

Cross Layer Signalling and Middleware: A Survey for Inelastic Soft Real-time Applications in MANETs

July 14, 2011

Abstract

This paper provides a review of the different Cross Layer Design and protocol tuning approaches that may be used to meet a growing need to support inelastic soft real-time streams in MANETs. These streams are characterised by critical timing and throughput requirements and low packet loss tolerance levels. Many Cross Layer approaches exist either for provision of QoS to soft real-time streams in static wireless networks or to improve the performance of real and non-real-time transmissions in MANETs. The common ground and lessons learned from these approaches, with a view to the potential provision of much needed support to real-time applications in MANETs, is therefore discussed.

Contents

1	Introduction	1
2	Background	3
2.1	Real-time Applications for MANETs	3
2.2	Cross layer Optimisation Defined	3
3	A Taxonomy of Cross layer Signalling Methods	4
3.1	Network-Wide Signalling	5
3.2	Local Node Signalling	7
4	QoS Control with Protocol Tuning	10
4.1	Network-adaptive Tuning	10
4.2	QoS-adaptive Tuning	12
4.3	Hybrid Network and QoS-adaptive Tuning	14
5	Conclusions	15

1 Introduction

Mobile ad hoc networks (MANETs) are emerging in all sectors as the vision for future communications. This vision has at its basis the belief that a mobile device, whatever its location and speed, should have the ability to connect to the rest of the world. For example, a cellphone user may require access to a video stream while at a distance from a cellular mast. In the military sector, the mobile device may take the form of an aircraft transmitting mission critical video data to a ground unit, on friendly vehicles in the vicinity. The usefulness of such connectivity is not limited to communications and there is growing interest in the transmission of command and control data over ad hoc links, for example in the operations of remote industrial or medical safety-critical devices.

The provision of such services to users is dependant on ability to guarantee a high level of performance or QoS. A MANET has several performance-limiting factors, stemming from the mobility of the infrastructure devices (or nodes) and the nature of the transmission medium. A MANET is a self-configuring wireless network where mobile devices connect to each other, when in range, creating a dynamic and somewhat unpredictable topology over which packets can be forwarded. Such networks may stand-alone or be connected to other wired and static wireless nodes or networks, although these are generally not considered part of the MANET. As a result of node mobility, connections are intermittently set up and torn down and there is also the potential for no end-to-end path to exist at a point in time. A single node will therefore discretely act as transmitter, receiver and router. In the latter case it is referred to as an intermediate node, on the path between packet transmission and receipt.

Problems faced in providing QoS guarantees in wireless networks are extended to ad hoc networks. Radio frequency transmissions, propagated from multiple transmitters into a real world environment, are subject to interference, multipath fading, Doppler effects and shadowing. In this environment, a channel, or the path of a packet from one node to another, varies in quality along its length and over time. With no spatial separation of channels low frequency radio transmissions can also interact with each other. All these conditions are characterised by varying Signal to Interference-plus-Noise Ratios (SINR) that decrease with inter-nodal distances. SINR determines channel and application performance. What results from low SINR is limited and varying available bandwidth, a high frequency of bit errors and packet loss and increased packet latency and jitter.

The traditional, layered protocol stack of wired networks institutes static relationships between modular protocol layers. These relationships are characterised by the encapsulation of functions in higher layer objects so that they are hidden from their underlying sources. Layered approaches to network QoS control perform poorly in in ad hoc networks due to the MANET-specific characteristics that do not appear in static networks [44]:

- Node mobility leading to random and possibly rapid topology changes
- Available throughput that is variable and asymmetric and lower than maximum transmission rates,
- Lack of centralised control
- Limited processing capacity, memory and energy resources.

This creates a challenge in providing guarantees of bounded jitter and latency to applications with real-time (RT) deadlines. However, when layer boundaries are blurred and layer information (representing the quality and availability of channel resources as well as QoS requirements) can be shared and tuned between layers, QoS guarantees can again be provided. Cross layer design, moves away from the oblivious layered approach, introducing layer interdependence.

Many cross layer models have been proposed in the two fields of performance improvement in MANETs and of soft RT (SRT) in wireless networks, with a few straddling both. Proposals concentrate either on the performance of the signal transfer mechanism or on the tuning of specific protocol parameters to improve network performance. The majority of these have been developed in order to meet highly specialised network performance goals such as video quality improvement in spite of changing channel conditions. Caution has therefore been suggested in avoidance of spaghetti design wherein the complexity of a cross layer interaction can reduce its re-usability [27].

The focus of this paper is to provide a survey of recent cross layer proposals in both of these fields, identifying the common ground and learning points garnered from both. This is done with a view to identifying the signalling and protocol tuning methods that can provide necessary QoS guarantees to delay critical SRT applications in MANETs. Key holistic or middleware proposals for cross-layer information exchange are first investigated, beginning with cross-layer implementations that rely on global, network-wide information such as the Contention-aware Admission Control Protocol (CACP) and Dual Carrier Sensing with Parallel Transmission awareness (DSCPT). Then proposals that utilise only local information contained within the node are evaluated, including the Mobile Metropolitan Ad hoc Networks architecture (MobileMAN) and the Efficient Cross Layer Architecture (ECLAIR). A taxonomy of these is created, with a view to their potential to support delay critical SRT.

Aside from architectural approaches to cross-layer design, numerous research proposals have also concentrated on adaptively fine-tuning certain protocol parameters according to QoS and network requirements. The QoS of delay critical SRT applications has a high sensitivity to channel quality changes, hence evaluation of this second group of proposals provides a grounding for reducing the optimisation requirements of a cross-layer model to only those parameters with a strong influence on network performance. Section V thus concludes on the signalling mechanism and tunable parameters that should be incorporated in developing an appropriate optimisation model for delay critical SRT in MANETs.

The rest of the paper is structured as follows, Section II presents the background of the paper, by identifying the requirements of typical RT applications from a MANET scenario. This is followed by a suggested definition of cross layer optimisation. A taxonomy of existing signalling mechanisms with representative examples is developed in Section III, first considering those that use network-wide information in their optimisation and second those that utilise only local nodal information. This is followed, in Section IV, by discussion of the literature on protocol tuning to compensate for failing network conditions or to respond to stringent QoS requirements and of some approaches that propose to do both. The learning points from the models discussed in Section III and the fine-tuning approaches from Section IV and also their usefulness to SRT applications running over MANETs is concluded upon in Section V.

2 Background

2.1 Real-time Applications for MANETs

Timeliness is key to RT flows, for which QoS depends strongly upon deadline achievement and high packet arrival rates. All application processes and transmitted packets can be categorised as RT or non-real-time (NRT). RT processes are time-triggered, based on an internal system schedule or event-triggered by environmental stimuli, and explicitly use global physical completion time constraints to manage their resources. QoS for RT packets is therefore often expressed primarily in terms of deadline achievement. There is no benefit in delivering RT packets early, as this in fact may be detrimental to the system due to the consumption of buffer resources by the storage of early arriving packets. NRT processes may perform computations which satisfy their timing requirements but resource management is not time or constraint driven.

The definition of RT is divided into hard real-time (HRT) and soft real-time (SRT) and the latter has further been subdivided to inelastic and elastic SRT [34]:

- HRT processes have strict end-to-end delay requirements, and late packets are considered unusable. This is because the completion of a related computation after its deadline will impede a systems ability to operate correctly or have a critical impact on the system. Hard deadlines are therefore used in safety critical systems to guarantee no damage to equipment or personnel. HRT packet deadlines must always be realised in order for minimum QoS guarantees to be met, for example all directions for the remote operation of a medical device must arrive on time. HRT systems are also highly loss intolerant. For example in transmission of an important video stream to a military aircraft, loss of part a video frame may be preferable to the loss of an entire frame.
- Elastic SRT processes require constraints on end-to-end delay in light of a computational deadline but can tolerate packet arrival at a suboptimal time to differing degrees. Certain RT applications, such as multimedia streams, may be able to compensate for delayed completion, translating this to a lower level of user service. Soft deadlines are generally used to ensure an optimally efficient rather than fixed reaction to a trigger. A SRT deadline has an explicit best case execution time (BCET) and worst case execution time (WCET), between which the usefulness of the output decreases. A WCET can be missed occasionally, typically with an upper bound on the number of misses within a defined interval. For example VoIP applications have an interval between packet arrivals and buffer these prior to playback to compensate for jitter in the stream.
- Inelastic SRT processes have more stringent delay and jitter requirements and low tolerance to packet loss. In comparison to traditional SRT packets these are subject to a low upper bound on acceptable WCET misses, have a smaller difference between BCET and WCET and require guaranteed throughput. In this way a high level of deadline achievement is stipulated.

HRT transmissions strongly depend on the provision of predictable and bounded network jitter and latency and low packet loss. Flow jitter can result within a node, due to variable queueing and processing delays at multiple protocol layers, as a result of over-subscription of resources. It can also appear during packet transit across the network, as a result of multi-hop multi-path routing. The requirements from a network that will provide HRT support are stringent and wired HRT architectures rely on underutilisation and static management of resources in combination with predetermined fixed routing to provide QoS. Such implementations are highly inflexible and suffer from lack of scalability. When the medium of transmission is changed, from a predictable wireline to a non-deterministic wireless link, the static definition of scheduling and resource management has a negative impact on deadline achievement. A requirement still exists to transmit HRT packets in wireless networks and applications must therefore be modified, treating HRT packets as inelastic SRT.

As evinced by the need for over-provisioning in wired HRT support and the high loss-tolerance required of SRT, the layered network's simple forwarding services do not support good RT performance. Wireless networks have not yet been able to meet the deterministic QoS requirements of inelastic SRT, to provide zero or negligible packet loss and guaranteed fixed deadline achievement. A great amount of research effort has, however, concentrated on elastic SRT support, particularly in the area of wireless multimedia streaming [55], [1], [47], [75], [28], [11], [41], [60]. The support of inelastic SRT applications in MANETs is still an open research issue therefore this survey will investigate the predominant approaches to elastic SRT support in MANETs, evaluating the possibility of implementing these for inelastic SRT applications.

