18

Connection Admission Management in ATM Networks Supporting Dynamic Multi-Point Session Constructs

P. Moghe and I. Rubin
Department of Electrical Engineering, UCLA, CA 90024-1594.
{pmoghe,rubin}@ee.ucla.edu

Abstract

A framework for admission management of session-level requests exhibiting space/time heterogeneity is developed. A single sub-threshold based link-level connection admission scheme for a mix of uni-point/static session and multi-point/dynamic session Virtual Channel Link requests (VCLRs) is designed and evaluated under different scenarios. Aside from external blocking, **internal loss** is introduced as an important QOS parameter for multi-point/dynamic session services. Concepts of **service-optimal** and **throughput-optimal sub-thresholds** are formulated. Finally, we outline a network algorithm that designs link-level sub-thresholds in accordance with end-to-end session-level QOS parameters.

Keywords

Multi-point and Multi-party Resource Allocation, Performance Management, QOS Management, Connection Admission Management in ATM Networks.

1 INTRODUCTION

Unlike traditional connection establishment protocols that treat a **call** as a monolithic end-to-end object (used for one service type, using one channel or connection), BISDN signaling needs to be tailored to incorporate an efficient mechanism to service multi-point and multimedia traffic [1][2][3]. In this context, we redefine a **call** as a high-level distributed network object that describes the communication paths connecting the clients. A **View** or a **Session** is the call-context of each client. In the most general case it represents a broadcast tree rooted at a client; its leaves comprising the recipient clients (also called *sink-clients*). Each session is implemented at setup time through end-end Virtual Channel Connection requests (VCCRs). A **VCC**, identified by a unique source VCI, is an end-end directional logical tree between source and sink clients. Each fork represents multicasting of information cells. A VCC itself is established through a sequence of Virtual Channel Link requests (VCLRs). A **VCL** is the basic logical component of our relationship model and represents a logical connection (and a single channel bandwidth allocation) between adjacent switching nodes.

Applications such as multi-media conferencing and information browsing/sharing can be built using the above constructs. As the ATM layer matures, it is our contention that

A. S. Sethi et al. (eds.), *Integrated Network Management IV*

the admission management of these constructs, at the connection layer (above the ATM layer), will pose future challenges. In this work, we formulate appropriate connection-level QOS vectors and design a simple threshold-based admission scheme to handle heterogeneous session constructs.

The paper is organized as follows: In section 2, the problem is motivated and an objective is formulated. In section 3, the single-link (SL) admission model is described, evaluated and tuned for the chosen optimality measures. Section 4 discusses some numerical results of the SL Model. In section 5, we outline a two-tiered network algorithm that uses the SL model to design distributed network-wide sub-thresholds.

2 PROBLEM DEFINITION

We recognize two important resource allocation tradeoff issues related to the bandwidth demand of session requests:

- Spatial Heterogeneity: Multi-point vs. Uni-point Session Requests
  Multi-point requests are susceptible to higher levels of blocking than uni-point requests in networks with limited multi-cast edge switches. The spatial issue thus requires that the multi-point requests be given special care, so that they are not blocked beyond tolerance.

- Temporal Heterogeneity: Static vs. Dynamic Sessions
  In static sessions, the number of member clients is constant and declared by the session request. Dynamic sessions are characterized by a variable number of clients during their life-time. Reservation of optimal number of VCLs for dynamic sessions is a challenging issue. If enough capacity is not reserved for a carried dynamic session, a secondary request for addition of a new user is liable to be blocked. This can adversely impact the carried users of the session. The resulting service degradation can, in certain applications, be severe enough to cause a subset of carried users to abort the session.

In general, session requests are of two types: *primary* (requests that initiate the session) and *secondary* (requests that add on to existing sessions, preferably reusing their resources). We combine the two heterogeneity issues into a single problem by defining two classes of session requests, $\mathcal{A}$ and $\mathcal{B}$. Class $\mathcal{A}$ requests initiate uni-point/static size sessions. Class $\mathcal{B}$ requests set up a multi-point session through a primary request. If admitted, this is followed by uni-point secondary class $\mathcal{B}$ requests for additional client connections. If secondary requests are blocked, a fraction r of the sink-clients are assumed to abort(*internal loss*). Class $\mathcal{A}$ and $\mathcal{B}$ session-requests generate lower-layer class $\mathcal{A}$ and $\mathcal{B}$ VCLRs at the link level. We assume that the required service quality is specified through session-level QOS vectors for both classes. For instance, class $\mathcal{A}$ and $\mathcal{B}$ applications declare worst-case session-level and link-level(VCLR) external blocking probabilities as Θ_{ex}^{max} and Φ_{ex}^{max} respectively. In addition, the worst-case internal loss probability $\Theta_{in-loss}^{\mathcal{B},max}$ (and corresponding link-level $\Phi_{in-loss}^{\mathcal{B},max}$)

defines the maximal acceptable probability with which a carried class $\mathcal{B}$ client aborts due to secondary blocking.

