Exploring the Challenges in MANETs

Hardik K. Molia
Master of Computer Engineering, Department of Computer Engineering *Atmiya Institute of Technology & Science, Rajkot- 360005, Gujarat, India*
hardik.molia@gmail.com
Prof. Rashmi Agrawal
Assistant Professor, Department of Computer Engineering *Atmiya Institute of Technology & Science, Rajkot- 360005, Gujarat, India*
rashmi.agrawal@aits.edu.in

----------**ABSTRACT**----------
Mobile Adhoc Networks (MANETs) are infrastructure less, self-configured, self-controlled and self-organized wireless networks. Because of some differences in the way MANETs work, some design issues are required to be handled carefully while implementing them in real life scenarios. This paper is a study paper of some of the most important design challenges for MANETs. This paper explores the challenges with respect to Medium Access, Network and Transport Layers..

Keywords - MANET, Collision, Hidden Station, Exposed Station, Routing Algorithms, Congestion.

I- INTRODUCTION

A MANET -Mobile Adhoc Network, shown in fig.1[9] is a wireless network which has no fixed infrastructure such as conventional wired/wireless networks have. MANETs are self-configured, self-organized, self-controlled, infrastructure less and autonomous in nature. In MANETs, communicating devices play dual roles, they are the hosts as well as they are the routers too. Conventional wireless networks have wireless routers called Access Points.[1] Access points are connected with each other to provide host devices to access the network. In MANETs host devices take part in the routing process. MANETs are becoming popular because of their infrastructure less nature. It is easy to build MANETs in emergency situations for urgent communications. MANETs are also useful for temporary networks. MANETs provides user mobility, users are allowed to be mobile, and they are allowed to change their locations while accessing the network without any disturbance. [1]

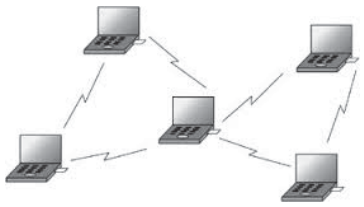


Fig. 1 – Adhoc Wireless Network

MANETs have changed the conventional definitions of the computerized networks and so several design issues are needed to be resolved for effective and efficient setup. A network is based on various layers. Every layer is responsible for providing some specific services for the purpose of effective communication. As the way devices take part in the network has been changed, it is highly required to change the way layers work inside. This is the exactly need of the researching new protocols which are more suitable to work with adhoc environment. This paper shows various issues related with Medium Access Control, Network and Transport layers which provide, node to node, host to host and end to end delivery respectively. This paper also describes some of the protocols which are used to resolve ill effects of such issues.

II- MEDIUM ACCESS CONTROL LAYER

A. Introduction

In wireless networks, nodes use radio signals for communication. Every node has a limited range of transmission. Multi-hop communication allows nodes to communicate which are not presently located in the same transmission and receiving range. The available bandwidth is shared by all the nodes. It is the essential need to avoid, detect and resolve conflicts-collisions across the signals, if any present. Collisions are the results of simultaneous communications performed inside the same transmission-receiving range by more than one communicating pairs. Collisions disturb the data and so subsequently degrade the network performance. MAC layer controls access to the medium to avoid the collisions.

The standard CSMA/CD protocol is not suitable to use because a node may not detect an ongoing communication just because it is out of the range of that communication. This leads to two issues, hidden station problem and exposed station problem.

B. Hidden Station & Exposed Station Problems

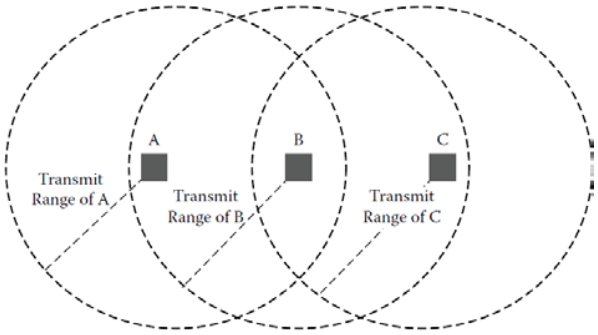


Fig 2- Hidden and Exposed Stations

In Fig 2. [9], Node B is in the range of nodes A and C, but A and C are not in each other's range. When A is transmitting to B. Node C cannot detect a signal and may send data to B, which will cause a collision at B. This is the "hidden-station problem," as nodes A and C are hidden from each other.

