes high protocol overhead and interference to the ongoing communication sessions. On the other hand the unreliability of a broadcast may obstruct the detection of the shortest path or simply cannot detect any paths at all even though there exists one.

(Refer Slide Time: 16:43)

Problem of efficiently constructing virtual backbone for ad hoc wireless networks

- In this lecture we study the “**problem of efficiently constructing virtual backbone**” for ad hoc wireless networks.
- The number of hosts forming the virtual backbone must be as small as possible to decrease protocol overhead. ✓
- The algorithm must be **time/message efficient** due to **resource scarcity**.
- We use a **connected dominating set (CDS)** to approximate the virtual backbone. ✓

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, in this lecture we will study the problem efficiently constructing the virtual backbone for ad hoc wireless networks, the number of hosts forming the wireless backbone must be as small as possible, to decrease the protocol overhead. Hence the core has to be of the smallest possible size. So, the algorithm must be also efficient due to the resource scarcity therefore, we must basically model this virtual backbone as the connected dominating set to approximate the virtual backbone for the wireless.

(Refer Slide Time: 17:29)

Assumptions (1)

- We assume a given ad hoc network instance contains **n hosts**. ✓
- Each host is in the ground and is mounted by an omni-directional antenna.
- Thus the transmission range of a host is a disk.
- We further assume that each transceiver has the same **communication range R**.
- Thus the footprint of an ad hoc wireless network is a unit-disk graph.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm


So, we will see how we can construct the connected dominating set using the distributed algorithm, which will be useful to create the virtual backbone for such applications.

So, let us take the assumptions. So, we assume that the given adhoc network instance will contain n hosts. Now each host is in the ground ; that means, these hosts can be dropped from an unmanned aircraft and when it reaches the ground, it has an antenna mounted by a on it that is a by de that is only directional antenna which will be useful to communicate and establish the network.

Thus the transmission range of a host is assumed to be a disk we further assume that each transceiver has the same communication range that is ours. Thus the footprint of an ad hoc network is nothing, but a graph which is called a unit disk graph.

(Refer Slide Time: 18:33)

Assumptions (2)

- In graph-theoretic terminology, the network topology we are interested in is a **graph $G=(V,E)$ where V contains all hosts and E is the set of links.**
- A **link between u and v** exists if their **distance is at most R .** In a real world ad hoc wireless network, sometime even when v is located in u 's transmission range, v is not reachable from u due to hidden/exposed terminal problems.
- Here, we only consider **bidirectional links.** 
- From now on, we use host and node interchangeably to represent a wireless mobile.

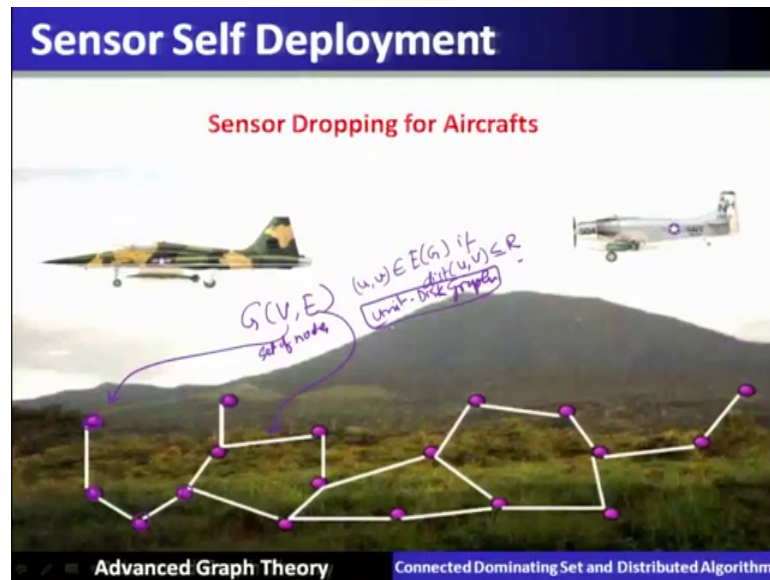
Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, in graph theoretic terminology the network topology we basically will assume is nothing, but a graph where V is the set of all hosts and E consists of the transmission links or when they are able to communicate with each other.

So, a link between the 2 nodes u and V exist if their distance is at most R . So, in real word the scenario is quite different, but for our study of the algorithm design we will assume that distance if it is at most R they are, basically established a link between the 2 knots.

Also we will consider that the links are bi directional that is if you is able to communicate to v. So, v also is able to communicate hence the link is bi directional between u and v in our model.

(Refer Slide Time: 19:32)



Now, let us see the scenario how basically the sensor nodes are dropped by an unmanned aircraft and once they will settle down using their only directional antenna they may be able to communicate hence, they will form a graph. Which is called a V comma E .? So, V is the set of nodes and the edges between these 2 nodes that is between u and v is an edge in a graph if the distance between u and v is at most R . R is the maximum transmission range. This graph we call it as the unit disc graph, the graph which is formed here after the deployment of the nodes this particular graph is called a unit disk graph and they will operate for a particular application on the field.