The *problem objective* is : Given an arbitrary session-request loading pattern, a network routing topology and multi-cast switch locations/specifications, design a threshold-based VCL-layer admission scheme on each link that can be tuned to satisfy the session-level QOS vectors (and possibly achieve connection-level optimality measures).

Since the network-wide problem is daunting to tackle on an end-to-end session basis, our approach is to build and solve exactly a flexible single-link (SL) model. This model makes natural sense since the admission scheme is on a link basis anyway. A network algorithm then approximates the end-to-end effect through its dependence structure.

3 SINGLE LINK MODEL

The link-level admission scheme is outlined next. It uses a sub-threshold (m_A) to reserve space for class $\mathcal{B}$ VCLRs. The SL analytical model is described in section 3.2. Parameters such as r (session dependence), D (initial session size), λ_s (secondary arrival rate per session) are formulated. Under the assumed traffic and service statistics, the VCL layer is analyzed for steady-state performance in section 3.3. Performance measures such as external blocking, internal loss, and aggregate throughput are computed in section 3.4. Feasibility and optimality sub-thresholds are defined in section 3.5.

3.1 VCL Connection Admission Scheme

Let m be the maximum number of VCLs on a link, capable of supporting cell-layer QOS. We assume m to be a known quantity; various studies such as [4][5] focus on admission at the ATM layer and indirectly compute it. Define a *sub-threshold* m_A ($0 \leq m_A \leq m$). Let D_{mc} be the maximum multi-cast gain of a switch (i.e. the maximum number of copies supported by the switch copy-network), and D be the instantaneous multi-cast demand of a Primary class $\mathcal{B}$ VCLR. Let N_t^{vcl} represent the aggregate carried VCLs on the link at time t. We employ the following admission policy for a VCLR arriving at time t :

3.2 Analytical Model Description

We treat each directional link as a multi-VCL resource. Under the homogeneity assumptions (i.e. each VCL represents equal bandwidth), the VCL layer can be modeled as a pure blocking

Class of VCLR	Characteristics	VCLR Admission Rule
$\mathcal{A}$ (Primary)	Initiates uni-point/static session	$N_t^{vcl} < m_A$
$\mathcal{B}$ (Primary)	Initiates multi-point/dynamic session	$N_t^{vcl} \leq m - D$, $1 \leq D \leq D_{mc}$
$\mathcal{B}$ (Secondary)	Uni-point VCLR, adds onto created session	$N_t^{vcl} < m$

Table 1: Connection Admission Policy.

system with 'm' maximum VCLs.

The concept of an end-to-end session extended to a link is defined as a Link-session (L-session). All VCLs of an L-session (VCL-members) share a unique L-session-id. Each VCL member normally holds for an exponentially distributed time (parameter μ). An L-session terminates when all VCL-members have terminated. The holding time of an L-session represents the interval from its initiation to its termination.

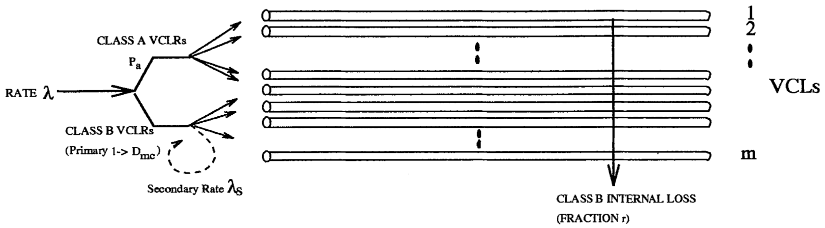


Figure 1: Single Link Model.

Primary VCLRs are assumed to arrive at a node-link User Request Manager, with a Poisson rate λ. A fraction p_a of the VCLRs are class $\mathcal{A}$ VCLRs, the rest class $\mathcal{B}$. Let $\lambda_a = \lambda p_a$, and $\lambda_b = \lambda(1 - p_a)$. Class $\mathcal{A}$ VCLRs represent requests for uni-point, static L-sessions. If admitted, they are allocated a single VCL. A primary class $\mathcal{B}$ VCLR initiates a multi-point, dynamic L-session by first demanding a multi-cast group of D VCLs. D is assumed to be a random number with a distribution $b_i = P\{D = i\}, 1 \leq i \leq D_{mc}$ (Section 4 assumes a uniformly distributed D, so that $b_i = 1.0/D_{mc}$). Each admitted L-session initiated by Primary class $\mathcal{B}$ VCLRs receives additional *secondary* class $\mathcal{B}$ VCLRs at a Poisson rate λ_s. If admitted, the secondary class $\mathcal{B}$ VCLR is allocated a single VCL and the VCL-member set of the corresponding L-session is incremented. Else, a fraction r of its carried VCL-members abort the L-session. Figure 1 illustrates the single-link model. The admission rule has been summarized in Table 1. Our immediate objective is to compute the steady-state VCL-size distribution.

