CHAPTER 3
CROSS LAYER DESIGN

3.1 PROTOCOL STACK PRINCIPLES

Currently, design of network architectures is based on the layering principle, which provides an attractive tool for designing interoperable systems for fast deployment and efficient implementation. B.Jain and A.Agrawala [46] developed ISO/OSI model to support standardization of network architectures using the layered model. The main concepts motivating layering are the following:

- Each layer performs a subset of the required communication functions.
- Each layer relies on the next lower layer to perform more primitive functions.
- Each layer provides services to the next higher layer.
- Changes in one layer should not require changes in other layers.

Such concepts were used to define a reference protocol stack of seven layers, going from the physical layer (concerned with transmission of an unstructured stream of bits over a communication channel) up to the application layer (providing access to the OSI environment). A protocol at a given layer is implemented by an (software, firmware, or hardware) entity, which communicates with other entities (on other networked systems) implementing the same protocol by Protocol Data Units (PDUs). A PDU is built by payload
(data addressed or generated by an entity at a higher adjacent layer) and header (which contains protocol information). PDU format as well as service definition is specified by the protocol at a given level of the stack.

The same concepts are at the basis of the de-facto standard protocol stack on the Internet, namely the TCP/IP protocol stack [71]. The main advantage derived from the layering paradigm is the modularity in protocol design, which enables interoperability and improved design of communication protocols. Moreover, a protocol within a given layer is described in terms of functionalities it offers, while implementation details and internal parameters are hidden to the remainder layers (the so-called “information-hiding” property).

3.2 THE CROSS LAYERING PARADIGM

Standardization of layered protocol stacks has enabled fast development of interoperable systems, but at the same time limited the performance of the overall architecture, due to the lack of coordination among layers. This issue is particularly relevant for wireless networks, where the very physical nature of the transmission medium introduces several performance limitations (including time-varying behavior, limited bandwidth, severe interference and propagation environments). As a consequence, the performance of higher layer protocols (e.g., TCP/IP), historically designed for wired networks, is severely limited.
To overcome such limitations, a modification of the layering paradigm has been proposed, namely, cross layer design, or “cross layering.” The core idea is to maintain the functionalities associated to the original layers, but to allow coordination, interaction and joint optimization of protocols crossing different layers.

Several cross-layering approaches have been proposed in S. Toumpis and A. J. Goldsmith [47], S. Pollin, et al. [48], L. Chen, et al. [49], X. Lin and N. B. Shroff [50].

In general, on the basis of available works on the topic, two approaches to cross-layering can be defined:

- Weak cross-layering enables interaction among entities at different layers of the protocol stack; it thus represents a generalization of the adjacency interaction concept of the layering paradigm to include “non-adjacent” interactions.
- Strong cross-layering enables joint design of the algorithms implemented within any entity at any level of the protocol stack; in this case, individual features related to the different layers can be lost due to the cross-layering optimization. Potentially, strong cross-layer design may provide higher performance at the expense of narrowing the possible deployment scenarios and increasing cost and complexity.
An alternative notation is “evolutionary approach” for the “weak cross-layering” and “revolutionary approach” for the “strong cross-layering” [51].

### 3.3 CROSS LAYER SIGNALING ARCHITECTURES

The large variety of optimization solutions requiring information exchange between two or more layers of the protocol stack raises an important issue concerning implementation of different cross-layer solutions inside TCP/IP protocol reference model, their coexistence and interoperability, requiring the availability of a common cross-layer signaling model [52]. This model defines the implementation principles for the protocol stack entities implementing cross-layer functionalities and provides a standardized way for ease of introduction of cross-layer mechanism inside the protocol stack.

In [53], Raisinghani, et al. define the goals the cross-layer signaling model should follow. They aim at rapid prototyping, portability, and efficient implementation of the cross-layer entities while maintaining minimum impact on TCP/IP modularity. In this framework, several cross-layer signaling architectures have been proposed by the research community. While the following paragraphs will provide an overview and comparison between the most relevant solutions, it is important to note that research on the topic is far from complete. In fact, up to now, just few of cross-layer signaling proposals
were prototyped and none of them is included in the current operating systems.

3.3.1 Interlayer signaling pipe

One of the first approaches used for implementation of cross-layer signaling is revealed by Wang, et al. [54] as interlayer signaling pipe, which allows propagation of signaling messages layer-to-layer along with packet data flow inside the protocol stack in bottom-up or top-down manner. An important property of this signaling method is that signaling information propagates along with the data flow inside the protocol stack and can be associated with a particular packet incoming or outgoing from the protocol stack.

Two methods are considered for encapsulation of signaling information and its propagation along the protocol stack from one layer to another: packet headers or packet structures.

Packet headers can be used as interlayer message carriers. In this case, signaling information included into an optional portion of IPv6 header [55] follows packet processing path and can be accessed by any subsequent layer. One of the main shortcomings of packet headers is in the limitation of signaling to the direction of the packet flow, making it not suitable for cross-layer schemes which require instant communication with the layers located on the opposite direction. Another drawback of packet headers method is in the
associated protocol stack processing overhead, which can be reduced with packet structures method.

Packet structures: In this method, signaling information is inserted into a specific section of the packet structure. Whenever a packet is generated by the protocol stack or successfully received from the network interface, a corresponding packet structure is allocated. This structure includes all the packet related information such as protocol headers and application data as well as internal protocol stack information such as network interface id, socket descriptor, configuration parameters and other.

