Admission Control Schemes for 802.11-Based Multi-Hop Mobile Ad hoc Networks: A Survey

Lajos Hanzo II. and Rahim Tafazolli

***Abstract*—Mobile ad hoc networks (MANETs) promise unique communication opportunities. The IEEE 802.11 standard has allowed affordable MANETs to be realised. However, providing Quality of Service (QoS) assurances to MANET applications is difficult due to the unreliable wireless channel, the lack of centralised control, contention for channel access and node mobility. One of the most crucial components of a system for providing QoS assurances is admission control (AC). It is the job of the AC mechanism to estimate the state of the network's resources and thereby to decide which application data sessions can be admitted without promising more resources than are available and thus violating previously made guarantees. Unfortunately, due to the aforementioned difficulties, estimating the network resources and maintaining QoS guarantees are non-trivial tasks. Accordingly, a large body of work has been published on AC protocols for addressing these issues. However, as far as it is possible to tell, no wide-ranging survey of these approaches exists at the time of writing.**

This paper thus aims to provide a comprehensive survey of the salient unicast AC schemes designed for IEEE 802.11-based multi-hop MANETs, which were published in the peer-reviewed open literature during the period 2000-2007. The relevant considerations for the design of such protocols are discussed and several methods of classifying the schemes found in the literature are proposed. A brief outline of the operation, reaction to route failures, as well as the strengths and weaknesses of each protocol is given. This enables patterns in the design and trends in the development of AC protocols to be identified. Finally, directions for possible future work are provided.

***Index Terms*—mobile ad hoc networks, admission control, quality-of-service, QoS-aware routing, 802.11**

I. Introduction

THE INTEREST in Mobile Ad hoc Networks (MANETs) [1] has grown immensely over the last 15 years. Much hope has been placed in MANETs to provide spontaneous, robust and ubiquitous communications in areas where the provision of central infrastructure is limited or lacking. A gateway node may provide Internet access, but MANET users are typically collaborators sharing messages and content with each other. Often-suggested applications of MANETs include battlefield communications, disaster recovery, temporary gatherings such as conferences [2] and highly mobile vehicle-to-vehicle networks (VANETs) [3]. The developing world, where a higher proportion of people live in areas with limited infrastructure, could also benefit from MANET technology. In fact, the "One Laptop Per Child" project[1] is enabling the establishment of what are, to the best of our knowledge, the largest real-world MANET-like networks to date.

Manuscript received 22 April 2008; revised 15 July 2008.
The authors are with the University of Surrey (e-mail: lh800@zepler.net).
Digital Object Identifier 10.1109/SURV.2009.090406.
[1]http://www.laptop.org, accessed 18th Nov. 2008

Significant factors in the increasing interest in MANETs were the improving capabilities and ubiquitous nature of mobile devices, as well as the development of the unifying 802.11 standard [4] for wireless networking. Most laptop computers and many personal digital assistants (PDAs) now come with 802.11-compliant air interfaces. With the option to operate them in ad-hoc mode, 802.11 is the primary enabling technology of MANETs. Indeed, an increasing portion of MANET research assumes 802.11-based physical (PHY) and medium access control (MAC) layer solutions. In multi-hop MANETs, as discussed in [5], contention-free MAC schemes such as time division multiple access (TDMA) or code division multiple access (CDMA) are difficult to implement due to the lack of centralised control and the dynamically changing network topology. For these reasons, we focus on 802.11-based MANETs.

For all but the least-demanding applications, which have no critical time-, reliability- or throughput-related constraints, a mechanism for providing Quality of Service (QoS) assurances is required. Such assurances may be relative, as in service differentiation, or absolute, as in guaranteed-throughput or bounded-delay services. A full system for providing QoS assurances requires QoS-aware routing, admission control (AC), resource reservation, traffic policing, perhaps traffic scheduling, and depending on the stringency of the guarantees required, possibly a QoS-aware MAC protocol.

In this survey, we focus on one of the most important functions: admission control (AC). In general, the purpose of AC is either to admit only those data sessions whose QoS requirements can be satisfied without violating those of previously admitted sessions, or only those that allow a required average network QoS to be maintained. Although the heart of the problem is the collection of information about the available network resources, a range of related mechanisms are additionally required in order to make admission decisions. Firstly, the AC protocol must establish if there are any nodes and/or links (node pairs) that have the necessary resources available. This can be done during route discovery, as in a QoS-aware routing protocol, or as a separate operation after routes have been discovered. In fact, the route discovery procedure of many QoS-aware routing protocols can be used as a simple form of AC: if no route satisfying the application's requirements can be found, the data session is rejected. However, in a contention-based 802.11 network it is not only the nodes along a session's route that have an effect on a session's achievable QoS, or which are affected by the session, but the nodes neighbouring the route as well. Hence it is also necessary to collect information about the resources of those nodes. Thirdly, while it is vital to make

1553-877X/09/$25.00 © 2009 IEEE

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

correct AC decisions when a data session begins, these may not mean much if the QoS assurances made to the session are soon violated. Hence, while dealing with link failures and QoS-assurance violations is not strictly part of AC, some protocols, discussed later, rightly consider these situations as well. In summary, a complete AC protocol may entail the following functions: QoS-aware routing, resource state information gathering or dissemination, resource reservation, admission or rejection of service-requesting data sessions based on the first two functions, maintenance of the state information and reservations, and possibly the management of sessions experiencing QoS-assurance violations and their re-admission if their packet-sending must be interrupted.

While the basic 802.11 distributed coordination function (DCF)-based MAC layer scheme is not QoS-aware, studies on QoS-aware routing and AC protocols designed for MANETs often assume the network to be DCF-based. This is firstly because it was part of the original standard which enabled infrastructure-less wireless networks to be supported. Secondly, end-to-end QoS can only be ensured by a network layer protocol provided with information about all of the nodes along a route. By utilising a non-QoS-aware MAC protocol, the effectiveness of the network layer QoS-aware protocol is exposed. Nevertheless, protocols based on the QoS-aware enhanced distributed channel access (EDCA) scheme [4] are also considered to fall within the scope of this survey on 802.11-based solutions. Finally, we also include protocols that do not require a particular type of MAC protocol for their operation, since they are, by definition, also compatible with 802.11.

Such a survey on AC protocols for multi-hop MANETs is required because the number of proposed solutions is now fairly large, and yet, to the best of our knowledge, no comprehensive survey on the subject exists. A survey of AC-related issues and approaches for wireless cellular networks was presented in [6]. Closer to the topic at hand, solutions for 802.11e-based WLANs were surveyed in [7]. However, [7] covered infrastructure-based WLANs only. A literature search for surveys on the topic of this paper identified only [8], which includes overviews of just three protocols. The abundance of diverse AC protocols together with the lack of wide-ranging surveys provide the motivation for the work in hand.

In this paper, we do not consider AC protocols that are designed for single-hop MANETs. This is because the problems in such an environment pose much less of a networking challenge and they can often be solved by centralised approaches, which are not applicable to multi-hop MANETs. Thus, a survey on multi-hop AC protocols requires a different section of research output to be studied. Note also that the focus is on unicast protocols and multicast is beyond the scope of this work.

The structure of the main body of this paper is as follows. Section II aims to provide an overview of the most important factors and choices involved in the design of AC protocols for multi-hop MANETs. Following this, Section III lists the relevant protocols that were found in the literature, tabulates their main features, proposes several methods of classifying them and illustrates their classification according to the method adopted for this survey. Descriptions of the operation of the surveyed protocols are given in Sections IV and V. To conclude, Sections VI and VII respectively summarise the patterns and trends in the field, and highlight potential areas of future work.

II. Protocol Design Considerations

A. QoS Metrics

Many metrics employed for specifying and measuring QoS were explained in [5]. A brief recap of QoS specification metrics is given in Section II-A1, while Section II-A2 provides a new list of performance metrics for AC protocols, which were not covered in the aforementioned survey.

1) Specifying QoS Requirements: An application's QoS requirements are usually derived from its traffic specification. The requirements can typically be expressed using one or more of the following metrics:

- Minimum average throughput (bps);
- Maximum packet delay bound (s): the accumulation of the queueing and MAC delays at each node plus the propagation delay, which is relatively short;
- Maximum delay jitter bound: can be defined as the difference between the upper bound on delay (including queueing delay) and the absolute minimum delay, which is determined simply by the cumulative propagation and packet transmission times [9]. A common alternative definition is the variance of the absolute packet delay [10];
- Maximum packet loss ratio (PLR) bound; the maximum tolerable fraction of the generated data packets that are lost en-route. The packet losses could be due to buffer overflow when congestion occurs, due to the retransmission limit being exceeded during periods of poor channel quality or after a node moves, or owing to timeout while waiting for a new route to be discovered.

2) Quantifying AC Protocol Performance: In terms of metrics, an AC protocol's task is essentially a balancing act. On the one hand, it aims to serve as many users, and therefore to admit as many sessions as possible, while utilising the network's resources fully and efficiently. On the other hand, any inaccuracy in the admission decisions can result in the pledging of more resources than are available, leading to *false admissions*. It is much easier to provide a high QoS to admitted sessions if the network is under-utilised, and resources are abundant, because then the risk of congestion is averted. However, this way results in low efficiency in terms of energy consumption and overhead and a wastage of network resources. Rejecting a session which could have been served without unduly degrading the QoS of previously admitted sessions may be termed a *false rejection*.

In summary, metrics for evaluating AC protocols should reflect this inherent balance, and possible trade-off between the probabilities of false admissions and false rejections. Thus, metrics can be categorised according to whether they measure the protocol's ability to utilise resources or its ability to satisfy applications' requirements. Although most AC protocol designers tend to demonstrate their protocol's effectiveness by showing traces of throughput and/or delay versus time, this only shows the small-scale performance of the protocol.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

Metrics are needed for quantifying the performance of large-(time and space) scale systems. Some suitable metrics are as follows:

- Capacity utilisation: the average fraction (over time) of the network's capacity[2], that is utilised by data traffic. A large number of false rejections leads to a low capacity utilisation. However, the capacity of wireless networks with random topologies can be difficult to quantify. Therefore, researchers often use the aggregate network throughput to reflect the level of capacity utilisation, e.g. [11]. Admittedly, the aggregate throughput is a subjective metric and thus cannot be used to compare results from different networks, only for comparing results for different protocols operating in the same network with the same offered traffic load;
- Session admission ratio (SAR): the ratio of data sessions admitted into the network to the total number requesting admission. This metric may be used because of the difficulty in estimating capacity utilisation efficiency. As opposed to the aggregate throughput, it reflects the number of data sessions served. It exposes the ability of the AC mechanism to discover available resources and utilise them. However, note that the ability of the underlying routing protocol to find suitable routes may also affect the SAR. For given network and traffic configurations, a protocol achieving a higher SAR, while not degrading the experienced QoS of applications, can be considered better. The weakness of this metric is that it depends on the offered traffic load and the absolute network capacity. A better measure would be the number of sessions admitted per unit of capacity, if the capacity can be quantified;
- False rejection ratio (FRR): the number of false rejections normalised by the number of rejected sessions or admission requests. In practice, the FRR is difficult to quantify, since whether a rejection is deemed false or not depends on the instantaneous states of resources and a session's requirements. Evaluating a protocol's FRR would require each admission decision to be compared to a global view of the network resources, and thus it cannot be accurately undertaken on a real system, only in simulation;
- False admission ratio (FAR): the number of false admissions normalised by the number of admitted sessions or admission requests. Akin to the FRR, this metric is difficult to quantify, but many other methods are available for indicating the level of resource over-pledging. One could measure the average proportion of packets (e.g. for delay) or the fraction of time (e.g. for throughput) for which the required QoS was not upheld. In [11], the FAR is illustrated through an "actual network throughput minus the total throughput promised to admitted sessions" metric. However, both the FAR and FRR metrics are also affected by conditions outside of the AC protocol's control, such as route failures and channel-induced bit errors;
- Session completion and dropping ratios (SCR/SDR): the ratio of the number of data sessions completed to the application's satisfaction, or dropped before being terminated by the application, to the number of sessions admitted into the network. Intuitively, $SDR = 1 - SCR$. The QoS requirements and experienced QoS of a session can be used to define the session completion and dropping conditions, for example, see [12]. The SCR and SDR can then easily be monitored and can partially reflect the accuracy of admission decisions. Unfortunately, again, these metrics are affected by factors outside of the protocol's control, but can be used to monitor how well the protocol copes with these.

As stated above, some metrics, especially those related to resource utilisation efficiency, are difficult to quantify. Aside from using simulations benefiting from global state information, as mentioned above, some information theoretic tools, such as those proposed in [13], may help to solve this problem. Such methods may be used to predict performance based on a particular traffic load, and thus the maximum load that is feasible while adhering to a given a set of QoS constraints may be found.

B. Network Resources

Again, the network resources relevant to QoS have been discussed in [5]. Therefore, the list here serves as a recap while highlighting the characteristics of each resource relevant to AC.

Channel capacity: this is the most important network resource [14]. If the channel around a node is always busy, no matter how abundant its other resources are, it cannot provide any level of service. A low level of residual channel capacity results in low throughput and long channel access delays for transmitting nodes. In the literature, most protocols have historically assumed a fixed transmission rate for nodes, both for ease of analysis and because the 802.11 standard [4] does not specify a rate-switching mechanism. Therefore, capacity is often expressed in terms of bits per second (bps). However, in some situations, such as in an environment with heterogeneous link rates, residual capacity may be more usefully expressed in terms of the fraction of idle channel time detected, as detailed in the next sub-section.

Buffer space: this is the second most important resource. The total buffer space determines the maximum queue size and the actual queue size at relay nodes is a major factor in the queueing delay and hence the total end-to-end delay of a packet. If a node has no buffer space remaining, it must drop any arriving packets for which it is not the destination. A larger maximum queue size means that fewer packets will be lost during periods of congestion, albeit the average end-to-end delay could increase due to longer queueing delays.

Battery charge: MANET devices, unlike sensor network nodes, typically have regular access to recharging facilities. Therefore, battery life does not have to span months at a time. However, overhead-heavy protocols may still have a significant impact on battery life and hence may necessitate frequent recharging, limiting the usefulness of devices. In the interest of fairness, protocols could also attempt to balance

[2]As opposed to the Shannonian notion of capacity, in this work, we use the term 'capacity' more loosely in the colloquial sense to mean the data-carrying capacity of a network or route in bits per second (bps), or the capacity of a node, in bps, to transmit to another node.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

traffic loads across different routes such that no single user's battery resources are unfairly burdened.

Processor time: usually, this is a non-critical resource for AC since most algorithms are computationally simple. However, some algorithms, such as QoS-aware routing-related optimisation problems, could benefit from abundant processor time.

C. Estimating Network Resources and Achievable QoS

A crucial part of any AC mechanism is the discovery of the state of the network resources. This information is required to make admission decisions. The following three broad categories of approaches have been considered in the literature:

- Test QoS-related states during route discovery;
- infer achievable QoS from that experienced by probe packets sent on particular routes;
- use the QoS already experienced by previously-transferred data packets as an indicator of future achievable QoS.

The majority of solutions in the literature fall into the first category, as this survey will show. More specifically, the network resources, and hence the achievable QoS in terms of particular metrics can be estimated in the following ways:

- Local residual channel capacity: monitor the fraction of time the 802.11 clear-channel assessment (CCA) and virtual carrier-sensing mechanisms [4] report the channel as idle e.g. [11]. Throughout this paper, this value will be referred to as the channel idle time ratio (CITR). Then, multiply this by the raw channel capacity. Alternatively, monitor the amount of capacity consumed by transmitting and receiving and subtract this from the raw channel capacity, e.g. [15]. Both of these methods assume a known maximum transmission rate.
- Link capacity: use the delay between transmitting probe packets of known sizes to estimate the capacity, e.g. [16]. Alternatively, use the minimum of the local residual capacities of the end nodes, estimated by one of the methods described above.
- End-to-end route capacity: use the minimum of the estimated local residual channel capacities of the nodes on the route, taking the intra-route contention into account e.g [11]. Alternatively, employ the same method but using the estimated link capacities. Another alternative is to probe routes end-to-end and use the interval between packet arrivals to calculate the route capacity, e.g. [17].
- End-to-end delay: in most solutions, this metric is estimated simply by probing a route and taking half of the average round trip time experienced by a series of probe packets, or by the route discovery packet, e.g. [18]. Alternatively, the traversal times of each hop can be estimated individually, and then summed e.g. [19].
- Delay jitter: jitter can be estimated based on the delay statistics of the existing data packets, or probe packets [20].
- Packet loss ratio (PLR): for a given link, the loss ratio of periodic beacons with a known frequency can yield an estimate of the PLR [16]. Alternatively, the loss ratio of probe or data packets can be monitored.

Note that these methods do not actually predict the future QoS for a requesting session. They tend to assume that the session will be able to use the residual capacity, and experience the same delay, jitter and PLR that was measured. In fact, this assumption is correct unless collisions, unexpected congestion, route failures or inaccuracies in the admission decisions occur.

D. Challenges Posed to AC Protocols by the MANET Environment

Since the goal of AC protocols is to avoid the attempted over-utilisation of network resources, thereby upholding QoS assurances, the challenges of interest are those that prohibit or burden this operation. Many of the challenges are the same as those posed to QoS-aware routing protocols, which were discussed in [5]. However, the impact may be different in the context of AC protocols and hence we recap those challenges in the current context.

The unreliable wireless channel: received signals are prone to bit errors due to interference from other transmissions, thermal noise, shadowing and multi-path fading effects [21]. Such errors may lead to packets being undecodable. Sometimes they can be mitigated by forward error correction, or 802.11's retransmission scheme. However, persistent packet errors can result in link failure being falsely detected, leading to re-routing, lapses in throughput, increased packet delays and possible congestion, causing more packets to be dropped.

Lack of centralised control: the major advantage of an ad hoc network is that it may be set up spontaneously, without planning and its members can change dynamically and be connected via multi-hop routes. This makes it difficult to provide any form of centralised control, exacerbating the MAC problem. It also makes admission decisions much more difficult because they must be made in a distributed manner. There is no central entity to collect state information and make a fully-informed decision. Instead, as this survey will show, nodes must make decisions based on a limited-scope "snapshot" view of the network resources, leading to potential inaccuracies.

Channel contention: In order to discover network topology, nodes in a MANET must communicate on a common channel, even if the MAC protocol in use is not the single-channel 802.11 scheme. However, this introduces the problems of interference and channel contention, which determine the fraction of the channel capacity available to a node, as discussed in Section II-B. This implies that, in order to make correct admission decisions, a node must know about all the traffic that could possibly interfere with the reception of its transmissions, as well as be aware of the interference any newly-admitted traffic would introduce to the nodes in its vicinity.

In order to reduce the chance of collisions at receivers, 802.11-compliant transmitters use a carrier-sensing threshold (*cs-thresh*) to detect interfering signals at a much lower power than at which they can decode them reliably (the receiving

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

threshold, which determines the average[3] reliable transmission range). The channel is deemed busy if any signal with power above this cs-thresh is detected. Depending on the signal propagation characteristics, and the transmission power, the cs-thresh results in a particular *cs-range*. In recognition of this fact, many works on AC have considered the need to evaluate the resources of all nodes within a transmitter's cs-range (the *cs-neighbourhood*), prior to session admission (e.g. [11], [22]). This is to ensure that they have sufficient available capacity to still be able to transmit at the rate required to uphold the QoS guarantees of sessions they are carrying, even if a new session was admitted and began imposing extra interference on them.

While decreasing the cs-thresh, thereby increasing the cs-range, reduces the chance of collisions, it also decreases the spatial reuse. The level of spatial reuse determines the number of possible concurrent transmissions in the network, which determines the network capacity. On the other hand, collisions not only waste resources due to packets not being decodable, but necessitate retransmissions. The retransmission count limit being exceeded can lead to falsely-detected route failures, as discussed above in the case of channel errors. Therefore, the collision probability must be carefully balanced against the level of spatial reuse [23].

Another consequence of channel contention is mutual contention and interference between the nodes on a route forwarding the packets of a data session. This means that, depending on the number of transmitters within the cs-range of each other, a data session consumes multiple times its stated capacity requirement at each node [11], [22]. We will refer to this phenomenon as *intra-route contention*. The number of nodes that are both within a selected node's cs-range and are transmitters on the route, is termed the selected node's *contention count* for that route [11], [22].