3.3 Analysis

Define the system size process $X = \{X_t, t \geq 0\}$, where $X_t = (X_t^{\mathcal{A}}, X_t^{\mathcal{B}}, X_t^{\mathcal{B}ls}) \triangleq$ Number of $\mathcal{A}$ VCLs, $\mathcal{B}$ VCLs, and $\mathcal{B}$ L-sessions carried at time t. Let $T_n = n$th transition time of X. Define the underlying state sequence $V = \{V_n, n \geq 0\}$, where $V_n \triangleq (V_n^{\mathcal{A}}, V_n^{\mathcal{B}}, V_n^{\mathcal{B}ls}) =$ Number of VCLs carried at time T_n^+. Thus, $X_t = V_n$, for $T_n \leq t < T_{n+1}$, $sup(T_n) = +\infty$.

<u>*THEOREM*</u>: X is a time-homogeneous continuous-time Markov chain over state space $S = \{(i, j, k), \; i = 0, 1, ..m_a; \; 0 \leq (i + j) \leq m; k \in K_j\}$, where $K_j = \{k \,|\, min(1, j) \leq k \leq j\}$, under conditions of session-homogeneity and assumptions of Section 3.2.

We *omit* the proof for brevity. The probability law of X is determined by its transition probability function: $P_t((ijk),(xyz)) \triangleq P\{X_{t+s} = (x,y,z) \mid X_s = (i,j,k),\ s \le t\} = P\{X_{t+s} = (x,y,z) \mid X_s = (i,j,k)\}$.
Let $S_{loss} = S \cap \{(i,j,k) \mid (i,j,k) \in S, (i+j) = m)\}$ be the state-space subset that represents a full system. The infinitesmal generator rates are derived next:

$\forall (i,j,k) \in S \backslash S_{loss}$,

$$
\begin{aligned}
q_{(ijk),(xyz)} &= \lambda_a, && x = i+1, y = j, z = k, && if\ i+j < m_a\\
&= \lambda_b b_D, && x = i, y = j+D, z = k+1, && if\ D \le (m-i-j)\\
&= k\lambda_s, && x = i, y = j+1, z = k, &&\\
&= i\mu, && x = i-1, y = j, z = k, && if\ i \ge 1,\\
&= \Psi_1(i,j,k), && x = i, y = j-1, z = k, && if\ j \ge 1,\\
&= \Psi_2(i,j,k), && x = i, y = j-1, z = k-1, && if\ j \ge 1,\\
&= 0 && else, &&
\end{aligned}
$$

where: $\Psi_1(i,j,k) = j\mu P_{nl}, \Psi_2(i,j,k) = j\mu(1.0 - P_{nl})$ and

$$
\begin{aligned}
P_{nl} &= 1.0 - (\tfrac{k-1}{k})^{j-1}, && for\ j,k > 1\\
&= 1.0, && for\ j > 1, k = 1,\\
&= 0.0, && for\ k,j = 1
\end{aligned}
$$

$\forall\, (i,j,k) \in S_{loss}$,

$$
\begin{aligned}
q_{(ijk),(xyz)} &= i\mu, && x = i-1, y = j, z = k, && if\ i \ge 1\\
&= \Psi_1(i,j,k), && x = i, y = j-1, z = k, && if\ j \ge 1\\
&= \Psi_2(i,j,k), && x = i, y = j-1, z = k-1, && if\ j \ge 1\\
&= \Psi_3(\alpha_j), && x = i, y = j-\alpha_j, z = k, && if\ \alpha_j \le \lfloor j\, r \rfloor\\
&= \Psi_4(\alpha_j), && x = i, y = j-\alpha_j, z = k-1, && if\ k \ge 1 \text{ and } \alpha_j \le \lfloor j\, r \rfloor,
\end{aligned}
$$

where: $\Psi_3(\alpha_j) = k\lambda_s[\{\sum_{l=\lceil \frac{\alpha_j}{r} \rceil}^{min(\lceil \frac{\alpha_j+1}{r} \rceil -1, j)} B(1/k,j,l) I(r < 1.0, k > 1)\} + I(k=1)]$,

$\Psi_4(\alpha_j) = k\lambda_s I(r = 1.0)[\{\sum_{l=\lceil \frac{\alpha_j}{r} \rceil}^{min(\lceil \frac{\alpha_j+1}{r} \rceil -1, j)} B(1/k,j,l) I(k > 1)\} + I(k=1)]$,

$B(p,j,l)$ is the binomial probability of j successes in l trials with success probability p, $I(exp) = 1$ if exp evaluates true, 0 else and $\alpha_j \in \mathcal{Z}^+$.

Assume that under appropriate conditions, steady-state distribution P (of X) and stationary distribution π (of underlying discrete-time Markov chain V) can be computed using balance equations [6].