When B is transmitting to A, Because C is within B's range, it senses a signal and decides to defer its own transmission to some other node D. This is unnecessary because there is no way C's transmission can cause any collision at receiver A. Here B is exposed station to C. MAC schemes are designed to overcome these problems.

C. MAC Protocols

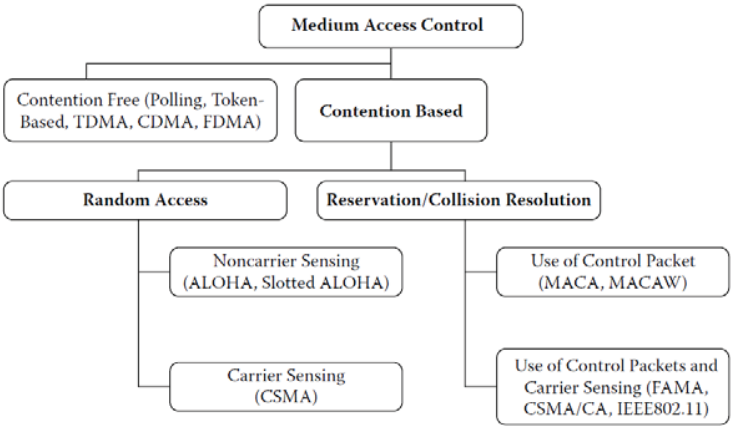


Fig 3- MAC Protocols

In Fig 3. [9], classifies various MAC Protocols. Contention is the ability of a node to initiate the communication whenever it wants. Contention based protocols allow nodes to start communication at any time they find the medium is free to use. Contention free protocols have some time based allocation of available bandwidth – medium in terms of slots. A node can communicate in its concerned slot-turn only. A brief introduction to a dynamic reservation based contention based protocol is given here.

D. Multiple Access Collision Avoidance (MACA)

The MACA overcomes the hidden and exposed station problems. MACA uses two short signalling packets.RTS – Request to Send, CTS – Clear to Send. Sender sends a RTS message to receiver. If receiver is ready for the communication, it sends CTS back to the sender indicating that now sender can start sending the actual data. RTS and CTS also carry approximate time for the completion of data transfer. RTS–CTS exchange enables nearby nodes to reduce the collisions at the receiver only. They don't avoid collisions at sender sides.

In a hidden station problem of Fig 2, C will not hear the RTS sent by A, but it will hear the CTS sent by B. C will defer its transmission during A's transmission. In the exposed station problem of Fig 2., C will hear the RTS sent by B, but not the CTS sent by A. Therefore C will consider itself free to transmit during B's transmission. Both RTS and CTS messages contain the duration of the data transmission. All stations receiving either RTS or CTS will keep silent during the data transmission period.

If two RTS packets collide, each sending node waits for a random interval before trying again. This process continues until one of the RTS transmissions gets the desired CTS from the receiver. RTS and CTS packets are significantly smaller than the actual data packets, and so collisions among them are less expensive compared to collisions among the longer data packets.