2.2 Cross layer Optimisation Defined

It has been widely concluded that layered, OSI-type architectures perform poorly with a wireless physical layer [71], [22], [65]. This performance deteriorates with the independent mobility and intermittent connectivity of nodes in

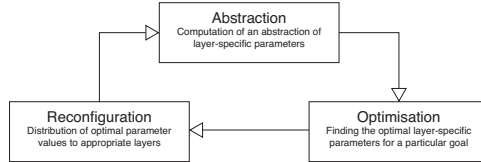


Figure 1: Stages of Cross layer Interaction defined in [28]

MANETs [10], [66], [33]. The practice of withholding internal layer parameters from other layers facilitates the fast development of interoperable systems. Conversely, this limits performance, due to a lack of coordination between the efforts of coexisting protocol layers. Oblivious layers are unable to distinguish between possible causes of packet losses and errors, or to estimate and fairly distribute available end-to-end bandwidth. Under unreliable conditions, oblivious layer use results in the inefficient use of network resources and duplication of efforts at multiple layers. For example, data transmission rates cannot be dynamically tuned according to varying channel quality. As a result, wireless networking is being extended to include the communication of signals between layers. This usually involves increased responsibilities at the lower layers that represent these scarce resources. The variety of QoS provisioning schemes that follow this paradigm sit under the title of cross layer optimisation or design.

QoS considers the ability of a network to provide a range of services, each suited to a certain class of flow. It is evaluated in terms of metrics such as bandwidth, latency, and jitter. QoS control measures are used to improve network performance in order to meet particular goals of the traffic flows being serviced. The QoS control measures, that make up a network service model, guarantee a minimum level of performance. When cross layer interaction is included in this model, QoS control becomes more sophisticated, but also more complex. Multiple control measures are mutually tuned, providing increased performance levels. In order to reflect this trade-off between performance and complexity in cross layer designs, Kliazovich et al [30] have proposed the classification of two categories of cross layering: weak and strong.

A scheme in the former category generalises the traditional interaction between adjacent layers of the protocol stack to non-adjacent interaction. The latter type of scheme takes a joined up approach to algorithm design, which may be implemented in any entity at any layer. This can possibly result in characteristics, formerly unique to a particular layer, moving between or merging across multiple layers. While strong cross layering provides higher performance, this is at the expense of cost, complexity, and flexibility in deployment scenarios. As interlayer interaction increases the complexity of the network stack, the ability to exchange protocols at one layer without considering other layers is lost. This is significant as this ability has supported innovation and widespread reuse of network protocols and architectures. Therefore, while moving away from the OSI model, it is still preferable for new architectures to maintain a balance between weak and strong. In placing minimal constraints on future modifications, the flexibility to support unmodified areas of the protocol stack is maintained.

There is a common format to all cross layer optimisation processes. This involves taking a set of parameter values from one or a subset of protocol layers and returning optimised parameter values to the same or other protocol layers. This commonality has enabled Khan et al [28] to define an overarching three-stage method of cross layer optimisation, illustrated in Figure 1. The first, abstraction, stage is the most critical to the reduction of communication and processing overheads. It decides whether a reduced number of parameters are to be circulated, and underlying technologies veiled. Optimisation and reconfiguration then enable protocol adaptation to current network conditions and QoS requirements, in order to maximise network performance. This is through the tuning of the abstracted, or other related parameters that are then returned to the network stack. These three steps can be repeated according to changing QoS requirements and resource capabilities. However, cross layer proposals have spanned a multitude of descriptive signal transfer methods and QoS control methods. While the former concentrate on how layers collaborate, the latter place more importance on which layers collaborate.

3 A Taxonomy of Cross layer Signalling Methods

Cross layer signalling has had increasing interest in wireless networking and numerous proposals exist. The authors in [57] have presented a survey of the methods of signal propagation across cross layer interfaces in static wireless networks. New designs have also abstracted the responsibility of cross layer optimisation away from the protocol stack into kernel space middleware also known as a cross layer management entity, plane or optimiser. This vertical entity avoids the use of direct interlayer coupling or communication. As many of these designs can be combined to

create more complex or sophisticated cross layering, those that would be relevant to providing QoS guarantees to inelastic SRT applications in MANETs are evaluated here.

The methods of signal propagation proposed to date have included the use of packet headers and structures, external profiles and even network servers. However, abstracted signalling information can only be of two types: locally or globally abstracted, though these are not always mutually exclusive. With the former, parameters are abstracted and optimised within a single node. The latter uses non-local, network-wide distribution of information from neighbouring nodes. The designs that fall under these two categories are discussed in detail here and a representative list of examples is provided.

3.1 Network-Wide Signalling

The earliest approaches to cross layer design have used packets for signal transmission. Packet headers are used by Explicit Congestion Notification (ECN) [49], a network-wide cross layer approach that is in commercial use. It is used to mark Transport Control Protocol (TCP) packet headers, to provide an early notification of congestion. Increases in RTT can also be taken as an early implicit notification of congestion, in contrast to packet loss, that is a late notification. ECN is a more reliable form of explicit notification than RTT, especially in MANETs where loss and delay can have numerous causes. Its use is appropriate for applications that are sensitive to packet loss. Unfortunately, ECN's hop-by-hop scheme requires notification to travel by increasing back-pressure through a congested network to reach the source. ECN is also not applicable with UDP, a protocol commonly used by RT applications. This is due to the requirement of application layer based congestion control, as well as API constraints on the appropriate header bits. The alternatives, Datagram Congestion Control Protocol (DCCP), Stream Control Transmission Protocol (SCTP) and TCP are highly inefficient in wireless networks [65], with connection oriented overheads, delayed congestion control and energy inefficiency.

Packet structures have also been used in the detection and signalling of network conditions. Numerous approaches have used packet probing in the estimation of available bandwidth and end-to-end delay [59], [45], [2], [7], [19]. While accurate measurements of propagation delay cannot be obtained under heavy traffic loads, bandwidth estimation has been more successful. Ahn et al [2] used UDP control packets Taleb et al [59] used low-priority dummy RTP packets, marked within the unused header bits to probe new network capability. Each intermediate node updated the packet if its available bandwidth was lower than that requested in the packet. The destination could then relay the minimum available bandwidth to the source. This approach required the modification of multiple nodes, to ensure recognition of the probes at end nodes. Dummy RTP packets were sent at a maximum streaming rate of the multimedia data for a fixed period of less than 1s, to which the receiver responded with reception quality feedback in a Real Time Control Protocol (RTCP) packet.

Nodes may also utilise globally distributed MAC layer information to estimate resources in their localisation of the network. For example, CACP [73] uses discovery of neighbouring nodes' available resources, as a result of broadcast querying or carrier sensing of idle nodes. Queried nodes respond if they measure resources to be insufficient and the source will attempt to transmit again after a backoff period. CACP is difficult to implement in MANETs, due to the requirement of periodic querying along a static end-to-end path. Perceptive Admission Control (PAC) [9] therefore avoids the use of queries and estimates bandwidth according to channel utilisation, for distances up to which two flows can be transmitted simultaneously, without collision. Multipath Admission Control for MANETs (MACMAN) [36] avoids the flow throttling common in admission control schemes, that often result in resource underutilisation. It extends PAC with multipath routing: enabling senders to transmit on whichever path has sufficient resources. However, the methods of resource estimation used by CACP, PAC and MACMAN require high frequency signalling, to ensure that transmissions are delayed or stopped if a threshold value is reached.

Other approaches have used passive monitoring in bandwidth estimation, such as that suggested by Vanhatup et al [63]. This begins with a node measuring utilisation and signal strength or throughput of neighbouring nodes, based on their periodic beacon signals. If a terminal receives a beacon from a neighbour, that neighbour is assumed to be in the collision domain of the terminal. This estimation assumes that the interference range is less than the carrier sense range but incorporates nodal distance into the throughput calculation. The available throughput of node AP_x is of inverse proportion to the total activity of AP_i nodes, which is in turn a value calculated relative to the channel activity of AP_x and weighted according to the channel distance to the other node.

Ergin et al [19] express the aggregate link utilisation ρ_{aggr} in proportion to the number of nodes in contention range, N_{cont} in Equation 1.

$$\rho_{aggr} = \sum_{i=1}^{N_{cont}+1} R \cdot \left[\frac{L}{B_i} + T_{oh} \right] \quad (1)$$

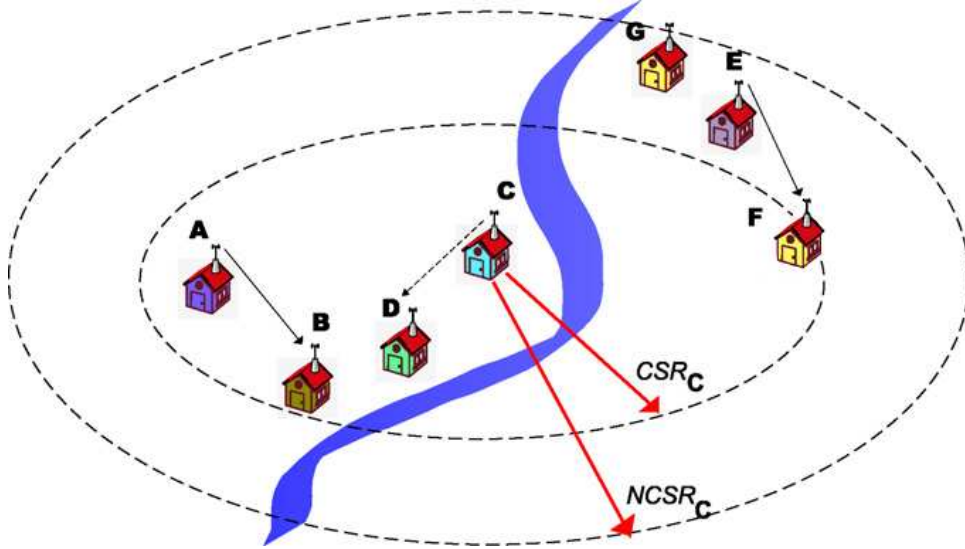


Figure 2: Self-interfering Flows [19]

Here R is the inverse of the packet generation frequency, L the length of the data packet, B is the transmitter link rate and T_{oh} is the link-rate/frame-size independent portion of the single-hop channel occupation duration. Such an implementation is subject to the exposed node problem [24], shown in Figure 2, where node C is throttled if it detects a CTS from F to E, as F is on the boundary of C's carrier sense range. This overly conservative bandwidth throttling leads to unused opportunities for spatially and temporally parallel transmissions.