3.4 Performance Measures

Primary class $\mathcal{A}$ and $\mathcal{B}$ VCLR Blocking Probabilities: Φ_{ex}^{A}, Φ_{ex}^{Bp}, and Φ_{ex}^{Bpg}
These probabilities can be determined by the PASTA property[7].

1. $\Phi_{ex}^{A} \triangleq$ P{Class $\mathcal{A}$ VCLR is blocked} $= \sum_{i=0}^{m_A} \sum_{j=m_A-i}^{m-i} \sum_{k \in K_j} P_{ijk}$

2. $\Phi_{ex}^{Bpg} \triangleq$ P{Primary class $\mathcal{B}$ VCLR (multi-cast group) is blocked}
$= \sum_{l=1}^{D_{mc}} b_l \sum_{i=0}^{m_A} \sum_{j=max(m-l-i+1,0)}^{m-i} \sum_{k \in K_j} P_{ijk}$

3. $\Phi_{ex}^{Bp} \triangleq$ P{Primary class $\mathcal{B}$ (individual) VCL is blocked}
$= \sum_{l=1}^{D_{mc}} \frac{lb_l}{\sum_{k=1}^{D_{mc}} kb_k} \sum_{i=0}^{m_A} \sum_{j=max(m-l-i+1,0)}^{m-i} \sum_{k \in K_j} P_{ijk}$

Secondary class $\mathcal{B}$ VCLR Blocking Probability: Φ_{ex}^{Bs}

$\Phi_{ex}^{Bs} \triangleq$ P{Secondary class $\mathcal{B}$ VCLR is blocked} $= \frac{\sum_{i=0}^{m_A} \sum_{k \in K_j} P_{ijk} k \lambda_s |_{j=m-i}}{\sum_{i=0}^{m_A} \sum_{j=0}^{j=m-i} \sum_{k \in K_j} P_{ijk} k \lambda_s}$

Class $\mathcal{B}$ Internal Loss Probability: $\Phi_{in-loss}^{B}$

$\Phi_{in-loss}^{B} \triangleq$ P{Admitted class $\mathcal{B}$ VCL aborts (is internally lost)}. We derive $\Phi_{in-loss}^{B}$ using busy-cycle arguments. Define the following parameters:

λ_{Bp} (λ_{Bs})	= Offered primary (secondary) class $\mathcal{B}$ VCLR rate
N_{AD}^{tot}	= Aggregate class $\mathcal{B}$ VCLs admitted per busy cycle (primary + secondary)
$N_{IN-LOSS}^{tot}$	= Number of class $\mathcal{B}$ VCLs internally lost per busy cycle
$N_{IN-LOSS}^{ijk}$	= Number of class $\mathcal{B}$ VCLs lost per busy cycle from state $(i,j,k) \in S_{loss}$.

Note that, $\lambda_{Bp} = \lambda_b \sum_{n=1}^{D_{mc}} nb_n$, and $\lambda_{Bs} = \sum_{i=0}^{m_A} \sum_{j=0}^{j=m-i} \sum_{k \in K_j} P_{ijk} k \lambda_s$

Then, N_{AD}^{tot} = Aggregate admission rate of class $\mathcal{B}$ VCLs x Busy Cycle Duration
$= \{\lambda_{Bp}(1-\Phi_{ex}^{Bp}) + \lambda_{Bs}(1-\Phi_{ex}^{Bs})\}(\lambda P_{000})^{-1}$

Also, $\forall (i,j,k) \in S_{loss}$, $N_{IN-LOSS}^{ijk}$ = Number of visits to (i,j,k) per cycle x losses per visit
$= \frac{\pi_{ijk}}{\pi_{000}} \sum_{\alpha=1}^{\lfloor jr \rfloor} \frac{\Psi_3(\alpha)+\Psi_4(\alpha)}{i\mu+\Psi_1(i,j,k)+\Psi_2(i,j,k)+\Psi_3(\alpha)+\Psi_4(\alpha)} \alpha$

Total VCL loss per busy-cycle $N_{IN-LOSS}^{tot} = \sum_{i=0}^{m_A} \sum_{k \in K_j} N_{IN-LOSS}^{ijk}|_{j=m-i}$

Finally, class $\mathcal{B}$ internal loss probability $\Phi_{in-loss}^{B} = \frac{N_{IN-LOSS}^{tot}}{N_{AD}^{tot}}$.