(Refer Slide Time: 20:50)

Existing Distributed Algorithms for MCDS					
	B. Das et al. [1997]-I	B. Das et al. [1997]-II	J. Wu et al. [1999]	K.M. Alzoubi [2001]	Mihaela Cardei et al. [2002]
Cardinality	$\leq(2\ln\Delta+3)\text{opt}$	$\leq(2\ln\Delta+2)\text{opt}$	N/A	$\leq 8\text{opt}+1$	$\leq 8\text{opt}$
Message	$O(n C +m+n\log n)$	$O(n C)$	$O(n\Delta)$	$O(n\log n)$	$O(n)$
Time	$O((n+ C)\Delta)$	$O((C + C)\Delta)$	$O(\Delta^2)$	$O(n\Delta)$	$O(n\Delta)$
Message Length	$O(\Delta)$	$O(\Delta)$	$O(\Delta)$	$O(\Delta)$	$O(\Delta)$
Information	2-hop	2-hop	2-hop	1-hop	1-hop

Table 1: Performance comparison of the algorithms. Here **opt** is the size of the given instance; **Δ** is the maximum degree; **C** is the size of the generated connected dominating set; **m** is the number of edges; **n** is the number of hosts.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

So, we will see the existing algorithms, which will construct the minimum size of connected dominating sets, which are compared on different parameters such as the cardinality, that is the size of the connected dominating sets how much is basically how many nodes are involved?.

Then how many messages are required to establish the algorithm and how much time is required to establish the algorithm. The messages which are exchanged what is basically the maximum size of them. And also basically the knowledge the local knowledge, which is required in the algorithm design that is called the information.

So, information up to 2 hop is required sometimes it is 1 hop is required. So, the little information is required to live better for the algorithm. So, here we will we will assume that we will see that this algorithm, we will study which will have the in information which is required is 1 hop information the message length is of the order big O of delta and the time required is big O of n delta. And the total message required is of the order n and the approximation is 8 approximation algorithms which we are going to discuss.

(Refer Slide Time: 22:22)

Preliminaries (2)

Minimum Connected Dominating Set (MCDS) in unit-disk Graph

- Computing an MCDS in a unit graph is **NP-hard**. Note that the **problem of finding an MCDS** in a graph is equivalent to the problem of **finding a spanning tree (ST) with maximum number of leaves**. All non-leaf nodes in the spanning tree form the MCDS. An MIS is also a DS.
- For a graph G , if $e = (u,v) \in E$ iff $\text{length}(e) \leq 1$, then G is called a unit-disk graph.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithms

Now, computing through an algorithm the MCDS that is minimum connected dominating set in a unit disk graph is NP hard note that the problem of finding and MCDS in a graph is equivalent to problem of finding or spanning tree with, the maximum number of leaves and here the non leaves nodes in the spanning tree will form the MCDS and MIS is also a dominating set.

Now for a graph G having an edge if and only if the length of the edge is less than or equal to 1 or at most 1, then that particular graph is called a unit disk graph that we will assume in our discussion.

(Refer Slide Time: 23:26)

Unit disk graph

- The topology of a wireless ad hoc network can be modeled as a $\sqrt{\text{unit-disk graph } G = (V, E)}$, a geometric graph in which there is an edge between two nodes if and only if their distance is at most one.

Figure 1: Modeling the topology of wireless ad hoc networks by unit-disk graphs.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Let us take the example how the unit disk graph is being formed. So, so a node is at the center and it is able to communicate in this range and that range is R , which is fixed for all the nodes to communicate. So, if this is the node this is this communication range if this is another node it is communication range is overlapping with this previous nodes communication range, hence there is an edge why because this particular disc is of only one unit length.

Hence they are able to communicate with each other and there exists an edge in the unit disk graph. So, the topology of a wireless ad hoc network can be modeled as a unit disk graph which is nothing, but a geometry graph in which there is an edge between the 2 nodes if and only if their distance is at most one.

Similarly we can see that there is an edge between these 2 nodes these nodes and these nodes. So, if we finally, see that this particular wireless network can be modeled as a graph and this graph is called a unit disk graph. So, from now on we will not call it as a unit disk graph, you will just call it as a graph and the graph we can obtain from wireless ad hoc networks and on this particular network we have to find out the connected dominating set minimum size connected dominating set.

For example, in this particular graph if we can see what is the minimum size of connected dominating set, we can take an example that here this particular node if we pick in a CDS and this node also in the CDS then this will become an independent set.

Why because all the nodes they are independent and all the nodes are which are not there in this particular set they are the neighbors of at least one of these nodes. For example, this node is the neighbor of both the nodes hence this particular blue one they are nothing, but they are the MISS or it is also called the dominating set.

So, these set of nodes are the dominating set together, but they are not connected to connect them we can use this node also to be included and if that is included then the size of the connected dominating set comprises of 3 nodes. So, if the 3 nodes are taken into an account then this will be a dominating set and since the induced sub graph induced by this dominating set is also connected hence it is also a connected dominating set if the 3 different nodes are taken to an to an account.