Consequently, cross-layer signaling information added to the packet structure is fully consistent with packet header signaling method but with reduced processing. Moreover, employment of packet structures does not violate existing functionality of separate layers of the protocol stack. In case the cross-layer signaling is not implemented at a certain layer, this layer simply does not fill / modify the corresponding parts of the packet structure and does not access cross-layer parameters provided by the other layers. Another advantage of packet structure method is that standardization is not required, since the implementation could vary between different solutions.
3.3.2 Direct Interlayer Communication

Q. Wang and M. A. Abu-Rgheff [54] proposed improvement of interlayer signaling pipe method by introducing signaling shortcuts performed out of band. In this way, the proposed Cross-Layer Signaling Shortcuts (CLASS) approach allows non-neighboring layers of the protocol stack to exchange messages, without processing at every adjacent layer, thus allowing fast signaling information delivery to the destination layer. Along with reduced protocol stack processing overhead, CLASS messages are not related to data packets and thus the approach can be used for bidirectional signaling. Nevertheless, the absence of this association is twofold since many cross-layer optimization approaches operate on per-packet basis, i.e. delivering cross-layer information associated with a specific packet traveling inside the protocol stack.

One of the core signaling protocols considered in direct interlayer communication is Internet Control Message Protocol (ICMP) [56, 57]. Generation of ICMP messages is not constrained by a specific protocol layer and can be performed at any layer of the protocol stack. However, signaling with ICMP messages involves operation with heavy protocol headers (IP and ICMP), checksum calculation, and other procedures which increase processing overhead. This motivates a “lightweight” version of signaling protocol CLASS [54] which uses only destination layer identification, type of event, and related to the event data fields.
However, despite the advantages of direct communication between protocol layers and standardized way of signaling, ICMP-based approach is mostly limited by request-response action - while more complicated event-based signaling should be adapted. To this aim, a mechanism which uses callback functions can be employed. This mechanism allows a given protocol layer to register a specific procedure (callback function) with another protocol layer, whose execution is triggered by a specific event at that layer.

### 3.3.3 Central Cross layer Plane

Central Cross-layer Plane implemented in parallel to the protocol stack is probably the most widely proposed cross-layer signaling architecture. In [58] K. Chen, et al. propose a shared database that can be accessed by all layers for obtaining parameters provided by other layers and providing the values of their internal parameters to other layers. This database is an example of passive Central Cross-Layer Plane design: it assists in information exchange between layers but does not implement any active control functions such as tuning internal parameters of the protocol layers.

Similar approach is presented by M. M. Khairy, et al. [59], which introduces a Central Cross-layer Plane called Cross-layer Server to communicate with protocols at different layers by means of Clients. This interface is bidirectional, allowing Cross-layer server to perform active optimization controlling internal to the layer parameters.
Another approach, called ECLAIR, proposed by Raisinghani, et al. in [53] is probably the most detailed from the implementation point of view. ECLAIR implements optimizing subsystem plane, which communicates with the protocol stack by means of cross-layer interfaces called tuning layers. Each tuning layer exports a set of API functions allowing read/write access to the internal protocol control and data structures. These API can be used by protocol optimizers which are the building blocks of the optimizing subsystem plane. This makes the optimizing system a central point for coordination of cross-layer protocol optimizers in order to avoid loops and other conflicts. Similar goals are pursued by Chang, et al. [60] with another architecture falling into Central Cross-Layer Plane category.

It assumes simultaneous operation of multiple cross-layer optimization approaches located at different layers of the protocol stack and aims at coordination of shared data access, avoiding dependency loops, as well as reduction of the overhead associated with cross-layer signaling. To this aim, an Interaction Control Middleware plane is introduced to provide coordination among all the registered cross-layer optimizers implemented in different layers. The main difference of this cross-layer architecture proposal with other proposals of this category is that signaling information propagates along the protocol stack with regular data packets, making it a unique combination of Central Control Plane and interlayer signaling pipe approaches.
3.3.4 Network-wide Cross Layer Signaling

Most of the above proposals aim at defining cross-layer signaling between different layers belonging to the protocol stack of a single node. However, several optimization proposals exist which perform cross-layer optimization based on the information obtained at different protocol layers of distributed network nodes. This corresponds to network-wide propagation of cross-layer signaling information, which adds another degree of freedom in how cross-layer signaling can be performed.

Among the methods overviewed above, packet headers and ICMP messages can be considered as good candidates. Their advantages, underlined in the single-node protocol stack scenario, become more significant for network-wide communication. For example, the way of encapsulating cross-layer signaling data into optional fields of the protocol headers almost does not produce any additional overhead and keeps an association of signaling information with a specific packet. However, this method limits propagation of signaling information to packet paths in the network. For that reason, it is desirable to combine packet headers signaling with ICMP messages, which are well suited for explicit communication between network nodes.

One of the early examples of cross-network cross-layering is the Explicit Congestion Notification (ECN) presented in [68]. It realizes in-band signaling approach by marking in-transit TCP data packet with
congestion notification bit. However, due to the limitation of signaling propagation to the packet paths, this notification needs to propagate to the receiver first, which echoes it back in the TCP ACK packet outgoing to the sender node. This unnecessary signaling loop can be avoided with explicit ICMP packets signaling. However, it requires traffic generation capabilities form network routers and it consumes bandwidth resources.

An example of adaptation of Central Cross-Layer Plane-like architecture to the cross-network cross-layer signaling is presented in [61]. The chapter suggests the use of a network service which collects parameters related to the wireless channel located at the link and physical layers, and then provides them to adaptive mobile applications.

A unique combination of local and network-wide cross-layer signaling approaches called Cross-Talk is presented in [62]. Cross-Talk architecture consists of two cross-layer optimization planes. One is responsible for organization of cross-layer information exchange between protocol layers of the local protocol stack and their coordination. Another plane is responsible for network-wide coordination: it aggregates cross-layer information provided by the local plane and serves as an interface for cross-layer signaling over the network. Most of the signaling is performed in-band using packet headers method, making it accessible not only at the end host but at
the network routers as well. Cross-layer information received from the network is aggregated and then can be considered for optimization of local protocol stack operation based on the global network conditions.