Class $\mathcal{B}$ Loss Prob. Φ_{loss}^{B}, Mean Holding Time HT^B, VCL Throughput TP

1. Class $\mathcal{B}$ (weighted) blocking probability $\Phi_{ex}^{B} = \Phi_{ex}^{Bp}\left(\frac{\lambda_{Bp}}{\lambda_{Bp}+\lambda_{Bs}}\right) + \Phi_{ex}^{Bs}\left(\frac{\lambda_{Bs}}{\lambda_{Bp}+\lambda_{Bs}}\right)$

2. Class $\mathcal{B}$ loss probability Φ_{loss}^{B} is the probability that an arbitrary class $\mathcal{B}$ VCL is externally blocked or internally lost. Then, $\Phi_{loss}^{B} = 1-(1-\Phi_{ex}^{B})(1-\Phi_{in-loss}^{B})$

3. Next, we compute the class $\mathcal{B}$ mean holding time HT^B through Little's law[7]:
HT^B = (Average class $\mathcal{B}$ utilization)/ (Aggregate admission rate of class $\mathcal{B}$ VCLs)
$= \left(\frac{\sum_{i=0}^{m_A} \sum_{j=0}^{m-i} \sum_{k \in K_j} jP_{ijk}}{\lambda_{Bp}(1-\Phi_{ex}^{Bp})+\lambda_{Bs}(1-\Phi_{ex}^{Bs})}\right)$ (used in section 5).

4. Finally, the aggregate VCL throughput (TP) is given by:
$TP = \lambda_a(1-\Phi_{ex}^{A}) + \{\lambda_{Bp}(1-\Phi_{ex}^{Bp}) + \lambda_{Bs}(1-\Phi_{ex}^{Bs})\}(1-\Phi_{in-loss}^{B})$

3.5 QOS and Feasible/Optimal Sub-thresholds

Assume a worst-case VCL QOS vector: $(\Phi_{ex}^{A,max}, \Phi_{ex}^{B,max}, \Phi_{in-loss}^{B,max})$. For simplicity, we combine the worst-case external blocking and internal loss of class $\mathcal{B}$ VCLs into maximum total loss probability $(\Phi_{loss}^{B,max})$ computed as: $\Phi_{loss}^{B,max} = 1 - (1 - \Phi_{ex}^{B,max})(1 - \Phi_{in-loss}^{B,max})$. Further, define $\Phi^{max} = min(\Phi_{ex}^{A,max}, \Phi_{loss}^{B,max})$.

The sub-threshold can be tuned to satisfy feasibility/optimality conditions. The sub-threshold scheme is said to be **feasible** at m_A^* iff $max(\{\Phi_{ex}^A\}_{m_A^*}, \{\Phi_{loss}^B\}_{m_A^*}) \leq \Phi^{max}$. In Figure 2, the set of feasible sub-thresholds $\mathcal{F}_{m_A}$ is, in general, the set of sub-threshold values bounded by the intersection of Φ_{ex}^A and Φ_{loss}^B with Φ_{max}.

From the *application* viewpoint, a **service-optimal sub-threshold** $(m_A^*)_S$ is defined such that, if it exists, $(m_A^*)_S \in \mathcal{F}_{m_A}$ and $\{\Phi_{ex}^A\}_{(m_A^*)_S} = \{\Phi_{loss}^B\}_{(m_A^*)_S}$. To satisfy the integral $(m_A^*)_s$ constraint, we allow for the nearest integer solution to the intersection of Φ_{ex}^A and Φ_{loss}^B. The sub-threshold $(m_A^*)_s$ defines the operating point at which the network provides the VCLs a service quality (QOS) independent of the higher-layer dependence(class $\mathcal{A}$ or $\mathcal{B}$). Also, note that if $(m_A^*)_s$ cannot be found at an offered load, it follows that there is no feasible solution to the admission scheme!

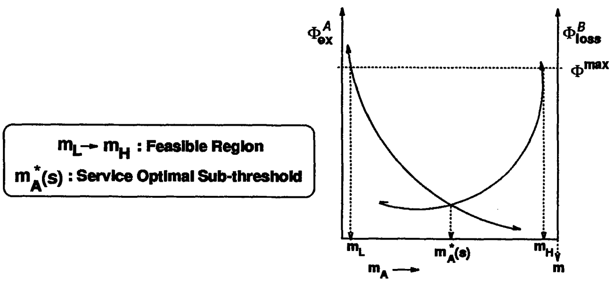


Figure 2: Feasibility and service-optimality issues.

From the *network operator* viewpoint, we select an optimality sub-threshold that maximizes aggregate throughput. Formally, a **throughput-optimal sub-threshold** $(m_A^*)_T \in \mathcal{F}_{m_A}$, such that $\{TP\}_{(m_A^*)_T} \geq \{TP\}_{(m_A^*)}, \forall\, m_A^* \in \mathcal{F}_{m_A}$.

4 RESULTS

4.1 Effect of r, m_A on Φ and TP

Note the parameters in the textual legends of Figures 3, 4, and 5. Figure 3 plots class $\mathcal{B}$ primary (individual/group) blocking, secondary blocking and internal loss probability with respect to m_A variation.

Figure 4 compares class $\mathcal{A}$ VCLR external blocking Φ_{ex}^{A} to the class $\mathcal{B}$ total loss probability Φ_{loss}^{B} formulated in Section 3.4. The service optimal point $(m_A^*)_S$ (assuming its feasible) is indicated. Note that r variation at a fixed offered load does not significantly change the performance measures. This is pleasing from the design point of view.