Node mobility: the nodes in a MANET may move completely independently and randomly as far as the communications protocols are concerned. Route failures thus induced can lead to the problems listed above in the case of channel-induced errors. Mobility can also cause QoS assurance violations without breaking routes. A transmitting node may move into the sensing range of another transmitter, thereby increasing its interference, and reducing its channel access time. A data session that was admitted based on the original level of available channel time may now be starved of transmission opportunities. The session would then need to be re-admitted on a new route.

E. Design Trade-offs

In [5] we discussed several of the general design trade-offs in MANETs that may also impact QoS-aware protocol design. In this section we focus only on the trade-offs specific to AC protocol design.

Routing protocol coupling vs. decoupling: as previously stated, many AC protocols directly involve a QoS-aware routing scheme in the admission decision. Often, especially in earlier proposals, AC is purely based on the route discovery process' ability to find a route with adequate resources. In the decoupled case, the AC protocol typically assumes that a route for a service-requesting data session has already been discovered, and its task is to evaluate the route's suitability for satisfying the session's requirements. The advantages of this approach are that any routing protocol may be employed, a more simple, modular design is enabled, and the storage of state information along the route may possibly be avoided. The advantages of the latter, coupled approach, are discussed later in this section. On the other hand, decoupling the AC mechanism from the routing protocol leads to the possibility of the discovered routes not being useful since they are discovered without regard to their residual resources. This also means that the resources consumed in discovering them could turn out to have been wasted. Furthermore, decoupled protocols often do not have such fine-grained control over the network resources, since they can only accept or reject whole routes at once, even if only a single node on a route has insufficient resources for supporting a session.

MAC protocol coupling vs. decoupling: as discussed in the context of QoS-aware routing in [5], protocols can benefit from directly accessing MAC layer information. Indeed, many of the protocols that will be discussed in this paper use the MAC protocol's information about the CITR for residual channel capacity estimation. However, this necessitates a cross-layer design, which complicates system development. Alternatively, protocols may utilise only network layer information, leading to a simpler, modular design, and inter-operability with different types of MAC protocols. Admittedly, this means that they can only infer the status of MAC layer resources, such as the raw link capacity and the residual capacity by using indirect methods such as probing.

Stateless vs. stateful protocol: commonly, a protocol that maintains information about the state of a process or transaction is termed "stateful"[4]. The term "stateless" is a slight misnomer when applied to AC protocols, for example in [24], since state information relating to data sessions is still stored at source nodes, and often at destination nodes. However, running stateless protocols, intermediate nodes on a route are spared from the burden of storing and managing such information. They save memory and their operation may be less complex due to being able to do without signalling functions to reserve and release resources. Furthermore, they are often suitable for decoupling from the routing protocol since they do not require knowledge of intermediate nodes on a session's route. However, without storing state information at all nodes on a route, intermediate nodes cannot make any decisions that require "memory" about the status or experienced QoS of individual sessions. Also, resource reservations cannot be made at intermediate nodes, and therefore there is a time window between a route being tested and the session beginning to use it, during which another session could be admitted to use the same resources.

Proactive vs. on-demand vs. passive resource state discovery: as discussed in Section II-D, due to the shared nature of the channel, admitted traffic impacts and is impacted by a

[3]We refer to the average, since fast- or slow-fading-induced signal power fluctuations can cause the effective range to vary about the location-dependent mean.

[4]see the definition on http://www.isp.webopedia.com, for example

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

larger set of nodes than merely those on a session's route. Therefore, information about the resources of the affected nodes should be collected. As with routing information, such state information can be collected only when needed, or disseminated proactively on a periodic basis. The benefits and drawbacks are similar to those that are well-known in routing protocol design [1]. In brief, proactive methods generally incur a greater overhead, using up network resources, but allow faster AC, since the information is already available when session admission requests arrive. In some particular cases, proactive protocols may also be able to respond better to changes in the states of network resources. For example, if a protocol proactively discovers that the capacity available to a downstream node is decreasing, it may be able to set up an alternative route before congestion occurs. An on-demand protocol would not be able to detect this change before a problem occurred. Thus, proactive resource dissemination schemes may be better for avoiding QoS assurance violations. On-demand, or reactive methods usually incur less overhead, but a greater admission delay. In general, for AC protocols, the scalability problems of proactive routing protocols [1] are not inherited to the same degree, since information gathering is only on a local scale, within a small number of hops of each node. A third option for implicitly determining the available capacity at neighbour nodes is to passively monitor the transmissions that affect their CITR. The pros and cons will be discussed below in the context of the protocols that employ such as scheme.

Size of impacted area to consider: precisely determining the radius within which to consider the impact of the interference, and thus the radius in which to discover neighbour resource state information requires accurate knowledge of the signal propagation characteristics and the distance to all possibly-affected nodes. Acquiring this knowledge is usually impractical and therefore the area must be approximated somehow. The size of the area which is considered to impact and be impacted by a data session is also a design choice. Considering a larger set of nodes may incur a greater overhead and/or may falsely include nodes which are not impacted, resulting in overly-conservative admission decisions and hence low network capacity utilisation. Including only a small set of nodes may ignore some sources of interference, resulting in the attempted over-utilisation of resources. Examples of these potential problems are provided later. Some aspects of the accuracy achieved by employing various resource discovery radii were studied in [25].

Global vs. individual requirements: AC protocols may be designed to maintain a particular average global level of QoS, and to block any sessions that would degrade it to below the required level. Alternatively, their goal may be to serve each user's or data session's requirements individually. This design choice depends purely on the intended purpose of the system. However, ensuring global goals may be more difficult in the distributed MANET environment, since it is not easy to determine which nodes will be impacted by an admitted session.

Approach to failures: naturally, this should depend on the requirements of applications. However, AC protocol designers often treat QoS requirements with various degrees of stringency. For example, some simply state that if there is a route failure, nothing can be done but to pause the affected sessions, even if their throughput guarantee is violated. At the other extreme, sessions may simply be re-routed to any known intact route. The former approach ensures that only those sessions are affected whose routes have failed. The latter approach, simply re-routing to any known route, involves doing the utmost to uphold those guarantees which are most endangered, even if it means introducing some risk to the QoS of other sessions. This could happen when the resources of the alternative route have not been recently tested, for example. Therefore, the trade-off lies in whether to adopt a low-risk strategy and accept the violation of those QoS guarantees for which it is highly likely anyway, or to try to "save" those guarantees but introduce some risk to others.

Another problem arises when, due to mobility, the reconfiguration of the network topology or resource availability results in there being no suitable routes to serve all previously-admitted sessions. In the case where there is no best-effort traffic whose packet-sending rate can be reduced to free up resources, one of a selection of QoS conflict resolution measures must be applied [26]. The related trade-offs lie in how much risk to impose on all sessions versus how many user sessions to sacrifice, as well as in which sessions to sacrifice. For example, the protocol may reject or reduce the capacity allocated to the session using the most resources, or the least resources [26]. Other rejection schemes may focus on the newest, or oldest session, or on buffering packets that do not have strict delay constraints.

III. Protocols and their Classification

In this paper, 28 unicast AC schemes for multi-hop MANETs, which either assume, or can operate with an 802.11-based MAC layer, are described and classified. All of the surveyed protocols can be found in the open literature. Where the protocols are named in the proposing paper, we use the names given by their respective authors. Otherwise, we name them based on their features. Their full and abbreviated names, which shall be adopted throughout this paper, and a summary of their main features are presented in Tables I, II and III. In the aforementioned tables, the meanings of the various columns, left to right, are as follows: Protocol - full protocol name, acronym and reference to proposing paper; Aim - the type of service the protocol aims to provide; Routing Scheme - whether the AC protocol is coupled with it and what type of routing protocol is required to operate; MAC scheme - coupled if the AC protocol uses information directly provided by the MAC or PHY layers, and the MAC scheme it assumes; MR - Whether or not the protocol considers the fact that multiple different link transmission rates may be in use; MP - Whether or not the protocol considers various classes or priorities of traffic, and if not, all data sessions are treated as equally important; S=Stateful - does the protocol store state information at all nodes on the session's route or not, and therefore only at end-nodes?; IR and CS- the assumed intra-route contention radius and the carrier-sensing range - the radii (expressed in hops) in which the protocol considers the impact of intra-route contention on the session's capacity requirement, and the impact of admitting the session on its

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

TABLE I
ADMISSION CONTROL PROTOCOL FEATURES COMPARISON PART 1/3. PLEASE CONSULT SECTION III FOR AN EXPLANATION OF THE MEANING OF EACH COLUMN.

Protocol	Aim	Routing scheme	MAC scheme	MR	MP	S	IR	CS	Reaction to failures	Innovations
Adaptive admission control (AAC) [25]	Guaranteed throughput	Coupled, AODV-like	Coupled, DCF	N	N	Y	2	2	Pause highest-rate session if congestion detected, else rely on routing protocol	Method of dealing with congestion; Consideration of various resource retrieval ranges
Admission control-enabled on-demand routing (ACOR) [27]	Guaranteed throughput and bounded delay	Coupled, AODV-like	Coupled, DCF	N	N	Y	0	1	Tear down reservations via route error packet; Destination initiates reverse QoS-aware route discovery	Definition of delay- and throughput-related cost functions; AC based upon them
Admission control and simple class-based QoS system (ACSCQS) [28]	Guaranteed throughput and bounded delay	Coupled, AODV-like	Decoupled, Any	N	N	Y	0	0	Notify source node which attempts re-admission	Improvements over QoS-AODV in estimating end-to-end delay and detecting QoS assurance violations
Admission control and reservation management protocol (ACRMP) [29]	Guaranteed throughput	Coupled, AODV	Coupled. DCF	N	N	Y	2	2	Quickly tear down obsolete reservations and attempt local route repair	Determination of contention counts by high-powered transmission of RRep packets; Management of capacity reserved at cs-neighbours through periodic high-powered forwarding of selected data packets
Ad hoc QoS on-demand routing (AQOR) [18]	Guaranteed throughput and bounded delay	Coupled, AODV-like	Decoupled, Any contention-based	N	N	Y	1	0	Destination initiates reverse QoS-aware route discovery	Consideration of intra-route contention between neighbour nodes for CSMA/CA-based MANETs; Source and destination clock offset-based delay estimation and QoS constraint violation detection based thereon; Method of recovery from failures
Contention-aware admission control protocol (CACP) [11]	Guaranteed throughput	Coupled, DSR-like source routing	Coupled, DCF	N	N	Y	2	2	Discover new route and/or downgrade session throughput requirement	Neighbour capacity-testing in cs-range (two hops) before session admission; Consideration of intra-route contention between cs-neighbours
Contention- and capacity-aware ad hoc on-demand distance vector (AODV) CCAODV [15]	Guaranteed throughput	Coupled, AODV-like	Coupled, DCF	N	N	Y	2	0	Quickly release reserved resources; discover new route	AODV HELLO message-based dissemination of channel usage information; Mechanism for quickly releasing correct amount of reserved resources upon route failure
Distributed admission control for MANET environments (DACME) [17]	Guaranteed throughput with robustness through use of multiple paths	Decoupled, variant of DSR	Coupled, EDCA	Y	N	N	0	0	Rely on pre-probed backup routes, else discover new routes	Use of the delay between the arrivals of multiple probes to establish route capacity in a stateless AC scheme; The sending of QoS-sensitive data over multiple disjoint probed routes for redundancy
Hierarchical routing-based admission control (HRAC) [30]	Guaranteed throughput	Coupled, hierar-chical routing	Coupled, DCF	N	N	Y	0	0	Rely on pro-actively maintained routing backbone and re-establish session state	Use of super-node structure-maintaining control packets to distribute neighbour capacity usage information for determining residual capacity and hence for performing AC
Interference-based fair call admission control (IFCAC) protocol [31]	Guaranteed throughput with fair share of channel capacity	Coupled, DSR or any reactive scheme	Coupled, DCF	N	N	Y	0	2	Discover new route	Method of allocating fair share of capacity to each transmitter within cs-range

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

TABLE II
ADMISSION CONTROL PROTOCOL FEATURES COMPARISON PART 2/3. FOR AN EXPLANATION OF THE COLUMN MEANINGS, PLEASE CONSULT SECTION III.

Protocol	Aim	Routing scheme	MAC scheme	MR	MP	S	IR	CS	Reaction to failures	Innovations
INSIGNIA and temporally-ordered routing algorithm (INORA) [32]	Guaranteed throughput	Partially coupled, TORA	Decoupled, Any	N	N	Y	0	0	Attempt local re-route, then increase scope of route repair search	Combination of INSIGNIA with TORA; AC over multiple routes with traffic splitting
In-band signalling in ad hoc networks (INSIGNIA) [33]	Guaranteed throughput	Any	Decoupled, Any	N	N	Y	0	0	Use in-band signalling to adapt resource reservations; depend on routing protocol if route fails	Setup and maintenance of QoS-sensitive session states and resource reservations using in-band signalling in MANETs
Interference- and QoS-aware optimised link state routing (IQOLSR) [34]	Guaranteed throughput	Partially coupled	Coupled, DCF	N	N	Y	0	2	Proactive mechanism automatically provides new route if one exists	Consideration of two-hop cs-neighbours' available capacity during AC running over a QoS-aware version of OLSR
Multi-path admission control for mobile ad-hoc networks (MACMAN) [35]	Guaranteed throughput	Coupled, DSR-like	Coupled, DCF	N	N	Y	2	2	Re-route to pre-tested backup path	Capacity-tested backup route maintenance in CSMA/CA network; Calculation of contention difference between primary and backup routes
Multi-priority admission and rate control (MPARC) [36]	Guaranteed throughput to highest-priority traffic	Decoupled, any source-routed scheme	Coupled, Variant of EDCA	N	Y	Y	3	3	Reject low-priority real-time sessions; depend on routing protocol if route fails	Analytical model for each node's allocable capacity based on heterogeneous traffic priorities; AC based thereon
Multi-rate- and contention-aware admission control protocol (MRCACP) [37]	Guaranteed throughput	Coupled, LUNAR	Coupled, DCF	Y	N	Y	2	2	Periodically refreshes routes anyway	Method of considering heterogeneous link-rates during AC; Consideration of parallel transmissions by admitting node and nodes outside the cs-range whose transmissions are being sensed as part of the passive capacity monitoring scheme
Perceptive admission control (PAC) [38]	Guaranteed throughput	Can be coupled with any	Coupled, DCF	N	N	Y	2	2	Pause session if available capacity drops; Rely on routing protocol if route fails	Consideration of maximum collision-causing interference range in a passive residual capacity monitoring scheme
Priority-based distributed flow AC (PDAC) [39]	Guaranteed throughput to highest-priority traffic	Decoupled, DSR	Can be coupled, DCF	N	Y	Y	0	0	Report failure to source which re-routes	Combination of DSR flow-state extension with AC features
Passive measurement-based admission control (PMAC) [40]	Bounded delay and packet loss ratio	Decoupled, any reactive scheme	Decoupled, Any	N	N	N	0	0	Report measured QoS, block admission for classes of traffic experiencing poor QoS	AC of heterogeneous traffic types based on observation of past QoS for the same traffic class
QoS admission control routing protocol (QACRP) [41]	Guaranteed throughput	Coupled, AODV-like	Coupled, DCF	N	N	Y	2	1	Discover new route	Re-balancing of accuracy/overhead compared to CACP and CCAODV
QoS-aware AODV routing-based admission control (QAODV-AC) [42]	Guaranteed throughput	Coupled, AODV-like	Coupled, DCF	N	N	Y	1	1	Discover new route	Method of considering node idle time and saturation throughput during AC

route's neighbouring nodes' resources; Reaction to failures - how the protocol deals with route failures and other causes of QoS assurance violations; Innovations - the features of the protocol that were not seen in proposals published before it.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

TABLE III
ADMISSION CONTROL PROTOCOL FEATURES COMPARISON PART 3/3. FOR AN EXPLANATION OF THE COLUMN MEANINGS, PLEASE CONSULT SECTION III.

Protocol	Aim	Routing scheme	MAC scheme	MR	MP	S	IR	CS	Reaction to failures	Innovations
QoS-aware AODV routing (QoS-AODV) [19]	Guaranteed throughput and/or bounded delay and/or bounded delay jitter	Coupled, AODV-based	Decoupled, Any	N	N	Y	0	0	Notify source node which attempts re-admission on new route	Framework for AODV-based RReq/RRep-utilising AC in MANETs
QoS protocol for ad hoc real-time traffic (QPART) [43]	High probability of guaranteed throughput and/or bounded delay	Decoupled, any	Coupled, DCF	N	Y	Y	0	0	Adapt session's virtual contention window and reject low-priority sessions	Virtual contention and contention window adaptation schemes; Rejection of sessions after admission based on age and nodes' available capacity
Robust flow admission and routing (RFAR) protocol [44]	Guaranteed throughput for session's duration; bounded delay	Coupled, any reactive scheme	Decoupled, Any contention-based	N	N	Y	0	0	Attempt to avoid admission on routes that are likely to break while session is active, so failures are not anticipated	Robust throughput metric; Network capacity and session blocking probability model; Mobility and SINR-related robustness-constrained AC
SoftMAC [16]	Guaranteed throughput	Coupled, DSR-like	Decoupled, DCF	Y	N	Y	1	1	Erase resource reservations; Discover new route if route fails	Link capacity estimation using probes of various sizes; Consideration of heterogeneous link rates and collision rate (ascertained via beacons) in the calculation of residual capacity and thus in AC
Staggered admission control (StAC) [12]	Guaranteed throughput	Partially coupled, DSR-like	Coupled, DCF	N	N	Y	2	2	Re-route to known route thought to support highest throughput	Consideration of unpredictable network conditions via gradual session admission; Both routing-coupled and decoupled AC
Admission control for stateless wireless ad hoc networks (SWAN-AC) [24]	Guaranteed throughput	Decoupled, any reactive scheme	Decoupled, DCF	N	N	N	0	0	Mark packets of sessions experiencing congestion, notify source which attempts to re-admit, else terminates	Stateless AC and congestion-based regulation of admitted sessions
Time-based admission control (TAC) [45]	Guaranteed throughput	Coupled, AODV-like	Coupled, DCF	N	N	Y	2	1	Notify source which pauses affected sessions and attempts to re-route and re-admit	Method of considering channel access time wastage due to back-off and collisions

As with most types of communications protocols, AC methods can be categorised in several ways. These often relate to the choices made regarding the design trade-offs discussed in Section II-E. Firstly, protocols can be classified based on whether they are coupled with or decoupled from the routing protocol. Secondly as stateful or stateless. Thirdly, AC approaches could be categorised by the QoS metric(s) of choice. Admittedly, this would yield rather uneven categories, since most protocols emphasise throughput as the most important QoS requirement. Fourthly, by their goal: whether it is to guarantee the requested QoS to data sessions, or just to block sessions that would take the average QoS below a particular threshold. As a fifth criterion for categorisation, we also propose to use the basis for making admission decisions. These decisions are typically based on a prediction of a session's achievable QoS relative to its requirements. Such predictions may, in turn, be based on the observed QoS of previously-admitted sessions, the QoS experienced by probe packets traversing a route, or the states of the resources of both the nodes on the route and those neighbouring nodes (cs-neighbours) that would be directly impacted by the session's admission. In the case of decisions being based on resource state discovery, this can be achieved either on-demand, periodically in a pro-active manner, or continuously in a passive manner. Details of these methods will be presented in the sections to come.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

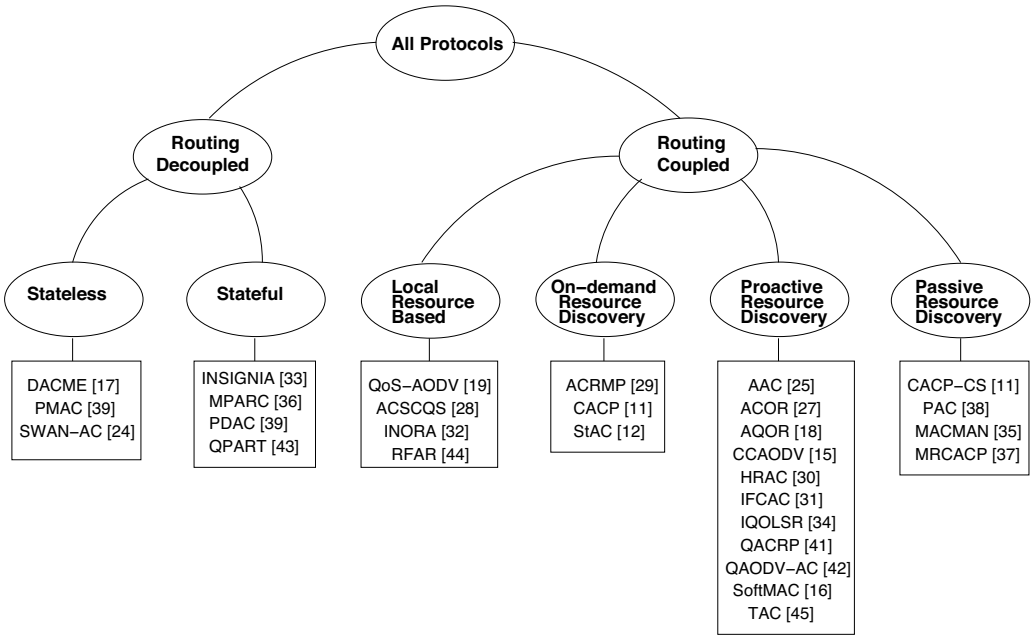

Fig. 1. Classification of the surveyed admission control protocols based on whether or not they are coupled with the routing protocol, their statefulness, and their approach to discovering network resources to aid in making admission decisions. Note that all routing-coupled protocols are stateful. The list of definitions of these acronyms is found in Section III.