So, through the algorithm you have to find out the connected dominating set of the minimum size and that will be our problem.

(Refer Slide Time: 27:21)

Preliminaries (3)

Lemma 1. The size of any independent set in a unit-disk graph $G=(V,E)$ is at most $4opt+1$.

Proof

- Let U be any independent set of V , and let T' be any spanning tree of \overline{OPT} . Consider an arbitrary preorder traversal of T' given by v_1, v_2, \dots, v_{opt} . Let U_1 be the set of nodes in U that are adjacent to v_1 . For any $2 \leq i \leq opt$, let U_i be the set of nodes in U that are adjacent to v_i but none of v_1, v_2, \dots, v_{i-1} . Then U_1, U_2, \dots, U_{opt} form a partition of U . As v_i can be adjacent to at most five independent nodes, $|U_i| \leq 5$. For any $2 \leq i \leq opt$, at least one node in v_1, v_2, \dots, v_{i-1} is adjacent to v_i . Thus U_i lie in a sector of at most 240 degree within the coverage range of node v_i . This implies that $|U_i| \leq 4$. Therefore,

$$|U| = \sum_{i=1}^{opt} |U_i| \leq 5 + 4(opt - 1) = 4opt + 1.$$

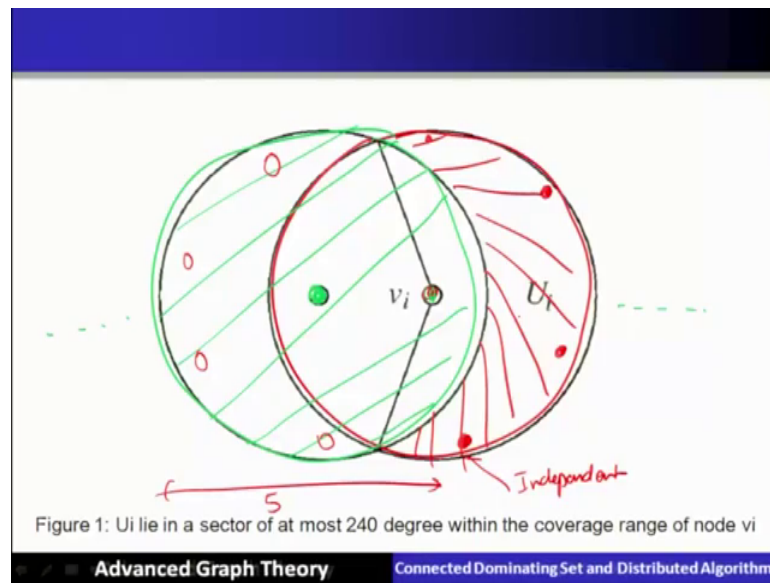
(MISS) = 4opt + 1

Advanced Graph Theory
Connected Dominating Set and Distributed Algorithm

In this particular lecture let us see some of the estimates that the algorithm which we are going to discuss is an approximation algorithm and what is the performance guarantees of this algorithm for that there are some backgrounds.

So, the lemma says that the size of any independent set in a unit disk graph is at most 4 off plus one for that let us see the picture.

(Refer Slide Time: 27:49)



Now, here this particular node v_i is having the transmission range and another node v that is able to communicate with v_i is having this communication range.

So, if all the nodes can be connected through a tree and we can list the nodes in a in some order of the traversal of this particular node, if let us say in that order these 2 nodes we have picked up this particular node, which is shown as the green color is able to communicate.

In this particular range now there can be at most 5 different independent nodes independent nodes means they are not able to communicate. So, 5 are there now if they overlapped then this will basically give a 5 the nodes and these are this particular space is overlapped with the previous space. Now we will count how many maximum number of independent nodes can be the neighbor of v_i or it can be lying in U_i set let I us say this node this node this node and so on these node all 5 node cannot be all 5 independent nodes when the nodes cannot be position here in U_i .

So, with this information let us see this particular fact let U be any independent set of V and let T prime be the spanning tree of an opt that is the optimal size of connected dominating set. Now consider an arbitrary preorder traversal of T prime given as $V_1 V_2$ and so on up to be opt U_i be the U_1 be the set of nodes in you they are adjacent to V_1 for any known I which b is basically more than 2, but less than opt let u_{iv} the be the set of nodes in u they are adjacent to V_i , but none of the other nodes. So, we will take only

these 2 nodes that we have shown in the picture and let us find out that this will partition these nodes into $U_1 \cup U_2$ and so on the partition of U .

So, v_1 is adjacent to at most 5 independent nodes. So, the size of U_1 is basically at most 5. Now for I is more than 2 that is the second this is U_1 . So, all 5 we have included there can be 5 different nodes this is V_1 . Now we will include 2 which will communicate with 1 and this particular portion we have to count how many more nodes can be there? So, this is at most 240 hence the coverage range of node V_i will imply that U_i must be less than or equal to 4.

