Scalability Evaluation of Multi-Protocol Over ATM (MPOA)

Indra Widjaja, Haining Wang, Steve Wright, Amalendu Chatterjee

Fujitsu Network Communications
4403 Bland Road
Raleigh, NC 27609
E-mail: indra.widjaja@fnc.fujitsu.com

Abstract

Multi-Protocol over ATM (MPOA) is being considered by the industry as an important short-cut technology that provides an efficient transfer of inter-subnet unicast data in a LANE environment. MPOA was initially considered to carry backbone traffic in the enterprise or campus networks. However, congestion in the public Internet provokes many to consider MPOA as a solution for the carrier or service provider networks as well. In this paper, we investigate the scalability issues of MPOA in the wide area network environment. We use a realistic simulation model driven by real Internet traffic traces to study crucial metrics such as the SVC setup rate, the number of VCs required, and the percentage of packets switched. Based on the simulation results, we find that MPC ingress cache size provides a three-way trade-off among the percentage of switched packets, the VC usage and the SVC setup rate requirement. We also find that the SVC setup rate is linearly dependent on the packet arrival rate.

Keywords: IP over ATM, MPOA, LANE, NHRP, scalability, performance.

1 Introduction

The Internet has grown to an unprecedented size with 1.3 million domains and 19.5 million hosts as of July 1997 [1]. The current situation is made more exciting not only because of the size of the Internet, but because of the continued and rapid growth that has been driving up the traffic demand exponentially. Vinton Cerf recently projected that traffic growth in the Internet would be roughly at 300 percent per year [2]. This growth in traffic demand is stressing existing network infrastructures, which may trigger a major bottleneck if not correspondingly upgraded. To maintain their network performance, service providers are constantly upgrading their links, typically with ATM technology, almost on a regular basis in order for the backbone links to keep up with the traffic demand. In turn, the routers have to be upgraded as well to keep up with increasing link speeds.

There is widespread concern that the limit in the conventional router architecture is approaching very quickly. The need to fix the problems with the conventional router architecture has generated many solutions which can be categorized as in Figure 1.

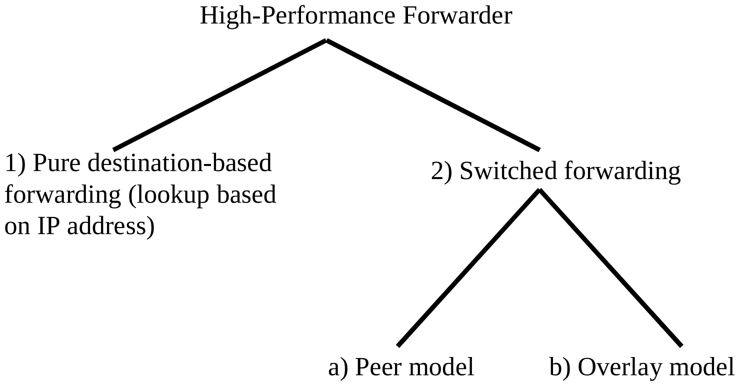
High-Performance Forwarder
1) Pure destination-based forwarding (lookup based on IP address)
2) Switched forwarding
a) Peer model
b) Overlay model

Figure 1 Forwarding taxonomy.

Category 1 retains the same forwarding paradigm as conventional router architecture, but solves the potential bottlenecks by providing multiple data paths. For example, this is done by replacing the bus backplane in the data paths with a switch backplane. Moreover, scalability is also improved by having an IP lookup engine at each interface. The best example is the gigabit router architecture.

Category 2 simplifies the lookup process by using short, fixed-length labels rather than long, variable-length addresses. The most typical way to do this is to run IP over ATM which uses VCI and VPI in the lookup process. Because label lookup uses direct indexing (rather than longest-prefix match), it can be easily performed in hardware. For example, label lookup in ATM simply uses the incoming VPI/VCI value as a direct index to a connection table entry to determine the outgoing VPI/VCI, the output port, and other information. ATM table lookup can be easily done in one cell time. On the other hand, IP address lookup requires implementation in software or firmware. Category 2 can be further classified into the *peer model* and the *overlay model*.

The peer model uses the existing IP addresses (or with algorithmically derived ATM addresses) to identify end systems, and IP routing protocols to setup ATM connections. One advantage of the peer model is that it does not require an address resolution protocol to interwork routable address spaces, and thus simplifies address administration. Switches are typically upgraded to ATM switching and IP routing, and can be viewed as “peers” to routers. The peer model maintains one network infrastructure. Examples of this model include Ipsilon’s IP switching [3] and Multiprotocol Label Switching (MPLS) [4][5].

In the overlay model, ATM switches are not aware of IP addresses and IP routing protocols. This model essentially overlays an IP network onto an ATM network, essentially creating two network infrastructures with two addressing schemes and two routing protocols. Each end system uses both IP and ATM addresses that are uncoupled. An address resolution protocol is required to map from one address to another. One advantage of this model is that the ATM infrastructure can be developed independently of the IP infrastructure. Examples of this model are Classical IP over ATM (CLIP) [6] and Multiprotocol Over ATM (MPOA) [7][8].

This paper focuses on MPOA which is intended to provide internetworking layer service such as IP, IPX and AppleTalk, over an ATM network. MPOA Version 1.0 Specification was approved by ATM Forum in July 1997. The main goal of MPOA is to reduce extra router hops by allowing internetwork layer communication over ATM VCCs (called *shortcuts*). To establish shortcuts, MPOA uses LANE [9] and NHRP [10] to locate the ATM device that is closest to the destination.

MPOA was initially considered to carry traffic in the enterprise or campus networks. However, congestion in the Internet has provoked many to consider MPOA as a possible solution for the service provider networks as well. In this paper, we investigate the scalability issues of MPOA to answer the feasibility of deploying MPOA in the service provider networks. We use a realistic simulation model driven by real Internet traffic traces to study crucial metrics such as the SVC setup rate, the number of VCs required, and the percentage of packets switched. MPOA is said to be feasible if the SVC setup rate and VC usage requirements can be fulfilled by existing ATM switches. The percentage of packets switched, which depends on the policy to establish a VCC, further relates to the effectiveness of MPOA. In general, increasing the percentage of packets switched increases the SVC setup rate and VC usage requirements.

Although MPOA has been developed by the ATM Forum for a few years, there is no related literature dealing with quantitative performance study, especially with its scalability. Lin and McKeown have studied the performance of IP switching [11]. However, fundamental differences in MPOA and IP switching require separate investigations. One particular difference is that MPOA performs flow detection at the ingress node while IP switching requires each node to perform flow detection. The other difference is that MPOA requires address resolution while IP switching does not.

