UNIT 4

Dissemination Protocols

Syllabus

Dissemination Protocol for Large Sensor Network. Data Dissemination, Data Gathering and Data Fusion; Quality of a Sensor Network; Real-time Traffic Support and Security Protocols.

Learning Objectives

- ✓ Introduction to Data Dissemination
- ✓ Various Dissemination Protocols
- ✓ Objective of Data Gathering
- ✓ Various Algorithms Used to Implement Data Gathering
- ✓ Brief Introduction to Data Fusion
- ✓ Coverage and Exposure Problems
- ✓ Various Protocols Used for Real Time Communication
- ✓ Various Security Attacks
- ✓ Various Protocols Used to Handle the Security Attacks.

Introduction

Data dissemination can be defined as a process through which data is routed in the sensor network. Different algorithms used for data dissemination includes, Flooding, Gossiping, Rumor routing, SAR, Directed diffusion, GHT, SPIN and Cost field approach.

The goal of data gathering process is to transmit the sensed data from every sensor to a base station. The process in which the base station collects data from all sensor nodes once is called one round. In this process, the algorithm tries to maximize the number of communication rounds of nodes prior to their death and before a network becomes inoperable. Some of the algorithms of data gathering process include Direct Transmission, PEGASIS, Binary Scheme and Chain based Three level Scheme.

The various security attacks are Spoofing, Wormhole attack, Blackhole attack and Hello flood attack and these attacks can be handled by using LEAP and INSENS protocols.

PART-B ESSAY QUESTIONS WITH SOLUTIONS

4.1 DISSEMINATION PROTOCOL FOR LARGE SENSOR NETWORK

4.1.1 Data Dissemination

Q9. Define the following,

(i) Data dissemination

(ii) Data collection

(iii) Data diffusion

(iv) Flooding

(v) Gossiping

(vi) Rumor routing.

Model Paper-II, Q8(a)

Answer :

(i) Data Dissemination

The process through which data is routed in the sensor network is called data dissemination. The node that generates data is called a source and this data is transmitted to interested nodes. An event is the information that needs to be reported and the node that is interested in an event is called a sink. Sensor networks have some traffic modules such as data collection and diffusion models.

(ii) Data Collection

This model involves the source that transmits the collected data to other entities based on a time period or on time basis.

(iii) Data Diffusion

This model has two steps – data propagation and interest propagation. An interest describes a particular kind of event or node such as intrusion and temperature. The interest is broadcasted by sink to neighbors and refreshed periodically. It is propagated across the network and a copy of cache is maintained by nodes in the network.

When an event occurs, it is informed to nodes that are interested in it. A data cache is also maintained by the intermediate nodes which are capable of reporting data by aggregating and modifying it. The data propagation path is changed by selecting the shortest path.

(iv) Flooding

A node that receives a data packet broadcasts it to its next node after verifying that itself is not an intended recipient of the packet. This broadcasting continues by nodes until the maximum hop count of the data packets is not reached. It is a simple routing technique that requires low maintenance. However, it has some disadvantages associated with it. It suffers from implosion i.e., when duplicate messages are sent to the same node from its multiple neighbors. Same event can be sent by multiple nodes which can happen due to coverage overlapping regions. This results in delivering duplicate copies to same node.

Further, the availability of every node is not considered by flooding protocol which also results in reduced messages to the recipients. All such disadvantages degrade the overall performance of the network.

(v) Gossiping

Gossiping is a modified form of flooding wherein a node does not broadcast the data packets, instead it transmits the packets in a random manner. This reduces implosion and lowers overhead. However, it consumes more time for propagation of messages and transmission of messages is not guaranteed.

(vi) Rumor Routing

Rumor routing algorithm is based on agents which are entities created by nodes on random basis. These agents are packets which are propagated across the network to create the shortest path for the events. When a node whose path is longer than its own path, it updates the routing table of the interested node.

A query generated at a sink is sent randomly on the network to find a path. When an event path is not found by this query, the sink times out and finally propagates the packets using flooding method.