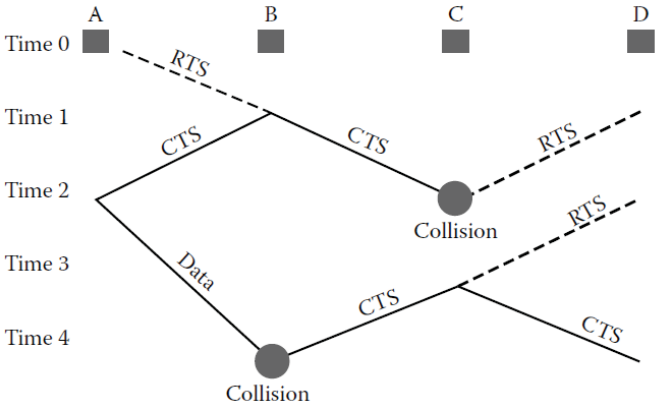


Fig 4- RTS and CTS

In In Fig 4. [9],, there are four nodes, A,B,C and D. Node A sends an RTS to B, and B sends a CTS back to A. At C, however, TS collides with an RTS sent by D. Therefore, C has no knowledge of data transmission from A to B. While the data packet is being transmitted, D sends out another RTS because it did not receive a CTS in its first

attempt. This time, C replies to D with a CTS that collides with the data packet at B. This problem will get solved due to random wait before resending RTS.

II- NETWORK LAYER

E. Introduction

Network layer performs routing of packets. The standard routing protocols developed for wired networks are not efficient and effective in MANETs. The main reasons behind such unsuitability are the dynamic topology feature and network partition issue in MANETs. In wired and infrastructure based wireless networks, routers are connected via a wired link allowing nodes to communicate.[2] In MANETs, every node has to act as a router too because two nodes may not directly communicate due to transmission range limitations. In this scenario, multi hop routing allows nodes to communicate which are out of each other's sensible ranges. Nodes which are not directly connected via layer 2-MAC layer can communicate via layer 3- network using concept of multi hop communication. [3]

F. Dynamic Topology and Network Partitioning

In a wired network, the physical topology changes rarely- infrequently. But as MANETs support user mobility, they inherently support dynamic topologies. New nodes get added, existing nodes get removed and few nodes may change their locations which will change the entire topology frequently. As the topology is not fixed for a fixed amount of duration, routing is very difficult. The conventional routing algorithms don't provide efficient performance under dynamic topologies.

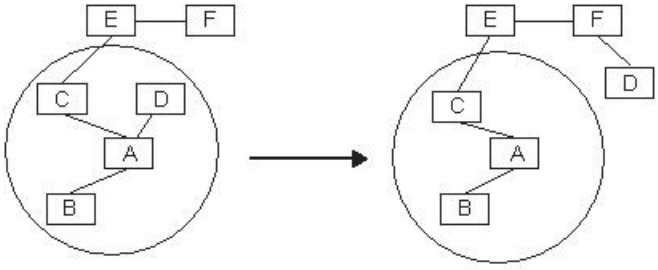


Fig. 5- Node D moves out of Range of A

Fig 5 and 6. [4] shows that sometimes, some nodes in a MANET become down, they shut down or damaged which will subsequently partition the network into two or more halves. In this situation, it is not possible for a node in one halve to communicate with any node in another halve. [2]

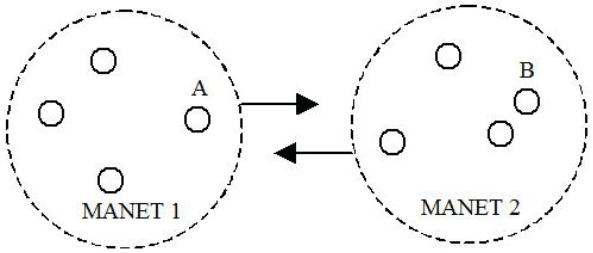


Fig. 6- MANET Partitions

Mobile nodes have limited power capacity, processing capacity and memory capacity so routing algorithms should be as light weighted as possible.

G. Design Issues

The Routing architecture can be either flat or hierarchical. Flat routing holds information about every other node in the network irrespective of their positions. For a smaller network, this architecture provides good performance, but it becomes unstable due to processing overhead as number of nodes increases. In hierarchical architecture, the entire MANET is divided geographically into set of clusters. Every cluster is assigned a set of nodes in that area and one of them is appointed as a cluster heads. Routing is performed across the cluster heads only. This architecture lets some devices don't take part in routing decisions.[3]

In general sense, we consider the route which is smallest is the best one. But from the quality of service aspects, hope count should not be the only criteria to measure quality of a route. Some other criteria like, security, stability, congestion condition should be considered while choosing a route to perform communication.