The rest of this paper is organized as follows. In Section 2, we provide a brief description of MPOA. We limit our discussion on the more important shorcut which is the inter-LANE shortcut. Readers interested in the intra-LANE shortcut are referred to the LANE specifications. In Section 3, we describe our simulation model and assumptions. In Section 4, we present the quantitative performance results and discuss the various implications. Finally, we conclude the paper in Section 5.

2 Description of MPOA

This section provides a brief description of how MPOA provides inter-LANE shortcuts. For simplicity, the discussion focuses only on the IP as the internetwork layer protocol. Readers interested in the details of the protocols are referred to the appropriate standard documents.

2.1 Overview

MPOA is based on a client-server architecture with two main components: *MPOA Clients* (MPCs) and *MPOA Servers* (MPSs). MPC and MPS communicate via an Emulated LAN (ELAN). MPC can reside in an edge device or an MPOA host, whereas MPS always resides in a router. Figure 2 describes how shortcuts are established in MPOA.

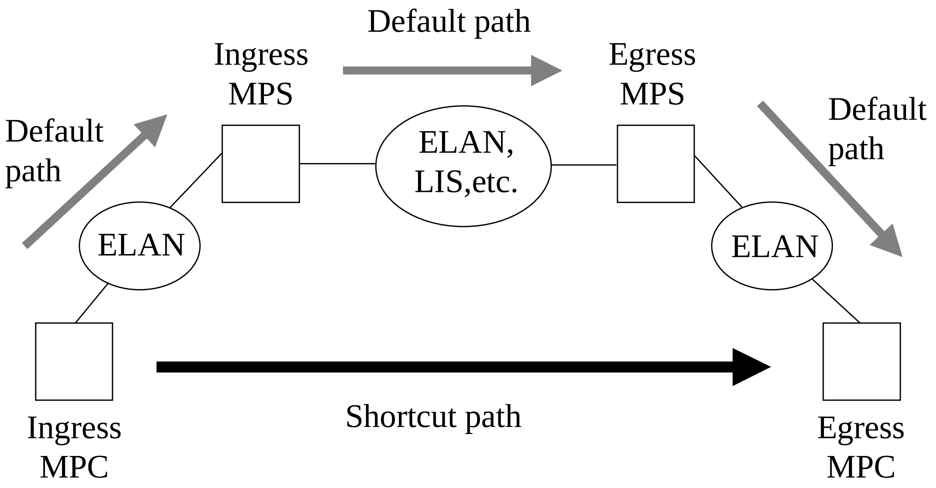

Figure 2 MPOA default and shortcut paths.

Suppose a flow of packets need to be sent from the ingress MPC to the egress MPC, and no VCC between them exists yet. The first few packets are forwarded via the default (or routed) path through the ingress MPS and finally through the egress MPC. This is not an efficient way to transfer data in an ATM network since each packet has to go through routers when the packet goes to a different subnet.

MPOA can make the transfer more efficient by using an ATM shortcut if the flow is deemed to be long-lived. Specifically, if the ingress MPC receives a given number of packets of the same flow within a window of time, it will try to establish a shortcut between the ingress MPC and the egress MPC through an ATM cloud. To do so, the ingress MPC has to know the ATM address of the egress MPC. If the ingress MPC does not know the ATM address, it would have to resolve the ATM address based on the IP address. First, the ingress MPC sends an MPOA Resolution Request to the ingress MPS containing the destination IP address. The ingress MPS forwards the resolution request message through the default path until it reaches the egress MPC which serves as the exit point to the destination[1]. If the resource for the shortcut is available at the egress MPC, it will insert its ATM address to a reply message sent to the egress MPS. Eventually the Resolution Reply containing the ATM address of the egress MPC arrives at the ingress MPC.

Once the ingress MPC knows the ATM address of the egress MPC, it can use standard ATM signaling to establish a VCC. From then on, the rest of the packets for that flow can be transmitted through the VCC. If there is no more data for that flow, the VCC will be eventually terminated through aging.

2.2 MPC

The main function of the MPC is to source and sink inter-LANE shortcuts. The MPC resides between a LAN Emulation Client (LEC) and its higher layer protocol (typically bridging functions), as shown in Figure 3.

[1] The forwarding of the Resolution Request/Reply messages between an ingress MPS and an egress MPS is usually done by NHRP servers.

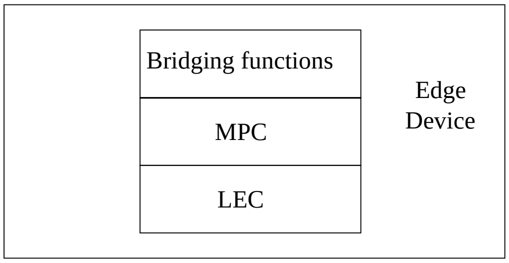

Figure 3 The structure of MPOA client.

An ingress MPC monitors the destination MAC address of each outbound packet that comes from the higher layer protocol to the LEC. If the packet is destined for the same subnet, the destination MAC address will not match the MAC address of the ingress MPS. This packet will be forwarded through LANE. If the packet is destined for a different subnet, the MPC will detect a match (or MPS MAC hit) and will try to find the <MPS control ATM address, Destination IP address> tuple in its ingress cache. The organization of the ingress cache is shown in Figure 4.

Keys		Contents		
MPS control ATM address	Destination IP address	VCC	Encapsulation information	Other control information

Figure 4 MPC ingress cache.

If the <MPS control ATM address, Destination IP address> tuple is not found, a new entry is created. The VCC is invalidated to indicate that the shortcut is not operational yet, and a count field in the control information is set to 1. The packet is then sent through the default path.

If the <MPS control ATM address, Destination IP address> tuple is found and the shortcut is invalid, the count field is incremented and the packet is sent through the default path.

When the count field exceeds a configured threshold (**MPC-p1**) within a configured time period (**MPC-p2**), the ingress MPC sends an MPOA Resolution Request to the ingress MPS requesting the ATM address of the egress MPC. The Resolution Request is propagated by NHRP (to be discussed below) to the egress MPS which then triggers an MPOA Cache Imposition Request to the egress MPC to determine if the necessary resources (e.g., cache and VC entries) are available to support a shortcut. If the resources are available, the egress MPC returns an MPOA Cache Imposition Reply containing its ATM address which will eventually be propagated back to the ingress MPC. While waiting for the MPOA Resolution Reply, the ingress MPC continues forwarding subsequent packets through the default path. Upon receipt of an MPOA Resolution Reply containing the ATM address of the egress MPC, the ingress MPC can establish a VCC to the egress MPC. The VCC information is cached and validated so that subsequent packets of the same flow will be sent over the shortcut.

