



Policy Framework for Introduction of Network QoS into JANET

Version 3.0

**Author:
Chris Cooper**

August 2004

Table of Contents

1. Preamble	2
2. Network model.....	2
3. Provisioning	3
3.1 Overview	3
3.2 Dimensioning interdomain provisioning.....	3
3.2.1 Absolute link capacity	4
3.2.2 JANET DiffServ limit for IP Premium	4
3.2.3 Operational allocation for IP Premium.....	5
3.2.4 Operational policing limit for IP Premium.....	5
3.2.5 Dimensioning model	5
3.3 Initial experimental provisioning	5
4. Service Level Specifications	6
5. Interdomain working.....	7
6. Authentication, Authorization, and Trust.....	8
7. Policing	8
8. Service monitoring	9

1. Preamble

The background, general requirements, technical rationale, and general recommendations relating to the introduction of QoS into JANET were set out in the *QoS Think Tank Report* [1]. In short, the technical approach is to use DiffServ within the network to support specific behaviour for a few classes of traffic aggregates.

The introduction of preferential service for a class of traffic not only requires technical support to achieve preferential treatment in the network, but raises issues relating to resourcing and allocation. Resourcing for higher performance quality is typically more expensive, since there are limits to how much the service is supposed to degrade under load. This has implications both in terms of control (to prevent everyone using it all the time), in terms of provisioning (how much such traffic the network should carry and in what proportion), and on more detailed aspects of limits on usage (who can use how much and of what sort). Some traffic will be denied service at higher quality. Accordingly, there will be a need for allocation or rationing. (In a commercial economic model this would be linked to a business plan and different classes of service would be priced differently. Such a model not only provides the elements of control but also the means to adjust provisioning to meet demand by investing revenue in the acquisition of resources.)

The linking of technical implementation with non-technical issues requires a variety of policy-related mechanisms and decisions to enable the service to be offered, managed and developed. It is the purpose of this document to define these areas, and to propose the approaches to be taken. Some parts of the document are intrinsically not static: as experience is gained, some policy issues will be resolved and will need to be recorded; others will arise and require study: the Framework will itself evolve.

2. Network model

JANET, and indeed the Internet as a whole, is a *multi-domain* network. The context for the usage of the term *domain* within this document is akin to that of the *autonomous system* (AS), which arose as the unit of routing policy, sometimes also referred to as a *routing domain*. It might consist of a single network (prefix) or a group of (by implication, interconnected) networks, but the important feature was of a single policy in respect of routing controlled by an administration or consortium of administrations. In the current context, the focus is on policy associated with the operation of network QoS, and by analogy a *QoS domain* is a set of one or more networks with a single policy regarding the operation and management of network QoS.

Within JANET there is at least a three-level hierarchy of domains co-operating to deliver end-to-end service: the *site* or *institution network*, the *regional access network*, and the SuperJANET backbone. Each of these networks is typically an AS. Considered as a set of routing domains, it is largely hierarchical (with a few minor non-hierarchical additions). For some institutions, the site network itself may well operate a further level of hierarchy, for routing and possibly management.

The international dimension is somewhat outside the scope of this document, especially as it remains unclear how or whether QoS based services will be offered or adopted widely. However, within Europe, GÉANT is evolving an IP Premium service, and in terms of QoS domains it is an instance of another routing and QoS domain. (The hierarchical aspect is somewhat less obvious or accurate, since a number of NRENs have independent peering arrangements beyond Europe.)

In developing a QoS policy framework, it will be helpful to distinguish logically between *edge* and *transit* networks. The former has hosts directly connected which originate and consume traffic; the latter transport traffic from one peer network to another. A network may be both. Site networks are typically predominantly edge networks. A regional or backbone network is certainly a transit network, but may have a (typically small) edge component.

The policy issues so far identified in association with QoS are the definition of services provided, and implementation oriented aspects such as authorizing, classifying, marking, provisioning, and policing. The definition of the services will inform the overall shape of the service level agreements (SLAs) needed between constituent JANET networks. In principle, each network would define its own service level specifications (SLSs) and would use these as the basis for SLAs with its neighbouring networks. Within JANET, it is suggested that the initial approach should be based on a single proforma SLS for transit networks and another for edge networks. These proformas will include a list of parameters by which the services are defined but specific values for some of them may be varied according to circumstances and within certain limits for each network, while still conforming to JANET SLS.