Hence the maximum independent set in this particular U_i , if we count is nothing, but 4 times this is an $opt - 1$. So, that comes out to be 4 or plus 1. Hence the size of the independent set maximal independent set is nothing, but 4 $opt + 1$ having proved this particular lemma you will see that the algorithm which we are going to discuss requires at most 4 the size of the MIS is 4 $opt + 1$.

(Refer Slide Time: 32:27)

Preliminaries (4)

- Lemma 1 relates the size of any MIS of unit-disk graph G to the size of its optimal CDS
- **Lemma 2.1** From lemma 1, the size of any MIS of G is **at most $4 \times opt + 1$** where opt is the size of any optimal CDS of G .
- For a minimization problem P , the performance ratio of an approximation algorithm A is defined to be

$$\rho_A = \sup_{i \in I} \frac{A_i}{opt_i}$$
- where I is the set of instances of P , A_i is the output from A for instance i and opt_i is the optimal solution for instance i . In other words, ρ is the supreme of A/opt among all instances of P .

Advanced Graph Theory
Connected Dominating Set and Distributed Algorithm

Now here we will show that the minimization algorithm the approximation is nothing, but is a ratio performance ratio, which is called it is nothing, but a supremum of this particular factor A_i upon opt_i ; that means, A_i is the output size of CDS and opt_i is the optimal size of that particular problem instance i .

So, for all instance of the problem you have to find out the supremum that becomes the approximation algorithm guarantees or the performance ratio.

(Refer Slide Time: 33:17)

An 8-approximate algorithm to compute CDS

- This algorithm contains two phases: ✓
- **Phase-1:** First, a maximal independent set (MIS) is computed;
- **Phase-2:** Then a Steiner tree is used to connect all vertices in the MIS.
- This algorithm has performance ratio at most 8 and is message and time efficient.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Let us see the algorithm which runs in 2 phases the phase one will construct the maximal independent set and in the second phase we will connect them through a tree, which is called Steiner tree and we will also show that performance ratio of this algorithm is 8 the algorithm is message and time efficient also.

(Refer Slide Time: 33:40)

Algorithm description

- Initially each host is colored **white**. ✓ *Colors (white, Gray, Black)*
- A **dominator** is colored **black**, while a **dominatee** is colored **gray**. ✓
- We assume that each vertex knows its **distance-one neighbors** and their **effective degrees d^*** . *no. of white neighbors*
- This information can be collected by periodic or event-driven **hello messages**. ✓
- The **effective degree** of a vertex is the total number of **white neighbors**. ✓

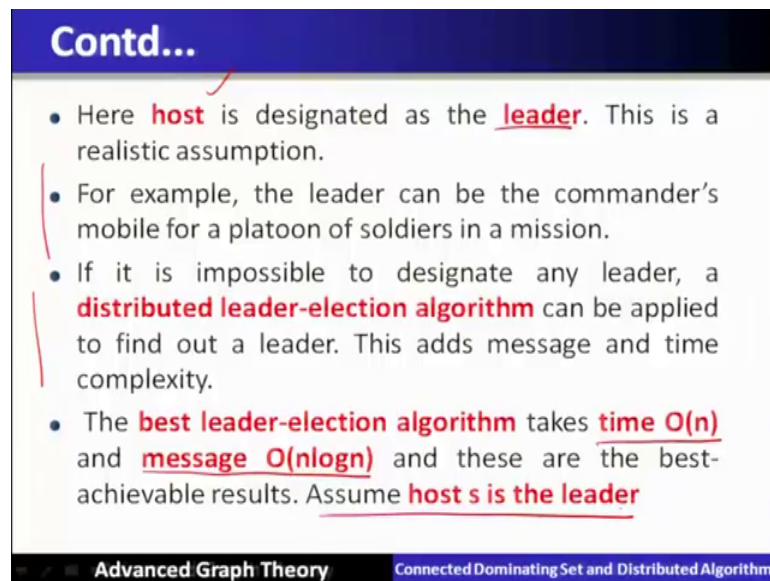
Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Let us start with the algorithm that all the nodes are initially colored as white the dominator is colored as black and dominatee is colored as gray. So, there are 3 different colors we are going to use in this algorithm white, then gray, and black.

Now we assume that each vertex has the knowledge of its distance one neighbors and also they have another information, which is called an effective degrees; that means, how many number of white neighbors are there at any point of time is called an effective degree.

Now this information can be collected by periodic event driven messages, which are called hello messages the effective degree of a vertex is the total number of white neighbor that I have already told.

(Refer Slide Time: 34:50)



Contd...

- Here **host** is designated as the **leader**. This is a realistic assumption.
- For example, the leader can be the commander's mobile for a platoon of soldiers in a mission.
- If it is impossible to designate any leader, a **distributed leader-election algorithm** can be applied to find out a leader. This adds message and time complexity.
- The **best leader-election algorithm** takes **time $O(n)$** and **message $O(n \log n)$** and these are the best-achievable results. Assume host s is the leader

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, one of these host nodes we had we have designated as a leader this particular assumption is also realistic, because if the leader is not because leader can be a commander mobile commanders mobile for the platoon of the soldiers in a particular mission. So, there is always a leader designated in such network and in many such applications.