While studying the operation of the various protocols, it became clear that they were not easy to categorise if the aim was to group those with the most similar operation. For this reason, a hybrid classification method is adopted for this paper, as shown in Figure 1. It has been found that the design choice that has the greatest effect on a protocol's features and operation is whether or not to couple it with a routing protocol. Therefore, the top-level categorisation is based on this. Decoupled protocols can either be stateless or stateful, while routing-coupled protocols are found to always store state information at intermediate nodes. The coupled protocols are classified based on whether and how they collect information on the state of the network resources in order to make admission decisions. Protocols that only consider the availability of resources at a route's constituent nodes, and not at the route's neighbours are categorised under "local resource-based".

In the following sections of this paper, the protocol descriptions are grouped into sections based on the classification method of Figure 1. Within each section, a subsection for each protocol briefly describes its AC mechanism, its method of dealing with mobility and QoS violations, as well as its particular advantages and shortcomings. Within each category, the protocols are described in a logical order that highlights links between their designs. At the end of each section, the common benefits and drawbacks of that category of approaches are summarised.

IV. Admission Control Schemes That Are Decoupled From The Routing Protocol

The AC procedures described in this section all assume that a route for a requesting session has been discovered prior to testing its resources.

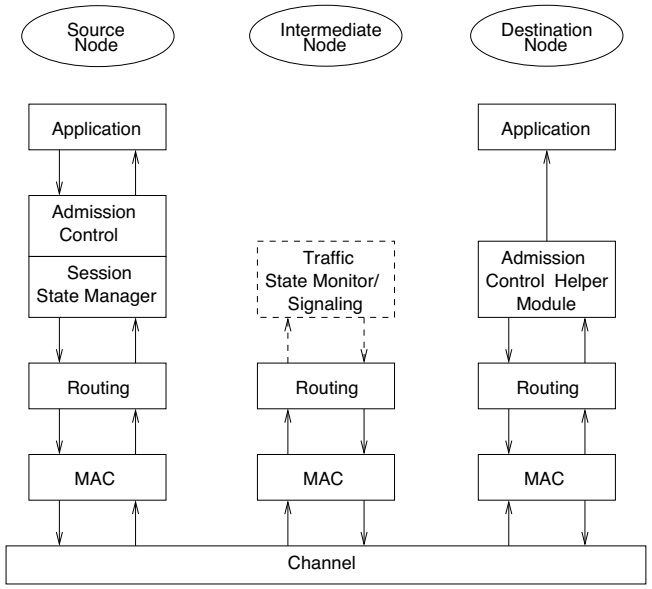

Fig. 2. Simplified functional block diagram of a routing-decoupled, stateless admission control scheme. Such schemes typically require no information storage at intermediate nodes, although some protocols may make use of some traffic monitoring and packet marking functions. The arrows represent the passing of data and control packets between modules.

A. Stateless Schemes

This sub-section deals with AC schemes that are stateless, and do not store any information regarding data sessions at intermediate nodes. They treat the route as a "black box" and admission decisions are made based on "probing" of the route by previously-admitted data traffic, or dedicated probe packets. Figure 2 illustrates the relationship between the functional blocks typically involved in such an AC scheme.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

1) Admission Control for Stateless Wireless Ad hoc Networks: The authors of [24] introduced (besides a rate control scheme), an AC mechanism for stateless wireless ad hoc networks (SWAN-AC). When a new data session requires admission, a probe packet is used to test a pre-discovered route. Each node forwarding the probe on the route estimates the amount of extra traffic (above its current load) it can support by using an analytical model to predict the level of traffic that would trigger excessive packet transmission delays at the MAC layer. The bottleneck achievable throughput on the route is stored in the probe packet header and returned to the source node, which admits the requesting session if the route can support its throughput requirement.

If a node detects congestion-related conditions, such as its buffer beginning to fill up due to re-routing or false admissions, it begins to mark data packets as having experienced congestion. The destination notifies the source on receiving such labeled data packets. The source then attempts to re-admit the session after a random amount of time.

Since all protocol decisions are made based on the current status of traffic and commands are delivered via packet headers, it is clear that the storage of state information at intermediate nodes is easily avoided. The remaining strengths and weaknesses of SWAN-AC are common to protocols in this category and hence are covered at the end of this section.

2) A Passive Measurement-based Approach to AC: The passive measurement-based AC (PMAC) protocol, presented in [40], makes decisions constrained by the measured PLR and end-to-end delay experienced by data packets. In this scheme, assuming a route is known, when a session request first arrives, the source node marks each data packet with a sequence number and a time-stamp (assuming a global clock). Initially, the packets are simply admitted and the destination node monitors the average end-to-end delay using the timestamps and the time of each packet's receipt. It also monitors the PLR by examining the sequence numbers to see which packets are missing. Newer measurements are weighted with greater importance than older ones. From these values a "path severity" metric is calculated. On any changes greater than a threshold, the severity value is reported to the source node. In this way, each source and destination pair knows the level of QoS it can expect on the route between them in terms of delay and PLR. Such path severity information is accumulated for various types of data sessions. A node chooses to reject or admit new sessions based on these values and the session's QoS requirements.

Advantages of this protocol which are not common to this whole category are the consideration of multiple QoS metrics and the fact that no extra overhead packets are introduced as part of the AC procedure. However, there is an inherent delay in reporting QoS violations. This can lead to false admissions and/or false rejections of new sessions. Furthermore, on network start-up, it takes some time for sufficient information to be collected for AC decisions to be made for each class of traffic and by each source-destination pair.

3) Probing-based Multi-Path Admission Control: The method referred to as distributed admission control for MANET environments (DACME) by its authors, was presented in [17], [20]. The DACME protocol assesses the achievable QoS on a given route by means of a set of back-to-back probe packets. In [20] methods are described for estimating route capacity, end-to-end delay, and delay jitter, when DACME is operating with a single-path routing protocol. In [17], the emphasis is on throughput-constrained applications only, but a multi-path routing protocol is considered. The authors state that the optimal operating environment for DACME is based on an EDCA MAC scheme, but it can also operate with a non-QoS-aware MAC [20].

In DACME, a route's residual capacity is estimated based on the average inter-arrival time between the probe packets sent from the source of a session. The end-to-end delay is determined to be half of a probe/probe reply's round-trip time. Jitter can be estimated, again, by the probe packets. For this purpose, the source explicitly notifies the destination of its packet sending rate, so that the expected inter-packet interval can be calculated. Alternatively, if there is already traffic on the path, jitter is estimated based on that experienced by data packets. In any case, once the achievable QoS in terms of a particular metric has been estimated, the destination informs the source node, which can make admission decisions.

The route-probing scheme of DACME makes it suitable for combination with a multi-path routing protocol. In [17], an extended version of the dynamic source routing (DSR) protocol [46], multi-path DSR (MDSR) is employed. This allows it to discover more routes than DSR by forwarding RReqs, even if they have been previously seen, as long as they arrive from different upstream nodes to earlier-seen copies. Multiple RReps are returned to a session's source and all discovered routes are cached, regardless of their nodes' resources.

The MDSR protocol splits the traffic of a session over at least two routes, increasing robustness. In order to be able to react quickly to failures, known routes are probed as soon as possible. This way, by the time a failure occurs, the achievable throughput on various routes is already known. A decision on whether an affected session can be re-routed, or should be dropped, can thus be made quickly. A weakness of this approach is that, until traffic is actually being carried on the two routes, there is no way to predict the effect of the inter-route interference on the achievable QoS. Owing to this inter-route interference, the achievable throughput could be much lower than that predicted by the probing scheme.

4) Common Advantages and Drawbacks of Stateless Schemes: The lack of state information storage at intermediate nodes means that they save memory and their operation can be less complex, as mentioned in Section II-E. With the exception of SWAN-AC, intermediate nodes do not even need to implement any of the protocol's functionality, since all protocol operations are performed at the source and destination nodes. It also makes protocol deployment easier if intermediate nodes require only the standard routing functionality. As disadvantages, the lack of reservations at intermediate nodes, and their reduced capabilities due to the lack of state information, were also mentioned in Section II-E.

B. Stateful Schemes

This section considers AC methods which are, again, decoupled from the routing protocol, but they do store state

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

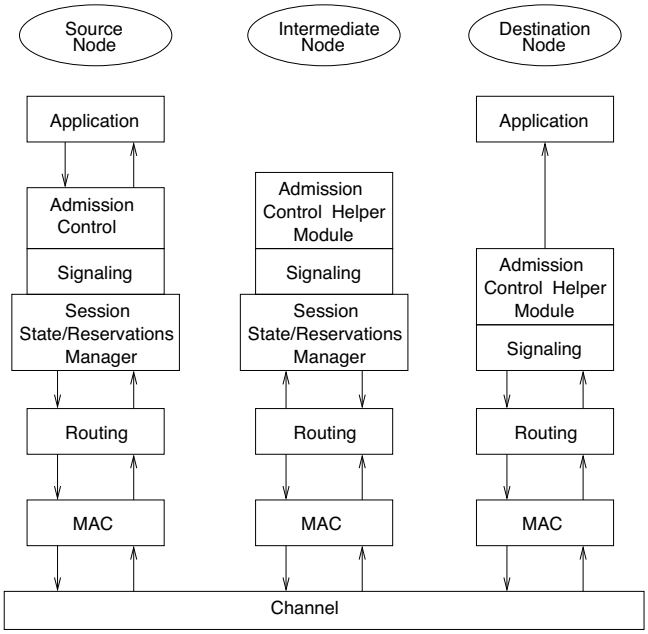

Fig. 3. Simplified functional block diagram of a routing-decoupled, stateful admission control scheme. Such schemes typically require reservation signalling functionality at most nodes, as well as session state management at intermediate nodes, but use pre-discovered routes. The arrows represent the passing of data and control packets between modules.

information at intermediate nodes. Figure 3 illustrates the relationship between the functional blocks typically involved in such an AC scheme.

1) In-Band Signalling-Based Admission Control: An in-band signalling in ad hoc networks (INSIGNIA) framework is presented in [33]. The framework specifies several traffic management mechanisms, but in keeping within the scope of this paper, we focus on the AC-related operation.

In-band signalling refers to the carrying of control information in data packets headers, as opposed to in separate control packets. With INSIGNIA, applications can specify a basic level of QoS, the minimum throughput they require and an enhanced level, the maximum they can benefit from. In fact, INSIGNIA admits all data sessions but its AC mechanism determines whether the admission is on a best-effort, basic QoS or enhanced QoS basis. A session's data packets carry resource reservation requests to intermediate nodes. Nodes reserve resources and mark packet headers according to their residual resources and the QoS they can provide. The destination notifies the source of the lowest common level of service provided by the nodes on the route. The source notifies all nodes of the level of service the session will then use, so that unused resources can be freed. Reservations are soft state and so they are also erased if not periodically refreshed by data packets.

Network dynamics are easily coped with by the in-band signalling mechanism. An increase in the interference due to mobility, which reduces the available capacity is detected by intermediate nodes, which notify the destination of the new QoS that they can support. This node in turn informs the source node that it should adapt its sending rate. Route failures are left to be dealt with by the underlying routing protocol, and data packet headers are again used to set up resource reservations at the new intermediate nodes.

This protocol requires little overhead to operate and can adapt quickly to network dynamics. However, it is rather lenient in terms of AC, as even "rejected" sessions are still admitted, albeit at a best-effort level of service. The problem is that best-effort traffic still consumes network resources. Therefore, INSIGNIA is not suitable for supporting applications with more stringent QoS requirements. Due to its decoupling from the MAC and routing protocols, the method of estimating available resources is not specified. It is therefore assumed that this could be done via CITR and/or queue size monitoring.

2) Dynamic Contention Window-Adapting Reverse Admission Control: A rather unique approach to AC is presented in [43]. The protocol it is part of is referred to as the QoS protocol for ad hoc real-time traffic (QPART). Initially, QPART admits all traffic automatically, and with a low priority. Each session's priority is increased periodically. Each node monitors the CITR at the MAC layer and each priority level is mapped to some specific CITR threshold level, with higher priorities being mapped to lower thresholds. If a given CITR threshold is reached, indicating decreasing node resources (residual capacity), all sessions that are being carried, which have the corresponding priority level, are rejected. This explains the term "reverse" AC since the states of resources are tested *after* admission.

The first data packet of a session informs intermediate nodes of its QoS requirements. QPART attempts to satisfy these requirements via an algorithm similar to the EDCA [4] scheme, except at the network layer. All data sessions have their own virtual queue and contention window size and data packets contend internally at each hop for the opportunity to be passed down to the packet scheduler which will send them down to the MAC protocol. For delay-sensitive sessions, the end-to-end delay bound is divided by the route length to determine the node traversal time limit. If the packets of a session are exceeding this limit, its virtual contention window size is decreased. If the delay is lower than required, the window size is increased. Similarly, for throughput sensitive applications, the packet sending rate relative to its requirement determines the session's virtual queue size and the contention window size is adapted based on this.

This protocol ensures that sessions which cannot be supported are quickly rejected and that older sessions are less likely to be rejected. The authors argue [43] the case for rejection after admission by saying that the achievable QoS is difficult to predict accurately and the guarantees made could be quickly invalidated by mobility anyway. Therefore, it is better to avoid creating the overhead used for node resource discovery. Indeed, QPART is completely overhead-free, except for some small header extensions on the first few data packets of a session. The virtual contention window adaptation scheme dynamically adjusts the chance of packet transmission based on a session's experienced QoS and its requirements. On the other hand, this scheme is more suited to applications which can tolerate frequent, albeit short-term, drops in QoS, since traffic is admitted without knowing if it will cause disruption to the QoS of other sessions.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

3) Multi-priority Admission Control: The work in [36], [47] detailed a unique capacity allocation model for a network handling sessions with heterogeneous priorities and a protocol that makes use of it called Multi-Priority Admission and Rate Control (MPARC). In a manner somewhat similar to the 802.11 standard's EDCA [4], the model assumes that different classes of data are differentiated by their minimum contention windows, and also by their unique frame sizes. It is assumed that each node carries only one class of data, though the classes may vary between nodes. The amount of capacity available to each node then depends on the traffic loads, minimum contention windows and frame sizes of all nodes in its cs-range.

Each node broadcasts periodic beacons containing all of this information. MPARC assumes that the cs-neighbourhood radius is three hops. The aforementioned information about all of a node's two-hop neighbours is included in its beacons. This way, each node learns all that is necessary to calculate the amount of capacity it can allocate for real-time sessions without unduly disrupting the QoS of the traffic carried by the nodes in its cs-neighbourhood.

The aim of MPARC is to ensure that no admitted session degrades the throughput of previously-admitted sessions that have an equal or higher priority. However, the throughput of lower-priority sessions may be degraded.

MPARC can utilise any ad hoc routing protocol, such as DSR [46] for route discovery. Also, existing signalling protocols, such as INSIGNIA (Section IV-B1) are suggested for use in reserving and freeing resources along a route. Once a route for a QoS-sensitive session has been found, MPARC attempts to make soft capacity reservations along it using a reservation request packet. The request only reaches the destination, and hence the session is only admitted, if the available capacity of each node, determined by the aforementioned model, is sufficient to support the session.

Since MPARC is not involved in routing decisions, it does not provide features for dealing with route failures. However, unlike the other decoupled protocols, it benefits from routing knowledge, such as the number of hops on the route, allowing it to factor the levels of intra-route contention into a session's capacity requirement. Also, it proactively discovers the resources of nodes neighbouring the session's route. Although this incurs significant overhead, it allows MPARC to ensure that an admitted session does not consume too much of the capacity of cs-neighbours. Another advantage of MPARC over the other protocols in this survey is its accurate consideration of the capacity that should be allocated to data sessions, depending on their priority. Its shortcomings that are common to this category are discussed later.

4) Priority-Aware DSR Flow State-based Admission Control: In [39], the priority-based distributed flow AC (PDAC) protocol is proposed. It builds upon the flow-state extensions of the latest version of DSR [48]. Once a route has been discovered by DSR, flow state information may be established at each node on the route to avoid the need to include the source route in each data packet header, thereby reducing overhead [48]. PDAC employs the flow state establishment packet to carry an admission request and test a route's available resources, namely capacity. Each node's capacity may be estimated by the MAC protocol, based on the CITR. Alternatively, if the designer wishes to avoid cross-layer interactions, the raw channel capacity can be shared equally between a node and its interference-imposing neighbours. In PDAC, each data session is assigned a priority based on its QoS requirements. Each node only forwards the admission request if it has sufficient available capacity, or if it can make sufficient capacity available by rejecting sessions that are of a lower priority than the new requesting session. If the request reaches the destination, a response is returned to the source. This triggers the rejection of lower priority sessions as required, whose source nodes are notified. Soft-state capacity reservations are also made at the intermediate nodes.

The main advantages of PDAC are its ease of implementation with the existing DSR protocol and its low overhead. The authors of [39] state that they are aware of CACP's [11] (discussed in Section V-B1) testing of the capacity of cs-neighbours, but chose not to implement a similar mechanism in order to reduce protocol overhead. This may result in false admissions. PDAC's further drawbacks are discussed below.

5) Common Advantages and Drawbacks of Stateful Schemes: For the most part, the positive and negative traits of stateful schemes, as opposed to stateless ones, are as discussed in Section II-E and the opposites of those stated in Section IV-A4. To elaborate with examples, the protocols described in this section were able to reserve resources by storing session state-related information at intermediate nodes. Especially in the case of the heterogeneous priority-aware protocols, MPARC and PDAC, this made them more versatile. For example, it allowed them to hold back resources from general network traffic, then transfer the right to their use when a more important, higher priority session required them. In the case of MPARC, state information storage allowed it to manage the amount of capacity allocated to nodes within each others' cs-neighbourhood sets. In the case of QPART, it enabled the dynamic fine-tuning of each session's QoS through its virtual contention window at each node on the route. Statefulness also allowed INSIGNIA-implementing intermediate nodes to remember what level of QoS had been promised to various sessions. Utilising a stateless approach, none of these features could have been implemented.

C. Common Advantages and Drawbacks of Routing-Decoupled Schemes

Due to them being decoupled from routing decisions, most of the AC procedures in this category benefit from the versatility of being able to operate in conjunction with any routing protocol. Furthermore, since AC decisions are based on the QoS experienced by packets on a pre-selected route, with the exception of MPARC, these protocols are not burdened by the consideration of the impact of using the route on the nodes surrounding it. This saves the often broadcast-natured overhead that is usually incurred by protocols testing the resources of neighbouring nodes. At most, protocol overhead consists of a few small probe packets or data packet header extensions per requesting session. However, not testing cs-neighbour nodes means that there is often no attempt to ensure that a newly-admitted session does not cause so much

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

interference that their traffic is starved of channel access opportunities.