An ingress cache entry is aged out if an MPOA Resolution Reply does not arrive within a certain timeout, called the Holding Time. To prevent an active cache entry from aging out, the MPC should refresh the cache entry by sending an MPOA Resolution Request to ensure that the MPOA Resolution Reply will return before the timer expires.

2.3 MPS

An MPS resides in an ATM-attached router, as shown in Figure 5. In addition of an MPS, the router must also have an NHRP server (NHS), one or more LECs, and some convergence functions (e.g., IP ARP).

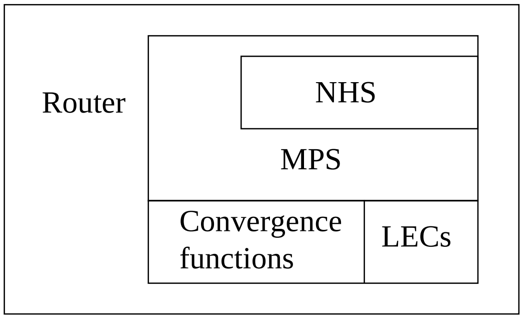

Figure 5 The structure of MPOA server.

The main function of an ingress MPS is to process MPOA Resolution Requests sent by ingress MPCs. If the destination is connected directly to one of its attached interfaces, the MPS will respond directly. Otherwise, the MPS will re-originate the request message through an NHS by sending an NHRP Resolution Request. The IP source address of the MPS is used in the NHRP Resolution Request since an MPC may not have an IP address.

When the NHRP Resolution Request eventually arrives at the router which is directly connected to the egress MPC, the egress MPS reformulates the message to an MPOA Cache Imposition Request before sending it to the egress MPC. Upon receiving an MPOA Cache Imposition Reply from the egress MPC, the egress MPS reformulates it to an NHRP Resolution Reply which will then be propagated back to the ingress MPS.

2.4 NHRP

MPOA relies on the Next Hop Resolution Protocol (NHRP) to perform IP-to-ATM address resolution across multiple subnets. NHRP is also based on a client-server architecture. An NHRP cloud may contain entities called *Next Hop Clients* (NHCs) which are responsible for initiating NHRP Resolution Requests, and *Next Hop Servers* (NHSs) which are responsible for processing the NHRP Resolution Requests and Replies. When used in conjunction with MPOA, the MPS rather than the NHC is the entity that triggers the NHRP Resolution Request.

The environment for NHRP is illustrated in Figure 6. It typically consists of IP subnets, called *Logical IP Subnetworks* (LISs), which are overlayed on top of an ATM network. In an overlay network, communication efficiency can be improved if data transfer between node S and node D takes place over an ATM VCC rather than a routed path. However, node S has to know the ATM address of node D before it can establish the VCC, and this is where NHRP is used to resolve the IP address of node D to its ATM address.

Referring to Figure 6, NHRP works as follows. When an ingress node S wants to resolve the ATM address of an egress node D, node S sends an NHRP Resolution Request packet along the routed path. The NHRP Resolution Request packet contains node D’s IP address, node S’s IP address, and node S’s ATM address.

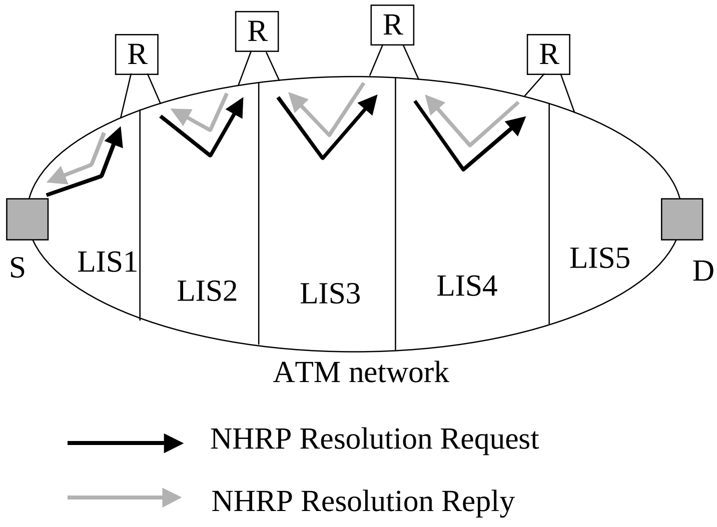

Figure 6 NHRP resolution process.

Each router along the path has an NHS. When an NHRP Resolution Request packet arrives at a router, the NHS determines if it is the "serving NHS" for node D. A serving NHS for node D has one of its router's interfaces connected to the same LIS as node D. If the NHS is not serving node D, it will re-initiate an NHRP Resolution Request packet to another NHS closer to the destination. The process continues until the NHRP Resolution Request reaches the NHS that is the serving node D. This NHS then resolves node D's ATM address and sends a positive NHRP Resolution Reply packet back to node S. The NHRP Resolution Reply packet contains node D's ATM address.

A transit NHS that relays the NHRP Resolution Reply may cache the IP-to-ATM address binding information. When a subsequent NHRP Resolution Request arrives at this transit NHS, it may reply using the cached information. However, such a reply must be identified as *non-authoritative*. Only the serving NHS can respond with an NHRP *authoritative* Resolution Reply. In general, non-authoritative reply speeds up the address resolution process. However, this service comes at the expense of increasing the cache size requirement at the NHS. Also, when the IP-to-ATM address binding at the destination changes, a transit NHS will respond with a wrong Resolution Reply.

NHRP does not require a destination node to be connected to the same ATM network as the NHC. For example, if node D is located in another network, NHRP will only resolve the ATM address of the egress router closest to node D. In such a case, the shortcut can only be established up to the egress router.

3 Simulation Model

We assume that the entire MPOA domain belongs to a service provider network. The interface to an external network is via an edge device which has an ingress and egress MPC. A large MPOA domain may have thousands of edge devices. To simplify the model, we focus on a particular ingress MPC. This allows us to arrive at a relatively simple simulation model which consists of a single ingress MPC, a corresponding ingress MPS, and intermediate NHSs/MPSs along the path toward the egress MPC, as shown in Figure 7. We assume that each egress MPC always has available resources to accept an MPOA Cache Imposition Request so that the detailed behavior of the egress MPC is not needed.