Main problems associated with deployment of cross-layer signaling over the network, also pointed in [72], include security issues, problems with non-conformant routers, and processing efficiency. Security considerations require the design of proper protective mechanism avoiding protocol attacks attempted by non-friendly network nodes by providing incorrect cross-layer information in order to trigger certain behavior. The second problem addresses misbehavior of network routers. It is pointed out that, in most of the cases, IP packets with unknown options are dropped in the network or by the receiver protocol stack.

Finally, the problem with processing efficiency is related to the additional costs of the routers’ hardware associated with cross-layer information processing. While it is not an issue for the low-speed links, it becomes relevant for high speeds where most of the routers perform simple decrement of the time to live (TTL) field in order to maintain high packet processing speed. The Fig. 3.1 shows the various approaches of cross layer signaling architecture.
3.4 DESIGN PARAMETERS OF CROSS LAYER ROUTING METRICS IN MCMR WMNS

Routing protocols use routing metrics which actually predict the weight of the link or the path in order to make efficient routing decisions. Parameters related to the design of cross layer routing metrics in MCMR WMNs can be categorized as Basic, Load aware, Interference aware, and QoS. In spite of these, selected path or route must be Isotonic in nature in order to carry out loop free routing. Furthermore, asymmetry of the wireless links (transmission behavior of wireless link is different in different direction), and route stability parameters are very critical in the design of efficient cross layer routing metric in MCMR WMNs. Theses parameters are explained in detail as follows:
Basic, load, interference, and QoS parameters related to physical layer, MAC layer, and network layer of multi hop wireless networks are very multifaceted in nature as compared to wired networks because of shared wireless medium. The characteristics of gateway router, mesh router and mesh client are summarized in Table 1. Basic (path length, packet loss ratio, delay), load (queue size, no. of flows), interference (intra-flow interference, inter-flow interference), and QoS (link capacity, overall throughput, power utilization) parameters related to design of cross layer routing metrics are summarized in Table 2, which not only predict the link quality of the network, but also facilitate in selecting the efficient route in multi-hop wireless network [63-70].

Isotonic aware property of the routing metric is an important design parameter for selecting optimum weight paths and avoiding routing loops. Isotonic property of the routing metric is defined as the order of weights of two paths is preserved if they are connected to a common third path. Isotonic property of the routing metric must be followed to calculate the optimum paths using Dijkstra’s algorithms in multi hop routing scenarios [71]. Fig. 3.2 shows the classification of routing protocols for multi-hop wireless networks.
Multihop Routing Protocols

Proactive (Table Driven)

- Distance Vector
  - DSDV
  - MDSDV

Reactive (On-Demand)

- Link State
  - OLSR
  - M-OLSR

- Source Routing
  - DSR
  - LQSR

- Hop Count
  - AODV
  - MAODV

Hybrid

- TORA
- HWMP

Fig. 3.2 Routing Protocols for Multihop Wireless Networks

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Gateway Router</th>
<th>Mesh router</th>
<th>Mesh client</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Topology</td>
<td>Always static</td>
<td>Normally static</td>
<td>Mostly mobile</td>
</tr>
<tr>
<td>Interfaces per Node</td>
<td>One or more</td>
<td>One or more</td>
<td>Normally one</td>
</tr>
<tr>
<td>Mobility model</td>
<td>Static in nature</td>
<td>Minimum Mobility</td>
<td>Highly mobile</td>
</tr>
<tr>
<td>Interference Intensity</td>
<td>Low intensity</td>
<td>High intensity</td>
<td>Low intensity</td>
</tr>
<tr>
<td>Power limitations</td>
<td>No limitations</td>
<td>No limitations</td>
<td>Limited power</td>
</tr>
<tr>
<td>Channel diversity</td>
<td>Present</td>
<td>Present</td>
<td>Not present</td>
</tr>
</tbody>
</table>

Table 1 Routing Characteristics Summary

Asymmetry of wireless Links actually define the propagation behavior of links which is quite different in different directions as
compared to wired links. Disseminate packets normally sent from a source node may successfully be received at the destination node, but the connection may fail when the destination node wants to send replay packets back to source node. This criterion is known as asymmetry of wireless link [72]. Asymmetry of the wireless links may reach up to 5 to 15% as mentioned by Ganesan, et al. [73]. Hence, asymmetry of link must also be taken into account while developing the routing metrics for WMNs.

**Route stability parameters** affect the overall throughput of the network. Since the overall performance of the network is highly dependent on the route stability parameters which actually minimize the fluctuation of the route after being declared as an efficient one. Frequent path oscillations result in the poor network performance because these frequent changes in path weight cause an increase in the number of route update packets. Route stability mechanism in wireless networks can be achieved by setting a limit of 10% throughput increase over the route which is currently being used by the routing protocol [74].

Design parameters shown in Table 2 reside at different levels of the network. Design of routing metric may consist of one or more parameters. However, it is a very challenging research problem to design an MCMR routing metric so that it will capture all the above mentioned parameters [75]. On the basis of above discussion,
taxonomy of available routing metrics for WMNs is explained in the following section.

3.5 CROSS LAYER ROUTING METRICS FOR MCMR WMNs

High link losses, asymmetric link, and MCMR functionality in WMNs have made the design of routing metric quite challenging. However, quite a good number of routing metrics are designed in the recent years for WMNs.

