SIDSP: SIMPLE INTER-DOMAIN QOS SIGNALING PROTOCOL

Paulo Pinto, António Santos, Pedro Amaral, and Luis Bernardo
Universidade Nova de Lisboa
Monte de Caparica, Portugal

ABSTRACT

This paper presents a study on a scalable inter-domain signaling protocol to provide quality of service to the Internet. The protocol uses a hard-state approach and separates data forwarding procedures, control procedures, and resource allocation procedures. Current solutions adopt a soft-state label switching approach where border routers are responsible for maintaining resource accounting and the path's next-hop information for every aggregated flow. Due to the Internet's almost hierarchical structure, the routers at the core can become a bottleneck to the system's overall scalability. Our protocol, SIDSP, reduces the core routers complexity by transferring most of the state data to the peripheral routers. This is accomplished by the usage of source routing for the flows. As a hard-state approach, it does not require periodical refreshing messages. However it puts more pressure on state coherence and state stabilization algorithms. SIDSP's performance was tested using ns-2 simulations, and was compared to BGRP's performance.

I. INTRODUCTION

The provision of Quality of Service (QoS) in the Internet has been based mainly on the usage of algorithms to differentiate the traffic, and signaling mechanisms to inform and reserve resources along the path [1]. Its deployment in real networks raises serious scalability problems: the amount of state information in routers along the path; and the weight of the signaling messages. This weight is felt both in terms of message exchange (to establish and maintain the path) and in terms of computational power (to process them). We assume in this paper that a two-tier approach to QoS signaling is used. I.e., there is an intra-domain QoS mechanism and an inter-domain QoS mechanism amongst domains. The paper focuses on the latter one.

The Internet is a concatenation of administratively and technologically different domains (Autonomous Systems – AS). Over the forty years of the Internet, the commercial relationships created an almost perfect hierarchy [2]. At the top of the hierarchy (*tier-1*) we find a backbone formed by transit ASes associated with the largest ISPs, interconnected in almost a full mesh. They compose the Internet dense core, where the scalability problems are mainly concentrated. On the lower layers we find [2] two national and regional transit ASes' layers (*tier-2* and *tier-3*), which interconnect the stub ASes.

The generalized use of multimedia services over the Internet (e.g. Voice of IP (VoIP), video broadcasting, etc.) has raised the importance of providing Quality of Service (QoS) on an end-to-end basis. Interestingly, recent results [3] show that the slow convergence of BGP (Border Gateway Protocol – the inter-domain routing protocol) after an update is responsible for 50% of the low quality best-effort VoIP connections. Unfortunately, most of the current inter-domain QoS mechanisms may also suffer from BGP limitations. Resource reservations must adapt to the topology changes. However, most of the times, during a large interval BGP [4] is unable to provide a new stable route, jeopardizing the reliability of virtual paths covering multiple ASes.

This paper makes two main contributions: (a) it assesses the costs and consequences of moving state information from the Internet's core routers to the lower layer AS's routers, which includes (b) the usage of a hard-state approach that reduces the dependency on BGP. One of our purposes is to assess the "costs" and consequences of these choices in order to consider them in the future.

The paper starts with an overview of inter-domain reservation protocols highlighting some of their drawbacks. Section 3 describes our system, and is followed by a section presenting some simulation results. Section 5 evaluates the systems and the final section presents our conclusions.

II. INTER-DOMAIN RESERVATION PROTOCOLS

Inter-domain reservation protocols use the routers of the network in two different ways. Certain routers, that we call here border routers (BRs), understand the inter-domain signaling messages (PROBE, RESV, etc.) and know exactly the flows passing through them. Other routers, the internal ones (IR), do not care about the signaling messages and differentiate the traffic per class or intra-AS flow. These latter ones need a different signaling mechanism to establish reservations, an intra-domain QoS signaling protocol. Examples of internal routers are the inner routers of a DiffServ domain. Examples of BRs are the edge routers of the same domain. Fig. 1 shows a two-tier QoS signaling architecture. The inter-domain QoS signaling protocol runs on the BRs along the path that connects the end hosts. BRs maintain the information about the inter-domain flows, and use the intra-domain protocol to allocate resources to the connections within the ASes.

A notorious example of a reservation protocol is RSVP (*Resource reSerVation Protocol*). RSVP keeps individual information per flow in routers in either its base form [5] or its aggregated form [6]. Therefore, it does not scale due to the information space required on the core ASes.