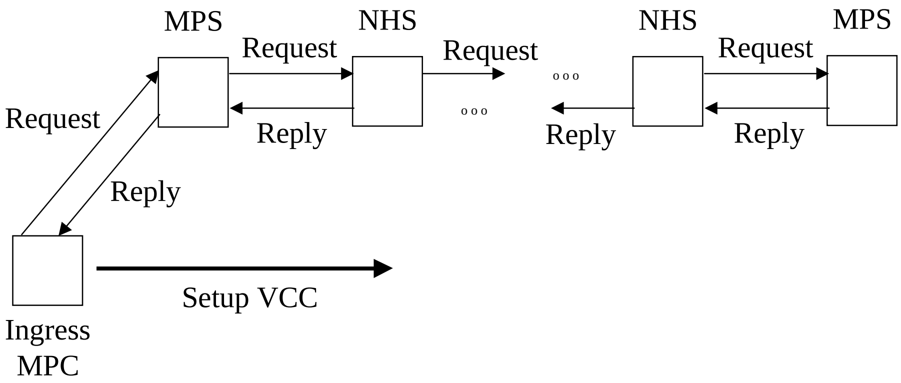

Figure 7 The structure of the simulation model.

The key component of the simulation model is the ingress MPC which processes arriving packets from an external network. The ingress MPC maintains a finite-sized cache of entries shown in Figure 4. When the cache is full, a request for a new cache entry will replace an existing entry that is least recently used (i.e., LRU policy is adopted in the simulation). MPSs/NHSs also maintain their own caches. Since each MPS/NHS could be associated with many MPCs, it also caches the resolution replies from other clients. The cache behavior of the MPS/NHS is described probabilistically. Specifically, a cache hit probability (P_{HIT}) to the MPS/NHS is assumed for each request independently. When a resolution packet follows a default path, the resolution delay depends on the number of router hops the packet has to follow. The maximum number of hops between the ingress MPC and the egress MPC is N_H hops. To decouple the resolution delay from the complexity of the physical topology, we assume that the number of hops from the ingress MPC to the egress MPC is a function of the destination address of the packet. Specifically, if a packet is assigned its destination address uniformly, then the packet would choose its hop number from one to N_H uniformly. This condition can be simply approximated by computing the hop number based on the result of modulo N_H of the packet destination address.

MPOA is a complex set of protocols whose operations include component configuration, device discovery, target resolution, connection management and data transfer. Since the goal of MPOA is efficient inter-subnet communication based on Emulated LANs, the address resolution mechanism of MPOA plays a key role in the success of MPOA. Our major concern lies in the operations of target resolution and connection management. The operations/functions that have no or negligible impacts on the relevant performance will be ignored. A short description of the assumptions in the simulation model is presented as follows:

- Appropriate MPOA configuration and device discovery are assumed.
- Intra-subnet traffic is taken care by LAN Emulation, and the correctness of Emulated LANs is assumed.
- Routing table in each router is stable (no route flapping) during the simulation period.
- Control VCCs between ingress MPCs and MPS are already established.
- Translations between MPOA resolution messages and NHRP resolution messages are always correct.
- MPOA Resolution Reply for each MPOA Resolution Request is guaranteed. Thus, the MPOA retry mechanism is not simulated.
- MPSs/NHSs in our simulation never fail. Thus, the Keep-Alive protocol is not simulated.
- Each egress MPC always has enough resources to maintain the cache entry and receive a new shortcut.

One critical point to consider in simulating MPOA is the traffic model at the ingress MPC. Unlike most simulation models which only rely on the packet arrival process, here we also have to consider the process that assigns the IP address to each packet. To make the traffic model as realistic as possible, we decided to use a real traffic trace which was measured at the FIXWEST West Coast Federal Interexchange Node which is suitable for a WAN traffic model. The trace can be downloaded from the National Laboratory for Applied Network Research (NLANR). The statistics of the trace is shown in Table 1.

Table 1 Traffic trace statistics.

Type	Date/time	Duration	# Packets	Avg Pkt Rate
Backbone	2/28/96/13:45	12.866 mins	11,550,348	37.62 Mbps

We focus on three main performance measures of interest: percentage of switched packets, maximum VC usage, and average SVC setup rate. The percentage of switched packets is defined as the long-term ratio of the number of packets that go through the shortcuts to the total number of packets (switched and forwarded packets). The maximum VC usage is defined as the maximum number of VCs that the ingress MPC requires in order to establish shortcuts successfully to its destinations. It is assumed that a shortcut is established by the ingress MPC per IP destination address. Finally, the average SVC setup rate is defined as the average number of SVC setups/second that the ingress MPC will generate to an ATM switch. Various default values of the parameters are given in Table 2.

Table 2 Simulation default values, unless otherwise specified.

Parameter	**Value**
MPC-P1	10
MPC-P2	1 sec
Holding time	20 mins
NHRP propagation and processing time	100 msecs
MPOA propagation and processing time	1 msec
Maximum number of hops, N_H	10

4 Simulation Results

In this section, we present the simulation results and discuss various behaviors and implications. We study the impact of flow detection policy, arrival rate, cache size, and hit probability on critical performance measures.

4.1 Impact of Flow Detection Policy

We first investigate the impact of the flow detection policy using the original traffic trace from FIXWEST. As described in Section 2.2, the flow detection policy depends on two parameters, **MPC-p1** and **MPC-p2**. Increasing **MPC-p2** or decreasing **MPC-p1** makes the policy more aggressive in the sense that more packets will be switched than forwarded. Due to space limitation, we will only concentrate on **MPC-p2** in this paper.

In Figure 8, we vary **MPC-p2** to see its impact on the percentage of switched packets for different values of P_{HIT}. A cache size of 10K entries is assumed. A P_{HIT} value of zero is equivalent to enforcing authoritative replies. Note that P_{HIT} does not have a significant influence on the percentage of switched packets. As expected, we observe that the percentage of switched packets increase as **MPC-p2** increases. However, beyond a certain point (say beyond 10 secs), the increase is negligible. This observation implies that most flows in the traffic trace are of relatively short duration. Note that using a default value of **MPC-p2** only allows about 60 percent of the packets being switched. Increasing **MPC-p2** even beyond 20 secs only increases the percentage of switched packets to about 80 percent. This indicates that about 20 percent of the packets are associated with flows of less than 10 packets.

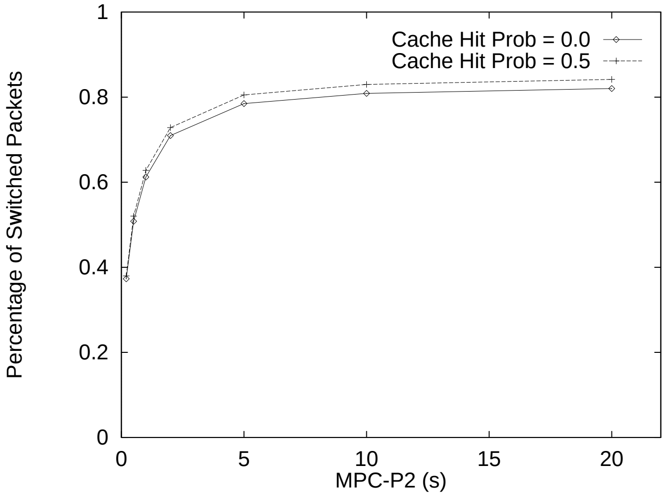

Figure 8 Impact of MPC-p2 on percentage of switched packets for different P_{HIT}.

