DATA RESILIENCE VIA DATA AGGREGATION: OVERCOMING OVERALL STORAGE OVERFLOW IN SENSOR NETWORKS

Bin Tang, Yan Ma

Presented by : Basil Alhakami

CONTENT

- Introduction
- Data overflow
- Data aggregation
- Formulation of Data Resilience via Data Aggregation (DRA)
- Multiple Traveling Salesman Walk Problem (MTSW)
- Solving DRA

INTRODUCTION

- Large amount of data
- Limited storage capacity
- Not feasible to install base station due to the challenging environment sensors are deployed in

DATA OVERFLOW

Data node : nodes with overflow data

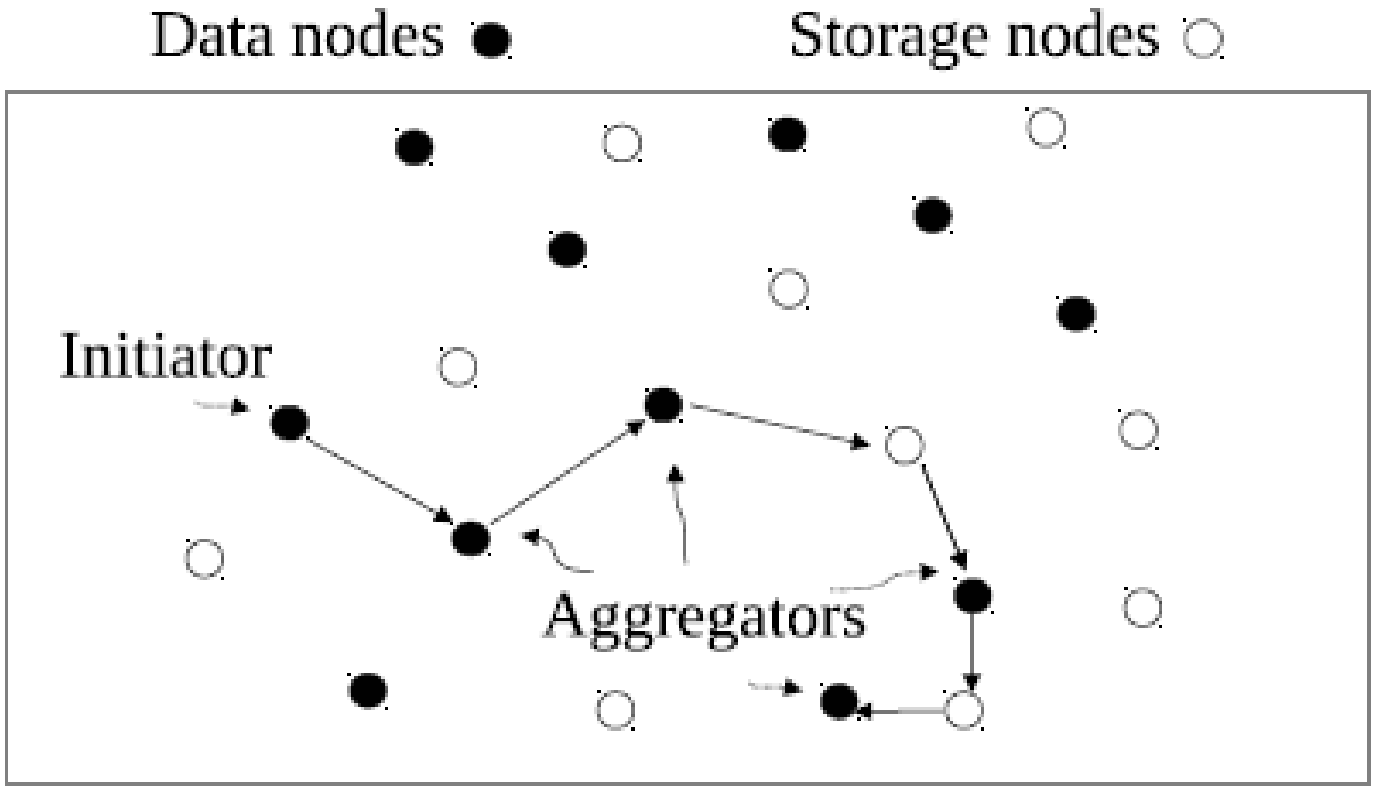
Storage nodes: nodes with available storage