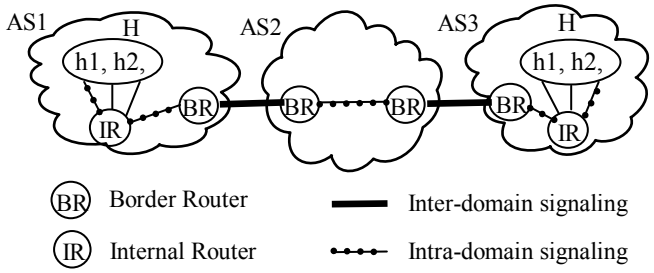

Figure 1: Two-tier QoS Signaling

Other three inter-domain signaling protocols use aggregation as the mechanism to reduce state information. They are: BGRP (*Border Gateway Reservation Protocol*); SICAP (*Shared-segment based Inter-domain Control Aggregation Protocol*); and DARIS (*Dynamic Aggregation of Reservations for Internet Services*).

2.1. BGRP

The BGRP [7] aggregates traffic using a sink-tree aggregation approach. Reservations from different domains destined to a certain domain are aggregated along the path, defining a sink tree rooted at the destination domain BR. BGRP flows are established by sending a PROBE message to the destination and getting a GRAFT message back. Resources are only reserved when the GRAFT message passes through the path discovered by the PROBE message. This option is potentially dangerous in loaded networks because resources can vanish before the GRAFT is received. ERROR messages are used to signal the allocation failure.

Each BR maintains a table with all the active allocations, indexed by the tree label ID. Each entry includes the next-hop BR address, and a list of previous-hop BR addresses and their branch aggregated reservations. Therefore, the table grows linearly with the number of destination BRs.

Aggregations start to be built when a new flow is accepted to an existing destination AS (e.g. BR S3 in Fig. 2 is responsible for the creation). The GRAFT message for the second flow contains the same tree ID as the aggregation to which it should be joined. BRs along the way (e.g. R5 and R4) add the bandwidth requirements to the one they already have and provide that enough resources are allocated inside their ASes using the intra-domain protocols. At a certain point in the tree (R3 in Fig. 2) a new branch of the tree is created for the second flow.

BGRP uses "soft-state". Flows are maintained within the aggregate with the use of periodic REFRESH messages per aggregate. Individual refreshes must also exist from the source until the aggregating BR. Flows are removed either explicitly by sending a TEAR message or implicitly by letting the state disappear by not including it in the REFRESH message.

The authors propose certain enhancements to reduce the signaling overhead. A first one is over-reservation performed by leaf BRs when they send the PROBE messages (they request more bandwidth than they really need). In result, intra-domain reservations are skipped in all traversed domains when a new flow is established exactly from the same leaf node to the same destination node. The authors also suggest the use of over-reservation by other BRs in the path, to support quiet grafting. By labeling the trees with the address prefix of the destination AS, BRs are able to identify trees when the PROBE arrives. This way the first BR in the tree to be contacted could answer the PROBE immediately with a GRAFT without the need to send the PROBE until the root. However, the generalized use of over-provisioning by all BRs on all trees in high load conditions can lead to false network resource exhaustions. Without over-provisioning, the PROBE and GRAFT always have to travel end-to-end.

BGRP uses BGP (and possible bilateral QoS agreements) to route the PROBE messages during flow setup, and after existing tree route changes. In order to handle BGP instability, it proposes delaying BGRP reservations changes in response to route changes. Unstable routes are delayed longer than previous stable routes. However, only bilateral agreements can provide a fast answer to a path failure because BGP exhibits the slow convergence problem [4].

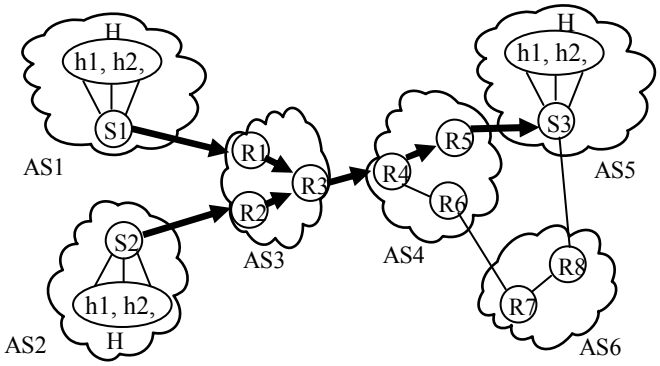

Figure 2: Example of a sink tree rooted at S3

2.2. SICAP

SICAP [8] combines the shared segment and the tree-based aggregation approaches. If some part of the path is identical, even if the destination domain is not, flows can be aggregated along that part. Certain domain BRs (called Intermediate De-aggregation Locations - IDL) will de-aggregate flows and not just aggregate.