Dual Carrier Sensing with Parallel Transmission awareness (DSCPT) and Packet Probing with RTS/CTS Handshake (PPRCH) [19] have been proposed to support parallel transmissions. DCSPT is an extension of PAC and CACP that relies on dynamic adjustment between higher ($NCSR_C$) and lower (CSR_C) carrier sense thresholds. As indicated in Figure 2, DSCPT use allows C to transmit to D, while considering the E to F transmission in its available bandwidth estimation. This approach gives throughput gains of up to 80%. However, in order to extend the carrier sense range of nodes, DCSPT assumes range modification capability in radio hardware. PPRCH was therefore suggested as this is not possible in most existing devices.

PPRCH utilises handshaking of only probe packets in order to avoid the hidden node problem. Two probe packets are sent out at highest and lowest priority respectively, so that the second does not compete with data packets and the dispersion of the probes is used to give the available bandwidth. However, this process utilises repeat probe packets with adjusted delays between pairs, in order to find a probe data rate that is equal to the available bandwidth. The problem is that this heuristic approach may take several RTTs to reach the correct measurement, which is not ideal for delay sensitive flows.

Packet probes have also been used in Ticket Based Probing (TBP), a QoS routing protocol. TBP uses tickets to narrow down possible paths and established routes in light of end-to-end delay requirements [15]. When a source needs a QoS path to a receiver, probes holding tickets are used for route discovery. Each probe accumulates the path delay, which is updated by intermediate nodes on the path. The total number of tickets available limits the possible paths probed. When a probe with N tickets arrives at a mobile node, the node will split the probe into $\frac{N}{x}$ probes, based on local state information, with a subset of the tickets. These are then forwarded onto x destinations. If delay restrictions are exceeded the remaining tickets are invalidated. Therefore, if multiple valid probes arrive at the destination the path with the least cost is selected and other paths kept as backup. A confirmation is then relayed to the source, using the list of mobile nodes along the path that is carried in the probe. Processing overheads are low for this mechanism. However, the accumulated link delay resulting from the use of probes can impact on performance in highly loaded networks. In addition, each node must store state information for each neighbour, entailing increased memory requirements for more complex network topologies [5].

As MANET nodes commonly have limited memory capacity, storage of signalling information in external servers can improve the performance of a cross layer approach. The Wireless Channel Information (WCI) [29] network service and CrossTalk [67] externally store signals, giving accessibility to the rest of the network. WCI abstracts

neighbours' data link and physical parameters, forwarding these via a proxy server. In contrast, CrossTalk [67], combines local with network-wide signalling. Each modified node coordinates parameter exchange in the local stack, which is then aggregated to a database of global network conditions. In using network-wide abstraction, processing overheads impact on end-to-end latency. Packet latency is also increased by the requirement for processing of cross layer parameters in intermediate nodes [50], [53], reducing the capability for RT support. The authors in [51] have also suggested a novel architecture that maintains global network status information and uses this to select and modify protocol parameters on a network-wide basis. The architecture is yet to be validated, but proposes the use of CTS packets to piggyback channel quality estimates from the receiver, triggering optimisation along the path to the sender.

3.2 Local Node Signalling

A local approach abstracts cross layer parameters from modified layers, storing them for access by other interested layers, for the appropriate optimisation of their services. In contrast with network-wide designs, these parameters can be associated with a particular packet or flow. As previously discussed, the earliest methods of local signalling used packet headers or structures for the encapsulation of parameters, providing accessibility to subsequent layers along the processing path. Packet headers are used by the Interlayer Signalling Pipe (ISP) [68], that modifies the IPv6 Wireless Extension Header (WEH) for in-band parameter propagation. The ISP does not require any add-on messaging protocols, as interested nodes can access signals if WEH aware. However, processing overheads increase when successive layers are required to access the network layer header. As packet based signal transfer is continual, the signalling benefit does not compensate for the resultant long-term increase in propagation latency.

Alternatively, signalling data can be inserted into a section of the packet structure, allocated to each layer, on transmission or receipt. Only layers that implement the cross layer modification can edit and access their corresponding packet structure segments, while non-participating layers need not perform any of these tasks. Similarly to ISP, the use of data packet structures limits the exchange of information to neighbouring layers in the direction of packet flow. In contrast, local out-of-band signalling shortcuts, using dedicated API, enable direct interlayer signalling between non-neighbouring layers. Both Internet Control Message Protocol (ICMP) and Real-Time Transport Control Protocol (RTCP), out-of-band control protocols, can be used for notification up the stack [68], [58], [40], [59]. IP uses ICMP messages for the same purposes as RTP uses RTCP: for the transfer of control information. These control messages are generated at any layer when convenient. For every parameter change, beyond a predefined threshold, a new control message must be generated, greatly increasing the competition with data packets for bandwidth.

ICMP messages must be encapsulated in IP packets, and RTCP in UDP packets. As a result, all signals are forced to pass the network layer, if ICMP is used, and transport layer, if RTCP is. This is even if the sending and receiving layers are not divided by those layers. CLASS [66] allows bidirectional message exchange between non-neighbouring layers. Signalling is higher speed, without the need to bypass adjacent layers. CLASS uses ICMP for general messaging and TCP/IP headers for shorter notifications. These out-of-band methods lose the benefit of signal association with a particular packet. They are not ideal for MANETS as they rely on control packet generation capability in all intermediate nodes, and are inflexibly limited to request-response procedures [30]. The result of the latter is increased network load, due to the large number of control packets on the network. Increased processing latency is also imposed by the heavy headers and checksum requirements as well as the storage of parameters in memory or a local hard disk in the node. Even if a node did have sufficient memory capacity, this local storage would not provide the direct access necessary in high-speed RT networks.

To avoid the drawbacks associated with external storage implementations, local profiles as well as functionality to manage their use, have been introduced to the field of cross layer optimisation. These profiles are also known as parameter databases, cross layer servers or planes. El Defrawy et al [18] have developed the Cross layer Server, to provide signal accessibility to all layers. Clients are used to communicate with layer protocols, requesting, optimising and controlling internal parameters. Similarly, the Central Cross layer Plane [14] uses local profiles of abstracted parameters, stored by an XML based mechanism, that are created on a per-layer basis for access by interested layers. The use of Callback functions by these profiles enables event based signalling: allowing one protocol layer to register these functions with a second layer for execution on a specific event's occurrence at the second layer. The benefit of an event-based method is that signalling traffic is reduced by at least half, as the need to request parameters is removed. Callback functions are defined and installed by a protocol, registered to the library at one point in time and only invoked when a parameter reaches a certain threshold value.

Callback functions can further be extended to contain instructions for encoding asynchronous private protocol data into a related abstraction with a local middleware entity. The related data repository can be accessed transparently across the protocol stack, as used in the MobileMAN [16]. Such a local entity, also known as an optimiser or interaction

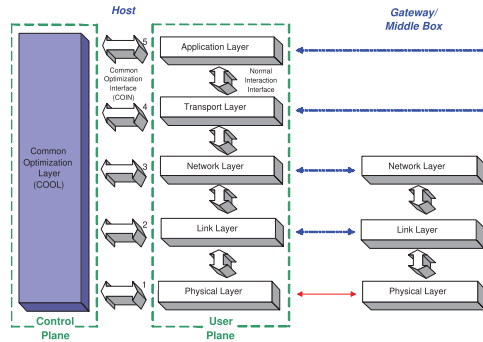


Figure 3: The POEM Model [23]

scheduler, can be made responsible for the abstraction of signal parameters, as well as the coordination of their usage. This entails a weakly cross-layered solution [30] that, while simple and flexible, reaps the performance benefits of strong cross layer interaction. The local entities proposed to date, can be divided into two classes, as defined by Foukalas et al [21]:

- Internal interlayer entity: associated with an individual protocol layer and acting in concert with other interlayer entities, associated with other layers.
- Internal intralayer entity: sitting between the application layer and the Operating System to coordinate signal propagation at and between all layers.

Carneiro et al [8] have used interlayer entities that are aware of the state of each protocol layer at any time, through notifications of layer specific events. These cross layer entities abstract parameters from protocol layers, thereby enabling interaction between heterogeneous technologies. Calculations are then made, based on comparison between a minimal number of parameters shared, to identify the optimised values for a particular function. For example the average Perceived Signal to Noise Ratio (PSNR) between the video stream at input and output, as well as the rate distortion factor can be calculated. These are then distributed to protocol layers, for comparison and amendment of their own related parameters [28]. Communication overhead is incurred in the transfer of parameters to and between optimisers and packet latency is increased. This is due to distributed calculations and protocol layer reconfigurations incurring high processing overheads.