Decoupling the AC decision from the routing protocol can also result in the discovered routes not being useful to any sessions due to a shortage of route resources. This would mean that the resources used during route discovery were wasted. The granularity of AC decisions made by routing-decoupled protocols is also generally less fine than with coupled protocols. This is because the former must discard whole routes if even a single node has insufficient resources. By contrast, coupled schemes could route around that node using this knowledge. Furthermore, methods relying on the type of "probe" packets that have been discussed for establishing a route's achievable QoS cannot take into account the level of intra-route contention that would occur. This obviously results in a session's capacity requirement being under-estimated, and thus possibly, false admissions. Finally, decoupled protocols are forced to rely on the routing protocol to recover from route failures, which usually means that the admission process must start from the beginning each time a session is re-routed.

V. Admission Control Schemes That Are Coupled with the Routing Protocol

In this section, we summarise the operation of AC schemes which are coupled with the routing protocol and thus route discovery is contingent upon adequate resources being deemed to be available at each node. Such schemes typically require almost full functionality at all nodes in order to manage session states, make admission decisions, and communicate these to the source and destination nodes. The relationships between the functional blocks typically involved in such an AC scheme are similar to those for routing-decoupled stateful schemes, shown in Figure 3, except that the routing module is not separate from the session state/reservations manager or signalling modules.

A. Local Resource Availability-Based Schemes

This sub-category of protocols make admission decisions based on evaluating only the locally-available resources of each node during route discovery. As with decoupled protocols, these schemes typically do not consider the impact a newly-admitted session may have on nodes that are not on its route.

1) Admission Control Employing In-Band Signalling and the Temporally-Ordered Routing Algorithm: The work in [32] proposes INORA, a combination of the temporally-ordered routing algorithm (TORA) [49], with the INSIGNIA framework, described in Section IV-B1. In INORA, routing information, modelled as an acyclic directed graph rooted at the destination, is assumed to have already been discovered by TORA. When a session request arrives, the data is automatically admitted and the INSIGNIA component attempts to set up soft-state reservations. The data follows the directed graph set up by TORA. If an intermediate node detects that it has insufficient available channel capacity (e.g. by comparison to the CITR) or its queue is full beyond a threshold level, it notifies the previous node on the route. The previous node then attempts to route the session via a different downstream node. If all of the intermediate nodes' resources are sufficient to support at least the session's minimum required throughput, reservations are set up along the path, as in INSIGNIA.

While INORA is trying to find a suitable route, the session's packets are forwarded at a "best-effort" level of service. The routing table at each node is updated in order to map each data session flowing through it to the next hop node that can support the session.

A more fine-grained AC and routing scheme is also proposed in the same paper [32]. In this scheme, the difference between an application's basic and enhanced capacity requirements (see Section IV-B1) is split into a number of levels or classes. This scheme is similar to the one described above, except that each node may inform the previous hop of the level of throughput it can support. The previous hop then attempts to split the session over as many next hops as are required to fully satisfy the end-to-end throughput requirement.

The main advantages of INORA over INSIGNIA are as follows. Firstly, it can re-route a session locally if any nodes are found to be unable to support the session's throughput requirement. Secondly, multiple routes can cooperatively support the session. However, this is only under a simplified interference model. Since inter-route interference is not considered, akin to DACME (Section IV-A3), a session could degrade its own QoS through interference when it is split over multiple routes. Additionally, even though, as opposed to decoupled protocols, INORA can re-route sessions based on individual nodes' states of resources, it does not stop the routing procedure from discovering routes without testing their resources in the first place. As stated in Section IV-C, this can cause routing information to go to waste.

2) QoS-AODV : The QoS-aware extensions [19], [50] to AODV [51] laid the foundations for the route request/reply-based admission decision procedure that prevails in most of today's AC protocols for multi-hop MANETs. These extensions specify that, if an application data session has constraints on the maximum end-to-end delay or delay jitter it can tolerate, or requires a minimum level of throughput, it must specify these requirements in a route request (RReq) header extension when seeking a route. An intermediate node receiving the RReq may only rebroadcast it if it can satisfy the QoS requirements specified in the header extension. Since a node may not have up-to-date information about the QoS-related states at downstream nodes, it should rebroadcast the RReq, even if it knows a route to the destination.

Delay-constrained route discoveries are handled by having each node forwarding a RReq subtracting its "node traversal time" from the maximum end-to-end delay bound, until the RReq either reaches the destination or the difference between the delay bound and the accumulated node traversal times reaches zero. In the second case, the RReq is dropped and the requesting session is not admitted. If the RReq reaches the destination, that node replies to the source with a route reply (RRep). Throughput-constrained route discoveries proceed in a similar manner, except that the RReq only reaches the destination if each forwarding node has sufficient available capacity to support the requesting session. In the RRep stage, the bottleneck residual capacity on the route is recorded in the RRep header. On receiving the RRep, the source admits the

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

session if the bottleneck achievable throughput is adequate. A jitter constraint is handled in a similar manner again.

Each intermediate node also stores the IP addresses of source nodes requesting various levels of QoS. If the node finds it can no longer support these requirements, an ICMP QOS_LOST message is sent to the sources of any affected sessions. Source nodes receiving such a message may attempt to re-admit the affected sessions by seeking an alternative route.

In fact, QoS-AODV is not considered to be an AC protocol, only a QoS-aware routing method. Indeed, the methods of estimating the node traversal times and residual channel capacities are not specified in [19], [50]. However, QoS-AODV provides a framework for RReq/RRep-based AC, since session admission is contingent upon finding a route that is able to satisfy its QoS requirements. The QoS-metric constrained route discovery mechanism described above shall henceforth be referred to as QoS-AODV-style route discovery.

3) An Improvement over QoS-AODV: The admission control and simple class-based QoS system (ACSCQS) proposed in [28] incorporates some simple extensions to QoS-AODV. We focus on its AC procedure. As in QoS-AODV, when searching for a constrained route for a new session, the RReq carries the session's throughput requirement. However, instead of also carrying its delay constraint, this is stored at the session's source. On sending out the RReq, the source sets a timer to expire after twice the session's delay bound.

The protocol performs QoS-AODV-style route discovery, except that each intermediate node checks only that its residual capacity is sufficient as the condition for forwarding the RReq. The method of estimating the residual capacity is again not specified in [28], although we assume that it can be based on the CITR. On receiving the RReq, the destination sends a RRep back to the source, which must arrive before the aforementioned timer expires, in order for the session to be admitted.

Once the session is admitted, each intermediate node monitors the rate at which it is receiving the session's data. If this is less than the session's specified minimum throughput requirement, a route error message is sent to the source, which must find a new route. The protocol also periodically verifies that the session's end-to-end delay requirement is being upheld. To do this, it sends a special type of RRep which must be acknowledged by the destination. Again, if it does not arrive within twice the delay bound, the session must be re-routed.

The ACSCQS provided some simple improvements over QoS-AODV, such as its initial method of testing the end-to-end delay and continuous testing of the experienced QoS. However, as highlighted by later approaches, the AC procedure was overly-simplistic. The method of establishing a node's available capacity was not specified. Other shortcomings are discussed at the end of this section.

4) Robustness-Constrained Admission Control: While many other works covered in this survey attempt to recover quickly from route failures, the authors of [44] argue that it is better to admit sessions only using routes that are likely to remain intact for the duration of the session. To this end, a route robustness metric is defined as a function of the expected time until one of the route's links fails due to signal fading or node mobility. The average time until fading is assumed to be an exponentially distributed random variable. The time until mobility-induced failure is expressed as a function of the weighted sum of the speeds of the nodes on the route, as well as the node transmission range.

The route robustness metric is incorporated into the robust flow admission and routing (RFAR) protocol, presented in [44]. The aim of RFAR is to maximise the network's "robust throughput" which depends on the notion that more credit should be given when a session is completed without interruption i.e. without violating its QoS requirements for its entire intended duration. To this end, QoS-AODV-style route discovery is employed, albeit conditions for forwarding the RReq are as follows. Firstly, the node's packet queue length must be below a threshold to aid in maintaining packet delay bounds. Also, for each class of data, a robustness threshold is set as the maximum tolerable probability that the route breaks before the requesting session ends. If, during route discovery, the cumulative robustness of the partially-discovered route indicates a route failure probability surpassing this threshold, the RReq is not forwarded. After receiving a RReq, a destination waits for a short period before replying on the route comprising the fewest hops.

Simulation results in [44] show that, due to the preference for robust routes, a much lower route failure rate is experienced than by protocols such as DSR. While this does not increase the overall throughput of the network, due to the careful admission control, it does increase the robust throughput, where gaining credit for data delivery is contingent upon session completion. However, a particular limitation of this protocol is that it relies on nodes being able to estimate their own speed. To achieve this, they must be equipped with GPS receivers or some location-determination system, and this may limit the application of this protocol. The RFAR protocol also does not explicitly check nodes' residual capacities, meaning that it could be unreliable in guaranteeing that traditional minimum throughput requirements are upheld.

5) Common Advantages and Drawbacks in this Category: The main advantages of protocols in this category are as follows. Firstly, since, with the exception of INORA, the resources of each node on a route are tested individually prior to propagating a RReq and prior to session admission, there is a higher chance that discovered routes will be able to adequately serve the requesting sessions. Secondly, no resource information is collected from node neighbourhoods, resulting in low-overhead operation. However, this also means that, as in the previous category, the potential impact of new sessions on nodes' cs-neighbours is not evaluated. Again, this could lead to QoS assurance violations. Although the protocols in this section are aware of routing information, they have not considered the effects of intra-route contention. This could be because they are designed without being coupled with a specific MAC protocol and hence do not assume 802.11's contention-based operation.

B. On-Demand Resource Discovery-Based Schemes

The protocols described in this section test the resources of not only the nodes on a route, but also those of neighbouring

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

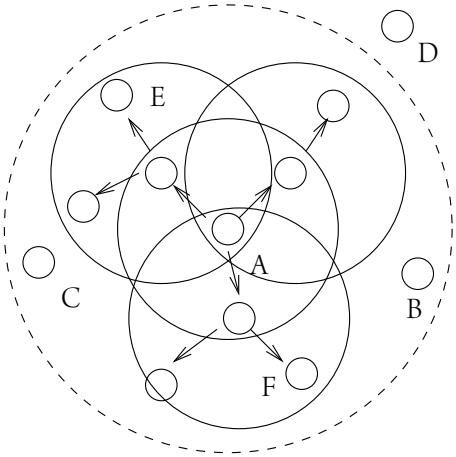

Fig. 4. An on-demand resource discovery method. The small solid-lined circles are nodes, the medium-sized circles represent transmission range coverage areas, and the largest, dashed circle represents node A's average cs-range coverage area when shadowing is assumed to be negligible. Note that all nodes have the same transmission and cs-ranges and therefore node A's transmissions cause a busy channel status to be detected by all nodes within its cs-range and vice-versa. Node A broadcasts an admission request (AdReq) which is received and forwarded by all of its neighbours. However, there are no suitable relay nodes to reach node B, and therefore its capacity cannot be queried by node A. Also, with a fixed time-to-live, in this case, two hops, the AdReq may not reach all cs-neighbours even if there are suitable relay nodes. An example of such a cs-neighbour is node C. If the time-to-live was increased to three hops, for example, too many nodes may be reached. In this case, node D is within three hops of A, but A is not within D's cs-neighbourhood (since D is not within A's). If node D had insufficient capacity to support the session, it would reject it. This decision would be incorrect since node D would not actually be significantly affected by node A's transmissions.

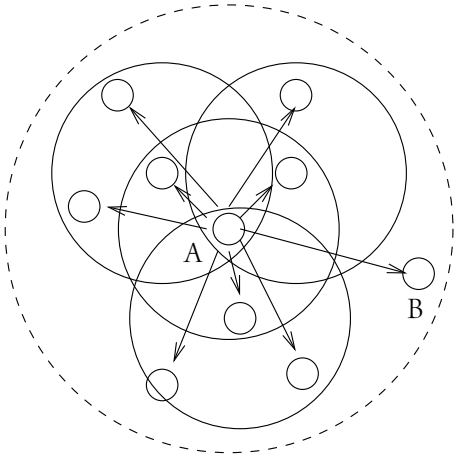

Fig. 5. A second on-demand resource discovery method. The small solid-lined circles are nodes, the medium-sized circles are transmission range coverage areas, and the largest, dashed circle represents node A's average cs-range coverage area when shadowing is assumed to be negligible. Node A broadcasts an admission request (AdReq) at a higher power, which is received by all of its carrier-sensing neighbours. This avoids the problems, depicted in Figure 4, of no relay node existing between nodes A and B, and of too few or too many nodes being reached by a query packet with a fixed time-to-live. However, a much higher level of interference is imposed on all nodes during the transmission, possibly causing collisions. Also, the signal propagation characteristics must be predicted accurately in order to know what power to transmit the AdReq with.

nodes that may be impacted by the admission of a new data session. This testing of the impacted region of each node, within its cs-range, is performed on-demand, and only if the resources of the node itself are adequate to support the requesting session.

1) Contention-aware Admission Control Protocol: The work in [11], [22] is considered something of a landmark in the design of AC protocols for MANETs, since it is cited in most papers in the field that were published after it. It is in this work, that, to the best of our knowledge, an AC protocol first tested the cs-neighbours of a route as a prerequisite for session admission.

The proposed protocol, the contention-aware admission control protocol (CACP) is combined with a source routing protocol similar to DSR [46]. Admission control takes place in two stages. When a session requesting admission arrives at a source node, a QoS-AODV-style route discovery is triggered. Nodes monitor the CITR and only forward the RReq if their capacity is sufficient, given the intra-route contention on the partially-discovered route up to this point.

On reaching the destination, the route in a RReq is cached for a short time. Thus, if multiple RReqs reach the destination on different routes, several routes are cached. One route is selected, such as the first one to be discovered, and a RRep is sent on this route back to the source. Each intermediate node receiving the RRep again tests its locally-available capacity, but this time with full knowledge of the level of intra-route contention.

The cs-range is assumed to be equal to the length of two hops. At each node, if the local capacity test is passed, the RRep is cached for a short timeout period, and the node's cs-neighbours' residual capacities must then also be tested. Three possible methods are proposed for this in [11]:

1) CACP-Multihop floods an admission request (AdReq) packet to a distance of two hops, assuming it will reach the nodes in the cs-neighbourhood. Figure 4 illustrates an example as well as the potential problems with this approach.
2) CACP-Power uses a higher power to transmit an AdReq packet to ensure it reaches all nodes within the cs-range with a single transmission. Figure 5 illustrates an example.
3) The third method, CACP-CS,employs a passive resource discovery-based approach and thus no explicit capacity query packet is issued. It is described under the relevant category, in Section V-D1.

With CACP-Multihop and CACP-Power, the AdReq carries a copy of the session's potential route that is stored in the RRep packet. On receiving an AdReq, a cs-neighbour calculates its contention count by checking its cs-neighbour cache to see how many of its cs-neighbours are also transmitters on the route that the session is requesting admission on. The cs-neighbour set is learnt by promiscuously listening to the channel, in a manner akin to the way DSR snoops on routes used by its neighbours [46].

Each of the above resource discovery/querying methods have their own advantages and disadvantages. CACP-CS incurs no overhead, but may underestimate the capacity available to cs-neighbours [11], as explained in Section V-D1. On the other hand, CACP-Multihop and CACP-Power both introduce some overhead. CACP-multihop's overhead depends on the

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

node density. CACP-Power only transmits once but produces a high level of interference while the AdReq is being transmitted. Further potential shortcomings are discussed in the captions for Figures 4 and 5.

If any AdReq-receiving node determines that its capacity is insufficient to admit the session on the selected route, it replies to the AdReq sender with an "AdReq denied" (AdDen) packet. If an AdReq sender receives an AdDen within the RRep-caching timeout period, it drops the RRep and notifies the session's destination node. The destination then selects one of the other cached discovered routes and attempts to send a RRep to the source node along it. If no AdDen is received before the timeout, the RRep is forwarded towards the session's source node, which admits the session if it receives a RRep. The session is blocked if none of the discovered routes have sufficient residual capacity in their cs-neighbourhood.

As discussed in Section II-D, mobility poses a few challenges to AC protocols. To deal with unexpected interference, the authors of [11] suggest reserving a portion of each node's capacity. The problem lies in knowing how much to reserve. If too much is reserved, then the capacity is wasted. If too little, the method fails to avoid throughput degradations anyway. Secondly, when route failures occur, CACP must search for an alternative route. This is because, after the initial route discovery, only one RRep is returned to the source, and since the AC procedure is coupled with the route discovery process, the session cannot simply be re-routed to another known route. This inevitably leads to a lapse in throughput, its duration depending on the existence of alternative routes to the destination and the levels of congestion affecting the RReq propagation and capacity tests. To deal with this situation, CACP must lower the session's throughput requirement while a new route is discovered and tested.

In summary, CACP was, to our knowledge, the first protocol to test the capacity of cs-neighbours as a prerequisite to session admission. However, CACP does not deal well with route failures. While searching for a new route, it reduces the data rate of affected sessions. This implies that CACP can only support applications with elastic throughput requirements. However, this is the same assumption implicitly made in many of the previously-discussed protocols, which simply rely on the routing protocol to find a new route and then re-admit a session once it has been re-routed. The advantage of pausing the affected sessions over such previously-discussed protocols is that congestion at the breaking point of a route may be avoided.

2) Staggered Admission Control: In [12], the staggered admission control (StAC) protocol was introduced. As opposed to most protocols discussed in this paper, StAC is neither fully-coupled to nor fully-decoupled from the routing protocol, as shall be explained below.

The basic routing functionality of StAC is based on DSR [46]. In brief, three stages of AC are employed. In the first stage, a capacity-constrained route discovery is conducted, akin to CACP (Section V-B1), considering the intra-route contention. However, in contrast to CACP, cs-neighbours are not tested at the RRep stage, and all discovered node-disjoint routes are returned to the source node. The second stage consists of testing the cs-neighbours of a route in a manner akin to CACP-Multihop (Section V-B1). However, note that information about the session is also stored at cs-neighbours via the AdReq packets. If no cs-neighbour rejects the session, it reaches the third AC stage, when it is partially admitted. The session is then allowed to gradually increase its packet sending rate over the first few seconds, during which it may still be rejected. If the desired throughput of the session is maintained in each second, for a few seconds, and no cs-neighbour reports its CITR dropping below a threshold level, the session is admitted. This phase tests to see if the unexpected increase in collision rate would make the session's QoS unacceptable or reduce the CITR of cs-neighbours to a critical level.

The rationale behind separating the first two stages is as follows. Nodes' local resources are tested in the RReq/RRep stage to see if they can support the requesting session, in order to avoid wasting resources discovering routes which are definitely not useful at the current time. However, DSR is able to discover routes in multiple ways other than the standard RReq/RRep-based procedure. Intermediate nodes may reply to RReqs. Nodes may overhear a packet of which they are not the intended recipient, and learn the routing information in its source route header. Also, such overhearing nodes may send route shortening information to source nodes. Having a separate route resource-testing procedure allows such routing information to be utilised so it does not go to waste. By contrast, fully routing-coupled AC protocols must always initiate a new route discovery procedure.

StAC handles the effects of mobility in two ways. Firstly, a portion of each node's capacity is reserved, akin to CACP, for routing packets and unexpected interference. Secondly, data packets carry a small header extension containing the source's view of the intermediate nodes' free capacity. On forwarding the data packet each node checks if this view needs to be updated and sends a small 'update' packet, if one has not recently been sent. When route failure occurs, there is no time to perform AC on an alternative route, since a session's packet transmissions must not be paused. Therefore, the source node of affected sessions re-routes them to the known route with the highest bottleneck throughput. If the route's bottleneck residual capacity is different to the value stored by the source node, the first data packet detects this, and an update packet is sent.