Within GÉANT it has been suggested that end-to-end SLAs/SLSs are needed. The issue of end-to-end service levels is currently open in JANET, but is discussed further under policing in Section 7.

3. Provisioning

3.1 Overview

At the network (service) level, four services (in descending order of quality) are under consideration:

- Premium;
- IP+;
- Best Effort (BE); and
- Less than Best Effort (LBE). [This is intended to be the same service as that referred to in the context of QBone/Internet2/Abilene as the “Scavenger Service” (QBSS).]

DiffServ differentiates services by treating packets differently according to service classification. The difference in treatment is relative, and applies to a class, which will generally consist of many ‘micro-flows’. In order to approximate a service defined in absolute terms when elements of the network are fully loaded, provisioning levels need to be defined for each of Premium and IP+. Correspondingly, within a given provisioning level it is necessary to limit the quantity of traffic in order to preserve the service. Use of admission control in turn requires some level of authorization in support of allocation or rationing.

Determining provisioning levels for Premium and IP+, relative to capacity for BE, will be partially a matter of policy in respect of capacity planning and allocation to handle specified amounts of traffic, and partly influenced by technology: the precise details of how the services are implemented in terms of scheduling, re-marking, and discard policies within routers.

A basic LBE service does not, by its nature, require any provisioning, nor does it provide any particular level of service: if other traffic consumes all capacity (at least somewhere on a path), then there is none left for LBE. A variation on this actually attempts to allow a low level of LBE to survive under loaded conditions in order, for example, not to destroy TCP connections and abort long transfers. It is suggested that if this can be configured this would be a more satisfactory form of LBE service to offer than one which can be completely starved.

3.2 Dimensioning interdomain provisioning

For the purpose of provisioning (and policing, see below), the model of JANET is of a set of (hierarchically) interconnected autonomous QoS management domains: the SuperJANET backbone, the regional access networks, and the campus networks. It is assumed that monitoring, provisioning, traffic engineering, dimensioning, etc, of the interior components of any of these constituent networks is a matter for the management of that network. The places where policy has to be agreed on a wider basis are at the interdomain

boundaries. Correspondingly, traffic levels, provisioning, monitoring, and policing needs to be on an agreed, JANET-wide basis on the interdomain links which constitute these boundaries.

In considering the provisioning of the links constituting such boundaries, several capacity limits can be distinguished:

- *absolute link capacity*: the fundamental transmission capacity of the link;
- *JANET DiffServ limit*: the maximum proportion of a given type of traffic allowed on any link;
- *operational allocation*: the operating proportion determined as being available for use by a given type of traffic on a given link; and
- *operational policing limit*: the operating policing parameters in force for a traffic type on a link.

Since initially the programme is concentrating on IP Premium, IP+ is not considered further in any detail in the initial provisioning model.

3.2.1 Absolute link capacity

This is the basic, physical transmission capacity or rate available for IP packet transmission. In considering what allocation of transmission capacity may be required by any particular application, due allowance needs to be made for IP packet overhead (and indeed any additional overhead imposed by higher level protocols supporting applications). The limits or allocations considered in the sequel apply to the capacity available for IP packet transmission.

This capacity is determined by the transmission capacity of the physical link. It is to be divided up amongst Premium, BE and LBE (and potentially IP+ later). No consideration is given here to any other potential additional subdivision of the link capacity, perhaps for link sharing amongst competing peer network transit traffic, for example. Although there are proposed hierarchical scheduling models applicable to such situations, such requirements are considered beyond the scope of this framework.

3.2.2 JANET DiffServ limit for IP Premium

The DiffServ model for EF is a relative one in the sense that EF traffic is given priority over non-EF traffic. (Precise details will vary between routers but this is expected to remain generically valid.) If too much EF traffic is allowed onto a link, then jitter and delay on the individual constituent flows will increase and the behaviour will tend to become more like BE. Indeed, in the limiting case, if all the traffic were allowed to be EF, the behaviour would revert to BE.