3.5.1 Expected Transmission Count

Expected transmission count (ETX) is defined as the number of expected transmission plus retransmissions required to successfully deliver a packet over a wireless link [76]. If forward delivery ratio \( d_{\text{fwd}} \), i.e., probability that the packet successfully received at destination node and reverse delivery ratio \( d_{\text{rvs}} \), i.e., probability that acknowledgment of the packet successfully received at source node, then ETX of the link is calculated as

\[
\text{ETX} = \frac{1}{d_{\text{fwd}} \cdot d_{\text{rvs}}} \quad (3.1)
\]

ETX metric has significantly improved performance over minimum hop count routing metric as shown by the test bed results [77]. ETX develops its design foundation on delivery ratios which truly affects the throughput as compared to minimum hop count metric.
Furthermore, ETX takes account of asymmetry of links in a duplex manner by considering the loss ratios.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Parameters</th>
<th>Monitoring</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Path length</td>
<td>Active probing</td>
<td>Per flow</td>
</tr>
<tr>
<td></td>
<td>Packet loss ratio</td>
<td>Active probing</td>
<td>Per link</td>
</tr>
<tr>
<td></td>
<td>Delay</td>
<td>Active probing</td>
<td>Per flow</td>
</tr>
<tr>
<td>Load aware</td>
<td>Queue size</td>
<td>Passive monitoring</td>
<td>Per link</td>
</tr>
<tr>
<td></td>
<td>No. of flows</td>
<td>Passive monitoring</td>
<td>Per link</td>
</tr>
<tr>
<td>Interference aware</td>
<td>Intra-flow interference</td>
<td>Channel diversity</td>
<td>Per flow</td>
</tr>
<tr>
<td></td>
<td>Inter-flow interference</td>
<td>Signal strength</td>
<td>Per node</td>
</tr>
<tr>
<td>QoS</td>
<td>Link capacity</td>
<td>Active probing</td>
<td>Per link</td>
</tr>
<tr>
<td></td>
<td>Overall throughput</td>
<td>Overall QoS</td>
<td>Per network</td>
</tr>
<tr>
<td></td>
<td>Power utilization</td>
<td>Power utilization</td>
<td>Per node</td>
</tr>
</tbody>
</table>

**Table 2 Parameters for Cross Layer Routing Metric**

The utilization of the spectrum is also minimized by ETX which is helpful in increasing the capacity of the network. ETX was actually designed for single-channel single radio multi-hop wireless networks; so it does not capture the channel diversity in MCMR multi-hop wireless networks. The design of ETX only predicts about the interflow
interference by considering the loss ratios in a static manner, but it
does not have any information about the intra-flow interference faced
by the links. In addition, active probing technique fails to predict the
queuing delay in the network and without load balancing mechanism
ETX may lead the traffic to bottleneck routes in the networks. Active
probing technique incorporated by this metric to capture the loss ratio
may result in underestimation or overestimation of losses because
data packets of IEEE 802.11 real transmission are of different sizes as
compared to probe packets of same size, i.e., 134 bytes [76].

ETX is based on average or mean loss ratio whereas in WMNs
burst losses exist which do not make off well by this routing metric
[78]. In addition, ETX does not take account of the option that
different communication links may possess different transmission
rates which has a critical effect on the network throughput. As
discussed earlier, the poor performance of ETX is mainly due to the
assumption that channel conditions are static in nature, i.e., average
or mean packet loss ratio whereas channel conditions in wireless
networks varies dynamically from time to time. To overcome the
drawbacks of ETX, modified expected number of transmission (mETX)
and effective number of transmission (ENT) are designed on the basis
of link variance in order to make ETX an quality aware routing metric
[78]. mETX is calculated as

\[
m\text{ETX} = \exp \left( \mu + \frac{1}{2} \sigma^2 \right)\]

(3.2)
where $\mu$ is average or mean packet loss ratio and $\sigma^2$ is variance of packet loss ratio. ENT is calculated as

$$\text{ENT} = \exp\left( \mu + \frac{1}{2} \delta \sigma^2 \right)$$  \hspace{1cm} (3.3)

where $\delta$ is the strictness of the loss rate requirement. Although mETX and ENT are improved form of ETX, they still fail to capture the link quality in terms of inter-flow and intra-flow interferences of the route [79]. Furthermore, they compute the losses on the basis of bit error rate which is quite infeasible due to its complex verification mechanism and MAC layer error packet drop mechanism.

### 3.5.2 Expected Transmission Time

IEEE 802.11 MAC layer protocols have multi rate transmission ability which has increased the throughput of the wireless networks significantly [80]. ETX design does not mention that different communication links may have different transmission rates. ETX was developed only by considering the average channel conditions. To solve the problems of ETX, Draves, et al. [81] designed the expected transmission time (ETT), which has significantly enhanced the performance of ETX by measuring the transmission rate of each individual link. ETT is defined as

$$\text{ETT} = \sum_j^n \text{ETT}_j = \sum_j^n \text{ETX}_j \times \left( \frac{S}{B_j} \right)$$  \hspace{1cm} (3.4)
\[
\text{ETT} = \sum_{j=1}^{n} \text{ETT}_j = \sum_{j=1}^{n} \left( \frac{1}{1 - p_j} \right) \times \left( \frac{S}{B_j} \right)
\]

(3.5)

where \(S\) is the packet size, \(p_j\) is the rate of packet loss and \(B_j\) is the transmission rate of link \(j\). The main idea behind the design of ETT metric is the use of multi radios in multi hop wireless networks to enhance the network performance. ETT is the amalgamation of packet loss rate and transmission rate of each individual link. ETT is an enhanced version of ETX with improved performance but still inherit the drawbacks of ETX being unaware of traffic load, intra-flow interference, inter-flow interference and channel diversity in MCMR WMNs. The design of ETT does not capture the losses due to contention caused by the traffic generated by the neighboring nodes. The traffic generated from the neighboring nodes contributes to the losses in two ways. First, it causes increase in collision which definitely increases the packet loss ratio. Secondly, it consumes the channel bandwidth.