The advantages of this protocol over AC schemes published earlier are largely two-fold. Firstly, the increase in collision rate that would occur upon session admission is considered during AC, which other protocols neglect to do. While other protocols assume that the QoS requirements of data sessions are elastic, StAC uses this property in a different way. It assumes that the sending rate of sessions may be gradually increased during admission, while the feasibility of supporting the session is evaluated. Once a session is admitted, the packet rate of a session may not be reduced below its original requirement. Secondly, the hybrid routing-coupling relationship of StAC allows opportunistically-discovered routing information to be utilised, possibly saving overhead, while still avoiding the disadvantages of decoupled protocols (Section IV-C). However, the staggered admission scheme can sometimes be overly-careful. For example, if a temporary burst of overhead causes a session's throughput to drop during the third stage

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

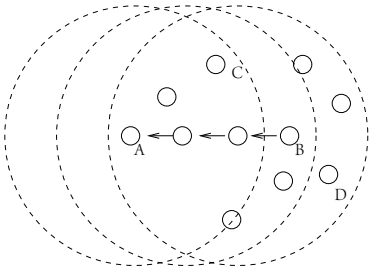

Fig. 6. Illustration of ACRMP's contention count determination and resource reservation scheme. The small solid-lined circles represent nodes and the arrows show the propagation of the RRep packet from node B to node A. The larger dashed-lined circles show the areas reached by the high-powered transmissions of the RRep by node A and the two intermediate nodes. For example, node C detects three such transmissions and thus infers that its contention count is three, while reserving the corresponding amount of capacity. Node D detects a contention count of one.

of admission, the session may be falsely rejected. Also, after a failure, by re-routing sessions to routes that have only had their local resources tested, StAC reverts to the operation of protocols described in Section V-A. If, after re-routing, a cs-neighbour has insufficient capacity for its own transmissions, further re-routings may be necessitated. The reserved portion of the capacity reduces the chance of this occurring, and use of the DSR route cache allows all sessions to quickly find new routes until the situation stabilises. The pay-off is that sessions do not need to be paused. This trade-off was discussed at the end of Section II-E. StAC adopts the higher-risk strategy that was discussed there. Finally, the scheme of flooding AdReqs to a distance of two hops has the weaknesses exemplified in Figure 4.

3) High-Power Transmissions-based Admission Control and Reservation Management Protocol: The work in [29] proposed an admission control and reservation management protocol (ACRMP) with some optimisations compared to the other protocols in this category. The first stage of AC consists of the same capacity-constrained route discovery process as in CACP (Section V-B1). Again, a RRep is only sent along one discovered route, while the other routes are cached at the destination for a short period. The innovations are introduced at the RRep stage. The RRep packet is transmitted with a higher-than-normal power, like the AdReqs in CACP-Power (Section V-B1). This allows all cs-neighbours of an intermediate node to learn which nodes on the route are in their cs-range. Each time a node hears a RRep transmitted with the higher power level, it increases its contention count for the corresponding session ID. Figure 6 provides an example. If it detects that, given the new contention count, its residual capacity is not sufficient to tolerate the interference the new session would impose, it sends a reject message to the RRep sender. This informs the destination node of the session rejection. The destination may attempt admission using another of the recently-discovered routes.

If the RRep is not rejected, soft-state capacity reservations are set up at the cs-neighbours and at each node on the route, and the session is admitted. Periodically, a bit set in a data packet header instructs intermediate nodes to forward it using the higher power which results in the capacity reservations at cs-neighbours being refreshed. The last data packet of a session is used similarly to erase all reservations.

The advantages of this protocol over similar schemes such as CACP and StAC are as follows. There is no need for a delay at each intermediate node while the cs-neighbours are queried, and overhead is saved compared to CACP-Multihop. Also, since the scheme is based on AODV, the source route is not carried in each data packet header. This, again, reduces the overhead, while the protocol is still able to calculate contention counts as accurately as CACP. The reservation management scheme is also fast and efficient. The major drawback of this scheme is the need for high-powered RRep and data transmissions which increase interference to cs-neighbours and can cause a burst of collisions.

4) Common Advantages and Drawbacks in this Category: As opposed to the previous two categories, these protocols explicitly query the residual capacity of nodes that would be impacted by a new session, thereby reducing the chance of false admissions. Compared to non-reactive resource discovery schemes, discovering the resources of cs-neighbours on-demand has several advantages. Firstly, this avoids needless overhead (compared to proactive approaches) at times when nodes are not receiving any new session requests and therefore do not require any resource state information. Secondly, it aids in avoiding false admissions in networks with many users, where a session admission request packet has passed through a neighbour node and reserved resources there, but the session has not yet begun using them, since the rest of its route is still being tested. Explicit querying of such neighbour nodes allows the reservations to be subtracted from their residual capacity values.

On the other hand, the disadvantages of these protocols are also obvious. Firstly, session or route request packets are delayed at each relay node while the resources of its cs-neighbours are queried. This results in increased session admission times. Secondly, querying can produce an unexpected burst of overhead, temporarily increasing interference and collision rates in the region. Unless such effects are averaged out in residual capacity estimations, some nodes may falsely report a decrease in residual capacity.

C. Proactive Resource Discovery-Based Schemes

The category of AC methods that are dealt with in this subsection are also coupled with the routing protocol. The state of the resources of neighbouring nodes is learnt through their periodic beacon transmissions. This information can reveal the impact each individual neighbour is having on a beacon-receiving node. Alternatively, it can be used to evaluate the impact the admission of a data session would have on the beacon senders.

1) Hierarchical Routing-based Admission Control: A hierarchical routing-based admission control (HRAC) protocol was proposed in [30]. A logical super-node network is established via periodic HELLO packet broadcasts. This structure is an approximation of the dominating set, such that each node is at most one hop away from a super-node. The HELLO packets also distribute node channel utilisation information. Each node estimates its available capacity in a simple manner.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

It first divides the raw channel capacity by the MAC overhead factor, which was estimated in [30] through simulations. From this result, it then subtracts the total channel utilisation of its neighbours.

When a session admission request arrives, a virtual route discovery procedure is triggered. This involves a RReq being propagated along the super-node structure until the destination is found. Each node only forwards the RReq if its residual capacity, calculated as described above, is sufficient to support the session. If the RReq arrives at the destination, this replies with a RRep to the source, and the session is admitted. Nodes forwarding the RRep add the requested channel capacity onto their utilised capacity values, which are then propagated on the HELLO packets.

In the event of a route failure, route error messages are generated, which inform nodes that they should release the reserved capacity for affected sessions. In this case, affected sessions must attempt to be re-admitted, as described above.

As is typical of earlier approaches to AC in multi-hop MANETs, this protocol does not consider the intra-route contention when calculating a session's capacity requirement. Moreover, the residual capacities of neighbour nodes are not checked prior to forwarding a RReq. This means that the impact of the new session on those neighbours is not considered. On the positive side, although overhead is required to disseminate channel usage information, this allows the super-node structure to be established, which helps to reduce overhead during QoS-aware route discovery.

2) Ad hoc QoS On-Demand Routing-based Admission Control : The ad hoc QoS on-demand Routing (AQOR) protocol, proposed in [18] also incorporates an AC mechanism as part of the routing process. Each node broadcasts a HELLO packet once per second, with a TTL of 1, akin to the HELLO packet-based neighbour discovery-employing version of AODV [51]. These packets are used for neighbour table maintenance. Moreover, through the HELLO packets, each node learns the amount of capacity that is reserved for QoS-sensitive data sessions at each of its neighbour nodes. This information is used in AC.

Route discovery, which forms the basis of the AC mechanism, is similar to QoS-AODV. A node requesting a route for a new session adds its maximum tolerable delay and throughput requirements to the RReq header. Each node considers the intra-route contention in a one-hop radius and subtracts the total capacity reserved for QoS-sensitive traffic at its neighbours from the (assumed to be fixed) raw channel capacity. The RReq packet sets up soft-state capacity reservations and reaches the destination only if each intermediate node has sufficient residual capacity to support the session.

A destination node performs similar capacity tests, and replies to all received RReqs. At the source node, the end-to-end delay of each route is estimated as being approximately half of the round-trip-time of the RReq/RRep packets. A simple analysis shows the difference between the uplink and the downlink times and avoids the need for global clock synchronisation. Finally, the source node is able to select the route with the lowest end-to-end delay from among those that satisfy the application's delay bound.

If no route is found within twice the application's maximum delay bound, the source node may back-off and re-initiate route discovery later, or reject the data session. On selecting an appropriate route, the source begins sending the data packets, which activate the reservation at the nodes on the selected route. The reservation times out if no data packets are received after an interval that is determined by the throughput requirement of the session.

Such reservation timeouts can be used as a fault-detection mechanism that is faster than relying on the absence of HELLO packets [18]. If no packet is received by the destination node for the reservation period, it can infer that a fault has developed. A route failure (as far as the QoS is concerned) can also be inferred if a threshold number of packets violate the session's end-to-end delay requirement. Packet travel times are calculated by a simple analysis that uses timestamps and an estimate of the source and destination clock offsets.

To recover from such failures, the destination initiates a reverse route discovery procedure. Apart from its direction, this is identical to the source-initiated procedure described above. On receiving the first in-time reverse-RReq, the source re-routes the violated session to the reverse of the discovered route. If no alternative route is found, the session may be switched to a best-effort service, or queued until later.

As far as it is possible to tell, this protocol was one of the first to consider intra-route contention during AC. Its advantages include the ability to consider both delay and throughput application constraints, as well as its method of QoS assurance violation detection. However, the intra-route contention is only considered in a one-hop radius, whereas the cs-range is typically larger than the transmission range for collision avoidance purposes, as previously discussed. Also, the HELLO packets are only used to collect neighbour information for calculating the locally available channel capacity. The impact of admitting the session on cs-neighbours is not considered.

3) Throughput and Delay-Aware Cost Function-based Admission Control: The admission control-enabled on-demand routing (ACOR) protocol is proposed in [27]. This protocol defines cost functions based on a session's delay and throughput requirements, and a node's residual resources. The throughput cost of a session at a particular node is defined such that it increases with an increasing throughput requirement and a decreasing CITR at the node. The delay of transmitting to any neighbour node is estimated via a probe packet/acknowledgement round trip time. The delay cost is defined such that it increases with increasing delay to the next hop and decreasing difference between the session's delay bound and the route delay accumulated so far. The total cost of a session on a route is then defined as the throughput cost plus the delay cost and these are defined such that they take a negative value if the requested QoS exceeds the route's current capabilities.

Each node broadcasts HELLO packets periodically. These contain its level of channel usage and its residual capacity. Using a QoS-AODV-style route discovery, the RReq is only forwarded if neither the route's cumulative delay nor throughput costs are negative. Although it is not stated in [27], it is assumed here that the residual capacity information of neighbour nodes, received in their HELLO packets, is used

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

to check that they have sufficient residual capacity before forwarding the RReq. In the RRep phase, nodes forwarding the RRep reserve resources for the session, which is admitted when the RRep is received by the source node.

The failure-handling capabilities of ACOR are inherited from AQOR (Section V-C2). Again, reservation timeouts are used to implicitly detect route failures or congestion. Once more, akin to AQOR, a reverse route discovery is triggered by the destination node.

The ACOR protocol is also coupled with the EDCA mechanism [4], by mapping different types of data sessions to different medium access categories. In addition to the varying access category inter-frame spaces, ACOR uses a smaller maximum retransmission count for less important classes of data.

This protocol contains some interesting innovations, most notable of which are the throughput- and delay-related cost function definitions. As an extension to the operation described in [27], these costs could also be used to rank routes in a multi-path version of the protocol. Furthermore, the utilisation of the mapping of data types to EDCA access categories can provide a higher average QoS to applications with more stringent QoS requirements. A notable omission is the lack of consideration of intra-route contention during AC.

4) Contention-aware Admission Control Based on AODV: Another AC protocol based on AODV, which we will refer to as contention and capacity-aware AODV (CCAODV), was introduced in [15]. The "cs-range = 2 hops" model is adopted here akin to CACP (Section V-B1). However, instead of monitoring the CITR, CCAODV-implementing nodes monitor the number of bits they transmit per second, i.e. their capacity usage. This information is piggybacked on AODV's HELLO messages (akin to HRAC, Section V-C1). Therefore, as long as the HELLO-based neighbour discovery-employing version of AODV is assumed, no additional control packets are introduced. Nodes also piggyback the addresses and channel usage values of their neighbours onto their HELLO messages, and thus every node eventually learns the channel capacity utilised by all of the nodes in its two-hop/cs-neighbourhood.

Finally, assuming that the channel capacity is known (i.e. it is fixed), a node simply needs to subtract the channel usage of its cs-neighbours from the channel capacity to obtain an estimate of the amount of the capacity that is available to it, again, akin to HRAC. As with CACP-Multihop (Section V-B1 and Figure 4), a possible weakness of this approach is that some nodes may not be able to receive information about all of their cs-neighbours due to the lack of a relay node between them. However, as long as a network is not too sparse, this is unlikely to be a major problem.

When a QoS-sensitive session requires admission, a QoS-AODV-style route discovery process is triggered. The RReq is forwarded to the destination node if the residual capacity of each intermediate node is adequate for supporting the session's throughput requirement. If the destination deems the route's capacity sufficient after considering an approximation of the intra-route contention on the discovered route, a RRep is sent to the source node, and the session is admitted.

In CCAODV, a route failure is detected when a HELLO message is not received from a node for a predefined interval. In this case, it is important to notify cs-neighbours of the newly-freed capacity. For this purpose, an "immediate HELLO" message is broadcast to neighbours, which forward the information of the new channel consumption on their HELLO messages immediately. Also, an error message is returned to the source node of active sessions being carried on the broken route, which triggers a new route discovery. As with CACP, this inevitably leads to an, admittedly possibly short lapse in throughput while a new route is discovered.

The primary advantage of CCAODV's AC scheme is its ability to quickly release reserved capacity when a route failure occurs. Other protocols discussed so far, which rely on monitoring of the CITR, cannot immediately release resources. If the re-admission process is triggered immediately after a route failure, CITR-monitoring nodes might not have updated their free capacity values yet. In CCAODV, the "immediate HELLO" packets inform nodes exactly how much capacity to free up. This is important when the affected session was consuming a large portion of the relaying nodes' resources. If it was not, the relaying nodes would be likely to have enough resources to re-admit the session anyway, hence the lack of this feature does not necessarily affect other protocols significantly. On the other hand, the main shortcoming of the CCAODV protocol is that the residual capacities of cs-neighbours are not considered before admitting a session. The information they broadcast is used by the receiving nodes only to ascertain the channel capacity available to themselves. Furthermore, there is no consideration of the fact that some two-hop neighbours of a given node may possibly transmit in parallel to each other, and hence simply subtracting the aggregate of their channel usage values from the raw channel capacity may yield an overly-conservative estimate of the available channel capacity.

5) Capacity-aware AODV-based Admission Control: As opposed to many of the previously-discussed methods, the protocol proposed in [42], QoS-aware AODV routing-based AC (QAODV-AC), states an extra condition on a node's channel being considered idle. Not only does the channel have to be sensed idle by both the physical and virtual 802.11 carrier-sensing mechanisms, as discussed in Section II-C, but also the interface (between link and MAC layer) queue must be empty. The saturation throughput of a node is calculated at the MAC layer as the average higher-layer packet size divided by the average time difference between enqueueing and receiving an acknowledgement for a packet. The benefit of this method is that it takes into consideration not only the back-off times, but also the capacity available at the receiving node, as well as the number of packets waiting to be transmitted at the sending node. A node's available capacity is then calculated by multiplying the saturation throughput by the idle time ratio defined above.

The AC protocol is based on QoS-AODV-style route discovery. Again, HELLO packets are employed for neighbour discovery and available capacity information dissemination. The saturation throughput is recalculated after every HELLO packet interval. A RReq is forwarded to the destination if all nodes, as well as their neighbours (checked using the information from the HELLO packets) have adequate capacity to admit the session. Again, during the RRep stage, soft capacity reservations are set up in a session information table

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

at each node. Route failures and unacceptable reductions in the availability of resources are handled via an ICMP QoS_LOST message, akin to QoS-AODV (Section V-A2).

In this protocol, the definition of node residual capacity, as described above, enforces a more careful AC mechanism than protocols that rely solely on the CITR. This theoretically results in fewer false admissions. However, a shortcoming of QAODV-AC is that it only tests the available capacity of the neighbours of a route, and only considers intra-route contention in one-hop radii prior to session admission. Typically, the cs-range is much larger than the transmission range, as discussed in Section II-D, and therefore, at least the two-hop neighbourhood must be capacity-tested. The lack of testing of all cs-neighbours reduces overhead but could result in false admissions.

6) Interference-based Fair Call Admission Control: The interference-based fair call AC protocol (IFCAC) was proposed in [31]. This approach is unique, in that, as opposed to previously-discussed protocols, the channel is not considered busy just because the sensed interference power exceeds the cs-thresh. In fact, [31] highlights that that definition of channel busyness provides the upper bound on the utilised channel time. At the other extreme, counting the channel busy only when the current node is transmitting or successfully receiving and decoding a packet yields the lower bound.

Re-visiting a previous example, CACP (Section V-B1) considers a lower bound on the available channel time by considering the channel busy if a signal is sensed with a power above the cs-thresh. In fact, if both the receiving (rxthresh) and cs-thresholds (csthresh) are used to separately monitor the channel busy time ratio at various ranges, an approximation of the relative positions of the interference sources can be obtained [31]. For example, if the channel is detected largely idle using the rxthresh, but busier using the csthresh, most interference sources are likely to be located outside the transmission range, but inside the cs-range. Each node should also monitor the level of noise, which in this case is defined as interference that is detected with a power below the cs-threshold. Other cases are explained in [31].

The word "fair" appears in IFCAC's name because each node allocates an equal amount of channel capacity to each of the transmitters in its cs-range. For each case of the possible relative interference source positions IFCAC determines the capacity to allocate to each transmitter within the cs-range in the most appropriate way, as detailed in [31]. HELLO packets, transmitted with a low frequency of 1 per 5s, maintain the identities of neighbour nodes and their status regarding whether they are a data transmitter or not. If interfering nodes are deemed to be located outside of the transmission range, a node transmits a high-powered beacon (as in CACP-Power) to contact those cs-neighbours. All nodes receiving the beacon, which are deemed to be outside the sender's transmission range and also have traffic to transmit reply using the same transmission power. This way, each admitting node learns the number of its interference sources, both inside and outside its transmission range, and can decide the fair amount of capacity that is made available to each interfering node. This value is recalculated whenever the number of interference sources changes. This is detected via the HELLO packets for the local neighbourhood, and a change in the sensed noise level for the cs-neighbourhood.

The described locally-available capacity estimation mechanism is combined with DSR [46] to provide an end-to-end AC protocol. In fact, only the RRep stage of DSR is modified. The RRep is used to discover the bottleneck node having the lowest fair share of available capacity. This information is used by the source node, once it receives the RRep, to determine whether or not the session can be admitted.

This protocol is obviously unique among the surveyed approaches in that it provides a fair share of the local channel capacity to each of its interferers. If all interference sources are successfully counted, this should avoid attempted over-utilisation of the channel. However, it is clear that, as highlighted in [31], IFCAC can under-utilise the network if not all nodes require their fair share of the available capacity. On the other hand, sessions that require more than their fair share will not be admitted, or will have to reduce their sending rate when new sessions arrive. A further cost of fairness are the probe packets that are broadcast with a higher-than-normal power, which can cause collisions at a much greater distance.

7) A CACP and CCAODV-Inspired Protocol: The QoS admission control routing protocol (QACRP) described in [41] combines features from CACP (Section V-B1), QAODV-AC (Section V-C5) and CCAODV (Section V-C4), and proposes some modifications. Nodes' local residual capacities are determined via the CITR. Again, periodic HELLO packets are broadcast to neighbours. Akin to QAODV-AC, QoS-AODV-style route discovery is employed, and before forwarding a RReq, each relay node checks that both its own, as well as its one-hop neighbours' residual capacities are sufficient for admitting the requesting session. Interestingly, the intra-route contention is considered in a two-hop radius, akin to CACP, even though the impact on cs-neighbours is only considered in a one-hop radius. Each intermediate node inserts the minimum value of its neighbours' residual capacity values, as well as its own free capacity into the RReq prior to forwarding it. On receipt of the RReq, the destination node re-checks the residual capacity at each node on the route, this time using full knowledge of their correct contention counts. If all nodes have sufficient available capacity, a RRep instructs the source node to admit the session.