A more aggressive flow detection policy can be obtained by decreasing **MPC-p1**. As can be seen from Figure 9, 90 percent of the packets can be easily switched by setting **MPC-p1** to 2 packets and keeping **MPC-p2** to 1 second. An even higher percentage may be achieved by increasing the value of **MPC-p2** further.

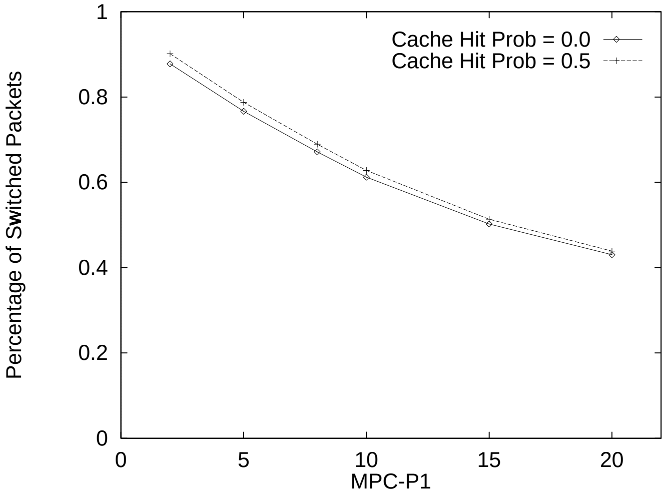

Figure 9 Impact of MPC-p1 on percentage of switched packets for different P_{HIT}.

Figure 10 shows the corresponding maximum VC usage. As we keep increasing **MPC-p2**, we observe that the maximum VC usage increases and eventually levels off at around 5K. This requirement can typically be met since many ATM switches provide at least 8K-16K VCs per interface. We have also tried to measure the average VC usage rather than the maximum value. We found that the average VC usage is close to the maximum one, indicating that the distribution of VC usage is heavily weighted toward the maximum value.

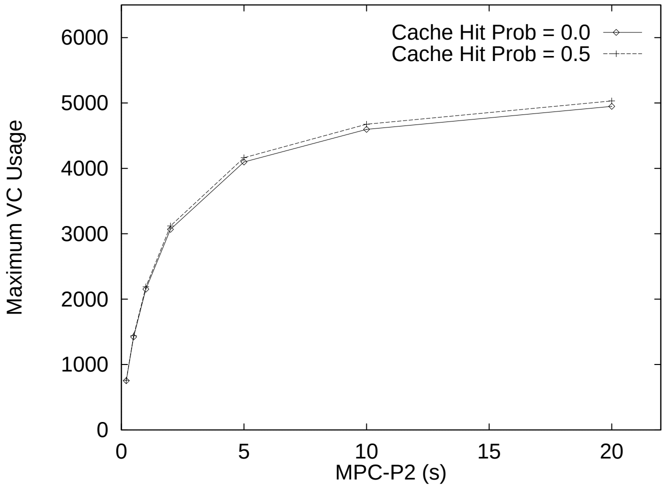

Figure 10 Impact of MPC-p2 on maximum VC usage for different P_{HIT}.

Using the same configuration, we plot the average SVC setup rate in Figure 11. It is worth clarifying that the SVC setup rate requirement is typically per ATM interface. Thus, if the SVC setup rate requirement is *x*, then an ATM switch with *N* interfaces must be capable of setting up *Nx* SVCs per second. From the figure, the SVC setup rate increases as the policy becomes more aggressive by increasing **MPC-p2**. The requirement appears to level off at around 120 setups per second. Note also that the SVC setup rate is virtually insensitive to the cache hit probability. This implies that the resolution delay has little effect on the setup rate.

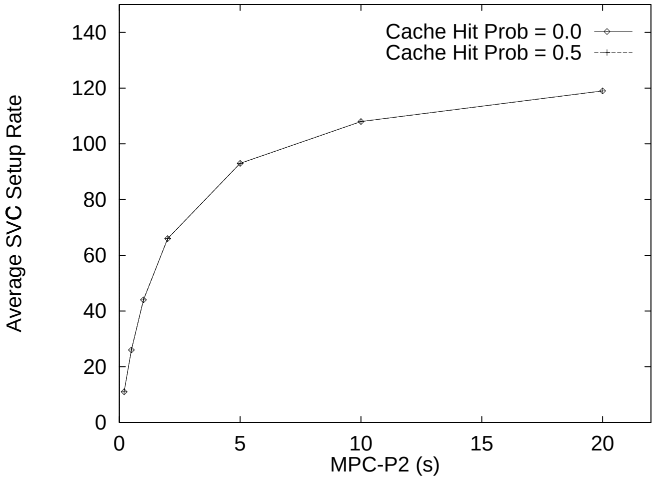

Figure 11 Impact of MPC-p2 on average SVC setup rate for different P_{HIT}.

We now fix the cache hit probability and vary the MPC cache size to understand the MPOA behavior in more detail. Again, we will plot the main performance measures against **MPC-p2**. In the figures that follow, we set P_{HIT} to zero.

First, the percentage of switched packets is shown in Figure 12 for cache sizes of 10K and 100K. One can generally conclude that a bigger cache size allows MPOA to switch more packets. This is also confirmed in the figure where MPOA switches more packets with a cache size of 100K than 10K. In particular, the percentage of switched packets increases roughly by 10 percent as we increase the cache size from 10K to 100K.

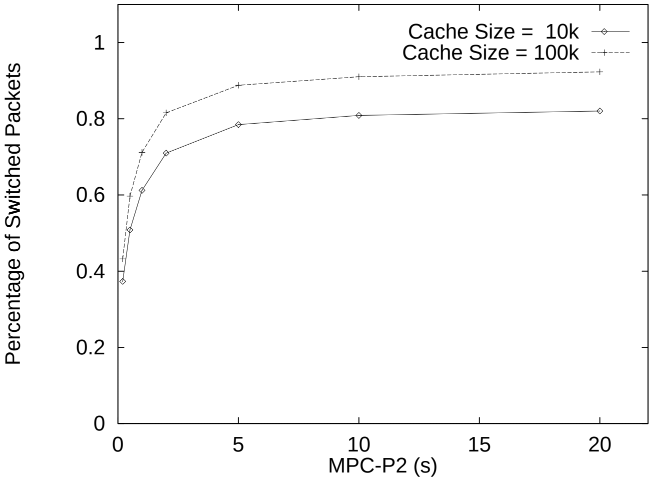

Figure 12 Impact of MPC-p2 on percentage of switched packets for different cache sizes.