When optimisation is centralised in an intralayer entity loops and conflicts between layers and processing and communication overheads of a distributed approach are avoided. A centralised optimiser has been used in the Interaction Control Middleware Plane [12]. This plane coordinates multiple optimisers operating simultaneously in different protocol layers. An intralayer entity controls the multiple interlayer entities. The Interaction Control Middleware Plane uniquely uses in-band signal propagation and suffers all of the drawbacks of its earlier in-band counterpart, the ISP.

The Performance-Oriented Model (POEM) [23] uses an internal intralayer entity that does not compromise normal protocol layer functionality. POEM is made up of two conceptual planes, or optimisers, as in Figure 3. The first permits normal non-optimised data flows and the second optimises interactions. An common interface between multiple protocol layers and an Optimisation Layer provide self-optimising services through a control protocol, the Common Optimisation Protocol (COP). POEM has been designed to execute concurrently with normal protocol interactions, without modification of the protocols but is still to be performance tested through simulation.

The ECLAIR [48], [47] architecture, in Figure 4, uses an internal intralayer entity: the Optimizing Sub-System (OSS). ECLAIR introduces cross layer interfaces, alongside the OSS, acting as tuning layers to support and control communication between the plane and layers. The tuning layers can manipulate protocol data structures at generic or operating system specific levels. Protocol layers themselves are not adapted but API functions are exported to layers, allowing read and write access to protocol control and data structures. Some examples of the APIs are presented in Figure 5.

Protocol Optimisers then use these APIs to manipulate protocol runtime behaviour. This optimisation is intelligently based on input from other layers and devices. Function call use does incur communication overheads but the overall processing overhead in the stack is negligible as the optimiser executes at the same time as the stack. A similar entity was also used by Kwon et al [32] in an OFDMA network, for cross layer scheduling, with showed the same benefits of reduced processing latency.

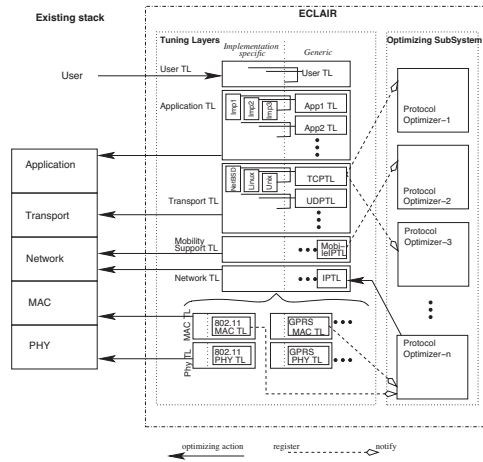


Figure 4: The ECLAIR Architecture [48]

<pre> register() set_application_priority() </pre>	User
<pre> get_delay_requirement() register() get_bandwidth_requirement() </pre>	App- lication
<pre> get_proto_block_head() register() get_rcv_win() set_rcv_win() get_tcp_state() set_tcp_state() get_rtt() set_rtt() get_retx_timer() set_retx_timer() </pre>	TCP
<pre> register() </pre>	Mob- ileIP
<pre> get_active_interface() register() set_active_interface() </pre>	IP
<pre> get_contention_window() register() set_contention_window() get_rts_cts_threshold() set_rts_cts_threshold() get_fragmentation_threshold() set_fragmentation_threshold() </pre>	802.11 MAC
<pre> get_transmit_rate() register() set_transmit_rate() get_transmit_power() set_transmit_power() </pre>	802.11 Phy

Figure 5: ECLAIR Tuning Layer APIs [48]

4 QoS Control with Protocol Tuning

A challenge in development of a cross layer design is in the selection of appropriate QoS-control measures for optimisation. More specifically it lies in the selection of protocol parameters that can signal relevant information to other layers or be dynamically tuned to enable improved application performance. Signalling latency and, proportionately, packet latency are dependant on the number of parameters abstracted and optimised. It follows that limiting the number of parameters abstracted is an aim of RT support. Additionally, the performance of a cross layer design is dependant on the selection of QoS control measures and appropriate parameters, in line with the optimisation goal. Establishing which should be optimised is not straightforward as there are indirect as well as direct linkages between protocol parameters. Khan et al [28] therefore defined the four key categories of parameters that can be abstracted by a cross layer model as:

- Directly tunable parameters: that can be reconfigured by cross layer optimisation, e.g. time slot assignment in TDMA.
- Indirectly tunable parameters: that cannot be themselves reconfigured but change as a result of reconfiguration of a directly tunable parameter, e.g. BER that depends on the coding and modulation scheme used.
- Descriptive parameters: that can be read but not tuned, e.g. channel quality estimates or picture size in video streams.
- Abstracted parameters: that are computed from the two types of tunable parameter but do not actually occur within the protocol stack, e.g. net transmission rate.

Many cross-layer models have been proposed in the fields of wireless SRT performance and NRT performance in MANETs. Each of these existing models has been holistically aimed at a specific performance target, such as improved TCP congestion control, delay or jitter reduction or fair sharing of available bandwidth. Research in the field of cross layer QoS guarantees to RT traffic in MANETs has been limited particularly to the support of high-quality multimedia. In order to meet the goal of guaranteed inelastic SRT QoS in MANETs, the commonalities and learning points from the former two aforementioned fields are of particular interest.

Cross layer approaches for MANETs predominantly deal with the bursty and congestion-prone nature of MANET traffic and corresponding packet losses. Rather than implementing congestion control, congestion may be avoided through the implementation of appropriate load control. Collaboration between devices is required to ensure that each flow satisfies its bandwidth requirements and does not surpass its allocation. Although the admission control process does not guarantee QoS, timing guarantees to applications are dependant on the path discovery phase of routing. This phase is, in turn, highly reliant on the stipulation of QoS requirements during admission control. The efficiency of distributed admission control depends on the accurate estimation of available resources, which is traditionally performed at the transport layer.

Models for wireless SRT have been particularly directed at the the need to guarantee and limit packet latency and loss. Common packet loss rates in wireless networks are between 10^{-3} and 10^{-1} . Compounded by the high bit-rate requirements of RT streams, when a number of streams contend for limited available bandwidth, packet loss can be increased by congestion and interference-related bit errors. Congestion occurs when data packets wait for service at a resource bottleneck, due to traffic flows exceeding the available capacity of buffers over the data path. As buffer occupancy increases, as does queueing delay experienced by neighbouring flows. When buffer resources are exhausted, uncontrolled packet dropping can result in network congestion collapse, with worst-case end-to-end latency and repeated, escalating packet loss.

The cross-layer models from both fields can be divided into three types that will be explored: network-adaptive, QoS-adaptive and what will be referred to as hybrid adaptive. In the first, higher layer protocols are tuned in light of variation in network resource conditions, whereas in the second lower layers are tuned to meet application-specified QoS requirements. Hybrid approaches combine both of the first two types of adaptation.

4.1 Network-adaptive Tuning

In wireless networks, congestion control can be implemented to detect and prevent congestion. However, congestion avoidance, if early notification is available, provides better network and RT performance. As previously discussed, both congestion control and notification are traditionally provided by TCP, a protocol that performs poorly in wireless networks and RT support [65], [43]. RT applications predominantly use UDP at the transport layer, but still have a congestion avoidance requirement.

Collisions and associated retransmissions are one of the key causes of congestion and therefore collision avoidance schemes have a large impact on congestion avoidance. Transmissions between two nodes in a shared medium consume

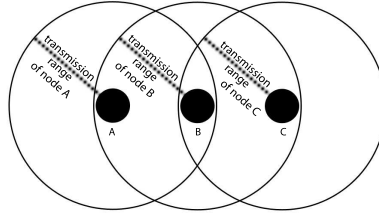


Figure 6: The Hidden and Exposed Node Problems [31]

bandwidth not only from those two nodes but also from all neighbouring nodes that are within carrier sense range. Some QoS mechanisms have therefore been designed to prevent contention for the shared medium and corresponding collisions. The aim of this is to maintain QoS guarantees during packet transmission. CSMA/CD cannot be implemented as radio broadcast nodes rely on a single antenna for transmission and receipt, therefore collision detection cannot occur at the same time as data transmission. Additionally, when collisions occur at a long-range receiver they may not be detected by other transmitters, as a result of signal fading. Carrier sensing can be used alone for collision avoidance, but introduces two problems: the hidden and exposed node problems, illustrated in Figure 6 where node B is in range of A and C which are not in each others range.

With synchronous contention-based schemes, collisions can occur but mechanisms for avoidance are implemented. For example, the use of control packet messaging, alone or in combination with carrier sensing. Multiple Access with Collision Avoidance for Wireless (MACAW) [6] employs a five-part control packet exchange, RTS-CTS-DS-DATA-ACK that exhibits high throughput and fast error recovery through use of the ACK. However, these extended control packet dialogues introduce a large amount of competition for resources and do not fully solve the hidden node problem. Control packets may still collide, leading also to eventual data packet collision. Some schemes that combine control packet messaging and carrier sensing utilise service differentiation [39], [64] or reservation [35] while others implement fair scheduling [62].

MACA with Piggyback Reservation (MACA/PR) [35] is a cross layer QoS-routing scheme with collision avoidance. It is designed for RT traffic and relies on MAC layer bandwidth reservation. Under this scheme the first data packet of a flow is used for MAC level reservation along the path. Source initiated control packet messaging is used to set up a reservation, with data packet transmission immediately following the receiver CTS and containing a piggybacked reservation for the subsequent packet. Data packet receipt is followed up by an ACK response, used solely to refresh the reservation. The protocol does not initiate loss recovery. Neighbouring nodes sensing data and ACK packets use these to maintain a reservation table of transmit and receive windows, for nodes in signalling range to backoff accordingly. If ACK receipt exceeds a timeout at the source, the channel is assumed to be insufficient for bandwidth requirements. The QoS routing protocol at the network layer is signalled accordingly.