Active probing mechanism implemented in the design of ETT to capture the transmission rate may lead to overestimation during the time when the communication channels are quite busy. Although minimum delay (MD) [82] and improved expected transmission time (iETT) [83] are delay based routing metrics, both of them inherit the basic drawbacks of ETT.
3.5.3 Metric of Interference and ChannelSwitching

WMNs have shared medium; hence, intra-flow interference, i.e., interference between nodes on the path of same flow and inter-flow interference, i.e., interference between the nodes, are very critical especially in MCMR networks. Due to interference, the load carrying capacity of the links is affected and the overall performance of the network is degraded. Metric of interference and channel switching (MIC) is developed to capture the interference aware parameters of the links [84]. It deals with both inter-flow and intra-flow interference to predict the quality of the path for efficient routing. Metric of interference and channel switching is based on ETT and is explained as

\[
\text{MIC}_p = \frac{1}{N \times \min \text{(ETT)}} \sum_{l \in \text{path}} \text{IRU}_l + \sum_{i \in \text{path}} \text{CSC}_i
\]  

(3.6)

\[
\text{IRU}_l = \text{ETT}_l + N_l
\]  

(3.7)

where IRU indicates interference aware resource usage that predicts inter-flow interference on the basis of ETT which is the minimum ETT available in the network measured with the help of minimum transmission rate of the interference card, channel switching cost (CSC) that predicts the intra-flow interference and N is the number of neighboring nodes interfered by the link l on the path p. CSC\textsubscript{i} is defined as follows:
\[ CSC_i = \begin{cases} w_1 & \text{if } CH(prev(i)) \neq CH(i) \\ w_2 & \text{if } CH(prev(i)) = CH(i) \end{cases} \]  

(3.8)

\[ 1 \leq w_1 \leq w_2 \]

Where CSC\(_i\) indicates the channel reserved for node \(i\)'s transmission and prev(\(i\)) denotes the previous hop of the node \(i\) through the route \(p\). Thus, CSC can capture the inference only between two successive links. MIC extends ETT by considering the intra and inter-flow interference required in MCMR WMNs but still lacks in load balancing and isotonic characteristics. To make MIC isotonic in nature, decomposition is carried out by transforming the real network into virtual networks which further increases its complexity [85,86]. Moreover, it only measures the interference in a static way which is actually the total number of interfering node that may or may not be causing interference at that time. Thus MIC prefers nodes having less number of neighbors, as a result of which traffic will be routed towards the edges or the boundary of the network [87]. MIC required dynamic information about ETT of each link in the network which introduces the overhead and degrade the performance efficiency of the network.

### 3.5.4 Load Aware Expected Transmission Time

Load aware expected transmission time (LAETT) has incorporated the load balancing and link quality component in ETT to remove the drawback of ETT [88]. Load aware ETT is a combination of ETT and remaining capacity (RC) on the node. RC of link is used as
load aware parameter to balance the traffic on the network. If two paths have same value of ETX, then LAETT will prefer the paths having high value of RC. Loop free route or isotonic nature of the LAETT is due to the fact that it calculates the weights on each link. RC

\[
RC_j = B_j - \sum_{\text{flow } k \in N_j} f_{jk} \cdot \gamma_{jk} 
\]

(3.9)

where \(f_{jk}\) represents the transmission rates of the \(N_j\) flows, i.e., total number of current flows passing through node \(j\), \(B_j\) indicates the transmission rate of node \(j\) and link quality factor of node \(j\) is represented by \(\gamma_{jk}\). LAETT is defined as

\[
\text{LAETT}_{ij} = \text{ETX}_{ij} \times \frac{S}{\left(\frac{RC_i + RC_j}{2 \cdot \gamma_{ij}}\right)} 
\]

(3.10)

where as \(\gamma_{ij}\) is a link quality factor, \(RC_i\) and \(RC_j\) are the RC of the node \(i\) and \(j\), respectively. Practically RC is calculated at layer 2 by measuring the free slots and completed slots provided by the modulation scheme in use. Transmission rate measurements in LAETT are carried out with the help of total number of flows passing across the node and are assumed to be of same data rate. This is actually not true in relation with the wireless networks as the data rates vary because of congestion and interference over the links from time to time. Moreover, different radios and applications utilizing the network have different transmission rate. Probing mechanism used in the
design of LAETT to measure ETX may result in underestimation of the link quality. Furthermore, Equation (3.10) does not predict any information about the intra-flow and inter-flow interference, which is very critical in MCMR environments.

3.5.5 Airtime Cost Routing Metric

IEEE 802.11 is an amendment of standard IEEE 802.11 for WMNs whereas Airtime Cost Routing metric is the default routing metric defined in IEEE 802.11 [89]. It is an interface aware (iAWARE) routing metric developed for communication between different IEEE 802.11 standards. Airtime cost captures the information related to the channel utilization during transmission, e.g., transmission rate, overhead, and frame error rate. It measures the load on each relay node in terms of mean delay faced by the transmission of packets having size equal to 1 kb [90]. Airtime channel cost of a link is calculated as

\[
C_a = \left[ O_{ca} + O_p + \frac{B_t}{r} \right] \times \frac{1}{1 - e_{pt}}
\]

where \(O_{ca}\) represents channel access constant, \(O_p\) indicates protocol overhead, \(B_t\) is the number of bits in the test frame, \(r\) is the node transmission rate in Mbit per second and \(e_{pt}\) frame error rate for the test frame having size equal to \(B_t\). The taxonomy of Airtime Cost routing shows that it is very close to ETT. Actually \((O_{ca} + O_p + B_t/r)\) indicates the transmission time and \(1 / (1 - e_{pt})\) indicates the number
of retransmissions same as ETT. No load balancing mechanism is defined in this metric which may lead the route to congested areas. Airtime metric is unaware of intra-flow interference which has a significant effect on the network performance in MCMR WMNs. Moreover, active probing mechanism to capture the data rate and losses cause overhead in the network depending on the traffic congestion. Therefore, airtime cost metric does not predict the actual quality of the link [91].