Finding the best IDLs along a path can be a difficult task. SICAP's proposes to elect the BRs in an AS with the highest number of neighbor ASes based on the assumption

that it is more likely that traffic will be spread over different paths at these ASes. In reasonably loaded networks SICAP's way of aggregating tends to attract flows to aggregations that are already in place for parts of the path, establishing in practice virtual trunks. These virtual trunks have no relation to ASes, i.e., they might cover more than one AS. Therefore, traffic engineering techniques are harder to implement because they may require inter-AS state information modification on new BRs, if the path for the aggregation is modified.

SICAP is also sender initiated with a message REQ probing the network and establishing a path without leaving any state. The last de-aggregator (generally at the destination AS) answers with a message RESV containing the identification of the aggregate and reserving the resources. The path that is established is in fact the concatenation of potential aggregations between the various IDLs that were identified (the message REQ contains the number of neighbor ASes for each AS in the route record list). So, the identifier of the aggregation in the RESV message is reset at intermediate aggregators and set again with the proper number. SICAP also defines a bundle REFRESH to maintain soft state for a set of flows.

SICAP tries to minimize the total number of aggregates in use and consequently minimize global state, compared to BGRP. This is true only if the total number of aggregations is smaller (the global state can also be smaller). It must be noted that IDLs have more state in the network compared to the BGRP approach of having only de-aggregators at the tree roots. As the BRs on the higher hierarchical layers of the Internet are the main IDL candidates, state is increased on the routers where more scaling restrictions exist.

2.3. DARIS

DARIS [9] aggregates flows between two arbitrary domains if it founds that there are more than k flows between them. DARIS uses a logically centralized agent/manager that controls the resources of a whole domain – DSDM (*Differentiated Services Domain Manager*). It has knowledge of the internal topology, resource capacities and current selected routes (obtained using BGP). A graph of the whole Internet is then created in each DSDM with the record of every established flow and its path (for flows that begin or traverse its domain). When a new flow request arrives the graph is checked to see if there are other flows downstream. If at least k flows exist, and the common path has at least two hops, a new aggregation is created. Over-reservation is used to reduce signaling overhead. Once the aggregation is formed the state of the individual flows is erased. So, these BRs switch from active to passive for these flows. However, it is not said how flows cease to exist and the way they do will have consequences on the state that must be preserved. If a soft-state approach is used then the individual states cannot be deleted and these BRs cannot become passive. If a hard-state approach is used, some component (e.g. DSDM) must maintain this information. Therefore, too much centralization is required in DARIS, jeopardizing the system's overall scalability. A second pitfall can result from basing its aggregation decisions on the Internet topology received from BGP, especially after network failures due to the slow convergence problems.

III. SIDSP

In the SIDSP system, Simple Inter-Domain QoS Signaling Protocol, three aspects are handled separately: data forwarding procedures, control procedures (establishment, tear down and self-healing of flows), and resource allocation. The active routers are the egress and ingress BRs in ASes and the end hosts (or their routers on their behalf). Each ingress BR in an AS has a MPLS [10] label to every egress BR in the same AS to identify a virtual trunk. Considering C classes of QoS and N BRs in an AS this will lead to (N*(N-1)*C) trunk identifiers.

The flow path is then defined by the sequence of ASes and virtual trunks crossed, assuring the QoS end-to-end. Each data packet carries the route on a shim header, avoiding the need for (aggregated) flow information in the core BR. Instead, a core BR only needs to store bandwidth allocation information. All route information is transferred to the flow's endpoint BRs.

The control procedures are the major part of SIDSP. Fig. 3 shows a typical topology with some flows. Some of the trunks between BRs are drawn and are identified by their MPLS numbers. For the purpose of the description let´s consider the following flows: A-B; A-D; B-D; and B-C. Note, for instance, that A-D and B-D use the same trunk in AS g, but it is not an aggregation in the way defined in the three systems above. They are simply source routed through the same virtual trunk. Resources are allocated to the trunk, and are shared by all the flows using the trunk. A hard-state approach is used to manage resources, avoiding reservation refreshment overhead. Self-healing mechanisms are proposed to handle possible trunk resource stale allocations, due to not having flow information on the BRs.