H. Routing Protocols

The limited resources (processing, memory, power) make design of routing protocols very difficult to design. The conventional methods of routing are distance vector (every node keeps information about next nodes and total distance to every node) and link state (every node keeps information about entire network). Both the concepts are required to be modified to adopt dynamic changes in the topologies. Routing algorithms are classified based on how routing information is acquired and maintained by the mobile nodes. [4]

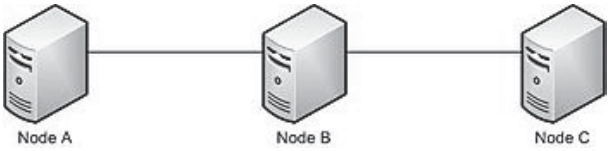
In proactive – table driven – global routing algorithms, every node continuously analyses all the nodes to which it can connect. So every node has complete information about to whom it can communicate at every time. The algorithm requires large amount of processing power, power consumption as well memory requirement. But it provides faster route while initiating a communication. WRP – Wireless Routing Protocol, DSDV – Destination Sequenced Distance Vector Routing Protocol are examples of proactive routing protocols in MANETs.

In reactive – on demand routing algorithms, routing paths are searched only when needed. Route discovery

performs using route determination procedure. Reactive protocols are more scalable and suitable for large networks. Nodes may suffer from large delay while discovery is performed. DSR – Dynamic Source Routing, Adhoc On-Demand Distance Vector – AODV are examples of reactive routing protocols in MANETs.

I. DSDV- Destination Sequenced Distance Vector Routing

DSDV routing protocol is a partially proactive, bellman ford algorithm based protocol. The standard DV – Distance Vector based protocol, RIP - Routing Information Protocol is based on finding shortest path among source node and destination node. RIP suffers from count to infinity and loop problems. In MANETs, improvised DSDV is used to avoid such issues. Every node's routing table stores all available destinations, the next node to reach to destination and the number of hopes to reach the destination. DSDV propagates the changes periodically or update- event based to all the neighbour nodes.



Destination	Next Hop	Number of Hops	Sequence Number	Install Time
A	A	0	A 46	001000
B	B	1	B 36	001200
C	B	2	C 28	001500

Fig 7- DSDV Table

To avoid loop problems, every destination assigns a sequence number to the update information of another node along with the time stamp. Even numbered updates are considered as alive (present) and odd numbered updates are considered as dead (not present). Node increments its own sequence number by 2 with every advertisement and by 1 for an unreachable node (time out basis) and sends the routing information to neighbours as advertised message. Advertised information is compared to own routing table.

1. Select route with higher destination sequence number (This ensure to use always newest information from destination)
2. Select the route with better metric when sequence numbers are equal.

Suppose in above scenario node B finds that the node C is dead because of time out issue, it increments Sequence number of C by 1 making it even. When node A will advertise its own routing table to node B, B will find an entry for node C with the older sequence number. As node selects the information of higher sequence number, B continues with considering C as dead. This logic prevents count-to-infinity problem.

DSDV responses to the change in the topology in two ways, immediate advertisements and Full/Incremental Update

1. Immediate advertisements: - Information on new Routes, broken Links, and metric change is immediately propagated to neighbors.

2. Full/Incremental Update:-
   Full Update:-Send all routing information
   Incremental Update: -Send only changed information.

J. AODV- Adhoc On Demand Distance Vector Routing

AODV routing protocol is reactive, on demand protocol. AODV is based on route discovery process to find a route. Sequence numbers are used for loop prevention and as route freshness criteria. As the protocol serve as an on demand protocol, it reduces unnecessary sharing of routing information which is a disadvantage of DSDV. At the same time, nodes may have to wait before they start transmission due to the delay of route discovery phase. AODV maintains routes as long as they are carrying on going communication.