The advantages of this scheme over those it is inspired by are as follows. Compared to CACP-Power and CACP-Multihop, it saves time and complexity by not testing cs-neighbour capacities on-demand at each intermediate node. At the same time, it is able to calculate the contention count at each node as accurately as CACP, which is a more accurate method than that employed by CCAODV and QAODV-AC. The HELLO packets are smaller than in CCAODV, since they only include a node's own residual capacity information and not that of its neighbours. However, this means that the impact of admitting the session on nodes outside the transmission range, but inside the cs-range, is not considered. This could lead to some false admissions. There is no discussion in [41] of how the protocol handles QoS assurance violations or route failures, although it may be assumed that the action undertaken is similar to CACP. In summary, this protocol does not contain any previously unseen features, but combines

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

features of earlier-published protocols in new ways. In this manner, it achieves a different balance between accuracy and overhead.

8) Adaptive Admission Control: The work in [25] investigates the accuracy of two of the parameter choices made in the design of previously-discussed protocols. These are the radius of the impacted region within which to consider cs-neighbours' resources, and the radius in which to consider intra-route contention, which were discussed in Sections II-D and II-E. Based on the model of [25], the most accurate estimate of the impacted region, in terms of the number of nodes, was obtained by considering nodes within a three-hop radius. The most accurate view of the intra-route contention was obtained using a radius of either two or three hops, depending on the smoothness of the routes, where the smoothest possible route has nodes positioned in a straight line. However, considering the impacted region as having a radius of two hops still produces an accurate estimation, and results in much lower overhead [25]. This was the value considered in several other protocols, such as CACP and StAC.

This knowledge was utilised in a protocol called adaptive admission control (AAC). The AAC protocol employs HELLO packets to disseminate nodes' residual capacity (CITR) values. Each HELLO packet only travels one hop, but akin to CCAODV (Section V-C4), the information of one-hop neighbours is also carried, and therefore each node learns the minimum capacity available in its two-hop radius/estimated cs-range. Two hops is used as the information retrieval range, as in CACP, for the reasons given above.

AAC utilises QoS-AODV-style route discovery, with residual capacity being determined as described above. Again, intra-route contention is considered, but it is only fully accurate at the RRep stage.

The method of dealing with potential QoS assurance violations due to mobility, is as follows. When data packets from a QoS-sensitive data session occupy a significant portion of a node's interface queue, one selected source node is notified. On receiving this notification, the source node pauses the sending of data packets for the session with the highest throughput requirement. This way, the smallest number of user sessions are disrupted, while freeing up the most resources. However, it might be difficult to re-admit this session in the future. Therefore, in order to reduce the chance of needing to pause data sessions in the future, every time a session must be re-admitted, AAC-implementing nodes increase the amount of capacity they request for QoS-sensitive sessions. This method achieves a similar effect to the reservation of a portion of each node's capacity implemented by CACP, for example. However, this scheme is slightly more robust, since the "spare" capacity is increased only when QoS assurance violations occur. On the other hand, one might argue that by pre-reserving a portion of the capacity, as in CACP, QoS violations could be made less likely to occur in the first place.

Its robustness in terms of dealing with congestion caused by route failures, while disrupting the minimum number of sessions, is the AAC protocol's primary asset. It also combines many of the beneficial features of previously-discussed protocols, such as the accuracy of CACP in residual/required capacity estimation together with proactive resource discovery. The protocol's drawbacks are discussed later, with those common to the whole of this category.

9) Time-based Admission Control: The time-based AC (TAC) protocol, inspired by the technical report version [52] of AAC (Section V-C8), was proposed in [45]. In TAC, instead of monitoring its CITR, each node calculates the average fraction of time it spends on transmissions and on backing off. The back-off time is estimated based on the 802.11 saturation throughput formula derived in [53]. Once the fraction of time spent on transmissions and back-off periods has been estimated, the available normalised channel capacity is calculated by subtracting this value from 1.

As in AAC, the AC procedure is coupled with QoS-AODV-style route discovery. Each node broadcasts periodic HELLO messages which contain its estimation of its available channel capacity. The AC procedure is then identical to AAC's, except that each node forwarding the RReq/RRep only compares the session's required capacity to the minimum capacity in its one-hop neighbourhood.

The handling of QoS requirement violations is similar to QoS-AODV (Section V-A2), employing ICMP_QoS_Lost packets. On receipt of such a packet, the source pauses the session's packet sending, and initiates a new route discovery.

In essence, the innovation of this protocol lay in the method of calculating the channel time required by a session by incorporating the average back-off period. However, although, according to [45], the same cs-range model is employed as in [11], [25], with the cs-range being greater than twice the transmission range, only the available capacities of one-hop neighbours are checked during AC. This choice is not justified in [45], though it can reduce the protocol's overhead. However, simulation results in [45] suggest that TAC achieves a lower PLR and average delay than AAC, in some scenarios, due to its more accurate consideration of the available channel capacity.

10) SoftMAC: In [16], [54], the softMAC architecture is presented. We again focus only on its AC mechanism. The softMAC architecture is so-called because it resides above the MAC layer, but below the network layer, at "layer 2.5". As opposed to the previously-discussed protocols, SoftMAC takes the auto-rate feature of 802.11 into account, meaning that link capacities may vary. Physical link capacities are established using the experienced delay between transmitting back-to-back probe packets of various sizes.

In SoftMAC, each node monitors the fraction of the channel time T_{tx} used by its transmissions of throughput-sensitive traffic. Periodic broadcast packets are transmitted by each node informing its neighbours of its T_{tx}. Each node calculates its nominally-available channel time T_{free} by subtracting all of its neighbours' T_{tx} values from one. The value of T_{free} is then also included on nodes' beacons. Finally, this allows each node to learn the minimum T_{free} among its neighbour nodes $T_{minfree}$ and it adds this value to its beacons as well. The available capacity on a link is estimated by the minimum of the $T_{minfree}$ values of its end-nodes. A node's channel time utilisation T_{tx} is calculated from the expected time spent on collisions and on successful transmissions. These in turn depend on the average packet collision probability and the link rates used by the node's various transmissions. The collision probability is estimated based on the loss rate of the known-

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

frequency beacon broadcasts. Note that, akin to CCAODV (Section V-C4) the estimate of T_{free} is overly-conservative, since it does not consider the possible overlap in time between the transmissions of a node's neighbours [16].

The AC procedure of SoftMAC is coupled with the DSR protocol and operates in a manner that is similar to CACP (Section V-B1). The main difference is that the forwarding of the RReq/RRep packets is contingent upon each node, as well as all of its out-going links having a $T_{minfree}$ that is sufficient to support the session requesting admission. This implicitly considers the impact that admitting the session would have on each node's one-hop neighbours.

This protocol has many strengths including the consideration of heterogeneous link rates and the inclusion of the estimated collision probability in the channel utilisation time of a node. However, note that this estimation cannot predict the increase in collision rate that occurs when a new session is admitted. Also, as in QAODV-AC, only one-hop neighbours' available capacity is tested prior to admission, which may not be accurate enough. Real testbed-based experimental results in [16] indicate that the protocol can make relatively accurate admission decisions for small numbers of data sessions.

A final important consideration is not a shortcoming of this protocol, but potentially affects all multi-rate-aware AC protocols. Consider that the bottleneck-capacity link on a route may support a rate of 11Mbps. Based on this rate, total traffic requiring 3Mbps is admitted. Now, if fading or mobility cause the supported link rate to drop to 2Mbps, the traffic can no longer be supported. If, like the previously discussed schemes, the protocol was not aware of rates higher than the basic fixed rate, the traffic of 3Mbps would never have been admitted in the first place. Therefore, misleadingly, the more capable multi-rate-aware protocol might actually appear to produce more false admissions.

11) Optimised Link State Routing-based Admission Control: The AC protocol proposed in [34] is based on an interference- and QoS-aware version (IQOLSR) of Optimised Link State Routing [55]. This protocol operates proactively, broadcasting HELLO packets to enable the construction of routing tables. Each HELLO packet contains the list of its source node's neighbours and thus each node discovers the identities of its one- and two-hop neighbours. Also, each node monitors its CITR with the aid of the MAC protocol and disseminates this information in its HELLO packets.

In OLSR, multi-point relay (MPR) nodes are selected to forward link state information so that non-MPRs do not have to, thereby reducing overhead and energy consumption. Nodes select MPRs with a heuristic that aims to allow them to communicate with each of their two-hop neighbours. In IQOLSR, from among several candidate MPRs, the one with the highest CITR is selected. The MPR selection algorithm is invoked upon any change in local topology or threshold change in an MPR's available capacity.

For a newly-arriving session, IQOLSR selects the shortest known route on which each node has enough locally-available capacity. A probe packet is then sent along this route and causes each node to check its two-hop/cs-neighbourhood's minimum residual capacity. If any node detects this is insufficient, a rejection message is returned to the session's source, and the session is blocked. Otherwise, the destination sends a message confirming the session's acceptance.

The entire route of a session is stored in each data packet header (akin to DSR) in order to avoid individual packets of the same session being routed along different paths. When a link on the path breaks, the detecting node notifies the source, which replaces the route selected for this session with another feasible one.

The proactive nature of IQOLSR theoretically enables fast recovery from route failures and QoS requirement violations. Also, there is inherent redundancy meaning that more than one MPR may be suitable for reaching a two-hop neighbour. Therefore, a session requiring re-routing may not even need to wait for the next HELLO interval. The protocol does incur a relatively large overhead since it relies on proactive OLSR routing. The MPR feature somewhat reduces this compared to earlier, purely proactive link state protocols. However, another shortcoming of this protocol seems to be that it does not consider the intra-route contention [34], potentially underestimating sessions' capacity requirements.

12) Common Advantages and Drawbacks in this Category: While obviously incurring periodic overhead, and thus permanently taking up a portion of each node's capacity and increasing the chance of collisions, beaconing has several advantages. Firstly, within the range in which the neighbour information is forwarded, beacon packets implement a form of proactive route discovery. Therefore, source nodes that are within this range of their destinations can avoid the route discovery procedure if they have separate route testing functionality. Secondly, there is no delay in testing the resources of cs-neighbours, resulting in lower session admission times compared to the previous category of protocols. Thirdly, beaconing-based protocols may be able to react to network dynamics more quickly. However, as stated, periodic overhead is incurred and much of the disseminated information might never be used for AC.

D. Passive Resource Discovery-Based Schemes

Into this final category, we place AC protocols that test the resources of each node on a route, and those of the nodes within their sensing ranges via passive monitoring.

1) Carrier-Sensing-based Contention-Aware Call Admission Control Protocol: Most functions of the CACP protocol were described in Section V-B1. However, [11], [22] also proposed a passive method of ascertaining the residual capacity of a node's cs-neighbours at the RRep stage. This method is referred to as CACP-carrier-sensing mode, or CACP-CS. Aside from the cs-thresh, it employs a second, lower, neighbour cs-threshold (the ncs-thresh) to sense all transmissions occurring to and from its two-hop cs-neighbours. CACP-CS then estimates the available channel capacity as the CITR detected by the ncs-thresh, multiplied by the raw channel capacity. The CITR measured with the ncs-thresh excludes all transmission periods that could possibly cause any of a node's cs-neighbours to detect a busy channel and thereby decrease their available channel capacity. This means that it is not necessary to transmit any AdReq packets. Based on the adopted "cs-range=two hops" model, this neighbour-cs-range (ncs-range) must be equal to four hops. This can

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

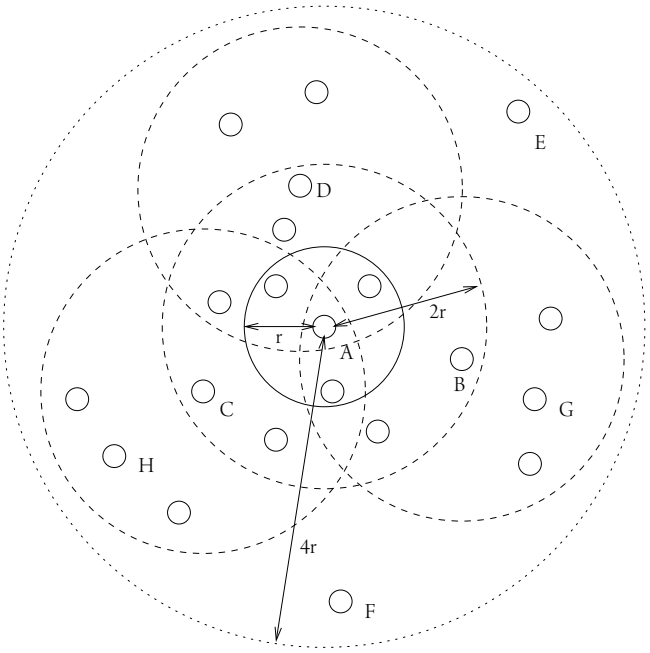

Fig. 7. A passive resource discovery method. The smallest solid-lined circles are nodes and the larger solid-lined circle of radius r represents node A's transmission coverage area. The second-largest dashed circles, of radius $2r$, represent the average cs-range coverage areas of nodes A, B, C and D, assuming negligible shadowing. Finally, the largest, dotted circle of radius $4r$ illustrates node A's neighbour-cs-range coverage area. Within the area of the dotted circle, node A senses all transmissions which could potentially decrease the available channel time of its cs-neighbours (such as nodes B, C and D), and thereby avoids having to explicitly query them. For example, if node G transmits, it reduces the channel time available to node B, which is one of A's cs-neighbours. Using the ncs-thresh, node A can take this into account. However, node E is not within the cs-range of any of node A's cs-neighbours, but node A still senses its transmissions. Therefore, node A incorrectly assumes that the channel time used by node E is made unavailable to all of its cs-neighbours. Another manifestation of this problem is the lack of consideration of the possibility of parallel transmissions. For example, transmissions from nodes G and H both affect at least some of node A's cs-neighbours. Therefore, their transmissions should be considered. However, node A cannot know that they can transmit at the same time because they are not within each other's cs-ranges. Therefore, unless their transmissions completely overlap in time, node A adds up the channel time they use and subtracts this from the estimate of the channel time available at all cs-neighbours. This leads to the cs-neighbourhood's available capacity being underestimated.

be implemented by exploiting the received signal strength indicator (RSSI) function provided by 802.11's various PHY specifications [4]. If the ncs-thresh is lower than the threshold at which the PHY reports a busy channel to the MAC, then the monitoring range can be increased without increasing the cs-range that is employed by the MAC protocol. Therefore, the amount of spatial reuse is not decreased. Figure 7 illustrates an example and the potential weaknesses of this approach.

In light of the examples provided in Figure 7, it is clear that CACP-CS estimates the lower bound on the capacity available in a node's cs-neighbourhood. Therefore, it can be overly-conservative in some situations. In fact, the use of any monitoring range that is greater than the cs-range has the potential to underestimate the residual capacity of some cs-neighbours, as demonstrated by Figure 7. Aside from its method of residual capacity estimation, CACP-CS operates in the manner described in Section V-B1.

2) Perceptive Admission Control: The work in [38], [56] introduced the perceptive admission control (PAC) protocol, which operates on a basis similar to CACP-CS, albeit with some modifications. PAC again uses passive monitoring to estimate the available capacity at the current node and its neighbours. However, PAC's monitoring threshold is set such that the average ncs-range r_{ncs} is less than that used by CACP-CS, on the following basis. Given a particular signal-to-interference-plus-noise ratio (SINR) requirement for a received signal to be reliably decoded, there is a minimum signal power with which interference can cause a collision with a packet being received. Knowing the signal propagation characteristics, this gives a maximum collision interference range (CIR) r_{ci}. Thus, for a transmitter to sense all of the transmissions that could cause a collision at its receiver, due to 802.11's ACKs, the monitoring range must be twice the transmission/reception range r plus the CIR, i.e. $r_{ncs} = 2r + r_{ci}$ (as between nodes A and D in Figure 8). The cs-neighbourhood's available capacity is estimated using the CITR detected with the ncs-thresh set to achieve the above-mentioned ncs-range.

Figure 8 illustrates the ranges employed by PAC's capacity estimation mechanism and highlights some of its drawbacks. The figure shows that the thinking behind the setting of the ncs-thresh is intelligent, but comes with an unavoidable trade-off. If the CIR is less than the cs-range, the monitoring range does not encompass all of the cs-neighbours of the current node's cs-neighbours which could decrease their available capacity. On the other hand, if the ncs-range is increased, the problems discussed in Section V-D1, with CACP-CS, are exacerbated. The conclusion is that there is no single "correct" setting for the ncs-threshold, instead, as expected, the level of spatial reuse can be traded off against the probability of false admissions.

PAC can be coupled with a QoS-aware routing protocol like CACP's in order to perform multi-hop AC. In the case of mobility causing imminent congestion due to unexpected interference, PAC detects if the CITR decreases below a threshold level and commands the source nodes of affected sessions to pause packet sending for a random back-off period. After this, source nodes can attempt to re-admit any paused sessions.

As stated above, PAC suffers from the same problems as CACP-CS, exemplified in Figure 7, albeit to a different degree. Its remaining pros and cons are common to this category of protocols and hence are discussed later in this section.

3) Multi-Path Perceptive Admission Control : A multi-path-aware extension to PAC was introduced in [35], which is referred to as Multi-path Admission Control for Mobile Ad hoc Networks (MACMAN). In MACMAN, the route discovery procedure follows a source-routing approach similar to CACP's (Section V-B1). However, local residual capacity at nodes is tested with PAC's mechanism, described in the previous sub-section. The intra-route contention is also taken into account in a manner akin to CACP. Multiple routes are discovered by ensuring that the destination replies to all arriving RReqs; a feature first seen in DSR [46]. The full routes are stored in the source's cache.

In order to ensure that only routes which satisfy a session's throughput requirement are stored, periodic Route Capacity

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

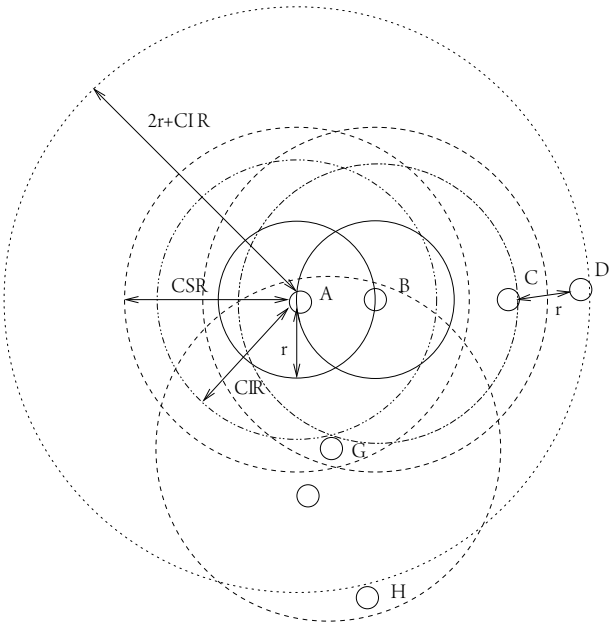

Fig. 8. Illustration of ranges of interest for PAC. The smallest, letter-labeled circles are nodes. The circles of radius r represent nodes A's and B's transmission range coverage areas. The dash-dotted circles of radius CIR show their collision interference range coverage areas within which transmissions can cause collisions with their received packets. Thirdly, the dashed circles of range CSR stand for the same nodes' cs-range coverage areas. Finally, the dotted circle of radius $2r + CIR$ represents node A's PAC monitoring coverage area. Again, equal signal attenuation in all directions is assumed. Consider an example. Node A's transmissions to node B cause B to reply with acknowledgement frames (ACKs). These can cause collisions at node C. Similarly, node D's transmissions to C potentially lead to C's ACKs causing collisions at node B. With the PAC monitoring range, A can sense all transmissions within $2r + CIR$ and hence can establish the fraction of the channel time in which it can transmit without fear of collisions at its receiver, node B. However, consider a second example. The channel capacity is fixed at 2Mbps. Node H is transmitting at 1Mbps. Node G is transmitting at 500kbps. Node A wishes to admit a session of 1Mbps. No other nodes are transmitting, so node A senses only node G's transmissions, resulting in its assumption that G has a residual capacity of 1.5Mbps. It cannot sense node H's transmissions. However, node H is within G's cs-range and so G senses a busy channel when H is transmitting, even though H is too distant to cause collisions at G (typically, $CSR > CIR$ for collision avoidance purposes). This means that node G's residual capacity is only 500kbps. When node A admits its session, node G will be starved of transmission opportunities.