The Premium traffic DiffServ limit is the maximum proportion of IP Premium traffic which should be admitted on a link, and it is determined by the need for EF to operate correctly for the traffic (mix) which IP Premium is intended to carry. The importance of this limit is that if more than the absolute amount represented by this proportion is needed on a given link, then *a larger capacity link is needed* in order to maintain the Premium service. The overall effect of not increasing the link capacity in these circumstances will be to increase the jitter experienced by the individual Premium ‘micro-flows’ within the Premium aggregate. Since limits on such jitter will necessarily form part of the Premium SLS, this will lead to non-conformance with the SLS under periods of congestion on the link, and higher-level services such as videoconferencing will degrade unacceptably.

The actual proportion represented by this limit depends in principle on the traffic mix, the parameters it is desired that the IP Premium service should meet, and the individual details of how specific routers implement the differential (preferential) queuing for EF. This needs to be determined in the light of experience, and it is suggested that H.323 traffic be taken initially as the benchmark traffic which it is desired to support using IP Premium. However, it is expected that this limit will be in the region of 20% (the origin of the figure in [1]), and unlikely to exceed 30% if videoconferencing service and VoIP are not to suffer. In an idealised model in which traffic mixes on all (interdomain) links were identical and EF scheduling was the same everywhere, then this limit would be a single figure, the same for all interdomain links. With experience, it may be that it will turn out to be similar in a variety of cases. In any case, in order to cater for the case where traffic passes over a series of interdomain links of different characteristics, it would appear to be necessary to adopt a policy of choosing a JANET-wide figure which is the ‘mini-max’ (least upper bound) of those found to be satisfactory.

In view of the implications of the above for link capacity (and hence ultimately funding) to support Premium service, it will be important to gain quantified experience of the effect of under-provisioning on jitter and application performance before a Premium service is offered.

3.2.3 Operational allocation for IP Premium

At any particular time, the amount of IP Premium which is provisioned, i.e., the amount which is admitted across a link, may be less than the JANET DiffServ Premium limit, since demand within a domain may not require the maximum allocation, or it may be decided to limit it for other reasons. This limit is here referred to as the operational allocation. It is determined by the management of the network domain in question by taking into account application traffic requirements. An initial model for estimating such allocations for supporting JANET audio/video traffic is proposed below (Section 3.2.5).

3.2.4 Operational policing limit for IP Premium

The purpose of policing traffic as it is received off a transmission link is to ensure that the agreed provision of capacity for traffic using a particular service is not exceeded. Even in an environment where networks trust each other, such policing is generally desirable if only as a first level of defence against DOS traffic starving lower level service traffic.

In setting policing limits, the immediate issue is how to parameterise the traffic rate in such a way as to be both capable of implementation and adequate in terms of its relationship to a provisioning capacity allocation limit such as the operational allocation above. The major issues relate to burstiness of the traffic and the period over which traffic rate is measured. The classic model for accommodating this in the context of policing or rate control is based on the notion of the token or leaky bucket (see Section 13.3.4 of [7] and Section 5.4.2 of [8]). An advantage of this model is that it may be used in abstract form to characterise behaviour. Implementations may be based on the algorithm directly or may provide an approximation. In the current context it is proposed for use to characterise policing behaviour in terms of a small number of parameters. Formulation of the token bucket parameters needs further study.

3.2.5 Dimensioning model

This model is intended primarily to aid in the calculation of the operational allocation for IP Premium on the assumption that the traffic for which Premium will be used is H.323 traffic. For some interdomain links it may be that only very specific usage is to be catered for (a few well-known videoconferencing streams, for example) and in consequence the necessary allocation is known quite well, at least in terms of what applications and how many simultaneous instances.

Another possibility is that the service is catering for some level of 'on-demand' service, in which case the simplest estimate is that needed to cover peak load: the sum of the maximum number of videoconferencing channels for which provision is to be made. Here, an initial estimate might be on the basis of 768kbps or 1Mbps or 2Mbps being needed (on average) for a single channel.

Of course, such an estimate takes no account of the finite duration of each conference call, so it will be an overestimate which will be increasingly gross as the number of calls increases. This is a well-known area in telephone engineering, in which for a given holding time and call volume on a link, it is possible to size the capacity of the link in order to meet a defined call blocking probability.