The collision avoidance in MACA/PR achieves lower end-to-end delays than synchronous schemes but also lower aggregate throughput. Reduced throughput results from the communication overheads incurred with reservation table update and exchange and the requirement of the source node to consult these prior to transmission [31]. Collision avoidance alone cannot replace congestion avoidance. At the same time the unmodified carrier sense capability of radios is still important in ensuring that nodes within carrier range gain timely and reasonable access to good quality channels.

If transmissions between two nodes consume bandwidth from all neighbouring nodes, the corollary of this is that contention-aware load control must be implemented in order to avoid over-subscription of resources. Admission control does not explicitly use bandwidth reservation but many approaches have suggested the joint allocation of capacity and flow. This is via the exchange of link capacity and flow requirements between network and link layers. It can result in greater link utilisation and reduced congestion. Supported data rates have been shown to increase when link layer signalling is employed [55] to adjust transmission rates according to the capacity of individual links. However, resource reservation relies on a large amount of control information being passed between and maintained by nodes, and as such is not applicable in MANETs.

Service differentiation in stateless Wireless Ad hoc Networks (SWAN) [2] is a wireless network model that provides traffic classification and servicing with different priorities, without requiring per-flow signalling. Rate control is automatically configured based on MAC delay, and available bandwidth is estimated according to neighbouring flow rates. However, RT sessions are stopped, not throttled, under congestion so high overheads would be incurred with

its use in highly dynamic networks requiring regular session re-establishment.

Scheduling also plays a key role in the support of jitter and delay sensitive wireless applications, given its impact on packet deadline achievement. When packets are dequeued for transmission it is scheduling algorithms that make the decision of which packet is the head of line, based on priority, delay requirements, nodal congestion and other QoS requirements. Queues exist at multiple protocol layers inside a node and must be serviced appropriately in order to ensure maximum and fair resource utilisation and that QoS guarantees are upheld. Wired network schedulers rely on traffic and queueing statuses, as a basis for prioritising packets, but in wireless networks with varying channel-capacities this is insufficient.

Wireless packet scheduling is utilised to ensure efficient link-utilisation; provision of delay bound guarantees, smooth service degradation and protection from non-conforming sessions. The fair redistribution of resources across sessions and guaranteed short-term and long-term throughput [20] may also be provided. A number of cross layer designs incorporate dynamic schedulers at the transport or MAC layers and occasionally, when optimising video transmissions, at the application layer. Kwon et al [32] used a MAC scheduler and resource controller that together increased achievable throughput according to Channel Quality Information from the physical layer. However, this required a Hybrid ARQ (HARQ) scheme, to support the selection of modulation and coding and improve throughput guarantees, at the expense of high retransmission overheads.

Several scheduling approaches have considered optimising the source bit rate according to channel conditions in order to minimise congestion and delay [55], [1], [75], [28], [41]. If the transmission rate selected is lower than the optimal transmission rate along a path, a large amount of jitter is introduced into the stream. If it is higher, packets will be dropped at intermediate nodes and the receiver. This is generally only useful to a select group of elastic SRT applications including media streaming, video conferencing and interactive network gaming, but has shown an increase in video quality of between 0.63dB [75] and 2dB [28]. In order to reduce distortion in video streaming a select set of parameters may be optimised. The application layer participates, as it contributes parameters relating to per-packet loss distortion effects. It can also adapt the source rate according to network capability information provided by the data link or physical layer.

4.2 QoS-adaptive Tuning

Many load control approaches for wireless RT do not directly address the effect of MAC layer contention or neighbour interference on bandwidth estimation. Therefore some QoS-adaptive models have moved to take a MAC layer approach to bandwidth reservation, combining this with QoS routing to support wireless SRT. It has been suggested [55], [1] that the MAC layer should perform dynamic capacity assignment, determining resource allocations for different flows that have undergone congestion-optimised multipath routing at the network layer. However, this requires all participating nodes to implement the same MAC layer or employ a bridging device between heterogeneous devices, which may introduce certain security issues.

In wireless ad hoc networking, contention within carrier sense range of a node must be considered in the bandwidth estimation before load control is performed, given the dynamic topology. When admitting a flow, the transport layer cannot provide an accurate estimate of current network conditions but lower layers can [26]. Yuhe and Jie [74] suggested the joint control of the physical and MAC layers for the estimation and prediction of channel variation based on packet error rates (PER). Modified RTS packets carry required PERs and data rates as well as training bits used by cross layer middleware in estimation and prediction. The MAC layer can then access physical layer parameters including available transmission rate that, combined with the SINR, can be used to improve scheduling decisions. As a result, higher rate transmissions can be prioritised on links of degrading quality.

The MAC layer continually monitors instantaneous signal strength (ISS) changes, and can provide better information on available resources in a MANET, where link performance is particularly dependant on SINR. A tunable MAC protocol, Congestion Reducing Medium Access Control (CRMAC), was proposed by Bag and Bassiouni [4], which could be adapted to the requirements of the application, on the basis of buffer status data from the network layer. A combination of the recent collision history and number of collisions for that node was then used to calculate an appropriate backoff value, prioritising congested nodes above others in the same collision domain. Although RT delay guarantees cannot be provided with a random backoff, the calculation utilises the useful concept of a weighted collision history to calculate the probability of a collision occurring. This is expressed through the sum of the collisions since last transmission α and the collision history β given by Equation 2. The calculation of collision history then utilises the constant μ to ensure that earlier collisions are weighted more than recent ones.

$$\beta = \mu\beta + (1 - \mu)\alpha \tag{2}$$

The performance of admission control is highly dependant on the accuracy of the bandwidth estimation method

implemented. While those approaches that use analytical modelling for collision prediction provide a more sophisticated estimate, they require nodes to store and maintain large amounts of data. For example a node may be required to compute transmission probabilities and identify traffic within carrier sense range, but outside of transmission range, or that may be hidden from the node. These requirements render the cross layer approach topology-dependant. They do not perform well in MANETs due to the difficulty in each node securing and maintaining up-to-date information, for a required period, on all receiving nodes.

Admission control can only take place once a route has been selected, as end-to-end bandwidth and latency are highly dependent on channel quality and node capacity at each of the hops along the chosen route. Routing protocols are responsible for the selection of the most appropriate of multiple available paths, from source to destination, over which traffic flows can be transmitted. In MANETs, routing protocols must compensate for the lack of centralised control of resource management, while also dealing with the exposed and hidden node problems [31]. As a source cannot know the whole network topology, distributed routing is usually implemented. Here each mobile node must select the next hop from among its one-hop neighbours.

Distributed routing protocols must refer to quantitative metrics to satisfy application QoS requirements, under the constraints of available resources [17]. The most common metrics used in the cross layer path selection process are bandwidth and delay, followed by hop count, jitter, energy, loss probability and signal strength or distance. Depending on the protocol used, one or several metrics are calculated for each discovered path and then each path is compared, to identify the best one. Metric selection can have an increasing impact on performance, depending on whether simpler additive or more complex multiplicative or concave calculations are used.

Additive computations aggregate the metric for all links, such as with delay, jitter and hop number calculations, whereas multiplicative computations such as reliability and packet loss probability multiply the per-link metrics. Calculations of bandwidth are concave metrics as minimum and maximum values are required for each link [5]. Link and MAC layer metrics can also affect the QoS of a session and several protocols exist to jointly optimise these lower layer metrics with the network layer. The per-nodal MAC delay; frame delivery ratio, a statistical measure of arrival probability; predicted link lifetime, or link stability; normalised MAC load, the ratio of transmitted control frame bits to user data frame bits and relative node mobility-stability ratio of neighbouring nodes could all be used by a QoS-adaptive protocol to gauge the best path.

QoS-adaptive routing protocols can be subdivided into two classes: proactive and table driven, or reactive and on-demand [52]. Proactive protocols maintain tables of routing information throughout the network, to which a self-replicating update is triggered according to topological changes. Destination-Sequenced Distance Vector (DSDV), Wireless Routing Protocol (WRP) and Cluster-head Gateway Switch Routing (CGSR) are examples of proactive protocols that provide greater efficiency through reduced spatial diversity of updates. Proactive routing in MANETs does ensure that routes are consistently available. However, high overheads are incurred by the regular, network-wide updates required that increase with the rate of topology change and rapidly degrade QoS provisioning.

Alternatively, reactive routing can be used, which is initiated only when a source requires a path to a destination that it cannot itself generate. This necessitates a path-discovery mechanism with an associated delay. The most basic reactive routing schemes are best effort as with Dynamic Source Routing (DSR), Ad hoc On-demand Distance Vector routing (AODV) and Temporally Ordered Routing Algorithm (TORA). These allow all nodes in range to compete for the medium. Similar to Open Shortest Path First (OSPF) and RIP, routing with these is often based around identification of the shortest path. Much like a road traffic scenario, the shortest path does not necessarily signify the path with the least delay. In fact the use of the minimum hop count metric alone for the identification of multiple source-receiver routes can be insufficient when a route must satisfy QoS constraints [57], [46]. This is primarily due to the variable, asymmetric and unreliable nature of a MANET channel. However, the most widely used routing algorithms in MANETs are still those that rely on shortest path routing: TORA, DSR, AODV and Zone Routing Protocol (ZRP). DSR employs each packet to carry routing information and AODV uses a method of partial routing table exchange while TORA discovers several end-to-end paths maintaining these until they have all failed [54].