Q10. Write

uting, Directed Diffusion and Geographic Hash Table.

Model Paper-II, Q8(b)

Answer :

Sequential Assignment Routing (SAR)

Sequential Assignment Routing (SAR) protocol generates large number of trees wherein every tree's root is a one-hop neighbor of the sink. Every tree grows from the sink in outward direction and the nodes which have high delay or low throughput are avoided. Further, every sensor node stores two parameters of path using additive QoS metrics and the energy resources available on the path. This algorithm also reduces the QoS metrics of the network and triggers an update on regular periods.

Directed Diffusion

The directed diffusion protocol is applied in circumstances where the sensor nodes are capable to generate the queries sensed by nodes. This is in contrast to all queries generated from a base station. Therefore, a sensor node or a base station can sink the query. This protocol enhances data diffusion by utilizing interest components (gradients). Every sensor node assigns a name to its data along with a single or multiple characteristics which are used to attract interest from other nodes. The interest components that are positive allow data to flow along a path while a negative component does not allow data to flow along a path.

Geographic Hash Table

Geographic Hash Table (GHT) is based on data-centric storage. It hashes keys into geographical edges and stores pair of key and value at the sensor node. Data is replicated and stored at multiple sites to ensure proper functioning even in case of node failure. A consistency protocol is applied to maintain the copies of data. Further, the data is propagated across the node along with consideration of scalability and data storage load.

Greedy Perimeter Stateless Routing (GPSR) protocol is applied to route data and queries.

The GHT protocol works best for large scale networks. In such networks, the data in stored across the network instead of storing at a central place. Queries are routed to the nearest node that holds a copy of data allowing uniform distribution of storage and data traffic.

Q11. Explain about SPIN Protocol and Cost Field Approach.

Model Paper-II, Q9(a)

Answer :

Sensor Protocol for Information via Negotiation (SPIN)

Sensor Protocol for Information via Negotiation (SPIN) protocol overcomes the disadvantages of flooding protocol by resource adaptation and negotiation. It enhances network lifetime by adopting resource-aware operations and reduces implosion and overlap through negotiation. It also lowers processing overhead by transmitting only meta data in place of raw data. SPIN protocol consists of three messages – ADV, REQ and DATA. ADV message is used to propagate meta data of actual data. REQ message is used by a neighbor if it is interested to receive data. Once a node receives REQ message, it transmits actual data to the requestor through DATA message. A node again sends ADV message to its neighbors and this process continues in the network. This protocol is data-centric routing based where a node informs about its data to its neighbors and then waits for data request from them.

SPIN-2 protocol enhances the performance by reducing participation of nodes by using energy or resource threshold. In other words, a node participates in ADV-REQ-DATA handshake only when it is coupled with enough resources to do so.

Cost Field Approach

The cost field approach sets path to a sink. It consists of two phases. In first phase, cost field is setup on the basis of metrics and in second phase, data dissemination is done using the cost.

The cost is defined from a node to the sink by every node. The sink broadcasts ADV packets with its own cost as '0'. A node $N1$ when receives a message from node $N2$, it assigns its packet cost to min (T_{N1}, $T_{N2} + D_{N1N2}$)

Where,

T_{N1} = Total cost of path from node $N1$ to sink

T_{N2} = Cost of node from $N2$ to the sink

$D_{N1\,N2}$ = Cost of nodes from $N1$ to $N2$.

When N_1 is upgraded, the new cost is broadcasted across the network using another ADV message. Further, a node defers its ADV messages instead of broadcasting them at the back-off time. The back-off time is $\gamma_1 \times D_{N1N2}$, where,

γ_1 = Algorithm parameter.

4.1.2 Data Gathering and Data Fusion

Q12. What is the objective of Data Gathering? Discuss various algorithms that implement data gathering.

Model Paper-I, Q8(a)

Answer :

Data Gathering

The goal of data gathering process is to transmit the sensed data from every sensor to a base station. The process in which the base station collects data from all sensor nodes once is called one round. In this process, the algorithm tries to maximize the number of communication rounds of nodes prior to their death and before a network becomes inoperable.