3.1. Establishing a Flow and Resource Management

SIDSP flows are established by sending a PATH message to the destination and getting a RESV message back. Let's look at the establishment of the flow B-D in Fig. 3 and assume that the other three flows are in use and also that resources are available throughout the path. Host B uses a PATH message to establish the flow. It will establish a path from B to the egress BR of b. This is an intra-domain procedure and it is outside of the scope of this paper. When the PATH arrives at the egress BR there is a decision to

forward it through AS *d*, based on BGP routing tables and bilateral QoS agreements with adjacent BRs. The egress BR forwards the PATH message to the ingress BR of *d* and pre-allocates the resources. Both BRs have to increase the used bandwidth of the path by the demanded one of the flow. The ingress BRs have an AS table where they store the bandwidth allocated per BR of the stub ASes – this is the only state information in the core BRs (and it will be useful to support multiple failures of stub ASes BRs). The ingress BR appends its identification and the MPLS label for that flow to the message and forwards it directly to the egress BR of its domain. The same will happen between egress BR of *d* and ingress BR of *e* and between *e* and *g*. Each BR appends its identification, as well as the identifiers of the MPLS trunks that the flow will use. When the PATH message arrives at the ingress BR of *i*, the BR stores the complete path and identifiers in a flow table and the message is sent to the host (or its router) with the identification of the MPLS label inside *i*. A RESV message is sent back confirming the allocation and containing the whole information about BRs and MPLS labels. All resources pre-allocated are turned permanent. When it arrives at the flow egress BR (at *b*) a record of this flow is also stored in a flow table. The RESV message ends at the host (or its router). The flow is established and if the resources exist all the way as assumed the establishment time is very small. At this moment the host (or its router) has a sequence of MPLS labels that it must stack on top of each other for each packet it sends. Fig. 4 illustrates this stacking for the B-D flow of Fig. 3 (fields such as TTL, the path index pointer, or addresses were omitted). If the trunks are bidirectional the receiver must do the same thing for the packets that it sends. At each AS egress BR on the path the stack is popped and the label pointed by the path index pointer identifies the next trunk in the following AS. Hence, the routing information is carried in the packet and the active BRs do not have any state information per flow.

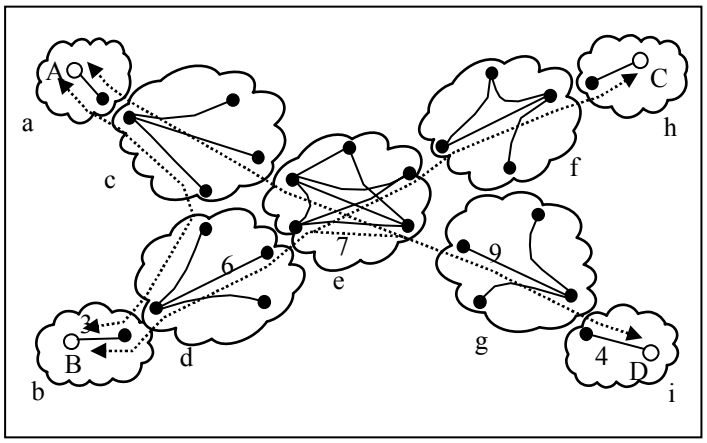

Figure 3. Multi-AS connections and the associated paths

Resource reservation during path setup is a heavy operation. All three works described previously tried to use the mechanism of overprovision an aggregation to easily integrate a new flow in the future. However, resources cannot be shared by two different aggregations because they are associated to individual aggregations. In SIDSP resources are associated to virtual trunks and to links connecting them at neighbor ASes, which can be shared by several flows. Therefore, we can reason about resources reservation in an independent way of the flow establishment. Of course, if resources are needed when an establishment is being set up the algorithm has to be run at that moment and the establishment is slightly delayed. But the algorithm can be run in advance in less critical times (not at the moments of flow establishment) keeping the trunk resources above the current needs in order to speed up the establishment. A prediction algorithm based on the past history can even be used (e.g. seasonal pattern of users throughout the day).