Now, if it is not designated then we can designate using an algorithm which is called a distributed leader election algorithm. So, if we run the distributed leader election algorithm it will take order n time and the number of messages will be $n \log n$ with the

best known such available. So, hence let us assume without loss of generality that a host s is the leader in the construction of CDS.

(Refer Slide Time: 35:50)

Phase 1:

- Host s first **colors itself black** and broadcasts message **DOMINATOR**.
- Any **white host u** receiving **DOMINATOR** message the first time from v colors itself **gray** and broadcasts message **DOMINATEE**. u selects v as its dominator.
- A **white host** receiving **at least one DOMINATEE** message becomes active.
- An active white host with **highest (d^*, id)** among all of its **active white neighbors** will **color itself black** and broadcast message **DOMINATOR**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now the phase one will construct the MIS. So, let us start with the leader node s is the leader. So, leader first color itself as a black now here this is the leader, which will color itself is let us use the black while all other nodes are white.

Now, this particular node will send a message which is called a dominator, because after becoming black this particular node is called a dominator node it will send a message which is called a dominator. And that will go in it is all the nodes in it is communication range which is nothing, but a disk within the disk now any white node u let us say this is the white node u receiving the dominator message for the first time from a vertex v it will color itself as a gray it will color itself as a gray and broadcast a message which is called a dominatee.

So, it will broadcast. So, this particular message when it will be broadcast will be broadcast in it is communication range and this you will select v as it is dominator. Now this particular white node here they are the white nodes when they will receive this particular message they become active. The active white host with the highest effective degrees D^* among all it is active white neighbor will color itself black and broadcast dominator.

Let us assume that this is the node having the highest degree and it has colored itself as a black and it will broadcast a message, which is called a dominator.

(Refer Slide Time: 38:29)

Contd...

- A **white host decreases its effective degree by 1** and broadcasts message **DEGREE** whenever it receives a **DOMINATEE** message.
- Message **DEGREE** contains the sender's current effective degree. A white vertex receiving a **DEGREE** message will update its neighborhood information accordingly.
- Each gray vertex will broadcast message **NUMOFBLACKNEIGHBORS** when it detects that none of its neighbors is white.
- **Phase 1** terminates when **no white vertex left**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

So, the white node decreases its effective degree by one and broadcast message the degree whenever it receives a dominatee message. The message degree contains the sender's current effective degree a white vertex receiving a degree message will update its neighborhood information.

So, each gray vertex will broadcast message number of black neighbors, when it detects that none of its neighbors is white. So, phase one terminates when there is no white neighbors left.

(Refer Slide Time: 39:05)

Phase 2: (Steiner Tree - Connect MIS (Phase 1))

- When s receives message **NUMOFBLACKNEIGHBORS** from all of its gray neighbors, it starts **phase 2** by broadcasting **message M**.
- A host is **"ready"** to be explored if it has **no white neighbors**.
- A **Steiner tree** is used to connect **all black hosts** generated in **Phase 1**.
- The idea is to pick those **gray vertices** which connect to many **black neighbors**.

(Minimize number of nodes in CDS)

Advanced Graph Theory Connected Dominating Set and Distributed Algorithms

Now, we will begin the phase 2 phase 2 is nothing, but it will construct a Steiner through a Steiner tree it will connect the nodes in MIS, which we have obtained in the phase 1 and hence MIS plus the connectors will become a connected dominating set. So, Steiner tree will be able to form a sub graph induced by the MIS nodes and that will be the connected dominating set together.

Now when S receives S here is basically the leader node when leader receives the message number of black neighbors from all of it is gray neighbors it starts phase 2 by broadcasting a message M. So, our host is ready to be exploring for hi as no white neighbors I Steiner tree is used to connect all the black host generated in phase 1 that I have told the idea is to pick those grey vertices, which connect to many of the black neighbors. Why because, we want to minimize the nodes in CDS so; that means, with the minimum number of nodes we want to establish a connection want to connect MIS nodes.

(Refer Slide Time: 40:56)

Contd...

- The classical **distributed depth first search spanning** tree algorithm will be modified to compute the **Steiner tree**.
- A **black vertex** without any dominator is **active**.
- Initially **no black vertex** has a dominator and all hosts are **unexplored**.
- Message **M** contains a field **next** which specifies the **next host** to be explored.
- A **gray vertex** with at least **1 active black neighbors** are **effective**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, we will apply a classical depth for search algorithm and that distributed version that is the distributed depth for search spanning tree algorithm to compute the Steiner tree here in this particular phase 2 algorithm. Now a black node without any dominator is active and usually no black vertex has the dominator and all hosts are unexplored message **M** contains the field **next** which specifies next has to be explored a gray vertex with a at least one active black neighbors are effective.

(Refer Slide Time: 41:33)

Contd...