Query messages are sent along each of the backup paths. These carry a copy of the session's current route. Each node on each backup path tests its residual capacity with PAC's mechanism. However, since the session has already been admitted, some nodes of the session's current path might already be imposing interference on the nodes on the backup path. If the session was re-routed, the capacity thereby consumed would be freed again. With this in mind, prior to comparing a node's residual capacity to the session's requirement on the backup path, the difference in the node's contention counts on the two routes is calculated. To facilitate this operation, each node must know which of the nodes on the session's current route are its cs-neighbours, and are therefore reducing its available capacity. Figure 9 provides an example. MACMAN employs periodic beacons transmitted at a higher power for cs-neighbour discovery [57].

If, in this manner, any node on a backup route detects that it no longer has adequate capacity, a "Route Capacity Failed" message is returned to the source. This then deletes the corresponding route. If no alternative routes to a destination are known, a new backup route search is initiated.

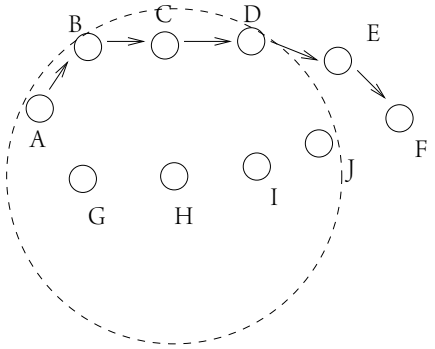

Fig. 9. The inter-route interference between a session's current route {A,B,C,D,E,F} and an alternative route from A to F. The dashed circle represents node H's cs-range coverage area. Node H is selected for illustration purposes. As part of the alternative route {A,G,H,I,J,F}, node H's contention count is five, since its cs-range encompasses five transmitters including itself. Notice also that node H's cs-range encompasses nodes A, B, C and D on the session's current route. Therefore, capacity equivalent to four times the session's end-to-end rate is already being consumed by the session at node H. This means that, when testing the route {A,G,H,I,J,F}, only the capacity equivalent to the difference in contention counts (i.e. one) times the session's rate must be currently free at node H in order to support the session if it was redirected to this alternative route.

The main obvious advantage of this protocol is that potentially several backup routes are known by a traffic source at any time. This means that lapses in end-to-end throughput can be avoided if a session's primary route fails. Moreover, each backup route is periodically ensured to have adequate end-to-end capacity for the requesting session. However, the periodic testing of every known backup route incurs extra overhead. The periodic high-power beacons employed for cs-neighbour discovery may also increase the chance of collisions. Note additionally that the residual capacity estimation scheme inherits PAC's strengths and shortcomings.

4) Multi-Rate-Aware Admission Control: Akin to softMAC (Section V-C10), the protocol considered in [37] considers heterogeneous link rates during admission control. However, the other features of this protocol, which we will term Multi-rate- and contention-aware admission control protocol (MRCACP), build on the ideas presented in the form of the CACP [11] (Section V-B1).

A method similar to CACP-CS, with two sensing thresholds, is employed for monitoring the CITR both inside a node's cs-range and inside its ncs-range. Recall from the description of CACP that CACP-CS is overly conservative when estimating the capacity available to cs-neighbours (this is also demonstrated in [25]). During the AC procedure for a new session, an innovation of the MRCACP mitigates this to some extent by considering the possible time overlap between the new session's transmissions and the transmissions originating outside of the current node's cs-range. This is done by considering the measured channel busy time ratios (CBTRs) within the cs-range T^{cs}_{busy} (the medium-dark grey-shaded ring in Figure 10) and the ncs-range T^{ncs}_{busy} (the light grey-shaded ring in Figure 10) as independent probabilities of transmission. Therefore, the chance of the current node's transmissions overlapping with the transmissions originating outside the cs-

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

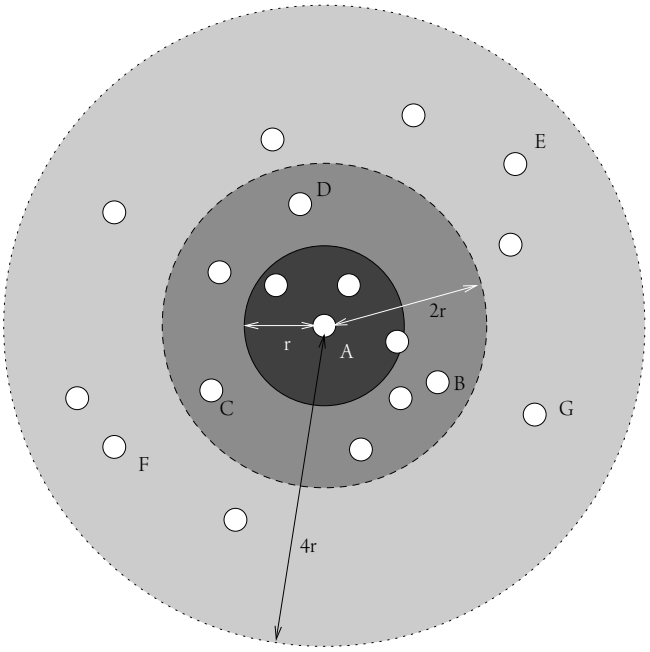

Fig. 10. Illustration of the ranges of interest for MRCACP, where r, $2r$ and $4r$ denote the average transmission, cs- and ncs-ranges, respectively. As an example, MRCACP considers the chance of node A, a session-admitting node, being able to transmit at the same time as nodes such as E, F and G, which are outside its cs-range, but inside its ncs-range. The transmissions of nodes E, F and G must be sensed by node A because they may decrease the available capacity of A's cs-neighbours, like node B. However, node A itself is not affected by any of them, meaning that it can transmit at the same time. This consideration is an improvement on the method of CACP-CS, exemplified in Figure 7. However, this model ignores the fact that ncs-neighbours interfering with node A's receivers could still stop node A's transmissions from being received. Therefore, the effective fraction of overlapping time in which node A can successfully transmit is less than that predicted by the described model.

range but inside the ncs-range is simply $\left(T_{busy}^{ncs} - T_{busy}^{cs}\right) T_{tx}^{A}$, where T_{tx}^{A} is the fraction of time required by the current node, A, for transmitting the packets of the new session [37]. The capacity available to the current node for admitting a new session is then not $1 - T_{busy}^{ncs}$ as in CACP-CS, but $1 - T_{busy}^{ncs} + \left(T_{busy}^{ncs} - T_{busy}^{cs}\right) T_{tx}^{A}$.

Figure 10 illustrates an example scenario and discusses some of the shortcomings of this model. Note also that MRCACP does not eliminate the inaccuracy caused by the assumption that all ncs-neighbours may reduce the available capacity of all cs-neighbours, as exemplified in Figure 7.

Having heterogeneous link rates means that the calculation of a node's contention count on a route is slightly complicated. A session's channel occupation time on each link is calculated by considering a rate-independent part (MAC headers and control frames and inter-frame spaces) and a link rate-dependent part (network and higher-layer headers and data).

The necessary routing functionality is built upon the lightweight underlay network ad hoc routing (LUNAR) protocol [58]. This protocol initially discovers routes on-demand but then periodically reconstructs all active routes from scratch. The initial route discovery procedure is again similar to the CACP-CS variant (Section V-D1) of CACP (Section V-B1). The difference is that each node only rebroadcasts the RReq if its cs-neighbourhood has sufficient residual capacity after adding on the potential overlapping transmission time, as described above, while considering the link-rate dependent channel occupation time on the two previous links and the next hop. The RRep stage is the same except with full knowledge of the session's channel occupation time in each intermediate node's cs-range. A destination replies with a RRep to each received RReq.

In the interests of route and QoS-assurance maintenance, MRCACP exploits the periodic route-refreshing operation of LUNAR. In each refresh period, preference is given to a session's current route such that it does not select a different route each time. If the refreshing mechanism detects that, for whatever reason, a node can no longer meet the QoS commitments it made to admitted sessions, [37] states that the protocol either rejects or finds alternative routes for the affected sessions.

One shortcoming of this method is that the neighbour cs-sensing mechanism still underestimates the available capacity at cs-neighbours, despite the above-described parallel transmissions probability model. Also, the overhead incurred by LUNAR's proactive refreshing of routes is only tolerable because LUNAR limits its route searches to a three hop radius. The motivation for this is that the authors believe that useful and feasible MANETs are limited to three hops in radius. However, since this is not always the case, this protocol's usefulness is limited to small MANETs. If this limitation was removed, the proactive route-refreshing operation would incur significant overhead.

5) Common Advantages and Drawbacks of Passive Channel Capacity Monitoring: Again, it is not difficult to see the appeal of passive resource monitoring. Firstly, compared to on-demand active methods, passive methods allow quicker session establishment. Also, compared to both active types of resource information gathering, they greatly reduce the protocol's overhead. However, as we have highlighted, passive monitoring is subject to a trade-off. As the descriptions of the individual protocols have demonstrated, the capacity-monitoring range must be greater than the cs-range in order to sense the transmissions of the cs-neighbours of cs-neighbours, whose transmissions can reduce their residual capacity. This allows the sensing node to estimate the fraction of the channel capacity that is available to its cs-neighbours and hence may be consumed by the sensing node's transmissions. If the sensing range is too short, not all of the transmissions that affect the residual capacity of cs-neighbours will be sensed. At the same time, *any* monitoring range may cause a node to sense transmissions that do not affect all of its cs-neighbours. Moreover, passive monitoring cannot tell the sensing node which of its cs-neighbours even have QoS-sensitive sessions to forward. Its cs-neighbours may not even require any transmission opportunities to be reserved for them. These factors all lead to a potentially significant underestimation of the available capacity. The greatest difficulty lies in the fact that the optimal monitoring threshold is different for each network topology and traffic pattern. Therefore, without explicitly querying each cs-neighbour individually, in order to find out their *own* assessment of their residual capacity, there is no way to reliably and accurately predict the impact that an admitted session would have on these cs-neighbours.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

Furthermore, without explicit querying, an admitting node cannot know whether any other sessions have reserved capacity at cs-neighbours, which they are not yet using. Such reservations should be subtracted from the estimate of the capacity that is currently available to those neighbours. Admittedly, this factor only becomes significant when more than one data session is beginning, ending or being re-routed within the same short time period in the same region of the network. Passive monitoring is also unable to ascertain the states of resources other than the available channel time, such as the residual buffer space.

E. Common Advantages and Drawbacks of Routing-Coupled Protocols

A major advantage of coupling the AC protocol with the route discovery process is that it is much less likely that network resources are wasted by discovering routes that could not serve waiting sessions anyway. On the other hand, this approach can be rather short-term oriented. In many cases, just because some routing information is not useful at the current time, does not mean it will not be useful later. This applies mainly to source-routed protocols such as DSR, where multiple full routes and sequences of hops can be stored. If there is a process for separately testing pre-discovered routes, as in decoupled protocols, opportunistically-discovered routing information can be made useful later on. For example, DSR [48] can snoop on routing information from overheard packet transmissions. Thus, in some cases, hybrid methods (such as StAC's, Section V-B2) that allow both coupled and decoupled AC-routing procedures may perform the best in terms of total overhead.

When no routing information is known yet, coupled methods may achieve shorter session admission times because decoupled protocols must wait for both the separate route discovery and route testing phases to be completed. Also, as discussed in Section IV-C, routing-coupled schemes can make more fine-grained admission decisions.

VI. Patterns and Trends in the Field

Several dominant patterns in the design as well as trends in the development of AC protocols can be identified.

A. Metrics and Methods of Estimation

Firstly, in accordance with the statement in Section II-B, that channel capacity is the most important network resource, the operation of most protocols in the literature focuses on the estimation and management of this resource. Also, most protocols consider throughput to be the most important QoS metric. This is reflected by the large proportion of protocols aiming to provide a guaranteed throughput service. The survey has revealed that there are generally three distinct categories of approaches to estimating the achievable throughput. Protocols may either:

1) monitor the channel idle time ratio (CITR), and then take the minimum value on a route (possibly including cs-neighbours as well) [11], [25], [27], [29], [31], [32], [34], [35], [37], [38], [39], [41], [42], [43], [12];
2) estimate the channel capacity used for communications and subtract this from the raw channel capacity. If the cs-neighbourhood capacity is considered, subtract the channel usage of cs-neighbours to obtain a node's residual capacity. Then, again take the bottleneck value on a route [15], [16], [18], [30], [36], [45];
3) or use the delay between transmitting or receiving probe packets of a known size to estimate capacity [17], [24].

In terms of the amount of attention in protocol design, delay-, PLR- and delay-jitter-related application requirements have been considered as being of secondary importance. However, this may be because managing the channel capacity is a requirement for avoiding collisions and queue build-up, which affect the other three main metrics. For these metrics, the protocols surveyed employed the methods already discussed in Section II-C.

B. Consideration of Channel Contention

Having discussed the estimation of residual channel capacity and achievable throughput, there is a related observable trend for the increasingly sophisticated consideration of the mutual channel contention between nodes.

As our detailed survey has revealed, earlier proposals typically completely ignored the effects of intra-route contention on a session's capacity requirement prior to admission. They also ignored any impact that admitting the session would have on nodes which are not on the session's route. Some protocols later began to consider intra-route contention and the effect of interference within the transmission range. More recently-published schemes considered the fact that a practical cs-range must be larger than the transmission range, and thus increased the range within which they considered the impact of interference. The protocols utilising the various approaches will be listed in the next sub-section.

Finally, some protocols went on to consider richer models of the interference between nodes, as opposed to the simple "two thresholds" model using the receiving and cs-thresholds. For example, PAC [38] considered the maximum collision-causing interference range as well, and IFCAC [31] and MR-CACP [37] utilised the extra information that could be gained by monitoring channel activity using the various thresholds independently.

C. Basis for Admission Decisions and Methods of Resource Discovery

The nature of the consideration of the channel contention, discussed above, manifested itself in various approaches to making admission decisions. Several approaches were briefly mentioned in Section III. However, the descriptions of individual protocols have revealed that they may be classified into even more accurately-defined categories of approaches:

- assume a route has been found and selected. Initially admit data sessions, observe their experienced QoS, or their effect on the network resources and reject some traffic later [40], [43];
- again, assume a route has been found. Initially admit traffic on a best-effort basis, and make QoS guarantees

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

if locally-available node resources permit it. Otherwise continue to serve traffic on a best effort basis [32], [33];
- once more, assume that at least one route to the destination is known, and send probe packets along known routes. Observe the QoS experienced by the probes at the destination node. Infer the achievable QoS from this and thereby make admission decisions [17], [20];
- send probe packets on pre-selected routes. Each node predicts the achievable QoS based on the current traffic or available resources. Admit sessions if the QoS prediction delivered by the probe is sufficient [24], [39];
- test the locally-measured resources of each node during route discovery. Only allow a route to be discovered if each of its nodes has sufficient resources to support the requesting session [28], [50], [44];
- as above, but also consider the flow of the traffic on the route and the intra-route contention this causes. Use this knowledge to calculate sessions' capacity requirements more accurately. Residual capacity may be determined by subtracting the capacity used by cs-neighbours, but the potential impact of the traffic on any nodes that are not on the route is not evaluated [15], [18];
- as above, but also test the resources of the nodes neighbouring the route. Only allow the route to be discovered if all of the nodes on the route and their neighbours have adequate resources to support the session; [27], [41], [42], [45]
- as above, but also test the resources of any nodes that traffic-forwarding nodes can impose interference on, even if they cannot communicate with them directly. Only admit the session if it would not impose too much interference on the nodes surrounding the route [11], [16], [25], [29], [31], [34], [35], [36], [37], [38], [12];

For schemes that consider the resources of more than just the nodes on a session's route, three methods for establishing cs-neighbours' resources were identified, as was discussed in Section III. From among these approaches, the survey has clearly revealed that the proactive approach (Section V-C) to neighbourhood resource discovery is the most popular. The rationale behind this approach is that some method of neighbour discovery is required either way, and therefore it incurs relatively little extra overhead to piggyback resource state information on the periodic HELLO packets. Also, session establishment then incurs relatively little delay. Based on the first rationale, many protocol designers argue that their protocol adds little overhead. However, this is something of a false argument, since the sending of periodic broadcast packets can be altogether avoided by on-demand methods using link-layer neighbour discovery. Inevitably, the optimal method is scenario-dependent, with a highly dynamic network probably benefiting more from the proactive approach.

D. Statefulness and Coupling with Routing

With the exception of DACME [17], all AC protocols published after 2004 have utilised state information stored at intermediate nodes. This is most likely to be due to the extra flexibility this offers in managing data sessions, as discussed in Section IV-A4. Also, mobile devices are continually advancing in terms of capabilities and specifications, and therefore storing and managing state information is becoming a less significant burden.

Furthermore, it is clear from Figure 1, and this survey, that the majority of AC protocols operate during route discovery and are coupled with the routing protocol. The strengths and weaknesses of this approach have been discussed in Sections II-E, IV-C and V-E. In order to limit their drawbacks and maximise their benefits, hybrid approaches may provide a good compromise (e.g. Section V-B2).

E. Approach to Coping with Mobility

Due to the dynamic and unpredictable nature of MANETs, most protocol designers have assumed that QoS assurances cannot be upheld in the face of mobility. For this reason, they either:

1) rely on the routing protocol to re-route affected sessions and simply restart the AC process each time a session is re-routed [15], [16], [19], [28], [29], [30], [31], [33], [32], [34], [36], [37], [39], [40], [41], [42], [43];
2) notify source nodes which then decrease the rate of, or pause packet sending for affected sessions. This at least avoids the development of congestion, but these sessions must then be re-admitted anyway [11], [25], [38], [24], [45];
3) or attempt fast local route repair to limit the QoS assurance violation time and the development of congestion [18], [27].

There are only a few protocols [12], [17], [35], [44] that make a serious attempt to improve the robustness of throughput guarantees in the face of route failures. However, these techniques come at a price. This is either overly-conservative methods of channel capacity estimation, and periodic overhead [35], reliance on location-awareness to estimate node speeds in order to select stable routes [44], or the risking of the QoS assurances of other sessions by not testing the capacity of cs-neighbours prior to re-routing [12], [17];

F. Consideration of Collisions

While most protocols at least consider the possibility of route failures, very few consider the effect of session admission on the collision rate. The more hops on which a session's packets are forwarded, and the higher its sending rate, the larger the likely increase in collision probability, and hence the unexpected waste of capacity that it causes [12]. As far as it is possible to tell, only two protocols have considered the collision rate during AC [16], [12], and of these, only StAC [12] considers the increase in collision rate that session admission would cause.

G. Channel Capacity

With the exception of SoftMAC [16], [54], MRCACP [37] and DACME [17], all of the protocols in this survey assumed a fixed channel capacity. As mentioned previously, this simplifies AC decisions and avoids the need to model a rate-switching mechanism, which is not specified by the 802.11 standard [4], in simulations. SoftMAC and MRCACP explicitly factor the current link rate into the amount of

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

channel time a session's traffic occupies at a node. On the other hand, DACME does not explicitly consider heterogeneous link rates, but the probing mechanism can estimate the capacity of any route.

VII. Future Work

In line with the ultimate aims of current protocols, the goals of future work on AC are two-fold. Firstly, to make the initial admission decisions more reliable, leading to decreased numbers of false admissions and rejections. Secondly, it is to make the QoS guarantees more robust in the face of network dynamics. Three dynamic aspects of network operation are of particular concern. Firstly, channel quality, secondly, node mobility, and thirdly, the changing of the state of network resources due to changes in the states of applications.

The accuracy of the initial decisions depends on how closely admitting nodes' views of the network resources match reality. This in turn depends on how well the protocol has adapted to the most recent topology and resource-related changes in the network. Therefore, the issues cannot be considered separately. As we have discussed, the most important resource is the residual channel capacity since it either directly or indirectly affects all QoS metrics. The usable capacity depends on the raw channel capacity as well as on the traffic at interfering nodes.

Even though the consideration of mutual interference between nodes has become more accurate and sophisticated over the years, the models are still somewhat simplistic. Most AC protocol designers still assume that the transmission, collision and cs-ranges are of fixed radius, whereas those ranges should be considered merely statistical averages [59]. For example, if shadow fading and multi-path fading were considered, these ranges may fluctuate about the mean and/or be direction-dependent [59]. Due to these fixed-range assumptions, all current protocols also approximate the area impacted by a node's transmissions as being of a fixed size. This may lead to incorrect admission decisions if realistic channel conditions are modelled. The use of real network test-beds can circumvent these modelling inaccuracies. However, development time is long and costs are high, especially for testing large networks. Simulation is a useful tool because it avoids those drawbacks and allows a range of deployment environments to be easily investigated. Thus, future work on AC with more accurate modelling of the physical layer is of great interest.