The maximum VC usage is shown in Figure 13. As can be seen from the figure, the 100K cache imposes more VC usage than the 10K cache. For example, the 10K cache imposes about 5K of VCs while the 100K cache imposes about 35K of VCs when **MPC-p2** is equal to 20. This result indicates that a bigger cache size is a disadvantage in terms of VC usage requirement.

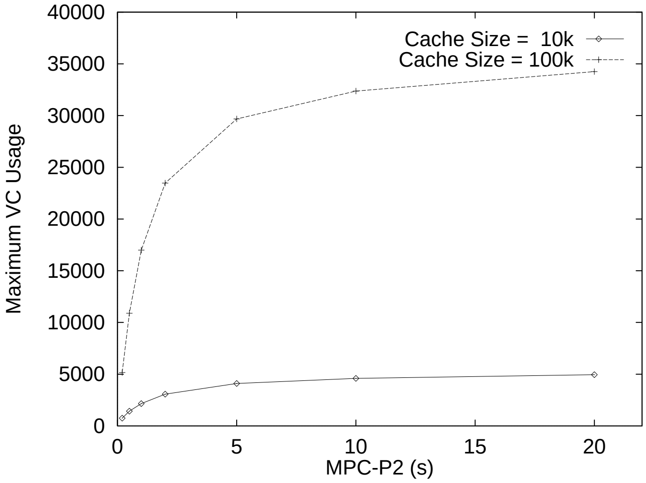

Figure 13 Impact of MPC-p2 on maximum VC usage for different cache sizes.

A more complete picture on the effect of the cache size is illustrated in Figure 14. Observe that the 10K cache generates more SVC setup rate requirement than the 100K cache, which may seem contradictory at first. To understand this behavior, first note that cache entry replacements occur more frequently with the 10K cache than the 100K cache. When a cache entry to destination D is being replaced, its associated VCC (if any) has to be torn down. However, if a new flow to the same destination D reappears in the future, a new cache entry to destination D will be recreated and its associated VCC will be established. This implies that smaller caches tend to tear down and setup SVCs for the same destination more frequently.

From Figure 12 and Figure 14, it is better to have a bigger cache. However, from Figure 13, it is more advantageous to have a smaller cache. This three-way relationship can be used to dimension the cache size to trade-off the percentage of switched packets, the VC usage, and the SVC setup rate.

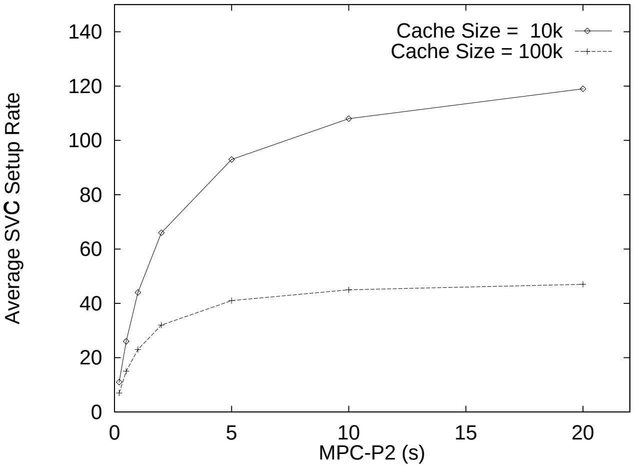

Figure 14 Impact of MPC-p2 on average SVC setup rate for different cache sizes.

Suppose that a service provider would like to switch about 70 percent of the packets (the other 30 percent would be forwarded through legacy routers) by employing an appropriate flow detection policy. From Figure 12, **MPC-p2** should be set to 2 seconds and to 1 second with the 10K cache and 100K cache, respectively. From Figure 13, the VC usage requirements are about 3000 and 11000 with the 10K cache and 100K cache, respectively. From Figure 14, the SVC setup rate requirements are about 70 and 25, respectively. This example shows that for a given percentage of switched packets, one can utilize a simpler two-way trade-off between the VC usage and SVC setup rate by dimensioning the cache size.

4.2 Impact of Packet Arrival Rate

The discussion so far used the original traffic trace which generates packets at the rate of 37.62 Mbps to the ingress MPC. It is of interest to see the effect of the packet arrival rate on the performance measures of interest. Unfortunately, there is no real data available to date that provides a higher speed traffic trace. In order to get around this problem, we generate higher speed traffic traces based on the original trace. To this end, we will generate two traffic models from the original trace.

In traffic model 1, we simply divide the arrival time of each packet by a constant factor called the *speedup* factor. If the packet arrival times of the original trace are $\{t_1, t_2, \ldots, t_n\}$, and the speedup factor is *S*, then the new packet arrival times of the new trace of rate *S* times the original rate are $\{t_1/S, t_2/S, \ldots, t_n/S\}$. Since the traffic trace is composed of an aggregate of many sources, traffic model 1 essentially assumes that each source increases its generation rate by *S*.

In traffic model 2, we assume that each source maintains its original rate. The speedup factor is emulated by multiplexing the *S* original traces, assuming that S is an integer. Since we only have one trace, we have to break the original trace into multiple subtraces. Suppose we want to multiplex two traces. We first break the original trace into two equal subtraces: {t_1, t_2, …, $t_{n/2}$} and {$t_{n/2+1}$, $t_{n/2+2}$, …, t_n}. This approach ensures that the correlation between the two subtraces will be minimized. Next we need to adjust the arrival times of the second subtrace so that both subtraces are being multiplexed during the same time period. One way is to subtract T_a from the arrival time of each packet in the second subtrace so that the first packet arrives near t_1. The resulting trace may be further broken into two subtraces if a speedup of four is desired. The number of breakups may be limited to a few times before the subtraces become correlated.

Figure 15 shows the impact of the packet arrival rate on the percentage of switched packets. It is assumed that the MPC cache size is 10K and P_{HIT} = 0. First consider traffic model 1. Note that we obtain a higher percentage of switched packets as we increase the packet arrival rate. As expected, traffic model 1 reduces the flow duration making the MPC more quickly to decide on shortcuts. This in turn tends to trigger more shortcuts as the packet arrival rate increases. An interesting point occurs between 150 Mbps and 300 Mbps where the percentage of switched packets actually decreases as the arrival rate is increased. The phenomenon is called *trashing* which is caused by too many cache entries being replaced before the arrival of MPOA resolution reply.