In order to achieve this goal, minimum energy should be consumed and the transmission should occur with minimum delay. Therefore energy × delay metric is applied to compare algorithms.

Some of the algorithms of data gathering process include,

1. **Direct Transmission**

In direct transmission algorithm, all sensor nodes propagate data directly to the base station. Due to direct transmission, it consumes more energy which makes it expensive. Media access delay also consumes more time as nodes have to wait for their turn in order to avoid collision with other nodes. Keeping in view energy × delay metric, this algorithm does not work properly.

2. **Power Efficient Gathering for Sensor Information Systems (PEGASIS)**

PEGASIS protocol assumes that one node possesses information about the location of every other node in the network. A node is selected as a leader among all nodes based on capability to reach the base station in only one-hop.

This protocol has certain goals such as,

(i) To reduce the number of messages that need to be sent to a base station.

(ii) To distribute energy consumption uniformly to all nodes.

(iii) To reduce the distance of every node transmission.

(iv) To reduce the overhead of broadcasting.

In this protocol, a chain of sensor nodes is formed. The chain starts from the node that is located at the farthest distance from the base station and at every step the next node that is not yet visited is added in the chain.

The chain is formed before starting data transmission and it is reformed if a node dies. Further, data is aggregated at every node to ensure that only one message in transformed between a node and its next neighbor. The leader node finally transmits one message at a time to the base station. The leadership is passed from a node to other in sequential order and when leadership is transferred, it is informed to all nodes so that they can route their messages to new leader.

3. **Binary Scheme**

Binary scheme is a chain-based scheme, wherein nodes are classified into various levels. A node that receives messages at a particular level is raised to next level and in this process the number of nodes are reduced to half from a level to the next level. This scheme works best when node uses CDMA technology wherein every level transmission takes place simultaneously.

4. **Chain-based Three-level Scheme**

Chain-based three-level scheme works best for non-CDMA based technology. This scheme uses chain similar to PEGASIS. However, the chain is segregated into several groups to reduce the interference. In a group, a node is elected leader. This leader collects data from different nodes which transmit data one at a time. There after, the leader raises to next level. In second level, two groups of nodes are formed and in third level, messages are exchanged between a node from every group of the second level. At last, a single message is transmitted to the base station by the leader node.

Q13. Write about data fusion and discuss how is it different from data integration.

Answer :

Data fusion is a process of integration of multiple data sources so as to get more consistent, accurate and useful information when compared to the provided data by individual source. It combines several sources of raw data to produce new raw data. They are categorized into following based on the processing stage.

1. Low
2. Intermediate
3. High.

There are different levels in sensor data fusion.

level 0 – Data alignment

level 1 – Entity assessment, Tracking and object detection/recognition/identification

level 2 – Situation assessment

level 3 – Impact assessment

level 4 – Process refinement

level 5 – User refinement.

The main aim of data fusion is to sense different environments and obtain data from multiple distributed sources which is highly reliable and error probability is low.

Comparison with Data Integration

Data fusion combines homogenous data where as data integration combines heterogeneous data. Data integration usually combines data from different sources to generate a unified view. However, it is not concerned about accuracy and consistency as provided by data fusion. Data integration combines data from technical and business processes from different sources to give meaningful and valuable information.

4.2 QUALITY OF A SENSOR NETWORK

Q14. Define coverage and exposure problems and briefly describe some mathematical techniques to solve them.

Answer :

Model Paper-I, Q8(b)

Coverage Problem

The measure of the ability of network to cover an event is referred to as coverage. It is evaluated based on various factors such as sensitivity, range, density and location of sensing nodes. Based on these factors, coverage can be worst-case or best-case at certain nodes. The best-case coverage is referred to the sensing nodes where coverage value is maximum or at its best whereas the worst-case coverage is referred to the sensing nodes where the coverage value is worst or least. By obtaining the worst-case coverage, the need for employing extra nodes can be determined. By obtaining the best-case coverage, Maximum Exposure Path (MEP) can be determined which is nothing but the best path along the best-coverage.