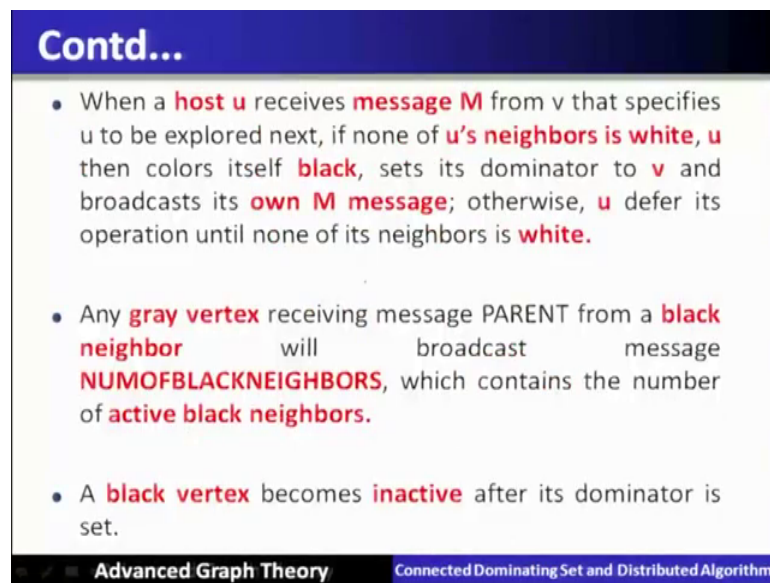
- If **M** is built by a **black vertex**, its **next field** contains the **id** of the **unexplored gray neighbor** which connects to maximum number of **active black hosts**.
- If **M** is built by a **gray vertex**, its **next** field contains the **id** of any **unexplored black neighbor**.
- Any **black host u** receiving an **M message** the first time from a **gray host v** sets its dominator to **v** by broadcasting message **PARENT**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

So, M is built by the black vertex it is next field contains the idea of unexplored gray vertices, which connects to the maximum active black host M is built by the gray vertex then the next field of id contains the unexplored black neighbors.

So, either M is built by the black vertex or M is built by the gray vertex. So, any black host receiving message M for the first time from the gray host sets it is dominated to that v by broadcasting the parent message.

(Refer Slide Time: 42:11)



Contd...

- When a **host u** receives **message M** from v that specifies u to be explored next, if none of **u's neighbors is white**, **u** then colors itself **black**, sets its dominator to **v** and broadcasts its **own M message**; otherwise, **u** defer its operation until none of its neighbors is **white**.
- Any **gray vertex** receiving message **PARENT** from a **black neighbor** will broadcast message **NUMOFBLACKNEIGHBORS**, which contains the number of **active black neighbors**.
- A **black vertex** becomes **inactive** after its dominator is set.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

So, when a host u receives message M from v that specifies u to be explored next, if none of u's neighbors is white, u then color itself as a black, sets it is dominated to v and broadcasts it is own message M; otherwise, u defer the operation until none of it is neighbors is white. Similarly this particular process will follow for the gray neighbors.

(Refer Slide Time: 42:38)

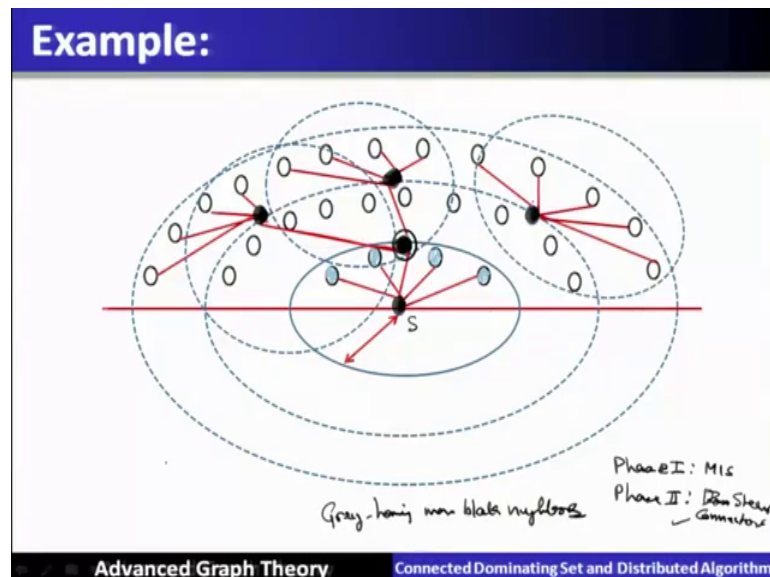
Contd...

- A **gray vertex** becomes **ineffective** if none of its **black neighbors is active**.
- A **gray vertex** without **active black neighbor**, or a **black vertex** without effective **gray neighbor**, will send message **DONE** to the **host** which activates its exploration or to its dominator.
- When **s** gets message **DONE** and it has no effective **gray neighbors**, the algorithm **terminates**.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, when there is no when the gray vertices becomes ineffective if there is no black neighbors active. So, when there is no active effective black neighbors remains the gray neighbors becomes ineffective similarly for the black nodes. So, when s gets a message done that is it has no effective gray neighbors and the algorithm will terminate.

(Refer Slide Time: 43:12)



Let us see this entire algorithm working through in this particular example this is s that is the leader node. So, in phase one it will broadcast first it will make itself as a black and broadcast message as a dominator in it is one hop neighborhood they will turn as a gray.