Figure 5 plots aggregate VCL throughput TP over similar conditions. Note that increasing r reduces TP slightly because the batch-loss increase dominates the external blocking reduction. Also, the dynamic variation of TP over m_A is small; increasing m_A increases the $\Phi_{in-loss}$ due to more frequent secondary blocking. This creates more space in the system and consequently reduces class $\mathcal{B}$ external blocking.

Figure 5 also indicates the simulated VCL throughput TP_{sim} for $r = 0.1$. The variation between the analysis and simulation results is no more than 5% (less than 1% for smaller systems). Thus, the session-homogeneity assumption is seen to perform well.

4.2 Throughput-Optimal Sub-threshold Trajectory

Figure 6 illustrates $(m_A^*)_T$ variation with traffic mix parameter p_a. This variation is plotted for two values of initial session size ($D_{mc} = 1, 5$). The secondary arrival rate per L-session is modified at each observation to keep a constant offered load = 0.6.

We observe that as p_a increases, $(m_A^*)_T$ reduces linearly over a significant range. This is equivalent to allocating more resources to class $\mathcal{B}$ VCLRs when the class $\mathcal{A}$ traffic dominates, since *goodput* per admitted class $\mathcal{B}$ VCLR is maximum under this condition.

Also, at a fixed p_a, $(m_A^*)_T$ is larger for larger D_{mc} values (refer to $p_a = 0.5$, where $(m_A^*)_T = 48, 49$ at $D_{mc} = 1, 5$ respectively.). Since secondary arrival rate λ_s is varied to keep offered load constant at both the points, the result offers an important interpretation. Consider the fixed abscissa $p_a = 0.5$. The shift of $(m_A^*)_T$ from 49 to 48 reflects the tradeoff between large initial-size static sessions and small initial-size dynamic sessions. Clearly, at $p_a = 0.5$, the dynamicity of secondary arrivals dominates the initial session size for the overall effect. At an increased value of $p_a = 0.9$, throughput becomes sensitively dependent on every large blocked primary class $\mathcal{B}$ VCLR. Hence, $(m_A^*)_T$ for $D_{mc} = 5$ converges with that for $D_{mc} = 1$. At this point, the initial session size completely counteracts dynamicity due to secondary arrivals.

5 NETWORK ALGORITHM

We present a distributed algorithm that designs network-wide service-optimal sub-thresholds on all the network links. Depending on the location of multi-cast switches and the routing scheme (stochastic routing), it is possible to encode each link (i.e. its offered primary and secondary VCLR traffic pattern, parameters $\lambda, p_a, b_i, D_{mc}, \lambda_s, p_s$) in the SL model format. However, solving independent SL models is inadequate because the offered rates at each link are dependent on the Φ vector of its neighbors.

The network algorithm presented here solves this problem by iteratively modifying the rates through a two-tiered structure. In the first tier, it computes the offered arrival rates

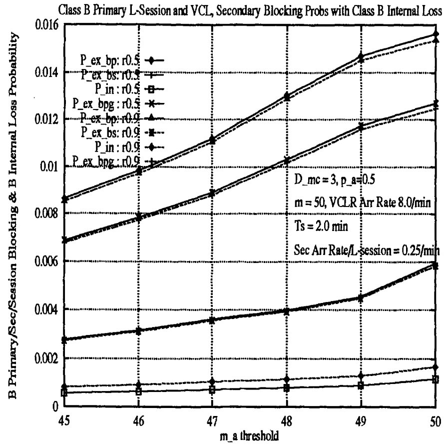


Figure 3: Effect of r, m_A on class $\mathcal{B}$ blocking and internal loss.

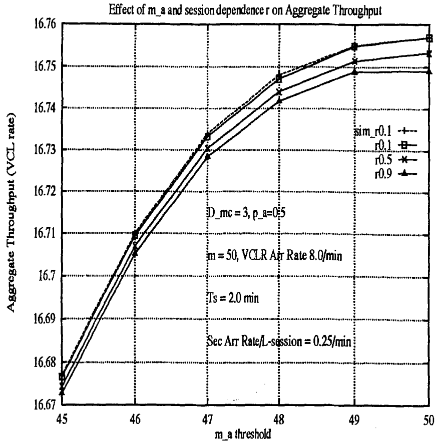


Figure 5: Effect of r, m_A on aggregate VCL throughput.

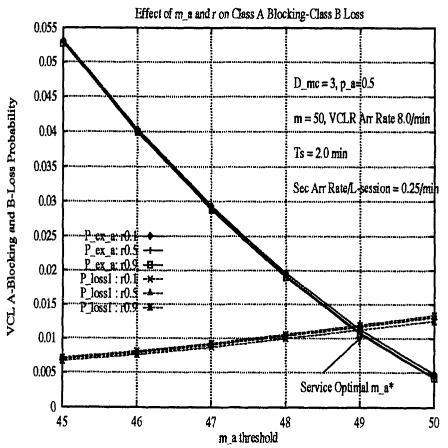


Figure 4: Effect of r, m_A on class $\mathcal{A}$ and $\mathcal{B}$ blocking and loss.