- Node storage overflow
- Overall storage overflow

DATA AGGREGATION

Initiator : Send the overflow data to other nodes

Aggregator: receives the overflow data and aggregates its own overflow data



FORMULATION OF DRA

$$q = \lceil \frac{p \times (R+m) - |V| \times m}{R-r} \rceil$$

- q : the number of aggregators needed
- |V| deployed sensor nodes
- m: the available storage space
- p: the number of data nodes
- R: the overflow data size at each data node before aggregation
- r: the overflow data size at each aggregator after aggregation

- At most (p-q) can be selected as initiators
- The number of aggregation walks cannot exceed the number of initiators

EXAMPLE

Sensor network of 9 nodes:

Data Nodes: B D E G I

Storage Nodes: A C F H

R = m = 1

r= ¾

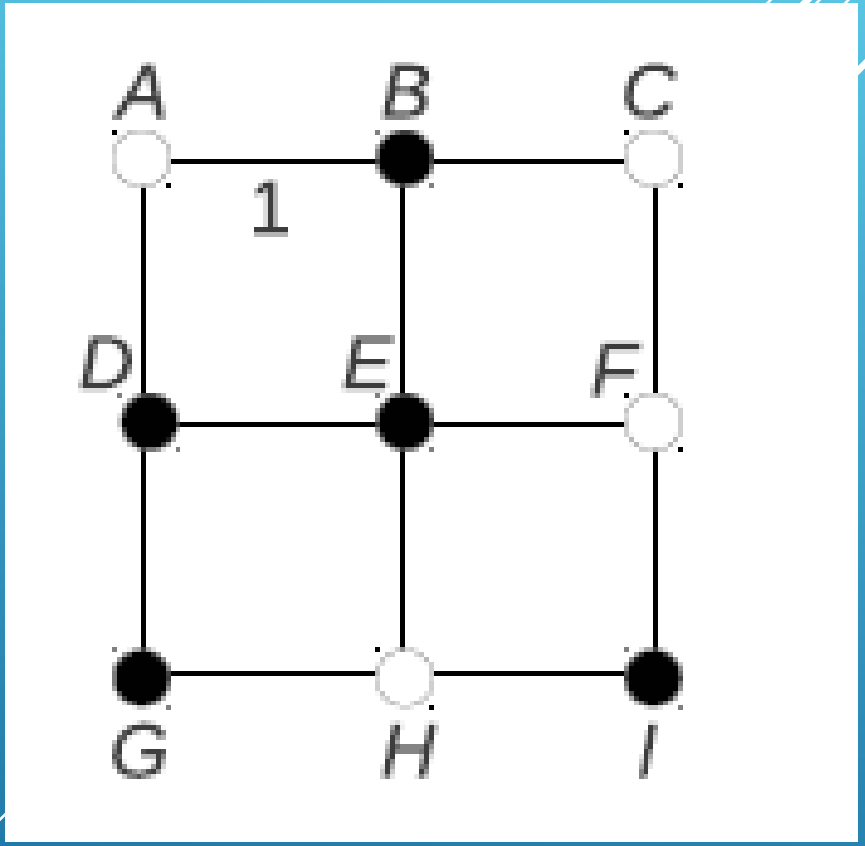
Energy consumption along any edge = 1

q=4 which means we have one initiator

Optimal Solution: B is the initiator

The walk is: B E D G H I

Cost : 5



OBJECTIVE OF MTSW

solving DRA in a sensor network is equivalent to solving MTSW in an aggregation graph transformed from sensor network.