To define the coverage problem, consider an area A carrying a set of sensors $S = \{S_1, S_2, S_3, \ldots S_n\}$. Let the location of any sensor S_i be (x_i, y_i) and I, E be initial point and ending point of an intruder respectively. The problem here is to determine the maximum breach path P_B which is the locus of points P within A.

One technique to handle this problem is to use Voronoi diagram. Using this technique, it can be proved that P_B coincides with the paths of Voronoi diagram. This diagram typically carries convex polygons in 2D view. Within these polygons, the sites residing are located as close as possible to the site enclosing the polygons. Moreover, these polygons are at similar distance from their neighboring sites.

The algorithm associated with this technique involves the steps,

(i) Firstly, a Voronoi diagram is developed by drawing perpendicular lines whose points of intersection act as vertices of convex polygons.

(ii) Secondly, a weighted graph is developed where weight of every edge is considered as the least distance value from all sensors. When this value is minimum, better coverage is obtained.

(iii) Breadth first search is used to evaluate the maximum cost path from I to E. i.e., from initial node to ending node.

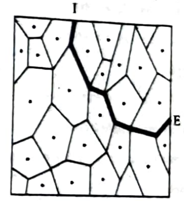

Figure: Voronoi Diagram

The breach path is considered as the maximum cost path whereas the area covered by this path is considered as vulnerable.

Another problem is determining the best-case coverage which can be resolved using Delaunay triangulation technique. In this technique, the maximum support path is identified by providing interconnection between the sites sharing a common edge.

Exposure Problem

In a sensor field, the ability of tracking a specific point is referred to as exposure. In mathematical terms, it can be given as the integral of sensing function with respect to certain from source to destination. It is given by the formula,

$$E[p(t), t_1, t_2] = \int_{t_1}^{t_2} I_{A\,or\,C}(F, p(t)) \left|\frac{dp(t)}{dt}\right| dt$$

Here, the elemental length of the arc is given by $\frac{dp(t)}{dt}$ and time instances are given by t_1, t_2 then the value of $I_{A \text{ or } C}$ can be calculated as,

$$I_A(F, p) = \sum_{i=1}^{n} S(s_i, p)$$

$$I_C(F, p) = S(s_{\min}, p).$$

Here $s_{\min}$ is the sensor which is atleast distance from p.

Exposure problem can be resolved with use of minimum exposure path. If the sensor node is located at point (0, 0), then the MEP is evaluated between (– 1, – 1) and (1, 1). In case if the sensor node is located at certain point (a, b), then an $n \times n$ grid of order m is created.

4.3 REAL-TIME TRAFFIC SUPPORT

Q15. Discuss the real-time traffic support and Qos parameters in ad hoc wireless networks.

Model Paper-II, Q8(b)

Answer :

Real-Time Traffic in Ad Hoc Wireless Networks

Real time applications are divided into two types. They are,

1. Hard real time applications
2. Soft real time applications.

1. **Hard Real-time Applications**

It needs precise QoS guarantees.

Example

Control system of Nuclear reactor, Air traffic control system etc.

2. **Soft Real-time Applications**

It can manage certain amount of degradation in QoS guarantees.

Example

Voice telephony, video conferencing etc.

These applications can generate dangerous results because there are certain critical factors such as loss of data, variation in delay and delay jitter etc.

Hard-real time guarantees are difficult to be imposed because of the mobility of nodes, time varying channel capacity, presence of hidden terminals etc in the ad-hoc networks.

QoS Parameters in Ad Hoc Wireless Networks

QoS parameters and services are different for different applications because of their varied requirements. The following are the examples which uses QoS parameters,

1. In multimedia applications, the major QoS parameter such as bandwidth, delay jitter etc are required whereas security is required in military applications.
2. In the applications which are based on emergency search and resure, the major QoS parameter is the availability of network.
3. In group discussion, the major QoS parameter is battery life which consumes minimum amount of energy for the transmissions among various node.

4.4 SECURITY PROTOCOLS

Q16. Discuss various security attacks that can be made on sensor networks and protocols used to handle them.