Now consider traffic model 2 which keeps the original flow duration the same. With infinite cache size, the percentage of switched packets would be roughly maintained at the same value as the arrival rate is increased. With 10K cache, however, trashing occurs much earlier with traffic model 2 because the number of flows per unit of time increases as the packet arrival rate is increased. The net effect that we see from the figure is that the percentage of switched packets decreases earlier with traffic model 2.

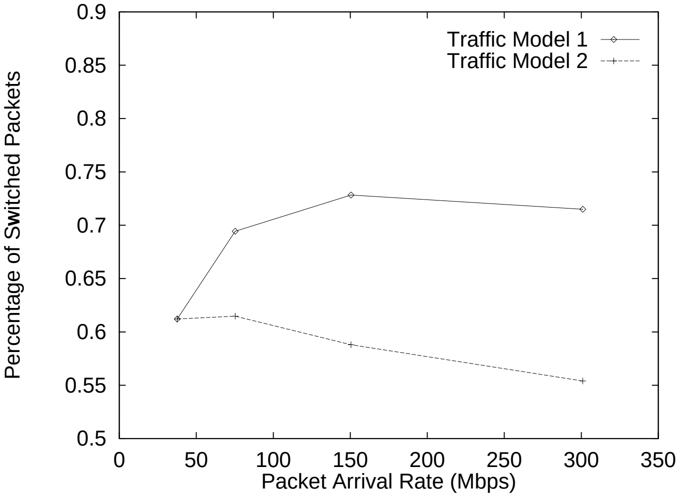

Figure 15 Impact of arrival rate on percentage of switched packets for different traffic models.

Figure 16 shows the same scenario for the maximum VC usage requirement. Cache entry replacements occur more frequently with traffic model 2 since it generates more flows per unit time. As a result, there are more cache entries with traffic model 2 that do not yet have the opportunity to establish the SVC before they are being replaced by new entries. Thus, traffic model 2 requires a lower VC usage requirement.

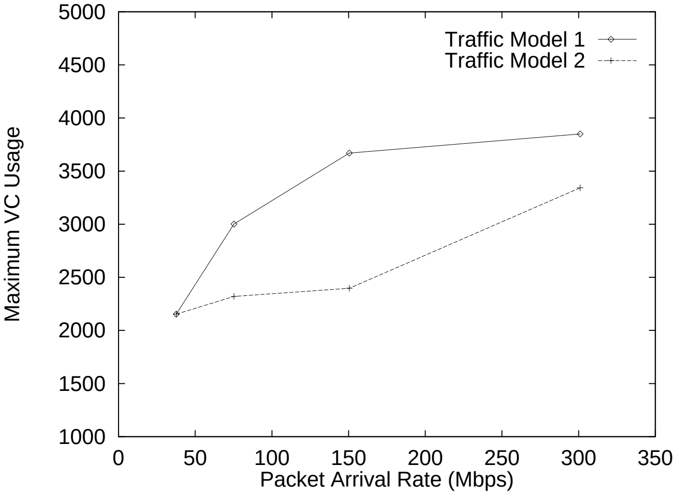

Figure 16 Impact of arrival rate on maximum VC usage for different traffic models.

We next display the SVC setup rate in Figure 17 to understand its dependence on the arrival rate. The figure reveals that the SVC setup rate is linearly dependent on the packet arrival rate for both models. This observation seriously questions the scalability of MPOA for the core networks where the arrival rates tend to be high. The case with 100K cache also exhibits similar behavior in terms of packet arrival rate, except that trashing does not occur because the cache never becomes full.

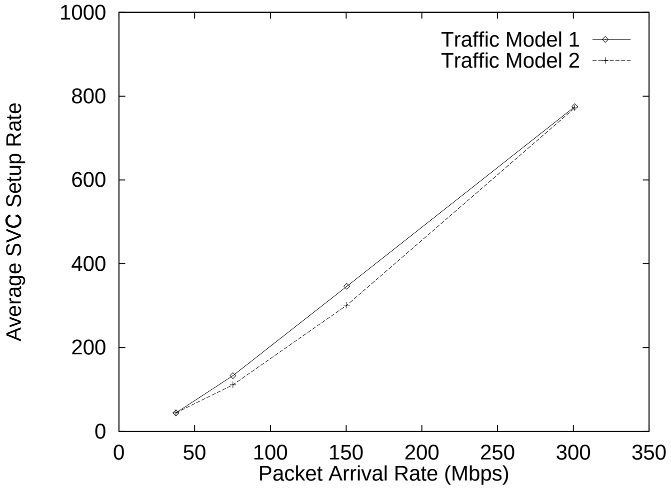

Figure 17 Impact of arrival rate on average SVC setup rate for different traffic models.

4.3 Cache Behavior

We now examine the ingress MPC cache behavior in more detail. We plot the cache utilization as a function of time in Figure 18 for two values of cache sizes. The original traffic trace is employed in this simulation. The cache utilization is defined as the percentage of cache entries being used out of the entire space. In this definition, it is important to note that a cache entry is being used as soon as a new packet arrives at the ingress MPC. Thus, a cache entry can be used even before the shortcut is established. Because of the way MPOA reserves a cache entry, the cache utilization is independent of the flow detection policy.

From Figure 18, we note that cache utilization approaches 100 percent quickly when the cache size is 10K. When the cache size is increased to 100K, we observe that many cache entries are empty. However, it is interesting to note that the cache utilization increases as a function of time as new destination IP addresses are being encountered. If longer traffic traces are available, it is possible that the cache utilization may eventually reach 100 percent. However, if shorter Holding Time is used, one may keep the cache utilization below a certain value by invalidating an entry more quickly.

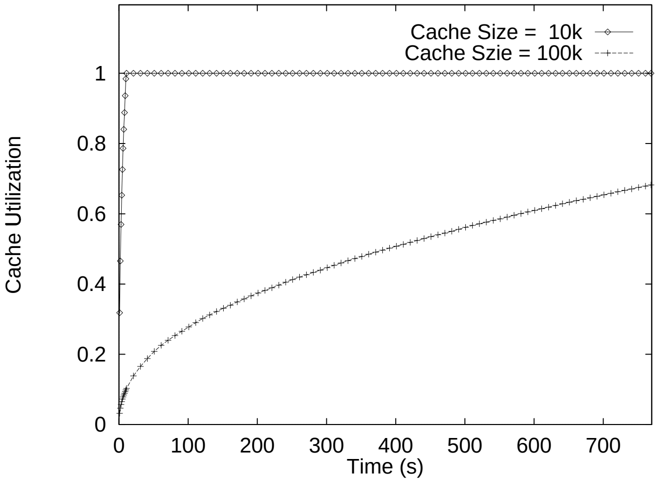

Figure 18 Cache utilization for different cache sizes.