### 3.5.6 Interference Aware Routing Metric

iAWARE routing metric addresses the problem of interference in MCMR WMNs by combining the interference ratio (IR) with the ETT metric. Predication of the link quality is based on the measurement carried out with the help of variation in link loss ratio and transmission rate parameters. Physical interference model is used to predict the interference faced by the links over the network using ratio between SINR (Signal to Noise Plus Interference Ratio) and SNR (Signal-to-Noise Ratio) at each node [87,92]. iAWARE is defined as

\[
iAWARE = (1 - \alpha) \sum_{i=1}^{n} iAWARE_i + \alpha \max_{i \leq j \leq k} X_j
\]  \hspace{1cm} (3.12)

where iAWARE measures the inter-flow interference in the network, \(X_j\) predicts the channel diversity and route towards the less intra-flow interference areas, \(k\) is the total number of channels, \(n\) is the total
number of links and \( p \) indicates the path of the network. \( \alpha \) is defined as the trade off parameter to tune between intra-flow and inter-flow interferences of the path. Inter-flow component of interference of the link is calculated as

\[
i\text{AWARE}_i = \frac{\text{ETT}_i}{\text{IR}_i}
\]  

(3.13)

where \( \text{IR}_i \) is the interference ratio of the link \( i \) and is defined as

\[
\text{IR}_i = \frac{\text{SNR}_i}{\text{SNR}_i}
\]  

(3.14)

\( \text{SINR}_i \) and \( \text{SNR}_i \) are calculated by

\[
\text{SINR}_i = \frac{P}{N}
\]  

(3.15)

\[
\text{SNR}_i = \frac{P_i}{N + \sum_{w \in N_i} \tau_w \cdot P_w}
\]  

(3.16)

where as \( P \) indicates the signal strength, \( N \) is the background noise and \( \tau_w \) is the fraction of time period for which node \( w \) makes the channel busy. \( X_j \) in Equation (3.12) can be calculated as

\[
X_j = \sum_{\text{conflicting links on channel } j} i\text{AWARE}_i
\]  

(3.17)

where \( 1 \leq j \leq k \).
Basically iAWARE is non-isotonic in nature like WCETT, thus cannot be used in link state routing protocols, i.e., OLSR. It only predicts the interference on the receiver side where as sender side interference component is also important for quality routes. Moreover, iAWARE has no MAC layer interference measurement mechanism, as it only captures the interference at a node level in terms of ratio between SINR and signal strength $P$ which is being received from other interfering nodes [85]. Lack of load balancing parameters may lead the traffic to congested route. When the value IR$_i$ of the link is greater then ETT$_i$ in Equation (3.13), then the value of iAWARE$_i$ becomes small causing the traffic to route towards the links having small value of ETT but may have higher level of interference causing performance degradation in the network.

### 3.5.7 Interferer Neighbors Count Routing Metric

Interferer neighbors count routing metric (INX) is a radio aware routing metric which actually the improved version of ETX by considering the interference parameters to optimize the radio resource utilization cost [93]. $INX$ of the link is calculated as

$$INX_j = ETX_j \cdot \sum_{k \in N_j} r_k$$  \hspace{1cm} (3.18)

where $N_j$ indicates the number of interfering links resulting from the transmission taking place on link $j$, $r_k$ represents the available transmission rate of the link $k$. Although INX is isotonic in nature, it
performs well only under low load scenarios because no load balancing mechanism is defined in the routing metric. As a result, it faces quick performance degradation as the network load increases. Moreover, it uses a probing technique to measure the interference parameters of the link in a static way which causes an overhead and also fail to predict the true quality of the link. Nevertheless INX behave in a better way as compared to MIC because it follows asymmetric links and isotonic behavior [69].

3.5.8 Resource Aware Routing for MESH

Resource aware routing metric (RARE) has a passive monitoring technique to capture the radio link quality parameters related to load and interference in order to overcome the overhead caused by active probing mechanism [94]. RARE is the combination of bandwidth, contention and signal strength and is defined by the following equation:

\[ \text{RARE}_i = \alpha \cdot \frac{C - \text{BW}_a}{\text{BW}_a} + \beta \cdot \frac{\text{RSSI}_{\text{max}} - \text{RSSI}}{\text{RSSI}} + \gamma \cdot N_c \]  

(3.19)

where \( \text{BW}_a \) is the available bandwidth, RSSI is the RSSI value, \( \text{RSSI}_{\text{max}} \) is the maximum value or RSSI, \( C \) is the link capacity, \( N_c \) is the average contention and \( \alpha, \beta \) and \( \gamma \) are weights associated with bandwidth, RSSI, and contention components, respectively. Available bandwidth component \( \text{BW}_a \) is further defined as
where \( TX_{rate} \) is the transmission rate, \( T_{idle} \) is the idle time interval and \( T_{busy} \) is the busy time interval, respectively, for the calculation of traffic load based on passive monitoring technique. Low overhead RARE can predict the inter-flow interference through contention component \( N_c \) as defined in Equation (3.19) in a passive manner but fail to predict the intra-flow interference and channel diversity in MCMR WMNs. Moreover, a passive measurement does not predict the brusty losses which normally occur in wireless links. As a result RARE may underestimate the link quality of the network. Furthermore, WMNs use Common-off-The-Shelf (COTS) products; so normally their network cards or drivers do not support passive monitoring while transmitting which may result in the performance degradation.