Every node increases its own sequence number with change in the neighbourhood topology, when two neighbouring nodes enter into communication ranges of each other, as well as when two neighbouring nodes drift out from each other's communication ranges.

Source node S begins Route discovery process with the creation of a Route Request (RREQ) packet. RREQ packet contains source node's IP address, source node's current sequence number, destination IP address, destination sequence number (previously known if any). S broadcasts RREQ packet to all its neighbouring nodes for further process and subsequently every neighbour will broadcast RREQ packet to all of its neighbours. This process is known as flooding. Every node saves the predecessor from which it has received the RREQ to reply back to the original sender.

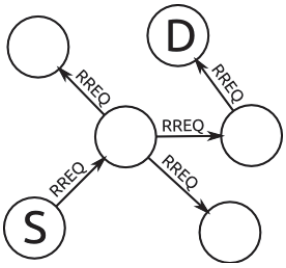


Fig. 8 RREQ Processing

Fig 8., fig 9[4] show that When the RREQ packet arrives at the destination, a Route-Reply (RREP) packet is sent back to the source node along the path that the RREQ has taken through the network as shown in figures.

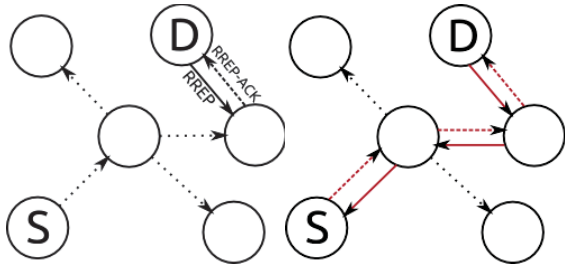


Fig – 9 RREP Processing

Route Requests are initially sent with small Time-to-Live (TTL) field, to limit their propagation.

A node may receivemultiple RREP for a given destination from more than one neighbors. AODV can be configured to take one of the following actions in this situation,

1. The node only forwards the first RREP it receives.
2. May forward another RREP if that has greater destination sequence number or a smaller hop count.
3. Rest are discarded. reduces the number of RREP propagating towards the source

IV- TRANSPORT LAYER

The TCP – Transmission Control Protocol at Transport Layer in TCP/IP protocol suite allows bidirectional communications between two end processes, guaranteeing that all the messages are delivered correctly and arrived in order. TCP supports process to process communication, stream communication, flow control and error control. TCP also handles congestion control without the need for a central control, which is very important in a large decentralized network such as the Internet. Standard TCP considers packet loss as a part of congestion only.[4] But the packet loss may occur due to random error loss caused by channel noise and interference. Because of the weaknesses of wireless medium of communication, error loss frequently disturbs the data transfer in wireless networks. Standard TCP is not able to distinguish between the congestion loss and the error loss, which degrades the overall performance because of unnecessary slow down of the sending rate even when actually not needed. As the wireless mediums introduce more errors because of interference, environmental losses, noise, attenuation, cross talks, it is highly required to distinguish congestion loss from the channel loss.[4]

K. Congestion and QoS

Congestion control and quality of service are inter related issues, improving one means improving the other and ignoring one means ignoring the other. Techniques used to prevent or remove congestion improve the quality of service in a network.

The main focus of congestion control and quality of service is data traffic. In congestion control we try to avoid traffic congestion in terms of reducing overflow at intermediate nodes (routers). In quality of service, we try to create a healthy environment for the traffic.

Congestion may occur when the load, number of packets sent to the network-is greater than the capacity of the network. Congestion control refers to the mechanisms and techniques to control the congestion used to keep load below capacity. Congestion occurs because routers and switches have queues buffers that store the packets before and after processing. A router has an input queue and an output queue for each interface. When a packet arrives at the incoming interface, it undergoes three steps before departing.