As an example, if on average 2 videoconference calls are set up per hour over a link, each of an average 2 hours duration, the link loading would be 4 Erlangs. If in addition the maximum capacity of the link were 10 such calls, then the blocking probability would be approximately 11%. This is an example of the application of the Erlang-B distribution. There are a number of assumptions made here (the most significant being that all the calls require the same capacity). Also, strictly speaking, the Erlang-B formula applies in the case of an infinite population of sources. Nevertheless, this can provide an initial estimate of the call blocking probability; or, alternatively, given a desired call blocking probability, an estimate can be obtained of the necessary link capacity, i.e., the operational allocation needed. More details about this (including references and Erlang calculator web sites) are given in Appendix A.

3.3 Initial experimental provisioning

Experimental trials of DiffServ technology are needed to establish a basis for the level at which the JANET Premium DiffServ limit should be set, and to gain experience with router configuration.

As a starting point, the following relative provisioning levels are suggested as starting points, to be varied as usage evolves and estimates of traffic are refined.

- i) *Premium* Four quantities have been identified above in Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4. The link capacity is absolute and may be used to determine absolute values for the other limits which are otherwise relative. As detailed above, the JANET Premium DiffServ limit needs to be determined by experiment. (Note The Think Tank report [1] suggested a figure of 20% for Premium service. This suggestion was made before the distinction was made between the JANET Premium DiffServ limit and the operational premium allocation limit, as described above in Sections 3.2.2 and 3.2.3, respectively. In effect, the 20% was based on a conservative estimate of the JANET Premium DiffServ limit.) For the purpose of gaining experience with configuring routers for DiffServ, an initial operational allocation for IP premium needs to be made: a figure of 5% of link capacity is suggested but it is only a suggestion and may be varied if appropriate during the QoS testing programme. This initial figure is based on two considerations: for the purpose of early testing, use of a sufficiently small fraction of the overall capacity of a link for EF should ensure that the mechanism is effective in providing low delay and jitter; moreover, since the test programme is being carried out on the production service network, ahead of any policy debate on what the level of provision should be for any such service on JANET, it seems prudent to provision only a small fraction of link capacity.

The operational policing limit, or rather the parameters to be set to ensure that no more traffic is admitted than specified in the operational allocation, is held over for discussion in Section 7 (Policing). The action to be taken when traffic violates the policing limit is discussed in Section 4 (Service Level Specifications).

- ii) *IP+* Development of an IP+ service will be delayed until experience has been gained with piloting Premium (and LBE). The service is intended as a priority service for elastic data applications needing better delay performance than that offered by BE but for which the tight constraints on delay and jitter offered by Premium are not necessary. Streaming delivery of non-interactive audio-video material and interactive data applications (remote logon or similar) are envisaged applications. Currently, traffic of this type is either relatively small as a proportion of total traffic (remote logon) or new (continuous media delivery). It is anticipated that there will be several limits along similar lines as for Premium. Apart from the general observation that any initial, experimental operational allocation would be small, on the same grounds as for Premium, it seems premature to propose any limits at the moment.
- iii) *BE* This service is the 'default' service which most users will normally use. Provisioning here has a different meaning, since it is not a maximum for which policing is implemented. Instead, under full load in each service category, BE would receive a minimum relative capacity determined by subtracting from the link capacity what has been reserved to the other services: the operational allocation for Premium (and for IP+, assuming this follows a similar model), and the floor allocation for LBE (see below). This derivative figure may be a useful one to keep in mind when considering upgrades in physical network component capacity.
- iv) *LBE* Although in general no specification of level of service is made for this service, it has already been pointed out that in the interest of encouraging use and avoiding total starvation of the service in the presence of temporary congestion, it may be prudent to allocate a 'floor' level of provisioning in support of the service. How difficult this may be to implement and monitor (since it only ever comes into action when the network is congested) should be the subject of experiment. As a starting point a floor of 1% is suggested, but this may need to be revised (down?) in the light of experience.

In any future service, it will be important to monitor, review and, if necessary, revise operational provisioning levels on a regular basis, and to use these decisions and the supporting figures to drive procurement of additional capacity. In particular, it is important that new services such as Premium (and IP+) are not seen as 'stealing' capacity from 'regular' (BE) use of the network. This will be just as important as ensuring that those using video services for a booked teaching session are not disrupted by, for example, a Grid ftp session.