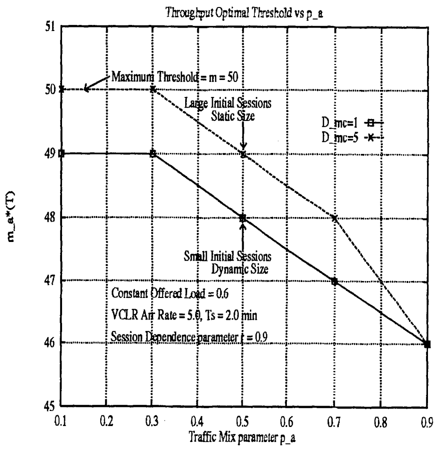


Figure 6: Throughput-Optimal threshold trajectory.

using only the external blocking component. It then calculates the sustained offered rates, as would be seen by the end-to-end connections. In the second tier, it computes the true internal loss on each link by accounting for reflected loss from other links. Finally, the sub-threshold is updated in the instantaneous direction towards service-optimality and the process is repeated.

The basic algorithmic framework follows. We assume for simplicity that sessions are independent of each other. The session QOS vector is $\Theta^{max} = (\Theta_{ex}^{A,max}, \Theta_{ex}^{B,max}, \Theta_{in-loss}^{B,max})$. A session-request is blocked if any component primary VCLR gets blocked. If a secondary VCLR is blocked at any node, a fraction r of the sink-clients of that session, downstream to that request terminate (assuming a topological dependence). $\Theta_{in-loss}^{B,max}$ represents the maximum internal loss probability that the sink-clients can tolerate.

Refer to Figures 7, 8, and 9 for the flow-charts. We qualify these with additional important comments:

1. Φ^{max} vector is derived on each link in the following steps:

   (a) Maximum primary VCLR blocking, $\Phi_{ex}^{Bpg,max} = 1-(1-\Theta_{ex}^{B,max})^{(H)^{-1}}$, and $\Phi_{ex}^{A,max} = 1-(1-\Theta_{ex}^{A,max})^{(H)^{-1}}$ where H = maximum hops traversed by VCCs over all sessions (conservative design).

   (b) $\Phi_{ex}^{Bp,max}$ is related to $\Phi_{ex}^{Bpg,max}$ through a simple bound (given the batch distribution b_i on the specific link): $\frac{\Phi_{ex}^{Bpg,max}}{\sum_{n=1}^{D_{mc}} nb_n} \leq \Phi_{ex}^{Bp,max} < \frac{\sum_{l=1}^{l=D_{mc}} b_l \Upsilon(l)(l-1)}{\sum_{n=1}^{D_{mc}} nb_n} + \frac{\Phi_{ex}^{Bpg,max}}{\sum_{n=1}^{D_{mc}} nb_n}$, where $0 \leq \Upsilon(l) = \sum_{i=0}^{m_A} \sum_{j=max(m-l-i+1,0)}^{m-i} \sum_{k\in K_j} P_{ijk}$. The derivation is omitted for brevity. We conservatively select the lower bound: $\Phi_{ex}^{Bp,max} = \frac{\Phi_{ex}^{Bpg,max}}{\sum_{n=1}^{D_{mc}} nb_n}$.

   (c) Assuming the same bound for secondary blocking, $\Phi_{ex}^{B,max} = \Phi_{ex}^{Bp,max}$. Also, it can be shown that $\Phi_{in-loss}^{B,max} = \Theta_{in-loss}^{Bmax}$ guarantees the sink-clients a feasible internal loss probability.

   (d) As before, $\Phi_{loss}^{B,max} = 1-(1-\Phi_{ex}^{B,max})(1-\Phi_{in-loss}^{B,max})$, $\Phi^{max} = min(\Phi_{ex}^{A,max}, \Phi_{loss}^{B,max})$.

2. In Figure 8, the Dependence Algorithm can be executed in parallel for all links incident on a single node, and sequentially node-wise. The algorithm modifies the holding time of a tagged link by reflecting the holding times of its neighbors on to it. This has the effect of modeling the system-size space effect due to internal loss.

3. The Threshold Guidance algorithm in Figure 9 updates the sub-threshold depending on the current Φ state with respect to the service-optimal threshold (see Figure 2) computed at the given load.

4. If the complexity of the single-link model is $0(SL)$ in an n-node network, the network algorithm can be shown to have a worst-case time-complexity of $O(SL.n^2)$, provided the iterations exhibit constant order. The algorithm has shown promising behavior on the examples tested.