When a cache becomes full, it becomes necessary to replace the information in a given entry with another. We define the cache replacement rate as the average number of cache entries being replaced per second. Figure 19 shows the cache replacement rate versus the packet arrival rate for the case where the cache size is 10K. The case with 100K cache never fills up using the available traffic trace. Recall that trashing occurs when the percentage of switched packets decreases as the arrival rate increases. From Figure 15, trashing begins to take into effect when the arrival rate is between 150 Mbps and 300 Mbps for traffic model 1, and between 75 Mbps and 150 Mbps for traffic model 2. Referring to Figure 19, we can conclude that trashing begins to occur when the cache replacement rate is approximately 2000 entries per second. In general, trashing should be avoided since it reduces the effectiveness of MPOA. It is important to note that trashing can not be solved by modifying the flow detection policy. It can only be solved by increasing the cache size.

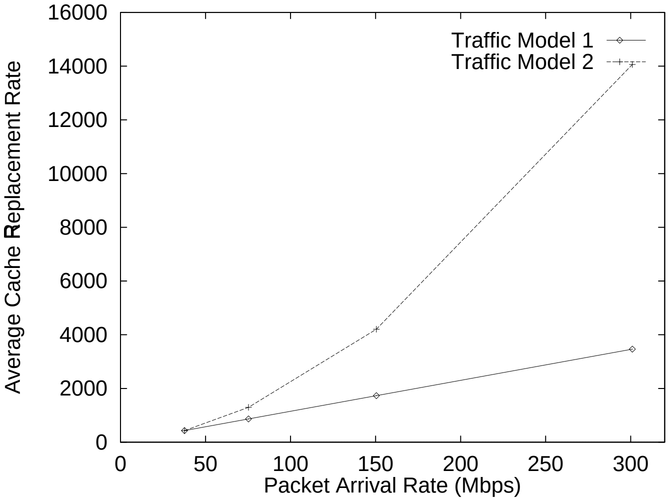

Figure 19 Cache replacement rate.

4.4 Transient Behavior

Most of the discussions so far focussed on the steady-state behavior. In this section, we investigate the effect of the initial startup on the transient behavior of the system in terms of the SVC setup rate. Figure 20 shows the SVC setup rate against time for different cache sizes. The plot with the 10K cache typically generates a higher SVC setup rate. Note that initially none of the SVCs associated with MPOA is established. As a result, the system tends to generate a very high SVC setup rate soon after the startup because each long-lived flow triggers a new SVC. After the system stabilized, the SVC setup rate settles to some equilibrium condition. Because of the high setup requirement during initial startup, MPOA policy may need to be made less aggressive at this period so that the ATM signaling process does not become overly burdened. Note that the SVC setup rate may also change if the traffic process changes significantly.

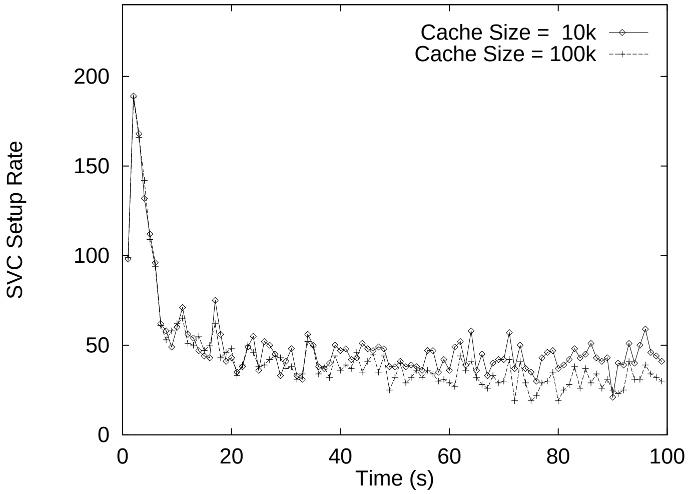

Figure 20 Effect of initial startup on SVC setups.

5 Conclusion

In this paper, we have investigated the scalability of MPOA by focusing on three important performance measures: percentage of switched packets, maximum VC usage, and SVC setup rate. We have also provided useful insight into the MPC cache behavior. A summary of important points that a systems designer should consider include:

- NHRP authoritative replies are recommended since it imposes smaller cache size requirement at the transit routers with minimal penalty on our performance measures (i.e., small dependent on P_{HIT}).
- Bigger cache size at the ingress MPC results in more switched packets.
- Bigger cache size at the ingress MPC imposes more VCs but less SVC setup rate.
- Ingress MPC cache trashing is a function of packet and flow arrival processes and is independent of the flow detection policy.
- SVC setup rate is linearly dependent on the packet arrival rate.
- Initial startup may cause a much higher SVC setup rate.

The most basic question on the applicability of MPOA for the core network can be answered by the fact that the SVC setup rate is linearly dependent on the packet arrival rate. This result suggests that MPOA with flow-driven SVCs will not scale very well in the core network supporting public Internet service. There are several approaches to mitigate the scalability problem, some of which are listed below.

There are several directions one can extend the current work. Some directions for further work include:

- Effect of Holding Time
- MPOA with QoS extension
- MPOA for long duration flows (e.g., IP telephony type of flow)
- Effect of flow aggregation
- MPOA for Virtual Private Network (VPN) support only
- Adaptive resource (e.g., signaling power, VC space, etc.) management

Reference

[1] Network Wizard, http://www.nw.com

[2] V. Cerf "Looking Beyond the Millineum", Keynote speech in INFOCOM'97, Kobe, Japan, April 1997.

[3] P. Newman et al, "IP Switching: ATM Under IP", accepted for publication in IEEE/ACM transactions on Networking, 1997.

[4] R. Callon et al, "A Framework for Multiprotocol Label Switching," work in progress, Internet Draft <draft-ietf-mpls-framework-01.txt>, July 1997.

[5] E. Rosen et. al., "A Proposed Architecture for MPLS," work in progress, Internet Draft <draft-rose-mpls-arch-00.txt>.

[6] M. Laubach, "Classical IP and ARP over ATM," RFC 1577, Jan 1994.

[7] Multi-Protocol Over ATM Version 1 Specification, the ATM Forum Technical Specification, July 1997.

[8] A. Fredette, "An Introduction to MultiProtocol Over ATM (MPOA)," Proceedings of SPIE Broadband Networking Technologies, Dallas, Vol. 3233, pp. 176-183, nov 1997.

[9] LAN Emulation Over ATM Version 2 Specification, the ATM Forum Technical Specification, July 1997.

[10] J. Luciani et. al., "NBMA Next Hop Resolution Protocol (NHRP)", RFC 2332, April 1998.

[11] S. Lin and N. Mc Keown, "A Simulation Study of IP Switching," Proceedings of SIGCOM'98, Cannes, France, Sep 1997.