4. Service Level Specifications

Ideally, the quality of each network service needs to be defined in terms of performance parameters which are meaningful in network terms, and which also satisfy the requirements of applications in terms of enabling specified performance levels for application services to be achieved. Following [1], a pragmatic approach is suggested here, whereby a few basic Service Level Specifications (SLSs) are developed which are agreed as being adequate for identified benchmark applications which will use each of the services.

All the service performance parameter values need to be given in statistical form, presumably eventually by quoting a percentile confidence level. Those already in use for the current single service are proposed for use in the BE service. (They may also form a partial starting point for IP+ but that is deferred for future study.) Those for Premium need to be agreed as suitable for supporting the benchmark services.

- i) *Premium* Interactive audio and video (audio and video conferencing; VoIP) will form the benchmark applications for Premium. Note that current experience with H.323 suggests that video is the more sensitive with respect to packet loss and jitter. Interactive audio on its own (particularly point-to-point with no MCU involved) presents potentially more stringent requirements in respect of delay (since the codec delays are small compared with those for video). Parameters of service need to include delay, loss, and jitter; and values of the parameters for the operational policing limit.

Exceeding the operational policing limit results in offending traffic being dropped or remarked. The applications using Premium are those sensitive to delay and jitter, attributes which will be amplified by any remarking. Intrinsically, continuous media applications can be resilient against occasional loss but are intolerant of delay. Re-marking can also introduce reordering, of which continuous media applications are also intolerant. Furthermore, for near-CBR applications with tight QoS requirements, and having sustained and peak rate values close to each other, dropping is likely to be more appropriate. In general, IP Premium needs to operate in a regime in which the Premium queue in any router is empty. It is anticipated that in view of the foregoing, dropping policed traffic will be appropriate but this needs confirmation by experience.

- ii) *IP+* Consideration of SLS issues for IP+ is deferred.
- iii) *BE* The existing specification in terms of loss and delay for the current JANET service will form the basis of the BE SLS. Jitter would remain unspecified.
- iv) *LBE* There is no SLS for LBE.

Technically, within each domain in JANET, it is the responsibility of that domain to determine what techniques to deploy within its routers in order to achieve its SLS. In particular, what DSCP value to use for the DiffServ per-hop-behaviour (PHB) associated with Premium (and IP+). (The DSCPs for BE and LBE are universal, see [4].) However, in practice, it may be beneficial to adopt a common set of prescriptions for achieving PHBs, with the caveat that there may be variations depending on the level of SLS and manufacturer's equipment. Details of this nature will need to be determined by experiment and documented in [4] as the services are developed and deployed.

5. Interdomain working

Much of the foregoing is implicitly couched in the context of a 'single' network. In practice, this means a network provisioned and operated by a single management. In principle, interconnected networks need not all operate to the same set of SLSs. In a commercial model, networks establish peering agreements for the services each will provide the other on the basis of their respective SLSs. What this does not necessarily establish is what end-to-end SLS may be delivered to end-users. Since loss, delay, and jitter are generally cumulative, a set of (global?) end-to-end SLSs may need to be established, together with corresponding single domain SLSs which on the basis of the number of domain hops within the UK would be adequate for the benchmark services.

Within JANET there is at least a three-level hierarchy of domains co-operating to deliver end-to-end service: the site or institution network, the regional (access) network, and the SJ backbone. (Actually, for many institutions, the site network itself may operate a two-tier management hierarchy. Addressing this directly is somewhat outside the scope of this document, though some of the ideas may be applicable to such layer 3 hierarchies.) Technical differences between the way in which networks at each level operate suggest that it may not be appropriate to seek to define a single SLS for a given quality of service to cover all levels in the hierarchy, but there seems no reason not to use a single proforma SLS at each level of the JANET hierarchy. It is proposed that two JANET SLSs be developed, covering the backbone and the regional networks. These would effectively cover all single domain transit networks within the wide-area segments of JANET. This would substantially simplify things. In the general case, where every domain defines its own SLSs, there is the potential for constructing pathologically incompatible SLSs which would render it essentially impossible, for example, to offer a workable video service.