So, all the other nodes are turned as gray nodes, these gray nodes will further communicate a dominating message in its neighborhood and the white nodes who will receive the gray or a dominating message become active. So, this particular every white node will count how many active white neighbors are there and the maximum one for example, having that particular value will basically become black. So, in the phase one when phase one will finish it will basically form the black nodes and they are nothing, but the MIS.

Now, in phase 2 when there is no white node left then phase 2 will begin. Now phase 2 will be initiated either by the gray nodes gray nodes will inform how many black neighbors are there. For example, this is the gray node which has 2 different black 1 2 and 3 3 different black nodes are there in its neighborhood. So, it will inform to these particular black nodes using this particular message M.

So, this particular black node will choose this gray node and it will send a parent message. So, it will establish a link between it now since this is the node which is having the maximum number of black neighbors. So, this will in the phase 2 through the Steiner tree this will become a black. So, this is the black node that is the gray node having the gray node having the maximum number of black neighbors will be basically elected as the connectors or through the Steiner tree it will be turned into a dominator.

So, when phase 2 will finish then there will not be; that means, all the black nodes will be having assigned a connector node through the gray node and also there is no gray node which is having an active or an effective black node without any connection with its neighbor dominating sets or it is neighbor MISS. Hence the algorithm will terminate with forming a connected dominating set, which is shown here with the black nodes.

(Refer Slide Time: 46:45)

Complexity

- Note that phase 1 sets the dominators for **all gray vertices**. Phase 2 may modify the dominator of some **gray vertex**.
- The main job for **phase 2 is to set a dominator for each black vertex**. All black vertices form a CDS.
- In Phase 1, each host broadcasts each of the messages **DOMINATOR** and **DOMINATEE** at most once.
- The message complexity is dominated by message **DEGREE**, since it may be broadcasted **Δ times** by a host, where **Δ is the maximum degree**.
- Thus the **message complexity of Phase 1 is $O(n \Delta)$** . The **time complexity of Phase 1 is $O(n)$** .

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

The complexity of the phase one and the complexity of phase 2 nodes which we will see the total of message and time complexity is given as of the order n and message complexities of the order n time's big delta.

(Refer Slide Time: 47:02)

Performance Analysis

- **Lemma 3.2** Phase 1 computes an MIS which contains all black nodes.
- **Proof.** A node is colored black only from white. No two white neighbors can be colored black at the same time since they must have different (d^*, id) .
- When a node is colored black, all of its neighbors are colored gray.
- Once a node is colored gray, it remains in color gray during Phase 1.
- From the proof of **Lemma 3.2**, it is clear that if (id) instead of (d^*, id) is used, we still get an MIS. Intuitively, this result will have a larger size.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, let us see the performance analysis phase one computes n MIS which contains all the black nodes.

(Refer Slide Time: 47:10)

Contd...

- **Lemma 3.3** In phase 2, at least one gray vertex which connects to maximum number of black vertices will be selected.
- **Proof.** Let u be a gray vertex with maximum number of black neighbors.
- At some step in phase 2, one of u 's black neighbor v will be explored.
- In the following step, u will be explored. This exploration is triggered by v .

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, another lemma says that in phase 2 at least one green vertex, which connects to the maximum number of black vertices, will be selected.

(Refer Slide Time: 47:20)

Contd...

- **Lemma 3.4** If there are c black nodes after phase 1, then at most $c-1$ gray hosts will be colored black in phase 2
- **Proof.** In phase 2, the first gray vertex selected will connect to at least 2 black vertices.
- In the following steps, any newly selected gray vertex will connect to at least one new black vertex.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Now, here in lemma 3.4 we will see that if c is the number of black nodes hosts after the phase 1, then it requires at most c minus 1 grey host which can be colored black in phase 2.

(Refer Slide Time: 47:37)

Contd...

- **Lemma 3.5** If there exists a gray vertex which connects to at least 3 black vertices, then the number of gray vertices which are colored black in phase 2 will be at most $c-2$, where c is the number of black vertices after phase 1.
- **Proof.** From **Lemma 3.3**, at least one gray vertex with maximum black neighbors will be colored black in phase 2. Denote this vertex by u . If u is colored black, then all of its black neighbors will choose u as its dominator. Thus, the selection of u causes more than 1 black hosts to be connected.

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Lemma 3.5 says that if there are gray vertex which connects at least 3 black nodes. So, the number of black vertices required is c minus 2.

(Refer Slide Time: 47:52)

Contd...

- **Theorem 3.6** This algorithm has performance ratio at most 8.
- **Proof.** From Lemma 3.2, phase 1 computes a MIS. We will consider two cases here.
- If there exists a gray vertex which has at least 3 black neighbors after phase 1, from Lemma 2.1, the size of the MIS is **at most $4 \cdot \text{opt} + 1$** .
- From lemma 3.5, we know the total number of black vertices after phase 2 is **at most $4 \cdot \text{opt} + 1 + ((4 \cdot \text{opt} + 1) - 2) = 8 \cdot \text{opt}$** .