- Given an undirected weighted graph G = (V;E) with |V |nodes and |E| edges
- a cost metric (which represents the distance or traveling time between two nodes)
- MTSW determine a subset of at most b starting nodes (i.e., the initiator in DRA)

salesman can be dispatched to visit a number of other nodes

following a walk, such that

a) all together q nodes (excluding starting nodes) are visited

b) the total cost of the walks is minimized

MTSW DECIDE

Set of starting nodes

Set of walks

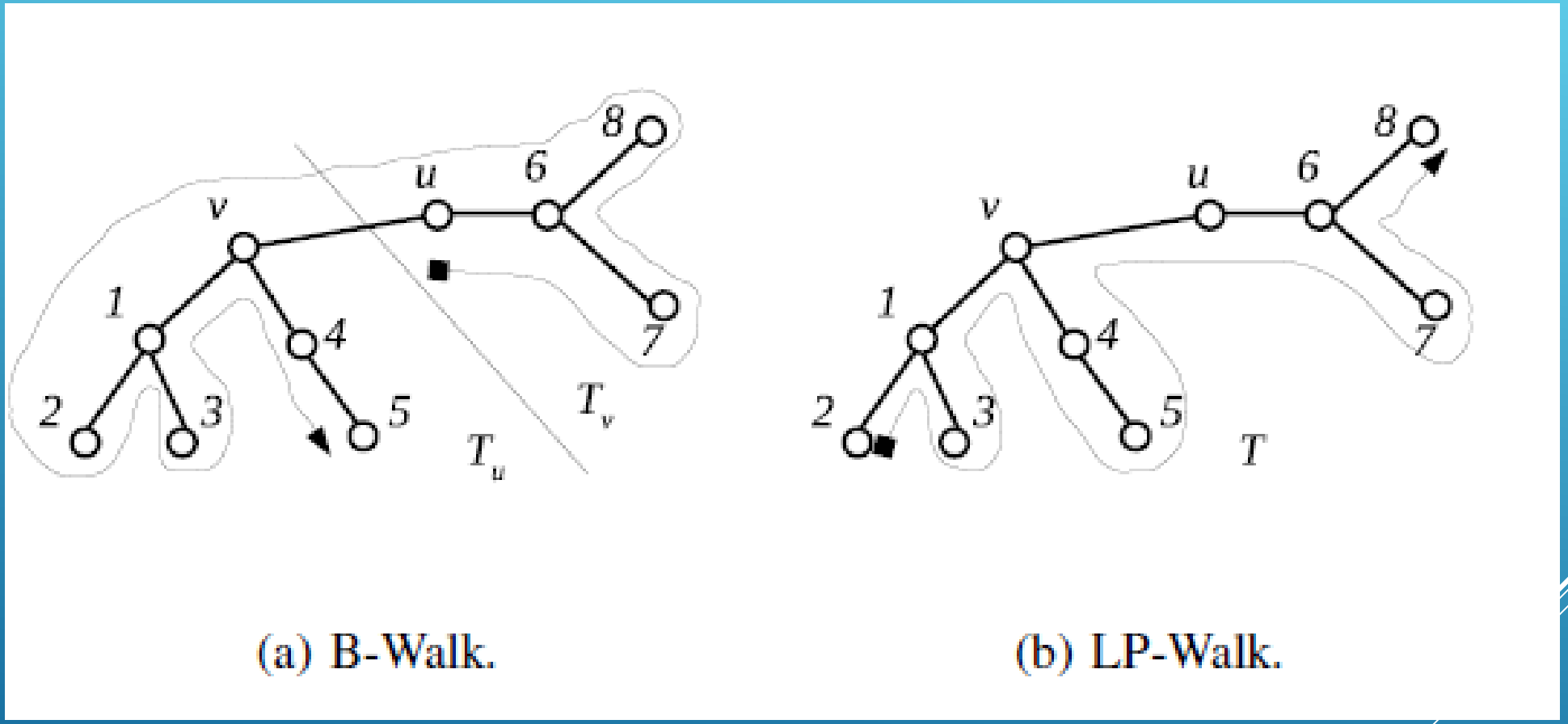
Such that

Walking cost is minimized

ALGORITHMIC SOLUTION OF MTSW (SOLVING DRA)

- Approximation Algorithm
  - B walk
- Heuristic algorithm
  - LP walk

We need better energy consumption (lower walk cost)



(a) B-Walk.

(b) LP-Walk.

THE APPROXIMATION ALGORITHM

yields a total cost of the walks that is at most (2 -1/q)times of the optimal cost.

1-sorts all the edges in E into nondecreasing order of their weights

2- initializes the set Eq to the empty set and creates |V |trees, each containing one node

3-checks each edge, if it is cycleless w.r.t. Eq. If yes, add it into Eq

4- repeat 3 until we have q edges

It then obtains:

all the connected components induced by these q edges.