1. The packet is put at the end of the input queue.
2. The processing module removes the front packet from the input queue and uses its routing table and the destination address to find the route for it.
3. The packet is put in the appropriate output queue and waits its turn to be sent.

If the rate of packet arrival is higher than the packet processing rate, the input queues become longer. If the packet departure rate is less than the packet processing rate, the output queues become longer. These two possibilities may introduce congestion in the network.

L. Congestion Control with TCP Variants

Various TCP variants have been designed to make congestion control efficient and effective. Some of them are Tahoe, Reno, New Reno, SACK, FACK, Cubic, Vegas, and Compound. TCP's congestion control has three phases: slow start, congestion avoidance, and congestion detection.

TCP works by using a byte oriented sliding window concept. The actual sliding window size is minimum of the congestion window size (cwnd) – determined by examining the patterns of acknowledgement and time outs and advertised windows size (rwnd) – size of window at receiver which was sent by the receiver as a part of flow control.

1. Slow Start: Exponential Increase:-

Initially, size of the congestion window (cwnd) starts with one maximum segment size (MSS). The sender keeps track of a variable named ssthresh (slow-start threshold). When the size of window reaches the threshold, slow start stops and the next phase starts. In the slow-start algorithm, the size of the congestion window increases exponentially after every round 1,2,4,8,…, after every ACK, 1,2,3,4,5…. until it reaches a threshold.

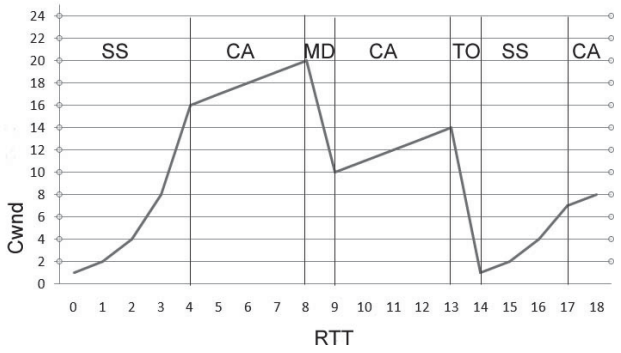


Fig – 10 TCP Congestion Control

SS-Slow Start, CA-Congestion Avoidance

TO-Time Out, FT- Fast Retransmission, MD-Multiplicative Decrease

2. Congestion Avoidance: Additive Increase

To avoid congestion before it happens, one must slow down this exponential growth by additive increase instead of an exponential increase. When the size of the congestion window reaches the slow-start threshold, the slow-start phase stops and the additive phase begins. In this algorithm, each time the whole window of segments is acknowledged (one round), the size of the congestion window is increased by 1. In the congestion avoidance algorithm, the size of the congestion window increases additively after every round 1,2,3,4,5…until congestion is detected.

3. Congestion Detection: Multiplicative Decrease

If congestion occurs, the congestion window size must be decreased. Retransmission can occur in one of two cases: when a timer times out or when three ACKs are received. In both cases, the size of the threshold is dropped to one-half, a multiplicative decrease. Most TCP implementations have two reactions:

I. If a time-out occurs, there is a stronger possibility of congestion; a segment has probably been dropped in the network, and there is no news about the sent segments.

In this case TCP reacts strongly:
a. It sets the value of the threshold to one-half of the current window size.
b. It sets cwnd to the size of one segment.
c. It starts the slow-start phase again.

If three ACKs are received, there is a weaker possibility of congestion; a segment may have been dropped, but some segments after that may have arrived safely since three ACKs are received. This is called fast transmission and fast recovery. In this case, TCP has a weaker reaction:

a. It sets the value of the threshold to one-half of the current window size.
b. It sets cwnd to the value of the threshold (some implementations add three segment sizes to the threshold).
c. It starts the congestion avoidance phase.