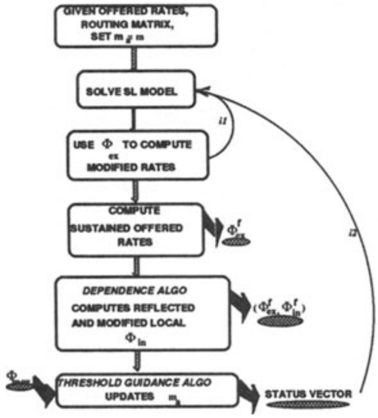


Figure 7: Two-tiered Network Algorithm.

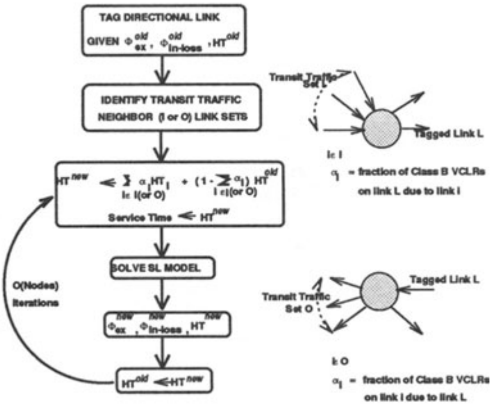
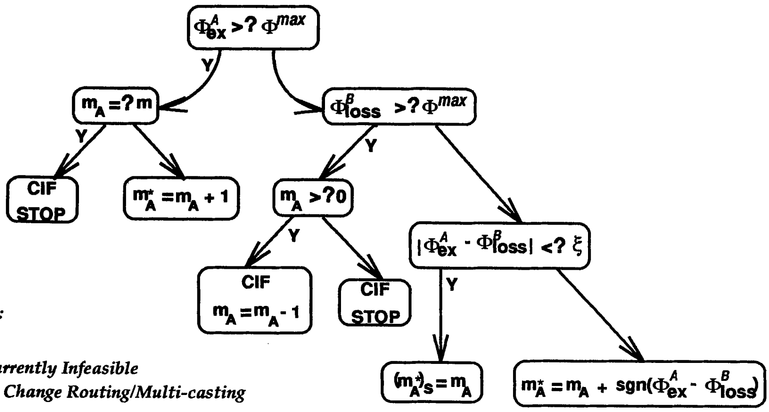


Figure 8: Dependence Algorithm for Internal loss.



Legend :

CIF : Currently Infeasible
STOP : Change Routing/Multi-casting

Figure 9: Threshold Guidance Algorithm for sub-threshold updates.

6 CONCLUSIONS

We contend that future multi-media/multi-point applications will require admission management at the connection layer (over and above the ATM layer). In this work, we have formulated a simple threshold-based distributed connection admission scheme for heterogeneous sessions. We have developed appropriate connection-level QOS measures for uni-point/static and multi-point/dynamic sessions. The threshold scheme can be tuned to attain service-optimality. A network algorithm extends this to incorporate end-to-end session requirements.

References

[1] M. Gaddis, R. Bubenik, and J. DeHart, "A Call Model for Multipoint Communication in Switched Networks," *ICC'92*, pages 609-615.

[2] S. Minzer,"A Signaling Protocol for Complex Multimedia Services," *IEEE Journal on Selected Areas in Communications,* 9(9):1383-1394, December 1991.

[3] ANSI T1S1 Technical Sub-Committee, "Broadband Aspects of ISDN Baseline Document," *T1S1.5/90-001*, June 1990.

[4] L. Gun and R. Guerin, "Bandwidth Management-Congestion Control Framework of the Broadband Network Arch.," *Computer Networks and ISDN Systems*, 26(1):61-78, 1993.

[5] H. Saito, "Call Admission Control in an ATM Network Using Upper Bound of Cell Loss Probability," *IEEE Trans. on Comm.*, 40(9):1512-1521, Sept 1992.

[6] E. Cinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, 1975.

[7] L. Kleinrock, *Queueing Systems: Vol I*, Wiley, New York, 1976.

PRATYUSH MOGHE is a graduate student researcher in the department of Electrical Engineering at UCLA. His current research interests focus on admission management of enhanced-calls supporting multi-media/multi-party dynamic applications. He earned the B.E (1988) from the department of Electronics and Telecommunications at College of Engineering, Poona, India, and an M.S (1990) in Electrical Engineering from Clemson University, SC. He was a technical member of the Network Architectures-Services group at GTE Laboratories, Waltham, MA (Summer 1990). He received the Best Student Award (1988 IAF Trophy, Univ. of Poona) and the UCLA Fellowship (1990).

IZHAK RUBIN received the B.Sc. and M.Sc. from the Technion, Israel, and the Ph.D. degree from Princeton University, all in Electrical Engineering. Since 1970, he has been a professor in the UCLA Electrical Engineering Department. He has had extensive research and industrial experience in the design and analysis of telecommunications, computer communications, and C3 networks. He has also been serving as chief-engineer of IRI Computer Communications Corporation. He is an IEEE Fellow, has served as chairman of IEEE conferences, and as an editor of the IEEE Transactions on Communications and of the journal on Wireless Networks.