### 3.5.9 Contention Aware Transmission Time

Contention aware transmission time (CATT) routing metric is a load aware and iAWARE routing metric which is actually based on ETT [85]. A key function of CATT is that it predicts location dependent contention and rate diversity of the links. Isotonic behavior of CATT makes it possible to work with link state routing protocols. CATT is calculated as

\[
CATT_i = ETX_i \cdot \sum_{j \in N_i} \left( \left( \sum_{k \in N_j} \frac{R_k}{L_k} \right) \cdot \tau_j \cdot \frac{R_j}{L_j} \right)
\]  

(3.21)
where $N_i$ is total number of links interfering the transmission taking place on the link $i$. Similarly $N_j$ is total number of links interfering the transmission taking place on the link $j$. $R_k$ and $R_j$ indicate the packet size of the links containing 1 and 2 hop neighbors, respectively. $B_k$ and $B_j$ measure the bandwidth of links in 1 and 2 hop neighbors, respectively. $\tau_j$ is defined as packet transmission attempt rate on link $j$. Although CATT captures the inter-flow and intra-flow interferences simultaneously [88]. Like MIC, CATT also assumes that all the neighboring nodes are participating in the inference parameters (weather or not they are involved in transmitting the data) which may overestimate the link quality. Another important drawback in CATT is that it uses active probing mechanism to measure the interference and delay which causes large overhead in the network. Hence, this metric is not suitable for triple play application networks where the traffic is quite congested [95]. Moreover, delay in transmission is used to measure the traffic load which does not predict the load in an accurate way.

### 3.5.10 Interference Load Aware Routing Metric

The interference load aware (ILA) is a hybrid metric based on the load, interfering neighbors, and transmission rates, specially designed for multi-channel WMNs [96]. Path weight of the metric is defined as

$$ ILA_p = \alpha \times \sum_{\text{link} \in p} MTI_i + \sum_{\text{node} \in p} CSC_i $$

(3.22)
where \( p \) is the path in the network, metric of traffic interference (MTI) and CSC which measures the efficiency of flows routed through the path \( p \). These two components of the metric measure the intra flow interference, inter flow interference, transmission rates, congested areas, and packet loss ratios. MTI is the first component which measures the quantity of traffic generated by the interfering neighboring nodes instead of number of interfering neighboring nodes as defined in MIC. MTI is defined as

\[
MTI_i(C) = \begin{cases} 
ETT_{ij}(C) \times AIL_{ij}(C) & \text{if } N_i \neq 0 \\
ETT_{ij}(C) & \text{if } N_i = 0
\end{cases}
\]  

(3.23)

where ETT measures the difference in transmission rate and packet loss ratio of the links in the Network. When node \( i \) and node \( j \) are transmitting over channel \( C \), average interfering load (AIL) is defined as

\[
AIL_{ij}(C) = \frac{\sum_{N_i} IL_{ij}(C)}{N_i(C)}
\]  

(3.24)

where \( N_i(C) \) is the set of interfering neighbor of the node \( i \) and \( j \) and is defined as

\[
N_i(C) = N_i \cup N_j
\]  

(3.25)

where \( IL_{ij}(C) \) is the interference load of the neighbors. CSC which is the second component of ILA, captures the intra-flow interference and is same as defined in Equation (3.8). \( \alpha \) which is a scaling factor to balance the effect of MTI and CSC is defined as
where \( \text{min}(\text{ETT}) \) and \( \text{min}(\text{AIL}) \) is the smallest ETT and average load in the network, respectively. In order to capture the difference in transmission rate, packet loss ratio, intra-flow interference, and inter-flow interference ILA uses an active probing mechanism which induces a large overhead in the network. However, it may not be suitable for congested traffic areas since as it is based on ETX and ETT, it inherits their drawbacks. Exposed node terminal problem causes the interference to occur in two hop range instead of one hop range as consider in ILA and MIC, results in the underestimation of the link quality. Furthermore, ILA does not consider the transmission delay in order to route the traffic efficiently [97].

### 3.5.11 Contention Window Based Routing Metric

Contention window based (CWB) routing metric routes the traffic by considering the channel utilization and average contention window used on the links. CWB is a load-interference aware routing metric which guides the routing protocol to balance the traffic load on the links and to increase the network throughput by routing towards less congested traffic areas [67]. The congestion level and channel utilization at a particular node of a network is given as

\[
\text{CWB} = \beta \cdot \overline{\text{CW}}
\]
In [61] congestion level is measured by the average value of contention window on link CW and is further defined as

\[ CW = \frac{1 - FER}{1 - FER^{r+1}} + \frac{1 - (2 \cdot FER)^{r+1}}{1 - (2 \cdot FER)} \cdot CW_0 \]  

(3.28)

where FER is the measure of frame error rate, CW_0 is the measure of minimum contention window and \( r \) captures the maximum back off stage. Channel utilization component \( \beta \) represents the channel busy time CBT, i.e., amount of time that a channel spends in transmitting, receiving, and occupying states. \( \beta \) is defined as

\[
\beta = \begin{cases} 
1, & \text{if } u \leq T \\
\min \left( \alpha \cdot (u - T_1) + \exp \left( \frac{u - T_1}{T_2 - u} \right), \beta_{\text{max}} \right), & \text{if } T_1 < u < T_2 \\
\beta_{\text{max}}, & \text{if } u \geq T_2 \end{cases}
\]  

(3.29)

So \( \beta \) is equal to 1 when channel utilization is quite small and \( \beta \) is equal to \( \beta_{\text{max}} \) when channel utilization is maximum, whereas \( T_1 \) and \( T_2 \) indicate the minimum and maximum threshold values of the channel utilization represented by \( u \) and the value \( \alpha \) will decide about the change in the value of \( \beta \) as channel utilization \( u \) passes the threshold value \( T_1 \). CWB can only capture the inter-flow interference and traffic load but fail to capture the intraflow interference which is a critical parameter for MCMR WMNs. Furthermore, this metric performs poor when the network conditions change quickly because calculations needed to find out the size of CW are quite sophisticated [69].
3.5.12 Metric for Interference and Channel Diversity