Other aspects of modelling that can be improved are traffic and mobility. With the exception of a few papers that utilised video codec-produced traffic, most of the AC protocols in this survey were evaluated with constant bit-rate (CBR) data sessions. While this allows the admission decision-making features of a protocol to be tested under various loads, it may not represent the performance that would be achieved in a real network sufficiently accurately [60]. Thus, more work on protocol evaluation with realistic traffic models is required. Secondly, while random movement-based models, such as the random waypoint mobility model [46] may do a satisfactory job of representing people mingling in a conference hall [61], they do not accurately represent mobility patterns for many networks. Again, evaluation of the QoS assurance-upholding capabilities of joint QoS-aware routing and AC protocols with more realistic mobility models for various scenarios would be useful.

Some work has been done on mobility-tolerant protocols, again as discussed in the previous section. However, future work should incorporate techniques for coping with both mobility and channel dynamics. Varying levels of interference and channel quality necessitate adaptation of the transmission modulation scheme and hence the transmission rate, thereby varying the raw channel capacity. As discussed in the previous section, some protocols already incorporate awareness of heterogeneous link rates, although the impact of frequently-varying link quality and rates was not studied. Future works should also consider heterogeneous QoS requirements, both in terms of type (throughput, delay, PLR etc.) and level of requirement. In summary of this paragraph, works considering many of these network dynamics individually already exist, but future protocols should be able to handle all of them concurrently.

As this paper has highlighted, the maintenance of network resource state information can be performed on-demand, proactively, or, in some cases, passively. However, all of these approaches have shortcomings. Future work on more intelligent hybrid methods, that adapt the approach to the state of the network resources, would be of practical significance.

The ideal scheme for the provision of QoS assurances would precisely know the area impacted by each transmission, and would be able to predict changes in topology and the availability of resources before they happen. To even begin to approach this ideal, accurate propagation and fading models, as well as node location awareness are required [62]. Accurate propagation and fading prediction models would foresee the impact of each transmission, while location awareness would provide distance estimates for the propagation model and could be used to infer node speeds and travel directions. The two models combined could be used to predict link failures [62]. Therefore, if mobile devices were equipped with location-determination systems, such as GPS receivers, and could activate an appropriate long-range propagation prediction model from a list of stored ones, more accurate admission decisions would be enabled. Investigations of the performance of such systems with existing 802.11 hardware would be interesting for establishing the limits on multi-hop MANET-based AC performance, and the cost of achieving them.

In order to alleviate the capacity limits of traditional omni-directional antenna-based systems, multiple antenna array-equipped MANETs have been envisioned. Already, a framework for AC based on such a system, has been published [63]. Such systems fundamentally alter the interference patterns among nodes as they can maximise signal strength towards an intended receiver and nullify it in other directions. This alters the nature of the tests that must be performed on a node's neighbourhood's resources prior to session admission. Although such systems are not directly compatible with the existing 802.11 MAC protocol, contention-based channel access schemes will remain relevant for the foreseeable future. Directional antenna-based systems also have other drawbacks, such as the introduction of a different type of hidden terminal,

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

higher directional interference, difficulty in maintaining the correct transmission direction in the face of mobility and deafness to certain transmissions [64]. Some MAC protocols for addressing these issues have been proposed [64], but higher-layer protocols still require attention. The design of AC and other communications protocols that tolerate these problems and effectively utilise directional antenna-based systems is thus an interesting area of future work.

VIII. Summary

This paper has firstly provided a thorough background on the field of admission control in IEEE 802.11-based multi-hop mobile ad hoc networks. The QoS metrics of interest, the types of network resources, and the challenges and trade-offs in protocol design were discussed. Secondly, a comprehensive survey of AC schemes found in the open literature, which can operate in multi-hop 802.11-based MANETs, was conducted. Several methods of protocol classification were proposed. The operation of 28 protocols was summarised, and the advantages and shortcomings particular to them, as well as common to their category of protocols, were highlighted. Finally, trends in the field were identified and possible avenues of future research were proposed.

References

[1] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, *Mobile Ad Hoc Networking*. Wiley-IEEE, 2004.

[2] R. Ramanathan, J. Redi, and B. Technologies, "A Brief Overview of Ad Hoc Networks: Challenges and Directions," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 20–22, 2002.

[3] H. Menouar, F. Filali, and M. Lenardi, "A Survey and Qualitative Analysis of MAC Protocols for Vehicular Ad Hoc Networks," *IEEE Wireless Commun.*, vol. 13, no. 5, pp. 30–35, 2006.

[4] IEEE Computer Society, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 2007. IEEE Std. 802.11-2007.

[5] L. Hanzo II. and R. Tafazolli, "A Survey of QoS Routing Solutions for Mobile Ad Hoc Networks," *IEEE Commun. Surveys Tutorials*, vol. 9, no. 2, pp. 50–70, 2007.

[6] M. Ahmed, "Call Admission Control in Wireless Networks: a Comprehensive Survey," *IEEE Commun. Surveys Tutorials*, vol. 7, no. 1, pp. 49–68, 2005.

[7] D. Gao, J. Cai, and K. Ngan, "Admission Control in IEEE 802.11e Wireless LANs," *Network, IEEE*, vol. 19, no. 4, pp. 6–13, 2005.

[8] J. Ratica and L. Dobos, "Mobile Ad-Hoc Networks Connection Admission Control Protocols Overview," in *Proc. 17th Int. Conf. Radioelektronika*, (Brno, Czech Republic), pp. 1–4, Apr. 2007.

[9] A. R. Bashandy, E. K. P. Chong, and A. Ghafoor, "Generalized Quality-of-Service Routing With Resource Allocation," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 450–463, Feb 2005.

[10] M. Wang and G.-S. Kuo, "An Application-Aware QoS Routing Scheme with Improved Stability for Multimedia Applications in Mobile Ad Hoc Networks," in *Proc. IEEE Vehicular Technology Conf.*, (Dallas, TX, USA), pp. 1901–1905, Sep. 2005.

[11] Y. Yang and R. Kravets, "Contention-Aware Admission Control for Ad Hoc Networks," *IEEE Trans. Mobile Computing*, vol. 4, pp. 363–377, Aug 2005.

[12] L. Hanzo II. and R. Tafazolli, "Throughput Assurances through Admission Control for Multi-hop MANETs," in *Proc. 18th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, (Athens, Greece), pp. 1–5, Sep. 2007.

[13] C. Skianis and D. Kouvatsos, "An Information Theoretic Approach for the Performance Evaluation of Multihop Wireless Ad Hoc Networks," in *Proc 2nd Int.Conf. Performance Modelling and Evaluation of Heterogeneous Networks*, (Ilkley, UK), pp. 1–13, 2004.

[14] L. Kleinrock and J. Silvester, "Spatial Reuse in Multihop Packet Radio Networks," *Proc. IEEE*, vol. 75, no. 1, pp. 156–167, 1987.

[15] L. Chen and W. Heinzelman, "QoS-Aware Routing Based on Bandwidth Estimation for Mobile Ad Hoc Networks," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 561–572, Mar. 2005.

[16] H. Wu, Y. Liu, Q. Zhang, and Z.-L. Zhang, "SoftMAC: Layer 2.5 Collaborative MAC for Multimedia Support in Multi-hop Wireless Networks," *IEEE Trans. Mobile Computing*, vol. 6, pp. 12–25, Jan. 2007.

[17] C. Calafate, J. Oliver, J. Cano, P. Manzoni, and M. Malumbres, "A Distributed Admission Control System for MANET Environments Supporting Multipath Routing Protocols," *Microprocessors & Microsystems*, vol. 31, no. 4, pp. 236–251, 2007.

[18] Q. Xue and A. Ganz, "Ad Hoc QoS On-Demand Routing (AQOR) in Mobile Ad Hoc Networks," *J. Parallel and Distributed Computing*, vol. 63, no. 2, pp. 154–165, 2003.

[19] C. E. Perkins, E. M. Belding-Royer, and S. R. Das, "Quality of Service in Ad hoc On-Demand Distance Vector Routing." IETF Internet Draft, Jul. 2000.

[20] C. Calafate, J. Cano, P. Manzoni, and M. Malumbres, "A QoS Architecture for MANETs Supporting Real-Time Peer-to-Peer Multimedia Applications," in *Proc. 7th IEEE Int. Sym. Multimedia*, (Irvine, CA, USA), pp. 193–200, Dec. 2005.

[21] S. Saunders, *Antennas and Propagation for Wireless Communication Systems: Concept and Design*. New York, USA: John Wiley and Sons, 1999.

[22] Y. Yang and R. Kravets, "Contention-Aware Admission Control for Ad Hoc Networks," tech. rep., University of Illinois at Urbana Champaign, 2003.

[23] J. Deng, B. Liang, and P. Varshney, "Tuning the Carrier Sensing Range of IEEE 802.11 MAC," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, vol. 5, (Dallas, TX, USA), pp. 2987–2991, Nov. 2004.

[24] G.-S. Ahn, A. T. Campbell, A. Veres, and L.-H. Sun, "Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN)," *IEEE Trans. Mobile Computing*, vol. 1, pp. 192–207, Jul. 2002.

[25] R. Renesse, V. Friderikos, and H. Aghvami, "Cross-layer Cooperation for Accurate Admission Control Decisions in Mobile Ad Hoc Networks," *IET Communications*, vol. 1, no. 4, pp. 577–586, 2007.

[26] R. de Renesse, P. Khengar, V. Friderikos, and H. Aghvami, "QoS Conflict Resolution in Ad Hoc Networks," in *Proc. IEEE Int. Conf. on Communications*, vol. 8, pp. 3826–3831, June 2006.

[27] N. Kettaf, H. Abouaissa, T. Vuduong, and P. Lorenz, "A Cross layer Admission Control On-demand Routing Protocol for QoS Applications," *Int. J. Computer Science and Network Security*, vol. 6, no. 9B, p. 98, 2006.

[28] M. Haq, M. Matsumoto, J. Bordim, M. Kosuga, and S. Tanaka, "Admission Control and Simple Class-based QoS Provisioning for Mobile Ad Hoc Network," in *Proc. 60th Vehicular Technology Conf.*, vol. 4, (Los Angeles, CA, USA), Sep. 2004.

[29] A. Derhab and A. Bouabdallah, "Admission Control Scheme and Bandwidth Management Protocol for 802.11 Ad hoc Networks," in *Proc. 4th Int. Conf. Innovations in Information Technology*, (Dubai, UAE), pp. 362–366, Nov. 2007.

[30] Y. Dong, D. Makrakis, and T. Sullivan, "Effective Admission Control in Multihop Mobile Ad Hoc Networks," in *Proc. Int. Conf. Communication Technology*, vol. 2, (Beijing, China), pp. 1291–1294, Apr. 2003.

[31] K. Sridhar and M. Chan, "Interference-based Call Admission Control for Wireless Ad Hoc Networks," in *Proc. 3rd Int. Conf. Mobile and Ubiquitous Systems: Networking & Services*, (San Jose, CA, USA), pp. 1–10, Jul. 2006.

[32] D. Dharmaraju, A. Roy-Chowdhury, P. Hovareshti, and J. Baras, "INORA-A Unified Signaling and Routing Mechanism for QoS Support in Mobile Ad Hoc Networks," in *Proc. Int. Conf. Parallel Processing Workshops*, pp. 86–93, 2002.

[33] S. Lee, G. Ahn, X. Zhang, and A. Campbell, "INSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks," *J. Parallel and Distributed Computing*, vol. 60, no. 4, pp. 374–406, 2000.

[34] D. Nguyen and P. Minet, "Interference-Aware QoS OLSR for Mobile Ad-Hoc Network Routing," in *Proc. 6th Int. Conf. Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS Int. Wksp. on Self-Assembling Wireless Networks (SNPD/SAWN'05)*, (Towson, MD, USA), pp. 428–435, May 2005.

[35] A. Lindgren and E. Belding-Royer, "Multi-Path Admission Control for Mobile Ad Hoc Networks," *Mobile Computing and Communications Review*, vol. 8, no. 4, pp. 68–71, 2004.

[36] Y. Yang and R. Kravets, "Throughput Guarantees for Multi-priority Traffic in Ad Hoc Networks," *Ad Hoc Networks*, vol. 5, no. 2, pp. 228–253, 2007.

[37] L. Luo, M. Gruteser, H. Liu, D. Raychaudhuri, K. Huang, and S. Chen, "A QoS Routing and Admission Control Scheme for 802.11 Ad Hoc Networks," in *Proc. Int. Conf. Mobile Computing and Networking*, (Los Angeles, CA, USA), pp. 19–28, Sep. 2006.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.

[38] I. D. Chakeres and E. M. Belding-Royer, "PAC: Perceptive Admission Control for Mobile Wireless Networks," in *Proc. 1st Int. Conf. Quality of Service in Heterogeneous Wired/Wireless Networks (QShine)*, (Dallas, TX, USA), pp. 18–26, Aug. 2004.
[39] Y. Pei and V. Ambetkar, "Distributed Flow Admission Control for Multimedia Services Over Wireless Ad Hoc Networks," *Wireless Personal Communications*, vol. 42, no. 1, pp. 23–40, 2006.
[40] C. McCann, G. Elmasry, B. Russell, and B. Welsh, "A Measurement-based Approach for Multi-level Admission of Heterogeneous Traffic in Wireless Ad Hoc Networks," in *Proc. Military Comms. Conf.*, (Monterey, CA, USA), pp. 1562–1565, Oct. 2004.
[41] S.-L. Su, Y.-W. Su, and J.-Y. Jung, "A Novel QoS Admission Control for Ad Hoc Networks," in *Proc. Wireless Communications and Networking Conf. (WCNC)*, (Hong Kong), pp. 4193–4197, Mar. 2007.
[42] H. Zhu and I. Chlamtac, "Admission Control and Bandwidth Reservation in Multi-Hop Ad Hoc Networks," *Computer Networks*, vol. 50, no. 11, pp. 1653–1674, 2005.
[43] Y. Yang and R. Kravets, "Distributed QoS Guarantees for Realtime Traffic in Ad Hoc Networks," in *Proc. 1st Ann. IEEE Conf. Sensor and Ad Hoc Communications and Networks*, (Santa Clara, CA, USA), pp. 118–127, Oct. 2004.
[44] R. Zhang and I. Rubin, "Robust Flow Admission Control and Routing for Mobile Ad hoc Networks," in *Proc. Military Communications Conf. (MILCOM)*, (Los Angeles, CA, USA), pp. 1–7, Oct. 2006.
[45] C. R. Cerveira and L. H. M. K. Costa, *Mobile and Wireless Communication Networks*, vol. 211, ch. A Time-based Admission Control Mechanism for IEEE 802.11 Ad Hoc Networks, pp. 217–228. Springer Boston, November 2006.
[46] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," in *Mobile Computing* (Imielinski and Korth, eds.), vol. 353, pp. 153–181, Kluwer Academic Publishers, 1996.
[47] Y. Yang and R. Kravets, "Throughput guarantees for multi-priority traffic in ad hoc networks," in *Proc. IEEE Int. Conf. Mobile Ad-hoc and Sensor Systems*, (Fort Lauderdale, FL, USA), pp. 379–388, Oct. 2004.
[48] D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR)." RFC 4728, Feb. 2007.
[49] V. .Park and S. Corson, "Temporally Ordered Routing Algorithm v1 Functional Specification." IETF Internet Draft, Nov. 1997.
[50] C. E. Perkins and E. M. Belding-Royer, "Quality of Service for Ad hoc On-Demand Distance Vector Routing." IETF Internet Draft, Oct. 2003.
[51] C. E. Perkins and E. M. Royer, "Ad hoc On-Demand Distance Vector Routing," in *Proc. 2nd IEEE Wksp. Mobile Computing Systems and Applications*, (New Orleans, LA, USA), pp. 90–100, Feb. 1999.
[52] R. de Renesse, M. Ghassemian, V. Friderikos, and A. H. Aghvami, "Adaptive Admission Control for Ad Hoc and Sensor Networks Providing Quality of Service," tech. rep., Center for Telecommunications Research, King's College London, UK, 2005.
[53] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distribued Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, Mar. 2000.
[54] H. Wu, X. Wang, Y. Liu, Q. Zhang, and Z. Zhang, "SoftMAC: Layer 2.5 MAC for VoIP Support in Multi-Hop Wireless Networks," in *Proc. 2nd Ann. IEEE Conf. Sensor and Ad Hoc Communications and Networks*, (Santa Clara, CA, USA), pp. 441–451, Sep. 2005.
[55] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot, "Optimized Link State Routing Protocol for Ad hoc Networking," in *Proc. IEEE Multi Topic Conf.*, (Lahore, Pakistan), pp. 62–68, Dec. 2001.
[56] I. Chakeres, E. Belding-Royer, and J. Macker, "Perceptive Admission Control for Wireless Network Quality of Service," *Ad Hoc Networks*, vol. 5, no. 7, pp. 1129–1148, 2007.
[57] A. Lindgren, "Multi-path Admission Control for Mobile Ad hoc Networks," in *Proc. 2nd Annual Int. Conf. Mobile and Ubiquitous Systems: Networking and Services*, (San Diego, USA), pp. 407–417, Jul. 2005.
[58] C. Tschudin, R. Gold, O. Rensfelt, and O. Wibling, "LUNAR: a Lightweight Underlay Network Ad-hoc Routing Protocol and Implementation," in *Proc. Next Generation Teletraffic and Wired/Wireless Advanced Networking Conf. (NEW2AN)*, (St. Petersburg, Russia), Feb. 2004.
[59] I. Stojmenovic, A. Nayak, and J. Kuruvila, "Design Guidelines for Routing Protocols in Ad Hoc and Sensor Networks with a Realistic Physical Layer," *IEEE Communications Magazine*, vol. 43, no. 3, pp. 101–106, 2005.
[60] E. M. B. Stefan Karpinski and K. C. Almeroth, "Wireless Traffic: The Failure of CBR Modeling," in *Proc. IEEE Broadnets*, vol. 1, (Raleigh, NC, USA), Sep. 2007.
[61] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002.
[62] J. Stine and G. de Veciana, "A Paradigm for Quality of Service in Wireless Ad Hoc Networks Using Synchronous Signalling and Node States," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 1301–1321, Sep. 2004.
[63] B. Hamdaoui and P. Ramanathan, "A Cross-Layer Admission Control Framework for Wireless Ad-Hoc Networks using Multiple Antennas," *IEEE Trans. Wireless Commun.*, vol. 6, no. 11, pp. 4014–4024, 2007.
[64] S. Kumar, V. Raghavan, and J. Deng, "Medium Access Control Protocols for Ad Hoc Wireless Networks: a Survey," *Ad Hoc Networks*, vol. 4, no. 3, pp. 326–358, 2004.

Lajos Hanzo (II.) (StM'05) earned an MEng degree in Computer Engineering from the University of Southampton, UK, in 2004 and a PhD in mobile ad hoc networking from the University of Surrey, UK, in 2009. He is currently working on various networking- and satellite communications-related projects in industry in Germany. His research interests include QoS solutions and MAC and routing protocols for wireless ad hoc networks, as well as network security.

Rahim Tafazolli (SM) is a Professor of Mobile/Personal communications and Head of Mobile Communications Research at the Centre for Communication Systems Research (CCSR), University of Surrey, UK. He has been active in research for over 20 years and has authored and co-authored more than 300 papers in refereed international journals and conferences. He is the Editor of Technologies for the Wireless Future (Vol.1 2004 and Vol. 2 2006). Professor Tafazolli is a consultant to many mobile companies, has lectured at, chaired and been invited as keynote speaker to a number of IET and IEEE workshops and conferences. He is currently Chairman of the EU Expert Group on Mobile Technology Platform, E-Mobility.

Authorized licensed use limited to: UNIVERSITY OF SOUTHAMPTON. Downloaded on February 1, 2010 at 17:38 from IEEE Xplore. Restrictions apply.