Part 4. Routing Protocols

1. Routing for Mobile Ad hoc NETworks (MANETs)

Routing in MANETs is difficult since mobility causes frequent network topology changes and requires more robust and flexible mechanisms to search for and maintain routes. When the network nodes move, the established paths may break and the routing protocols must dynamically search for other feasible routes. With a changing topology, even maintaining connectivity is very difficult. In addition, keeping the routes loop free is more difficult when the hosts move. Besides handling the topology changes, routing protocols in MANETs must deal with other constraints, such as low bandwidth, limited energy consumption, and high error rates, all of which may be inherent in the wireless environment. Furthermore, the possibility of asymmetric links, caused by different power levels among mobile hosts and other factors such as terrain conditions, make routing protocols more complicated.

1.1 Categories of Existing Routing Protocols for MANETs

Many protocols have been proposed for MANETs. These protocols can be divided into three categories: *proactive, reactive, and hybrid*. Proactive methods maintain routes to all nodes, including nodes to which no packets are sent. Such methods react to topology changes, even if no traffic is affected by the changes. They are also called table-driven methods. Reactive methods are based on demand for data transmission. Routes between hosts are determined only when they are explicitly needed to forward packets. Reactive methods are also called on-demand methods. They can significantly reduce routing overhead when the traffic is lightweight and the topology changes less dramatically, since they do not need to update route information periodically and do not need to find and maintain routes on which there is no traffic. Hybrid methods combine proactive and reactive methods to find efficient routes, without much control overhead.

1.2 Proactive Routing Protocols

As stated earlier, proactive routing protocols maintain routes to all destinations, regardless of whether or not these routes are needed. In order to maintain correct route information, a node must periodically send control messages. Therefore, proactive routing protocols may waste bandwidth since control messages are sent out unnecessarily when there is no data traffic. The main advantage of this category of protocols is that hosts can quickly obtain route information and quickly establish a session. For example, GSR introduced below is a proactive routing protocol.

Global State Routing (GSR) is based on the Link State (LS) routing method. In the LS routing method, each node floods the link state information into the whole network (global flooding) once it realises that links change between itself and its neighbours. The link state information includes the delay to each of its neighbours. A node will know the whole topology when it obtains all link information. LS routing works well in networks

with static topologies. When links change quickly, however, frequent global flooding will inevitably lead to huge control overhead.

Unlike the traditional LS method, GSR does not flood the link state packets. Instead, every node maintains the link state table based on up-to-date LS information received from neighbouring nodes, and periodically exchanges its LS information with its neighbours only (no global flooding). Before sending an LS packet, a node assigns the LS packet a unique sequence number to identify the newest LS information. LS information is disseminated as the LS packets with larger sequence numbers replace the ones with smaller sequence numbers.

The convergence time required to detect a link change in GSR is shorter than in the Distributed Bellman-Ford (DBF) protocol. The convergence time in GSR is O(D*I) where D is the diameter of the network and I is the link state update interval. The convergence time is normally smaller than O(N*I) in DBF, where N is the number of nodes in the networks and I is the update interval. Since the global topology is maintained in every node, preventing routing loops is simple and easy.

The drawbacks of GSR are the large size of the update messages, which consume a considerable amount of bandwidth, and the latency of the LS information propagation, which depends on the LS information update interval time. ``Fisheye" technology can be used to reduce the size of update messages. In this case, every node maintains highly accurate network information about the immediate neighbouring nodes, with progressively fewer details about farther nodes.

1.3 Reactive Routing Protocols

Reactive routing protocols can dramatically reduce routing overhead because they do not need to search for and maintain the routes on which there is no data traffic. This property is very appealing in the resource-limited environment.

1.3.1 Dynamic Source Routing (DSR)

The Dynamic Source Routing (DSR) protocol uses the source routing approach (every data packet carries the whole path information in its header) to forward packets. Before a source node sends data packets, it must know the total path to the destination. Otherwise, it will initiate a route discovery phase by flooding a Route REQuest (RREQ) message. The RREQ message carries the sequence of hops it passed through in the message header. Any nodes that have received the same RREQ message will not broadcast it again. Once an RREQ message reaches the destination node, the destination node will reply with a Route REPly (RREP) packet to the source. The RREP packet will carry the path information obtained from the RREQ packet. When the RREP packet traverses backward to the source, the source and all traversed nodes will know the route to the destination. Each node uses a route cache to record the complete route to desired destinations.

Route failure is detected by the failure of message transmissions. Such a failure will initiate a route error message to the source. When the source and the intermediate nodes receive the error message, they will erase all the paths that use the broken link from their route cache.

The path calculated in DSR is loop-free since loops can be detected easily and erased by the source routing. A few optimisations are proposed for DSR. For example, a flooded route query can be quenched early by having any non-destination node reply to the query if that node already knows a route to the desired destination; the routes can be refreshed and improved by having nodes promiscuously listen to the conversations between other neighbouring nodes.

DSR is simple and loop-free. However, it may waste bandwidth if every data packet carries the entire path information. The response time may be large since the source node must wait for a successful RREP if no routing information to the intended destination is available. In addition, if the destination is unreachable from the source node due to a network partition, the source node will continue to send RREQ messages, possibly congesting the network.

1.3.2 Ad hoc On-Demand Distance Vector (AODV) Routing

Since DSR includes the entire route information in the data packet header, it may waste bandwidth and degrade performance, especially when the data contents in a packet are small. Ad hoc On-Demand Distance Vector (AODV) Routing tries to improve performance by keeping the routing information in each node. The main difference between AODV and DSR is that DSR uses source routing while AODV uses forwarding tables at each node. In AODV, the route is calculated hop by hop. Therefore, the data packet need not include the total path.