Metric of interference and channel diversity (MIND) captures interference and load aware parameters on the basis of passive monitoring technique thus reducing the overhead usually caused by active probing mechanism to support internet traffic and many other applications like VoIP and video streaming in multi-casting and peer-to-peer service models [47] and is defined as

$$\text{MIND} = \sum_{\text{link}_p}^{n} \text{INTERLOAD}_j + \sum_{\text{node}_p}^{m} \text{CSC}_j \quad (3.30)$$

where INTERLOAD component captures the inter-flow interference and load, CSC captures the intra-flow inference and is same as calculated in Equation (3.8), $n$ measures the total number of wireless links and $m$ indicates the total number of nodes over the path $p$. The INTERLOAD component is defined as

$$\text{INTERLOAD}_j = \text{CBT}_j \cdot \text{IR}_j \cdot \text{CBT}_j \quad (3.31)$$

where $0 \leq \text{IR} \leq 1$ and $0 \leq \text{CBT} \leq 1$. IR is the interference ratio and is used to measure the interference between the links and is calculated as follows:

$$\text{IR}_j = \frac{\text{SINR}_j}{\text{SNR}_j} \quad (3.32)$$

MIND uses a passive mechanism to capture the single strength values in Equation (3.32) through wireless cards. Channel busy time (CBT) is
also calculated passively to predict load by using the following equation:

\[
C_{BT_j} = \frac{\text{Total time} - \text{idle time}}{\text{Total time}} \tag{3.33}
\]

Total Time is the measure of time between the first attempt to send the packet and the reception of its acknowledge. Idle Time is the measure of back off times and the time in which the radio nodes sense that the medium is free for access. Thus CBT is the measure of time spent during transmission, receiving, and occupying states. Overall, MIND captures the inter-flow interference and intra-flow interference in intelligent manner by considering physical and logical interference models. The major limitation of MIND is its non-isotonic nature which induces complexity in its implementation through virtual networks. IR component of MIND is quite different from the IR designed for iARWE because the design principle of MIND focuses on the node parameters whereas design of iARWE is based on link parameters.

Furthermore, MIND does not judge the asymmetry of the links which cause erroneousness in channel quality measurement parameters. MIND considers back off time period as an idle time. Therefore, it may underestimate the channel interference. Unlike MIND, interference-delay aware routing metric for multi-interface mesh networks balances the load by using multi-interface multi-channel capabilities of the node but the interference is measured in a
static manner which actually underestimates the quality of link in the network [97]. Similarly channel utilization and contention window based (CWB) metric is an interference and load aware metric which capture the inter-flow component of interference and congestion but unaware of the intraflow component of interference and channel diversity in multi-interface WMNs [48].

Routing metrics play a critical role in path selection and in route optimization in MCMR WMNs. Routing metrics are composed of set of parameters capture at different layers of the OSI model to predict about the quality of the link, e.g., path length, delay, packet loss ratio, queue size, link capacity, and interference. On the basis of these parameters routing metrics can be classified as simple, interference aware, load aware, and load & interference aware.

**Simple** routing metrics for WMNs utilize transmission rate, packet loss ratio, and delay parameters to capture the link quality of the link, e.g., ETX, mETX, ETT, ENT, MD, and iETT. These routing metrics are simple in design and easy to implement in the routing protocol but they lack in capturing the load and interference aware parameters of the links. Furthermore, these routing metrics are unaware of channel diversity in MCMR scenarios. As a result, these routing metrics lead the traffic towards the congested areas. Hence it is not suitable for efficient routing in MCMR WMNs.
**Interference aware** routing metrics capture the interflow and intra-flow interference parameters along with the transmission rate and packet loss ratio to predicting the quality of the link. Thus, increasing the intelligence behavior of the cross layer routing metrics as interference has a critical effect on the delay and overall throughput of the network in MCMR WMNs. WCETT, MIC, iAWARE, INX and Airtime Cost are the examples of iAWARE routing metrics. Although theses routing metrics perform quiet efficiently as compared to ETX and ENT but still lack in load balancing features. Furthermore, some of them only capture the single component of interference, i.e., inter-flow or intraflow interference although both are mandatory for quality links in real time applications. Non isotonic behavior of some of these routing metrics make their implementation in the routing protocol quite complex because they demand virtual networks to produce loop free routing.

**Load aware** routing metrics capture the traffic concentration and congestion parameters at node level to introduce load awareness in the routing which has a significant effect specially in multi casting and real time applications, i.e., VoIP and Video over IP in community networks. Examples of theses routing metrics are LAETT and WCETT-LB. Since they are based on ETX and ETT, so they inherit their drawback and make the traffic to route towards boundaries of the network. Moreover, they are unaware of inter-flow interference in MCMR WMNs.
Load & interference aware routing metrics, e.g., RARE, CATT. ILA, CWB, MIND, and C2WB are the most recent development in routing metrics as they incorporate the transmission rate, packet loss ratio, congestion, channel diversity, and interference parameters into the quality aware cross layer routing metric for MCMR WMNs. They actually interrelate traffic load and interference in the network and lead the network traffic towards efficient routes. The key benefit of MIND is that it uses a passive monitoring technique to overcome the overhead caused by active monitoring. Moreover, it does not inherit the drawbacks of ETX or ETT as it is not based on them. In spite of these routing metrics, cross layer routing metric design is still an open research issue in MCMR WMNs for QoS specially in real time applications.