IETF OSPF-MANET routing is one of the only commercially used cross layer designs [42] that includes a MANET-specific cross layer interface for signalling from the data link layer to the network layer. This implementation reduces packet loss, resulting from signal loss, by 60%. The cross layer interface tracks incoming frames and then receiving-link quality is assessed for use by the routing protocol. This enables a distinction to be made between physical link failure and congestion, for signalling to upper layers. This scheme enables the assignment of higher priorities to higher link qualities, reducing the re-routing delay that results from link failure. OSPF-MANET also relies on flooding and hop-by-hop acknowledgement and exploits the broadcast efficiency of the underlying radios. However, a cross layer processing overhead is incurred for each signal, as the routing protocol must use an address mapping function to map the MAC address to an interface IP.

QoS-adaptive scheduling has been implemented in wireless networks for RT application support. Differentiated-

Time Urgency Based Algorithm (D-TUBA) [60] utilises a cross layer scheduler to schedule packets according to class and global delay information abstracted from the network layer in participating nodes. A modified Weighted Deficit Round Robin (WDRR) is then used to decide to where packets are de-queued. WDRR was modified using a counter flag to indicate whether a queue remainder was larger than the next packet so that at most one packet would be serviced whenever a queue was polled. This scheme adopts IP packet header signalling. The next packet is enqueued according to the estimated remaining time of the packet (the latency bound less the estimated time to destination). Using a scheduler to estimate time to destination in lieu of more precise measurement avoids the introduction of large communication overheads, but this assumes that the traffic en route is of a uniform distribution.

An and Song [3] have also developed a priority-based wireless scheme, with routing and MAC scheduling working to meet RT latency requirements. The concept of packet urgency is used to give greater priority to packets where the accumulated delay to required maximum delay ratio is larger. With packet priority at one node dependant on the priorities of packets at other nodes, implementation of this network-wide approach in a MANET would entail high bandwidth consumption and increased latency overheads.

4.3 Hybrid Network and QoS-adaptive Tuning

A few cross-layer approaches have adopted hybrid adaptation, with parameters shared from and tuned at multiple layers in order to provide QoS guarantees without overloading the network. For example the congestion minimisation scheme proposed by Setton et al [55] supports the highest data rates and yields minimum end-to-end latency, by guaranteeing a given data rate between RT transmitter-receiver pairs. This requires the MAC and network layers to identify the set of network flows that minimize congestion, through the iterative exchange of possible suboptimal solutions. However, such an heuristic approach requires extensive signalling and would not be ideal for transmissions with critical timing requirements operating over a MANET. The authors in [56] have investigated the effect of jointly tuning application layer packet size, physical modulation and MAC retry limits with reference to received multimedia performance. This has been done offline, without implementation of a particular signalling method. Over a single-hop with low SINR lowering modulation and increasing the retry limit reduced delay, but at the expense of greatly lowered throughput. Lowering the packet size according to reduction in channel conditions was also suggested in order to provide minimum goodput and delay guarantees.

Congestion-aware physical rate selection and allocation has been suggested in [76] and [38]. [76] uses network-wide updates of video source rate and link congestion price in local rate allocation. Loiacono et al [38] suggest that such approaches fail in attributing PLR to channel conditions rather than collisions. Instead they propose consideration of the application codec type, collision probability and physical channel conditions to estimate received video quality. Tuning physical rate according to this estimate results in increased throughput of up to 2.4Mbps.

Liu et al [37] proposed a RT scheduler utilising a per-connection priority function that is updated dynamically according to wireless channel quality as well as QoS requirements. The scheme offers the highest priority and guaranteed QoS to CBR connections, such as VoIP, and a lower priority with some packet loss to RT traffic that can tolerate it. Generally the larger the delay satisfaction the lower the priority, but if it drops below a threshold packets are sent immediately. The use of Channel Quality Information (CQI) ensures that, within a single class, a large normalised received SNR translates to a higher priority, but that channels experiencing severe fading are not serviced at all. However, even when channel quality is low, if the delay satisfaction is low, the connection will still be serviced with a relatively high priority. The model developed by Chen et al [13] controls packet loss rate, resulting from link errors, using local channel conditions to determine transmission power level and media encoding rate. In the situation of buffer overflow, non-local coordinated scheduling is also initialised. Overall, this allowed for a 70% increase in parallel session support and reduced delay and packet loss when implemented in a collision-free network.

In MANETs, QoS-adaptive routing protocols have also taken into account the durability of the channel. That is, when channel durability is highly likely, a node can be offered a better connection with a low coverage range. When it is unlikely, the channel should be offered to connectors with larger coverage ranges, or ones that move with the same speed and direction as the node. Associativity Based Routing (ABR) and Signal Stability Routing (SSR) route reactively while considering link quality. The former prefers hop stability and the latter chooses routes based on the Received Signal Strength Indication (RSSI). Hybrid proactive-reactive protocols have also been developed. ZRP uses local proactive routing and non-local reactive routing, while LUNAR [61] combines reactive path discovery and proactive path rebuilding at a fixed frequency to deal with topology changes.

A number of approaches [25], [72], [69], also including the Extremely Opportunistic Routing (ExOR) protocol, have suggested different degrees of coordination between MAC and routing. ExOR is a routing protocol that selects the best next hop, after each per-hop transmission. It uses an average of one-hop link metric information to do this, therefore the performance improvement is limited until sufficient link metric measurements have been received. Wu

Table 1: Comparison of Cross layer Signalling Mechanisms

Signal scope	Signalling mechanism		Transmission latency	Communication overhead	Processing overhead	In-band or Out-of-band
Local to node	ISP	IP packet header	Medium	High	Medium	In-band
		Packet structure	Medium	High	Medium	In-band
	Direct interlayer	CLASS	Low	Medium	High	Out-of-band
		ICMP packets	Low	Medium	High	Out-of-band
	Callback functions		Low	Low	Low	Out-of-band
	POEM		Low	Medium	Low	Out-of-band
	Cross layer Server		Low	Medium	Low	Out-of-band
	Central Cross layer Plane		Low	Low	Low	Out-of-band
	ECLAIR		Low	Low	Low	Out-of-band
Control Middleware Plane		Low	Low	Low	In-band	
Network-wide	Packet header		High	Low	Medium	In-band
	ICMP		High	High	High	Out-of-band
	WCI		High	Low	Low	Out-of-band
	CrossTalk		High	Low	Low	Out-of-band

Table 2: Parameter Tuning for SRT in MANETs

Protocol Layer	Abstractable Parameters	Tunable Parameters
Application Layer	Delay tolerance, acceptable delay, acceptable jitter, required bandwidth, acceptable PLR Use: Avoiding low SINR triggered congestion response; QoS-adaptive routing and scheduling	Source rate, encoding format, compression Use: Network-adaptive rate control
Transport Layer	RTT, Recovery Time Objective, MTU, total packet loss and actual throughput Use: Avoiding low SINR triggered congestion response; QoS-adaptive routing and scheduling	Sending rate, MTU Use: Network-adaptive rate control
Network Layer	Timestamps of mobility events, route and network interface used Use: Network-adaptive admission control	Route selected, network interface selected Use: Network-adaptive routing; QoS-adaptive scheduling
MAC Layer	FEC scheme, retransmission totals, frame lengths, time stamps of transmission and handoff events, transmission rate and PER, ISS, nodes in transmission and carrier sense range, channel coherence time Use: Network-adaptive admission control; Network-adaptive routing	TDMA time slots, FEC scheme Use: QoS-adaptive scheduling

and Wu [69] have also proposed the joint use of QoS requirements, MAC queue length and physical SINR to distribute traffic over multiple paths to the receiver. In this network-wide framework, routing decisions are made at each hop, based on link status calculated from SINR and queue length information that is added to RREPs by intermediate nodes. At each hop the modulation mode is tuned to adapt transmission rate in proportion to SINR: a PSNR increase of around 1dB results. Wu and Wu further develop their protocol in [70] to include network congestion-awareness. This is implemented via global signalling of MAC layer utilisation over multiple paths, providing up to a 1.7dB PSNR increase over a protocol without the cross-layer signalling.

Hong et al [25] proposed further merging between MAC and routing with the use of virtual links to avoid processing delays between these layers. They require the link layer to both select the next hop and re-encapsulate packets. This resulted in a 7-10% throughput improvement and 50% reduction in processing time. Similarly, in [72] the MAC layer selects and prioritises paths based on physical link quality and route information. Yamao et al [72] state that the minimum hop-count route chosen by AODV results in the use of long low SINR links that fail under fading conditions. They suggest the use of multi-hop path selection, using shortcut paths and novel control messages to prevent transmission redundancy. For node distances of less than 250m this does result in a transmission delay reduction. However, in moving traditional network functionality to the lower layers the modularity and re-usability of such an approach is low.

5 Conclusions

Many Cross Layer models exist either for the provision of QoS to SRT streams in static wireless networks or to improve the performance of MANETs. This paper has surveyed the common ground and lessons learned from

structural models of cross-layer signalling and then protocol-tuning approaches in these two fields. This has been with a view to advising the open research area of high performance service provision to loss and delay intolerant real-time applications in MANETs.

It is the protocol parameter abstraction and tuning of a cross-layer design that provides QoS guarantees and Table 2 indicates some of the parameters available at each layer. With event-based optimisation, when a protocol parameter arrives at a pre-specified threshold value it is abstracted. The pre-optimisation threshold value must therefore correspond to minimum QoS requirements or network provisioning and the aim of optimisation process is to ensure that this value is always exceeded. Tuning models that use either or both of network-adaptive and QoS-adaptive approaches have been considered in Section 4.

QoS-aware routing is commonly implemented in wireless and ad hoc networks. However QoS-adaptive scheduling for inelastic SRT flows and for MANETs is an area which requires further research. The deterministic QoS guarantees required by these flows depend on timely, ordered and guaranteed packet arrival, to which packet scheduling is key. QoS-aware cross layer scheduling approaches can benefit from the use of priority scheduling and timestamping for separate packet and slot queues. These scheduling methods, combined, generally provide the best assurances to delay and jitter-sensitive applications. However, this is at the expense of reduced response rates due to higher node computation and processing requirements, with a corresponding impact on packet latency. For multi-hop networks this is an area that requires further performance analysis and testing as multiple priority queueing requirements result in increased packet latency.