These SLSs would be proformas in the sense that the service, all associated parameters, and actions would be defined, but specific values for some parameters would be defined at each instance of an interdomain boundary, as part of the relevant SLA.

Institutional networks have not traditionally operated SLAs, so an institutional SLS may not be appropriate. Whether the use of an informal 'proforma' SLS may be helpful in this context needs to be determined through experience. One aspect of where this may turn out to be relevant is in policing in-bound (into a site) levels of Premium (or IP+) traffic.

This area is further complicated by the need for international working. At this time, it is not generally clear how international support for QoS will evolve. Within Europe, GÉANT offers an IP Premium service, and the GÉANT SLS will effectively form another, international-level SLS in the hierarchy.

In this context, it is noted [3] and above that provision of IP Premium (for example) is an end-to-end service, and that this implies the need for end-to-end SLAs. While it is clear that the provision of end-to-end service does have, for example, end-to-end provisioning implications, this is potentially unscalable. The need for this within JANET needs investigation.

6. Authentication, Authorization, and Trust

While anyone may use both BE and LBE, uncontrolled use of Premium and IP+ will destroy them and lock out BE (and LBE) service. Thus use of these in principle requires authorization, which in turn generally requires authentication.

If no domain trusts any other, then all domains must do this for themselves. This is not generally practical or desirable for transit networks. The model proposed for JANET is that each domain is responsible for authentication and authorization of its own subscribers, but that where it is acting as a transit network it trusts its neighbouring networks in this respect. One consequence of this is that all classification and marking is done in edge domains (but note that as stated in Section 2, a domain may be both an edge domain and a transit domain). Another is that a transit domain does not need to re-classify, though it may need to re-mark, (because a different value is used to represent a given PHB, or a slightly different PHB is used, or as a result of policing packet admission, see Section 4). Within JANET, given that a unified set of SLSs is brought into use, then the only reason for re-marking would be as a consequence of policing.

An initial model is that marking in edge domains is statically applied to particular service traffic identified as having fixed known origin: one obvious early example is the H.323 gatekeeper.

Again, this will need reviewing in the context of international operation, with GÉANT in particular.

Notice that although the points in the network at which traffic is authorized has been defined in a domain-oriented fashion in the model proposed, the means of authentication and authorization need not be domain-oriented. While some communities may naturally form according to network domain boundaries, there may well be others (disciplined-oriented or project-oriented virtual organizations, for example). This appears to be more a matter for the structure of authentication and authorization services and the ways in which they are consulted by the network entities requiring their services.

7. Policing

Introduction In order not to oversubscribe provision within any domain on any given route, it is necessary in principle to police traffic on admission. Oversubscribing can come about either through *bona fide* demand or malicious action. Regardless of cause, protection is necessary in order to maintain not only the service at which oversubscription occurs but also all services at a lower priority or grade which will otherwise be starved of service. While in future dynamic provisioning (within limits) for particular services may be possible with the aid of inter-domain signalling and domain-level bandwidth brokers, the initial service will be much more static.

All domains will in principle need to police incoming traffic from neighbouring domains against agreed provisioning allocations for Premium (and IP+). This implies that SLAs will be needed between neighbouring networks, and since the traffic volumes and hence allocations are unlikely to be identical across all interdomain boundaries, these SLAs will need to be dimensioned individually. A goal of the QoS development programme is to explore the possibility that a small number of SLS templates or proformas may be used as the basis of all such interdomain SLAs within JANET, perhaps with only the dimensions of various parameters varied (within limits) to suit specific needs.

Shaping In general, wherever packet admission policing is in effect at the ingress side of a boundary, the question arises as to whether shaping should be used by the network on the complementary egress side of the boundary. Typically, temporary out-of-spec peaks of traffic may occur at multiplex points as a result of transient network congestion. This is not in general under the control of the end-user, who should not correspondingly be penalised. The current view for delay sensitive traffic is that shaping is undesirable since it implies use of additional buffering which in turn gives rise to increased delay. A question to be studied is whether acceptable service can be achieved by a sufficient degree of provisioning alone, avoiding the need for shaping.