M. TCP for MANETs

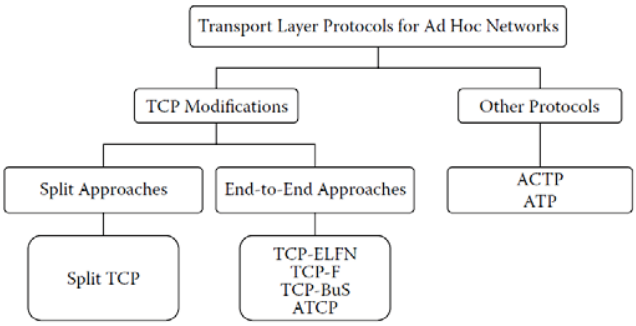


Fig – 11 TCP Variants of MANETs

1. Split TCP:-Large TCP connections that have a large number of hops suffer from frequent route failures due to mobility and network partitions. The Split TCP scheme was introduced to split long TCP connections into shorter localized segments.

2. TCP Feedback (TCP-F):- Routers may fail during the communication. Standard TCP may misinterpret the packet loss due to router failure as a part of congestion. Subsequently sender slowdowns the rate as well as starts retransmission. As soon as an intermediate node detects a broken route, it explicitly sends a Route Failure Notification (RFN) packet to the sender and records this event. Upon reception of the RFN, the sender goes into the snooze state, in which the sender completely stops sending further packets and freezes all of its timers and the values of state variables such as RTO and congestion window

size. Meanwhile, all upstream intermediate nodes that receive the RFN invalidate the particular route to avoid further packet losses. The sender remains in the snooze state until it is notified of the restoration of the route through a Route Reestablishment Notification (RRN) packet from an intermediate node.

V- CONCLUSION

MANETs provide access to the network from "anytime, anywhere" to "all the time, everywhere". Due to easy setup and less infrastructure requirements, MANETs are becoming popular rapidly. Other than various challenges explored in the paper based on the layers, some more challenges resist effectiveness and efficiency of MANETs. Few of them are listed below.

1. Limited bandwidth and wireless communication issues.
2. Low power devices.
3. Limited computational capabilities with some devices like mobile phones.
4. Limited memory.
5. Security issues.

Lots of sub fields are under research to make MANETs more and more efficient, effective and secure.

REFERENCES

[1]. Priyanka Goyal, Vinti Parmar and Rahul Rishi, *"MANET: Vulnerabilities, Challenges, Attacks, Application"*, IJCEM International Journal of Computational Engineering & Management, Vol. 11, January 2011.

[2] Mohseni, S.; Hassan, R.; Patel, A.; Razali, R, *"Comparative review study of reactive and proactive routing protocols in MANETs"*, 4th IEEE International Conference on Digital Ecosystems and Technologies, 304-309, 2010.

[3] Sunil Taneja and Ashwani Kush, " *A Survey of Routing Protocols in Mobile Ad-Hoc Networks*", International Journal of Innovation, Management and Technology, Vol. 1, No. 3, 279-285, August 2010.

[4] Gary Breed Editorial Director, "*Wireless Ad-Hoc Networks: Basic Concepts*", High Frequency Electronics, March 2007.

[5] C. Perkins, E. Belding-Royer and S. Das, "*Ad-Hoc On-Demand Distance Vector (AODV) Routing*", RFC3561, July 003.

[6] Humayun Bakht, " *Survey of Routing Protocols for Mobile Ad-hoc Network*", International Journal of Information and Communication Technology Research, 258-270, October 2011.

[7] Hongmei Deng, Wei Li, and Dharma P. Agrawal, *"Routing Security in Wireless Ad Hoc Networks"* IEEE Communications Magazine • October 2002

[8] Mohit Kumar and Rashmi Mishra "*An Overview of MANET: History, Challenges and Applications*" , Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3 No. 1 Feb-Mar 2012.

[9] Subir kumar sarkar,, T.G. Basavaraju "*Adhoc Mobile Wireless Networks*" Auerbach Publications, Taylor & Francis Group