Model Paper-4, Q8

Answer :

Security Attacks on Sensor Networks

Various security attacks that can be made on sensor networks are,

1. **Spoofing (Impersonating Address)**

A mode in sensor network can be treated as an address. Hence, all the routes may be blocked when a common mode is used for various paths.

2. **Wormhole Attack**

In wormhole attack, the attacker tunnels the selected packets to some other location different from the location from which it received the packets. The tunnel which collaborates two attackers is called as wormhole and it is generated using a single long-range wireless links or through a wired link. These is a possibility that attackers could create a wormhole even for the packets which do not belong to them. This attack can overcome by making use of some powerful security mechanisms.

3. **Blackhole Attack**

In this type of attack, contaminant node provides incorrect information to the destination node in the path finding process. It suggests the destination node fake shortest paths or stable path. The main aim of the contaminant node is either to stop the path finding process or to obstruct the data packets being forwarded to the destination node.

4. **Hello Flood Attack**

This attack increases the delay of messages to reach its destination. It does so by broadcasting a hello packet from a specific node with high power.

Protocols used to Handle Security Attacks

There are many protocols that are adopted to avoid attacks in sensor networks. The two most prominent among them are LEAP and INSENS.

(i) **LEAP**

LEAP stands for Localized Encryption and Authentication Protocol which is dedicated for the management of keys. It provides sensor nodes to have four different keys.

(a) A single key which can be shared only with the beacon signals.

(b) A common key which is shared by all the nodes in the network.

(c) A cluster key which can be shared among neighbours.

(d) A pairwise key which is also shared with neighbours.

Using these keys attacks on sensor networks can be avoided.

(ii) **INSENS**

INSENS stands for Intrusion tolerant routing in SENS or networks. It exploits the features of routing tables and routing mechanisms to avoid security attacks. It does not eliminate the attacks but can reduce the damage. It carry out its operation in two phases i.e., route discovery and data forwarding.

(a) **Route Discovery**

In this phase, a multi hop forwarding mechanism is used to send a request messages to all the nodes present in the routing table. On receiving the request, nodes forward this message to all its neighbours along with its identity. This eliminates the inclusion of attackers in the routing tables. However, there is still a possibility that malicious nodes can enter the network through spurious messages.

(b) **Data Forwarding**

This phase is responsible for forwarding the data to the respective nodes based on the data present in forwarding tables.

FREQUENTLY ASKED QUESTIONS AND IMPORTANT QUESTIONS

SHORT QUESTIONS

Q1. Write a note on data dissemination.

Ans: For answer refer Unit-IV, Q1.

Important Question

Q2. What is meant by GHT?

Ans: For answer refer Unit-IV, Q3.

Important Question

Q3. Define Exposure Problem.

Ans: For answer refer Unit-IV, Q4.

Important Question

Q4. What is Wormhole Attack?

Ans: For answer refer Unit-IV, Q6.

Important Question

Q5. Write a brief note on INSENS.

Ans: For answer refer Unit-IV, Q8.

Important Question

ESSAY QUESTIONS

Q6. Define the following,

(i) Data dissemination

(ii) Data collection

(iii) Data diffusion

(iv) Flooding

(v) Gossiping

(vi) Rumor routing.

Ans: For answer refer Unit-IV, Q9.

Important Question

Q7. Explain about SPIN Protocol and Cost Field Approach.

Ans: For answer refer Unit-IV, Q11.

Important Question

Q8. What is the objective of Data Gathering? Discuss various algorithms that implement data gathering.

Ans: For answer refer Unit-IV, Q12.

Important Question

Q9. Define coverage and exposure problems and briefly describe some mathematical techniques to solve them.

Ans: For answer refer Unit-IV, Q14.

Important Question

Q10. Discuss the real-time traffic support and Qos parameters in ad hoc wireless networks.

Ans: For answer refer Unit-IV, Q15.

Important Question

Q11. Discuss various security attacks that can be made on sensor networks and protocols used to handle them.

Ans: For answer refer Unit-IV, Q16.

Important Question