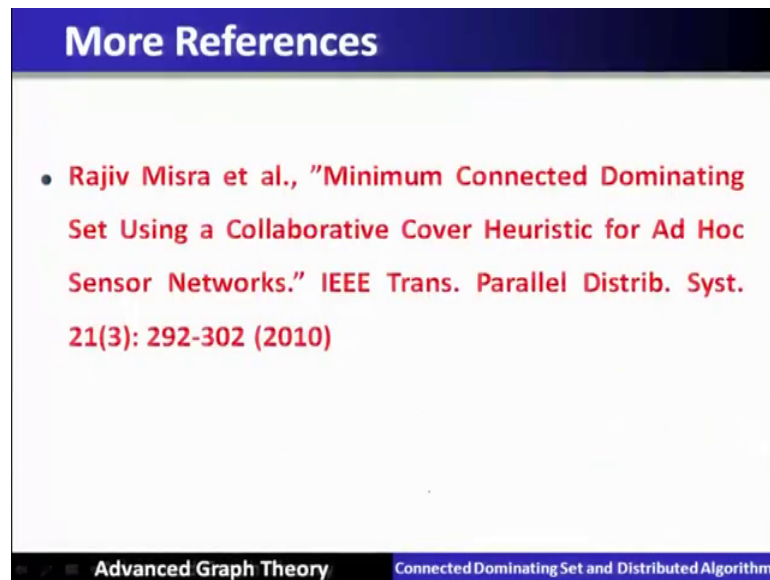
$|MIS| = 4 \cdot \text{opt} + 1 + ((4 \cdot \text{opt} + 1) - 2)$
PR. = $\underline{8 \cdot \text{opt}}$

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Therefore if we check the performance ratio of this algorithm in the phase 1 the size of the MIS is let us say $4 \cdot \text{opt} + 1$.

Now, in phase 2 we require let us say M is size minus 2 that is plus $4 \cdot \text{opt} + 1$ and minus 2 these are the connectors. So, totally if we count it becomes $8 \cdot \text{opt}$. So, $8 \cdot \text{opt}$ and this 1 and 1 will go out. So, hence the performance ratio of this algorithm is $8 \cdot \text{opt}$.

(Refer Slide Time: 48:43)



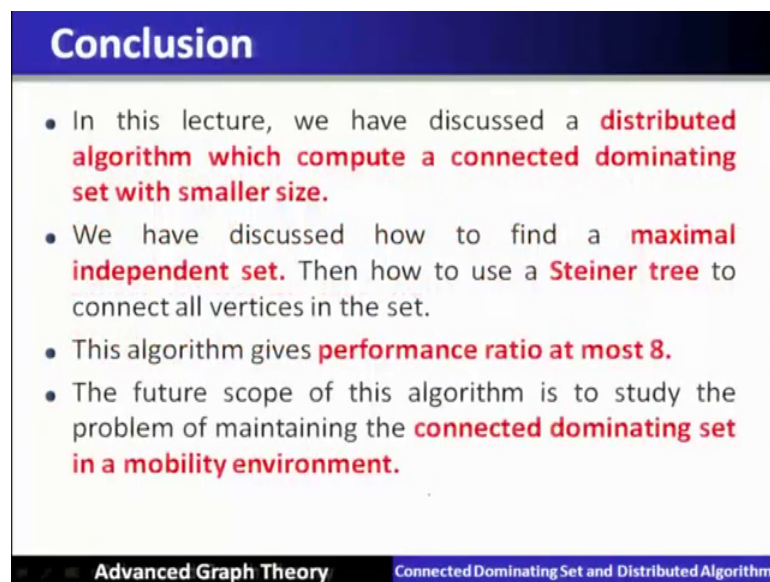
More References

- **Rajiv Misra et al., "Minimum Connected Dominating Set Using a Collaborative Cover Heuristic for Ad Hoc Sensor Networks." IEEE Trans. Parallel Distrib. Syst. 21(3): 292-302 (2010)**

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

More references on this you can find the paper by Rajiv Misra “Minimum Connected Dominating Set Using a Collaborative Cover Heuristic for Ad Hoc Sensor Networking” published in I triple e transaction parallel distributed computing 2010.

(Refer Slide Time: 49:00)



Conclusion

- In this lecture, we have discussed a **distributed algorithm which compute a connected dominating set with smaller size.**
- We have discussed how to find a **maximal independent set.** Then how to use a **Steiner tree** to connect all vertices in the set.
- This algorithm gives **performance ratio at most 8.**
- The future scope of this algorithm is to study the problem of maintaining the **connected dominating set in a mobility environment.**

Advanced Graph Theory Connected Dominating Set and Distributed Algorithm

Conclusion in this lecture we have discussed the algorithm distributed algorithm for connected dominating set of smaller size, we have seen that this construction requires first to construct maximal independent set and then Steiner tree to connect all these vertices and the performance ratio we have shown that it is 8.

In future scope of this algorithm is to study the problem of maintaining serious in the mobility environment.

Thank you.