The major difference between traditional wireless networks and the MANET is the dynamically varying resource conditions seen in the latter. Responsively, many cross-layer designs have elected parameters from lower layers, such as received packet power or optimal transmission rate to signify these conditions to higher layers or moved resource allocation to these layers. Such notification and tuning is useful but not always possible. Tuning the MAC layer or requiring it to access physical parameters entails modifications to radio firmware or hardware that have limited accessibility, primarily to vendors. This can reduce the transparency and interoperability of modified nodes. However, network-adaptive routing and admission control based on purely MAC layer information is key to the support of inelastic SRT applications in MANETs.

The high packet loss rates, jitter and varying latency common in a MANET must be addressed rather than compensated for. The low and time varying SINR conditions of a MANET adversely effects routing and admission control. Using MAC layer, transmission and retransmission rates, PER, channel coherence time and packet or ACK timestamps cross-layer middleware can characterise these varying conditions. Managed collaboration between QoS-adaptive routing and admission control that respond to this information can then provide dynamic optimal path selection (identifying alternatives to a congested shortest path) and smooth performance degradation. While the tuning of these parameters can improve performance, through opportunistic resource use, it is also essential that a cross-layer approach should be forward thinking, in terms of re-usability and modularity. This is highly dependant on the signalling method implemented.

Predominant network-wide and local-parameter based signalling methods were compared in Section 3. Table 1 notes the conclusions made on their relative overheads when applied in support of inelastic SRT applications in MANETs. Network-wide models are not appropriate to MANETs due to the reliance on maintained signalling contact with all intermediate nodes. Higher bandwidth overheads and transmission latency result from the addition of signalling traffic to network load. Cumulative signal transmission and computation delays also impact on packet latency. It is local signalling that better suits MANET nodes that are prone to intermittent loss of contact with other nodes and that are likely to be highly and randomly spatially distributed.

Local middleware models from POEM to the Control Middleware Plane outperform packet based approaches. This is because the latter incur increased processing and communication overheads that commute to increased jitter, as per-node processing is no longer possible at line speed. The adverse side of local middleware lies outside of run-time: in being a non-standard kernel component both implementation and porting can be complex. Conversely, localisation in the kernel enables high-speed, execution-concurrent optimisation. Middleware also avoids the resultant packet bursts and corresponding queueing delays that are not ideal for inelastic SRT flows or mobile nodes with limited storage capacity. Among the higher-performance middleware schemes, intralayer optimisers, such as ECLAIR, that use event-based signalling also leave the protocol stack intact. This enables adaptable rapid prototyping, transparency, portability and lightweight design. The lower overheads mean better packet-timeliness guarantees can be provided and in optimising the stack from a single, external location, signalling loop errors are avoided.

In line with these suggestions a cross-layer design for inelastic SRT in MANETs should avoid spaghetti design, being instead modular, transparent and reusable. However, given the singular characteristics of an ad hoc network this does not necessitate re-usability in other types of wireless network. Therefore MANET-specific conditions must be taken into account in developing a design, for example, scheduling must not assume uniform traffic distribution

or routing, symmetrical links. As learnt from the development of internet protocols, a codified approach to development as well as clear publication of the conditions under which a cross layer design will fail can better ensure its sustainability.

Acknowledgment

This work is sponsored by the Engineering and Physical Science Research Council (EPSRC) and BAE Systems, UK.

References

- [1] S. Adlakha, X. Zhu, B. Girod, and A. J. Goldsmith. Joint capacity, flow and rate allocation for multiuser video streaming over wireless ad-hoc networks. In *IEEE International Conference on Communications (ICC)*, pages 1747–1753, 2007.
- [2] G.-S. Ahn, A. T. Campbell, A. Veres, and L.-H. Sun. Swan: service differentiation in stateless wireless ad hoc networks. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 457–466 o.2, 2002.
- [3] K. J. An and H. Song. An effective cross-layer packet scheduling and routing algorithm for delay-sensitive media transmission over manet. In *Proceedings of the 2009 IEEE international conference on Communications (ICC)*, pages 124–127, 2009.
- [4] A. Bag and M. A. Bassiouni. A tunable cross-layer congestion reducing medium access control (crrmac) protocol for wireless networks. In *Military Communications Conference, 2007. MILCOM 2007. IEEE*, pages 1–7, Orlando, FL, USA, 2007.
- [5] P. Becker. Qos routing protocols for mobile ad-hoc networks - a survey. Technical report, University of Kaiserslautern, August 2007.
- [6] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang. Macaw: a media access protocol for wireless lan's. In *SIGCOMM '94: Proceedings of the conference on Communications architectures, protocols and applications*, pages 212–225, New York, NY, USA, 1994. ACM.
- [7] C. T. Calafate. Supporting soft real-time services in manets using distributed admission control and ieee 802.11e technology. In *ISCC '05: Proceedings of the 10th IEEE Symposium on Computers and Communications*, pages 217–222, Washington, DC, USA, 2005. IEEE Computer Society.
- [8] G. Carneiro, J. Ruela, and M. Ricardo. Cross-layer design in 4g wireless terminals. *Wireless Communications, IEEE*, 11; 11(2):7–13, 2004.
- [9] I. D. Chakeres, E. M. Belding-Royer, and J. P. Macker. Perceptive admission control for wireless network quality of service. *Ad Hoc Netw.*, 5(7):1129–1148, 2007.
- [10] M. C. Chan and R. Ramjee. Tcp/ip performance over 3g wireless links with rate and delay variation. In *MobiCom '02: Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 71–82, New York, NY, USA, 2002. ACM.
- [11] Yee Sin Chan and J. W. Modestino. A joint source coding-power control approach combined with adaptive channel coding for video transmission over cdma cellular networks. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, volume 5, pages 3415–3419 o.5, October 2003.
- [12] Z. Chang, G. Gaydadjiev, and S. Vassiliadis. Infrastructure for cross-layer designs interaction. In *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on*, pages 19–25, August 2007.
- [13] G. Chen, M. Song, Y. Zhang, and J. Song. Cross-layer adaptation with coordinated scheduling for heterogeneous wireless networks. In *VTC Fall*, pages 1–5, 2010.
- [14] K. Chen, S. H. Shah, and K. Nahrstedt. Cross-layer design for data accessibility in mobile ad hoc networks. *Wirel.Pers.Commun.*, 21(1):49–76, 2002.
- [15] S. Chen and K. Nahrstedt. Distributed quality-of-service routing in ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 17(8):1488–1505, August 1999.
- [16] M. Conti, G. Maselli, and G. Turi. Design of a flexible cross-layer interface for ad hoc networks. In *Challenges in Ad Hoc Networking*, volume 197 of *IFIP International Federation for Information Processing*, pages 189–198. Springer Boston, 2006.
- [17] E. Crawley, R. Nair, B. Rajagopalan, and H. Sandick. A framework for qos-based routing in the internet. RFC 2386 (Informational), August 1998.
- [18] K. M. El Defrawy, M. S. El Zarki, and M. M. Khairy. Proposal for a cross-layer coordination framework for next generation wireless systems. In *IWCMC '06: Proceedings of the 2006 international conference on Wireless communications and mobile computing*, pages 141–146, New York, NY, USA, 2006. ACM.
- [19] M. A. Ergin, M. Gruteser, L. Luo, D. Raychaudhuri, and H. Liu. Available bandwidth estimation and admission control for qos routing in wireless mesh networks. *Computer Communications*, 31(7):1301, 2008.
- [20] H. Fattah and C. Leung. An overview of scheduling algorithms in wireless multimedia networks. *Wireless Communications, IEEE [see also IEEE Personal Communications]*, 9(5):76–83, 2002.
- [21] F. Foukalas, V. Gazis, and N. Alonistioti. Cross-layer design proposals for wireless mobile networks: a survey and taxonomy. *Communications Surveys and Tutorials, IEEE*, 10(1):70–85, 2008.
- [22] A. J. Goldsmith and S. B. Wicker. Design challenges for energy-constrained ad hoc wireless networks. *Wireless Communications, IEEE*, 9(4):8–27, 2002.
- [23] X. Gu, X. Fu, H. Tschofenig, and L. Wolf. Towards self-optimizing protocol stack for autonomic communication: Initial experience. In *Autonomic Communication*, volume 3854 of *Lecture Notes in Computer Science*, pages 186–201. Springer Berlin / Heidelberg, 2006.