The route discovery mechanism in AODV is very similar to that in DSR. In AODV, any node will establish a reverse path pointing toward the source when it receives an RREQ packet. When the desired destination or an intermediate node has a fresh route (based on the destination sequence number) to the destination, the destination/intermediate node responds by sending a route reply (RREP) packet back to the source node using the reverse path established when the RREQ was forwarded. When a node receives the RREP, it establishes a forward path pointing to the destination. The path from the source to the destination is established when the source receives the RREP.

[Example here]

Dealing with path failures in AODV is more complicated than in DSR. When a node detects the link failure to its next hop, it propagates a link failure notification message (an RREP with a very large hop count value to the destination) to each of its active upstream neighbours to inform them to erase that part of the route. These nodes in turn propagate the link failure notification message to their upstream neighbours, and so on, until the

source node is reached. A neighbour is considered active for a route entry if the neighbour sends a packet, which was forwarded using that entry, within the active_route_timeout interval. Note that the link failure notification message will also update the destination sequence number. When the source node receives the link failure notification message, it will re-initiate a route discovery for the destination if a route is still needed. A new destination sequence number is used to prevent routing loops formed by the entangling of stale and newly established paths.

AODV saves bandwidth and performs well in a large MANET since a data packet does not carry the whole path information. As in DSR, the response time may be large if the source node's routing table has no entry to the destination and thus must discover a path before message transmission. Furthermore, the same problems exist as in DSR when network partitions occur.

1.4 Location-Aided Routing (LAR)

The Location-Aided Routing (LAR) protocol is an on-demand scheme. It utilises location information to limit the route query flooding area. The prerequisite is that every host knows its own location and the global time, which can be provided by a Global Positioning System (GPS).

LAR defines the concepts of ``expected zone" and ``request zone."

[Examples here]

When node S wants to send messages to node D, it will broadcast a route query message, which is forwarded only by the nodes in the ``request zone." When a node forwards the route query, it appends its node ID to the head of the packet. After node D finally receives the route query, it sends a route reply back to the source node S using the reverse path which is recorded in the head of the route query packet. The route from S to D is established when the source node S receives the route reply packet.

LAR can efficiently reduce the RREQ flooding cost. The main problem with this method is that obtaining accurate location information may be difficult in some environments (for example, GPS does not work well indoors, and proximity does not guarantee connectivity).

1.5 Hybrid Routing Protocols

A typical hybrid routing protocol is Zone Based Routing (ZBR). ZBR combines the proactive and reactive routing approaches. It divides the network into routing zones. The routing zone of a node X includes all nodes within hop distance at most d from node X. All nodes at hop distance exactly d are said to be the peripheral nodes of node X's routing zone. The parameter d is the zone radius. ZBR proactively maintains the routes within the routing zones and reactively searches for routes to destinations beyond a node's routing zone. Route discovery is similar to that in DSR with the difference that route requests are

propagated only via peripheral nodes. ZBR can be dynamically configured to a particular network through adjustment of the parameter d.

ZBR will be a purely reactive routing protocol when d = 0 and a purely proactive routing protocol when d is set to the diameter of the network.

ZBR discovers routes as follows. When a source node wants to send data to a destination, it first checks whether or not the destination is within its routing zone. If it is, then a route can be obtained directly. Otherwise, it floods a route request to its peripheral nodes. The peripheral nodes in turn execute the same algorithm to check whether the destination is within their routing zone. If it is, a route reply message is sent back to the source. Otherwise, the peripheral node floods the route request to its peripheral nodes again. This procedure is repeated until a route is found.

1.6 Proactive vs. Reactive vs. Hybrid Routing

The tradeoffs between proactive and reactive routing strategies are quite complex. Which approach is better depends on many factors, such as the size of the network, the mobility, the data traffic and so on. Proactive routing protocols try to maintain routes to all possible destinations, regardless of whether or not they are needed. Routing information is constantly propagated and maintained. In contrast, reactive routing protocols initiate route discovery on the demand of data traffic. Routes are needed only to those desired destinations. This routing approach can dramatically reduce routing overhead when a network is relatively static and the active traffic is light. However, the source node has to wait until a route to the destination can be discovered, increasing the response time.

The hybrid routing approach can adjust its routing strategies according to a network's characteristics and thus provides an attractive method for routing in MANETs. However, a network's characteristics, such as the mobility pattern and the traffic pattern, can be expected to be dynamic. The related information is very difficult to obtain and maintain. This complexity makes dynamically adjusting routing strategies hard to implement.

2 Mobile IP/Micro Mobility

C. Perkins has two very good tutorial papers for mobile IP (check the reading list). Cellular IP is proposed by A.T. Campbell's group at Columbia University.

Relationship between Mobile IP and Cellular IP: Mobile IP can efficiently provide mobility support on a global scale, for instance migrations between LANs. Cellular IP is optimized to support host mobility in a Cellular Wireless Access Network. Cellular IP shows great benefit in comparison to existing host mobility proposals for environments where mobile hosts migrate frequently. It can interwork with Mobile IP to support migrations between Cellular IP Access Networks. In short, Cellular IP can be thought of as a protocol supporting micro mobility while mobile IP is for macro mobility.

3 Multicast Routing

Compared to unicast routing protocols, there are relatively few multicast routing protocols. Note that supporting multicast service is not easy even in wired networks.

Group management will be very challenging in mobile environment. The cost involved in dynamic group management may make most proposals infeasible for wireless mobile networks.