Traffic violation & action In determining when to drop (or re-mark) as a result of policing the aggregate flow, the question arises of how to determine when the traffic level is being violated. This is a question of how

much burstiness is tolerated, for which the best-known algorithm is the leaky bucket, which is proposed in section 3.2.4. Since the premium service is intended for applications which are delay and jitter sensitive, for which queues of any appreciable size are inappropriate, the depth of the bucket should be small. The precise relation of the policing parameters to the Premium operational allocation will be a matter of experiment and measurement. It seems likely that any policing rate parameters may be expected to be less than the rate determined by the operational allocation, if only to take account of burstiness.

A problem At present, the model for policing incoming traffic at a network domain border only recognizes the gross aggregate. It does not distinguish traffic in any way: all classification (and, by implication, authorization) has been done at the edge where the traffic was first admitted from a subscriber, and policing at transit borders is on aggregates defined only by DSCP value. By analogy with basic switching at any level, the situation which this does not protect against is 'output port contention', that is, where traffic from a variety of sources contends for access to the same egress point. Assuming the use is genuine, and the traffic pattern is genuine, then there is a case for altering the provision on this interface. At issue is the question of what level of provision will be necessary in order that contention on a particular boundary (and consequent potential disruption of all traffic in the class on that interface) is rare for known, anticipated traffic.

Study of this area is needed in the light of experience. Simple policing of inbound aggregates against, in effect, overall provision-related rate limits has the advantage of simplicity. It depends, however, on knowing sufficiently well the traffic patterns. It may not be adequate for protecting authorized traffic from unexpected peak loads. An operational issue somewhat related, is whether protection can be devised against Premium denial-of-service attack, particularly a DDOS attack which makes use of the output port contention property at network level to attack a particular boundary between networks.

Within GÉANT there has been discussion of the need for end-to-end SLAs in order to be able satisfactorily to provide a Premium service. It is possible that this might in principle offer an approach to avoiding Premium congestion at an outbound interdomain boundary. However, it also appears to threaten unscalable increase in complexity of policing and number of SLAs, both of which contributed to RSVP & IntServ being abandoned for core use and gave rise to the DiffServ approach in the first place.

Signalling and dynamic provisioning and policing The approach outlined above to what is termed here the *egress port contention problem*, is really little more than a 'laissez faire' or 'hope for the best' attitude, and is not likely to survive serious deployment. Several approaches have been suggested, though none is deployed or tested in service at this time [11]. The following are put forward for discussion.

VoIP & telephony This traffic is composed of individual flows which are very small compared to link capacities (at least 3 orders of magnitude?). Substantial numbers can be accommodated by over-provisioning. For this traffic, the approach outlined above may be sufficient.

Substantial or critical applications Included in these categories are interactive video and any Premium flow of a sufficiently critical nature that it needs to be handled with better regard to protecting provisioning commitment. In order to do so, it is likely that policing will have to have regard to more than solely DSCP value. At the time when a flow is authorized, provisioning needs to be made aware of the commitment to this flow, both to ensure that it is adequate and to adjust policing levels. This implies that edge routers need to be aware of such flows, both to identify them and to police them (as part of an aggregate). This is familiar ground. There are two approaches discussed: dynamic bandwidth brokers (which have to include the facility in this context to adjust policing, or at least have the ability to alter the criteria identifying an aggregate), and signalling, typically by use of RSVP. Note that in the current context it is only domain edge routers which would need to participate in this process, and so would, for example, need to be RSVP-aware or interact with brokers.

Of the two, RSVP may be potentially more viable, particularly since it is an existing protocol already available in many end-systems, and so potentially capable of handling the signalling needed between host and first-hop router for any dynamic, on-demand service.

8. Service monitoring

A general feature of monitoring is that both sides of any service boundary need to be assured that the service is operating as it should. In multidomain networking, there are (at least) two orthogonal sorts of boundary. One obvious one is that between domains. Another is between levels in the service. An example of this is afforded by a videoconferencing service: it will expect to carry out its own monitoring but will need to be able to resolve problems with the network service, which will have its own set of performance measures. An important part of

the programme is to develop not only the appropriate measures, but to ensure that both sides of each boundary are monitored. In particular, this will require co-operation of all domains, including sites.