- [24] Rafael Guimaraes, Llorenç Cerda, Jose M. Barcelo, Jorge Garcia, Michael Voorhaen, and Chris Blondia. Quality of service through bandwidth reservation on multirate ad hoc wireless networks. *Ad Hoc Networks*, 7(2):388 – 400, 2009.
- [25] K. Hong, S. Sengupta, and R. Chandramouli. Cross-layer mac enabling virtual link for multi-hop routing in wireless ad hoc networks. *Communications Society*, 2010.
- [26] C. Huang, M. Loiacono, J. Rosca, and J.-N. Hwang. Distributed cross layer congestion control for real-time video over wlan. In *Communications, 2008. ICC '08. IEEE International Conference on*, pages 2270–2276, 2008.
- [27] V. Kawadia and P. R. Kumar. A cautionary perspective on cross-layer design. *Wireless Communications, IEEE*, 12(1):3–11, 2005.
- [28] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer. Application-driven cross-layer optimization for video streaming over wireless networks. *Communications Magazine, IEEE*, 44(1):122–130, January 2006.
- [29] B.-J. Kim. A network service providing wireless channel information for adaptive mobile applications. i. proposal, 2001.
- [30] D. Kliazovich, M. Devetsikiotis, and F. Granelli. *Chapter I: Formal Methods in Cross Layer Modeling and Optimization of Wireless Networks: State of the Art and Future Directions*, pages 1–24. Handbook of Research on Heterogeneous Next Generation Networking: Innovations and Platforms. IGI Global, January 2008.
- [31] S. Kumar, V. S. Raghavan, and J. Deng. Medium access control protocols for ad hoc wireless networks: A survey. *Ad Hoc Networks*, 4(3):326–358, May 2006.
- [32] T. Kwon, H. Lee, S. Choi, J. Kim, D.-H. Cho, S. Cho, S. Yun, W.-H. Park, and K. Kim. Design and implementation of a simulator based on a cross-layer protocol between mac and phy layers in a wibro compatible ieee 802.16e ofdma system. *Communications Magazine, IEEE*, 43(12):136–146, 2005.
- [33] D. Le, X. Fu, and D. Hogrefe. A review of mobility support paradigms for the internet. *Communications Surveys and Tutorials, IEEE*, 8(1):38–51, March 2006.
- [34] R. Li, A. Eryilmaz, L. Ying, and N. B. Shroff. A unified approach to optimizing performance in networks serving heterogeneous flows. *IEEE/ACM Trans. Netw.*, 19:223–236, February 2011.
- [35] C. R. Lin and M. Gerla. Asynchronous multimedia multihop wireless networks. In *INFOCOM '97: Proceedings of the INFOCOM '97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution*, page 118, Washington, DC, USA, 1997. IEEE Computer Society.
- [36] A. Lindgren and E. M. Belding-Royer. Multi-path admission control for mobile ad hoc networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 8(4):68–71, 2004.
- [37] Q. Liu, X. Wang, and G. B. Giannakis. A cross-layer scheduling algorithm with qos support in wireless networks. *Vehicular Technology, IEEE Transactions on*, 55(3):839–847, 2006.
- [38] M. Loiacono, J. Johnson, J. Rosca, and W. Trappe. Cross-layer link adaptation for wireless video. In *IEEE International Conference on Communications (ICC)*, pages 1–6, Piscataway, NJ, USA, May 2010. IEEE Communications Society.
- [39] B. S. Manoj and C. Siva Ram Murthy. Real-time traffic support for ad hoc wireless networks. In *Networks, 2002. ICON 2002. 10th IEEE International Conference on*, pages 335–340, 2002.
- [40] S. Merigeault and C. Lamy. Concepts for exchanging extra information between protocol layers transparently for the standard protocol stack, 2003.
- [41] P. Navaratnam, H. Cruickshank, and R. Tafazolli. A link adaptive transport protocol for multimedia streaming applications in multi hop wireless networks. In *MobiMedia '07: Proceedings of the 3rd international conference on Mobile multimedia communications*, pages 1–6, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering; ICST, Brussels, Belgium, Belgium, 2007. ICST.
- [42] G. Pei, P. A. Spagnolo, S. Bae, T. R. Henderson, and J. H. Kim. Performance improvements of ospf manet extensions: A cross layer approach. In *Military Communications Conference, 2007. MILCOM 2007. IEEE*, pages 1–7, 2007.
- [43] F. Peng, S. Cheng, and J. Ma. Proposal to apply ecn to wireless and mobile networks. In *The 10th Annual Internet Society Conference*, July 2000.
- [44] D. D. Perkins and Herman D. Hughes. A survey on quality-of-service support for mobile ad hoc networks. *Wireless Communications and Mobile Computing*, 2(5):503–513, 2002.
- [45] R. S. Prasad, M. Murray, C. Dovrolis, K. Claffy, Ravi Prasad, and Constantinos Dovrolis Georgia. Bandwidth estimation: Metrics, measurement techniques, and tools. *IEEE Network*, 17:27–35, 2003.
- [46] L. Qin and T. Kunz. Survey on mobile ad hoc network routing protocols and cross-layer design. Technical report, Carleton University, August 2004.
- [47] V. T. Raisinghani and S. Iyer. Cross-layer design optimizations in wireless protocol stacks. *Computer Communications*, 27(8):720–724, May 2004.
- [48] V. T. Raisinghani and S. Iyer. Cross-layer feedback architecture for mobile device protocol stacks. *Communications Magazine, IEEE*, 44(1):85–92, 2006.
- [49] K. Ramakrishnan, S. Floyd, and D. Black. The addition of explicit congestion notification (ecn) to ip, 2001.
- [50] M. A. Razzaque, S. Dobson, and P. Nixon. Cross-layer optimisations for autonomic networks. In *Advanced Autonomic Networking and Communication*, Whitestein Series in Software Agent Technologies and Autonomic Computing, pages 127–148. Birkhuser Basel, 2008.
- [51] X. Ren and J. Zhang. A novel cross-layer architecture for wireless protocol stacks. In *International Conference on Multimedia Technology (ICMT)*, Ningbo, China, 2010.
- [52] E. M. Royer and C.-K. Toh. A review of current routing protocols for ad hoc mobile wireless networks. *Personal Communications, IEEE*, 6(2):46–55, April 1999.
- [53] P. Sarolahti and S. Floyd. Cross-layer indications for transport protocols, October 2006.

- [54] S. Sesay, Z. Yang, and J. He. A survey on mobile ad hoc wireless network. *Information Technology Journal*, 3(2):168–175, 2004.
- [55] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod. Cross-layer design of ad hoc networks for real-time video streaming. *Wireless Communications, IEEE*, 12(4):59–65, 2005.
- [56] M. Seyedebrahimi and X.-H. Peng. Investigation of phy, mac and app layers for adaptive and cross-layer optimization in ieee802.11 wlan. In *Proceedings of the 10th IEEE International Conference on Computer and Information Technology, CIT '10*, pages 2581–2588, Washington, DC, USA, 2010. IEEE Computer Society.
- [57] V. Srivastava and M. Motani. Cross-layer design: a survey and the road ahead. *Communications Magazine, IEEE*, 43(12):112–119, 2005.
- [58] P. Sudame and B. R. Badrinath. On providing support for protocol adaptation in mobile wireless networks. *Mob.Netw.Appl.*, 6(1):43–55, 2001.
- [59] T. Taleb, K. Kashibuchi, A. Leonardi, S. Palazzo, K. Hashimoto, N. Kato, and Y. Nemoto. A cross-layer approach for an efficient delivery of tcp/rtp-based multimedia applications in heterogeneous wireless networks. *Vehicular Technology, IEEE Transactions on*, 57(6):3801–3814, 2008.
- [60] H. Tian, S. K. Bose, Choi Look Law, and Wendong Xiao. Cla-qos: A cross-layer qos provisioning approach for mobile ad-hoc networks, 2005.
- [61] C. Tschudin and R. Gold. Lunar: a lightweight underlay network ad-hoc routing protocol and implementation. In *In Proceedings of the 4th International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking (NEW2AN, 2004*.
- [62] N. H. Vaidya, P. Bahl, and S. Gupta. Distributed fair scheduling in a wireless lan. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 167–178, New York, NY, USA, 2000. ACM.
- [63] T. Vanhatupa, M. Hannikainen, and T. D. Hamalainen. Evaluation of throughput estimation models and algorithms for wlan frequency planning. *Computer Networks*, 51(11):3110, 2007.
- [64] V. Vishnumurthy, T. Sandeep, B. S. Manoj, and C. S. R. Murthy. A novel out-of-band signaling mechanism for enhanced real-time support in tactical ad hoc wireless networks. In *Real-Time and Embedded Technology and Applications Symposium, 2004. Proceedings. RTAS 2004. 10th IEEE*, pages 56–63, May 2004.
- [65] C. Wang, K. Sohraby, B. Li, M. Daneshmand, and Yueming Hu. A survey of transport protocols for wireless sensor networks. *Network, IEEE*, 20(3):34–40, 2006.
- [66] Q. Wang and M. A. Abu-Rgheff. Cross-layer signalling for next-generation wireless systems. In *Wireless Communications and Networking, 2003, IEEE*, volume 2, pages 1084–1089 o.2, March 2003.
- [67] R. Winter, J. H. Schiller, N. Nikaein, and C. Bonnet. Crosstalk: Cross-layer decision support based on global knowledge. *Communications Magazine, IEEE*, 44(1):93–99, January 2006.
- [68] G. Wu, Y. Bai, J. Lai, and A. Ogielski. Interactions between tcp and rlp in wireless internet, 1999.
- [69] Z. Wu and J. Wu. Cross-layer design for packetized video communications over wireless ad hoc networks. In *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010.
- [70] Z. Wu and J. Wu. Cross-layer routing optimization for video transmission over wireless ad hoc networks. In *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010.
- [71] G. Xylomenos, G. Polyzos, P. Mahonen, and M. Saaranen. Tcp performance issues over wireless links. *IEEE Communications Magazine*, 39(4):52–58, 2001.
- [72] Y. Yamao, Y. Kida, and Y. Kadowaki. Cross-layer multi-hopping scheme for efficient and reliable transmission in fading environment. In *VTC Fall*, pages 1–5, 2010.
- [73] Y. Yang and R. Kravets. Contention-aware admission control for ad hoc networks. *IEEE Transactions on Mobile Computing*, 4(4):363–377, 2005.
- [74] S. Yuhe and X. Jie. A new cross-layer interface for manet. *Circuits, Communications and Systems, Pacific-Asia Conference on*, pages 175–178, 2009.
- [75] L. Zhou, B. Zheng, B. Geller, A. Wei, and J. Cui. Joint qos control for video streaming over wireless multihop networks: A cross-layer approach. *AEU - International Journal of Electronics and Communications*, 63(8):638, 2009.
- [76] X. Zhu and B. Girod. Distributed media-aware rate allocation for wireless video streaming. *Information Systems Journal*, 20(11):1462–1474, 2009.