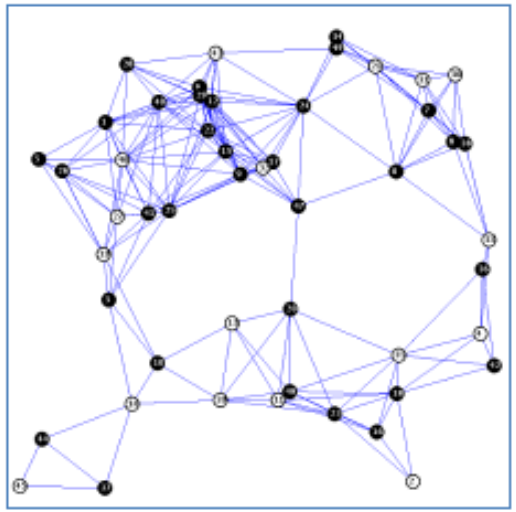
If linear topology : start from one end visits the nodes in the linear topology exactly once

If it is a tree : B walk along the tree

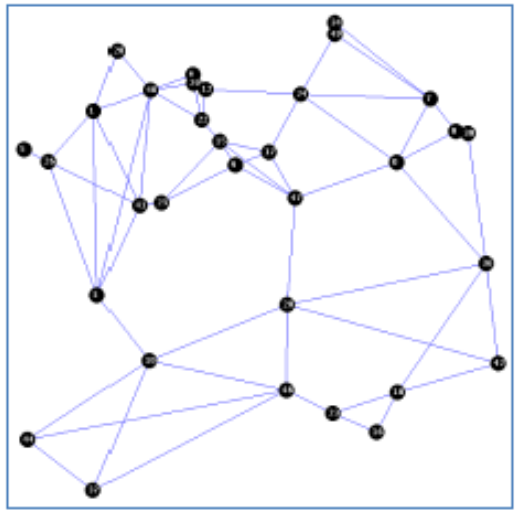
HEURISTIC ALGORITHM

- Improve the performance of the approximation algorithm by using LP walk instead of the B walk

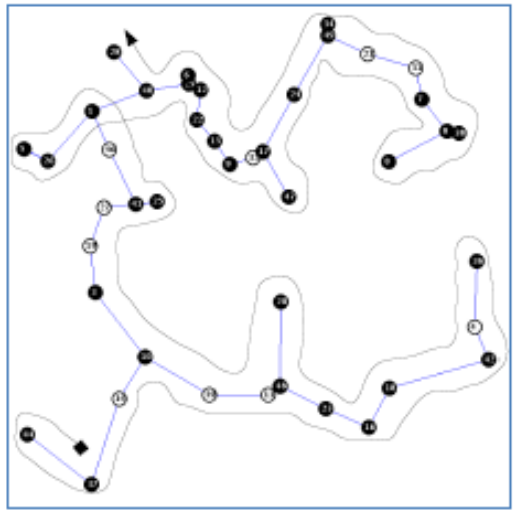
COMPARISON



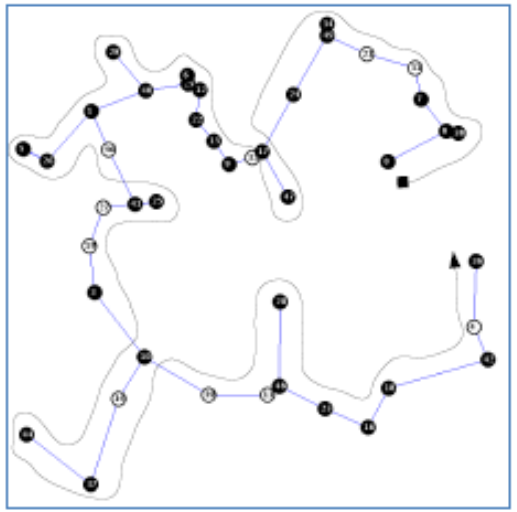
(a) Sensor network graph.



(b) Aggregation graph.



(c) B-Walk (cost=381.2J).



(d) LP-Walk (cost=290.6J).

THANK YOU