In association with volume traffic measurements in each class there need to be measures of packet drop rates. From the SLS point of view these figures are needed for all but the LBE service. From the pure performance point of view, figures for LBE may also be of interest. Delay and jitter figures are needed for Premium (and IP+).

These requirements for monitoring need to be incorporated into a network monitoring programme to support the introduction of these new services. It may also be that in some cases, measurement and monitoring techniques may have to be developed and agreed. For example, while traffic volume is intrinsically a passive observation of passing traffic, measurements of the other parameter values may require intrusive monitoring by the introduction of a (small relative) quantity of specialised probe traffic.

An important aspect of the monitoring programme will be the ability to reconcile such network-level performance measures with application service performance figures, in order to gain a full understanding of the behaviour of the network, as well as information to diagnose end-to-end performance problems.

Appendix A. Use of Erlang-B distribution to estimate Premium operating allocation

The standard texts which cover queuing theory and telecommunications cover Erlang's work. Apart from sizing trunks, the theory is equally applicable to sizing call centres, which is probably the commoner application at the present time. The most accessible account in the current context is probably the book by Keshav: see [7] Chapter 14, Traffic Management; Section 14.9, Admission control, Effective bandwidth; Section 14.11, Capacity planning (including Erlang-B distribution). [N.B. p.487, formula (14.8) is wrong: the N in the denominator should be N! (factorial).] An older but enduring authoritative text in telephone and packet network performance is Schwartz' book: see [6] Chapter 2, Section 2.4, p.55, Equation (2.55), Erlang-B distribution; Chapter 10, Section 10.3, p.522, application of Erlang-B distribution to call blocking in circuit network. The classic text on queuing is Kleinrock's, but this is for the more mathematically inclined: see [5] Chapter 3, Birth-Death Queueing Systems, Section 3.7, p.106, Equation (3.46), Erlang-B distribution.

Traffic loading is measured in terms of the busy hour (or minute, etc.), which is the number of hours (minutes) of call traffic present in an hour (minute). Some calls are already in existence at the start of the hour and continue throughout the hour; some exist at the beginning and end during the hour; others start and end during the hour; and yet others start during the hour and continue beyond the end of the hour. The unit of (average) call load is the Erlang: 1 Erlang is equivalent of an average of 1 call being present.

If the average call arrival rate is λ calls per unit time, and μ is the holding time, then the average call load, A , is given by $A = \lambda \times \mu$ calls. If N is the maximum number of calls on a trunk link, then the blocking probability, B , is given by

$$B = \frac{A^N}{N! \sum_{i=0}^N \frac{A^i}{i!}},$$

which is Erlang's B distribution.

Calculators [9] and tables [10] are available for this, for calculations either in terms of capacity required to achieve a given blocking probability for a given traffic loading or the blocking probability resulting from a known traffic (call) loading on a link of given capacity.

References

- [1] *Report of Quality of Service Think Tank*, UKERNA, July 2001.
- [2] *Specification and implementation plan for a Premium IP service*, GÉANT, Deliverable D9.1, April 2001.
- [3] *Implementation architecture specification for a Premium IP service*, GÉANT, Deliverable D9.1 - Addendum 1, April 2001.
- [4] *QoS Technical Implementation Plan*, UKERNA, March 2002.
- [5] Kleinrock, L. (1975). *Queueing Systems. Volume 1: Theory*, John Wiley & Sons; ISBN 0-471-49110-1.
- [6] Schwartz, M. (1988). *Telecommunication Networks: Protocols, Modeling and Analysis*, Addison-Wesley; ISBN 0-201-16423-X.
- [7] Keshav, S. (1997). *An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network*, Addison-Wesley, ISBN 0-201-63442-2.
- [8] Tanenbaum, A.S. (2003). *Computer Networks* (4th edition), Prentice Hall/Pearson Education, ISBN 0-13-038488-7.
- [9] *Traffic calculators*, Westbay Engineers Ltd: <http://www.erlang.com/calculator/>
- [10] Watson, R. (2003). *Free Erlang-B charts*: <http://www.quantumportal.com/erlangb.htm>
- [11] Private communications, particularly with Mark Handley, UCL.