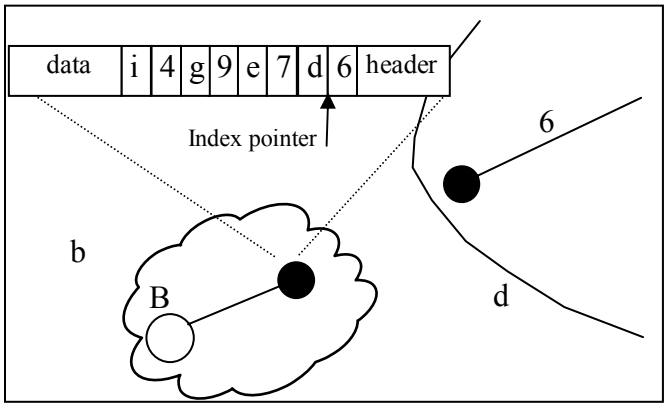

Figure 4. Data forwarding in SIDSP

When an ingress BR finds out that it has no resources and an attempt to allocate more fails, it sends back an ERROR message to the egress BR. This BR can then choose an alternative AS at the proper level if it is connected to others. If it is not connected to others (or all attempts have failed) then the ERROR message is sent back to the ingress BR of its AS. The ERROR message can arrive at the source router provoking the failure of the establishment. Obviously, the precise recover algorithm must converge and because it is not the main topic of this paper we will not enter into more details.

An ARQ (automatic repeat request) protocol is used to send control messages between BRs to avoid incoherent states. Each message carries a unique identifier, allowing multiple retransmissions. However, stale reservations can still occur due to BR failures or long lasting link failures during message processing. This failures are detected when a RESV (or an ERROR) message is not received in response to a PATH. A state enforcing message called BANDW (presented below) is used in these (and other) circumstances.

3.2. Tear Down of Flows and State Information

Flows have to be torn down explicitly because there is no information per flow in the active BRs of the core of the network. The TEAR message carries the traffic parameters of the flow to free resources along the path, as well as the path identification including the identification of the egress BR of the stub AS, the MPLS labels and ASes used. The TEAR messages are acknowledged and reliably sent using an ARQ algorithm. Therefore, SIDSP uses a hard state approach.

Each egress and its adjacent ingress BRs (it can have more than one) know exactly how much bandwidth the egress BR is using for every trunk. When a flow is accepted they both increment the overall bandwidth according to a common algorithm based on the traffic parameters of the PATH/RESV messages. When a flow is torn down they do the inverse procedure.

The state information maintained by BRs are the values of bandwidth per trunk they are using (and this will be object of billing between ASes), and the bandwidth per stub AS's BR (and not per flow, making it more scalable). Only the egress BRs of the stub ASes have complete information of the flows initiated and terminated at their ASes. This information contains the sequence of ASes and the corresponding MPLS labels together with the traffic parameters and the identification of the BR of the stub AS initiating the flow. The state information is logged into stable storage to be resilient to failures.

To manage the possible problems associated to hard state environments SIDSP defines an additional mechanism to handle stale resource allocation and assure state coherence at the core BRs: egress BRs in stub ASes can send a BANDW messages multicast along the spanning tree defined by all flows starting on that BR, to update the used bandwidth entry on each core BR onto a consistent view. The BANDW message contains the summation of bandwidth per AS and a timestamp. The summation is used to define the multicast tree and the bandwidth on each AS, and the timestamp synchronizes this message with PATH/RESV messages. Therefore, any incoherence on the per-stub AS's BR resource allocation table can be corrected using BANDW messages.

Stale resource allocation has mainly a billing consequence because resources are shared amongst the flows of that class.

3.3. Traffic Engineering and Self-healing

Inside a domain the AS is free to move trunks from a certain route to another for reasons of traffic engineering or for recovering from link failures. From the point of view of inter domain signaling this is transparent and the MPLS label has to be maintained. Packets can be lost in the process but this is a concern for the applications.

When a certain domain cannot recover from a failure the whole trunk will fail. This is considered a serious fault. The BRs at each side of this trunk inform the adjacent BRs with a TRUNKFAILURE message and freeze the MPLS label identifying the link for a time T in the domain to avoid misrouting of packets and to let the network stabilize. During this time the BRs related to the failed trunk discard the packets they receive. The TRUNKFAILURE message contains the MPLS label and the AS identifier, and a list of BRs at the stub ASes that were using the trunk. This message is sent upstream and downstream for informational purposes. Each egress BRs of the stub ASes checks if it has flows passing through the failed trunk. If it has not, it discards the message silently. If it has, it stops forwarding packets that pass through that trunk; sends a TEAR message until the AS of the failed trunk (to free the resources in between); and still does one of two things: tries to reconnect again following the usual procedures; or sends a TEAR message to its host to disconnect the flow. The choice of the alternatives is done by the BR belonging to the connection initiating AS.

Multiple failures can break the path between a stub BR and some BRs along the path leaving stale reservations on inner BRs. Repetitions of TEAR messages directed to specific BRs that did not respond can help to solve the problem.

A more serious problem can occur in result of the simultaneous failure of two (or more) egress stub BRs (because they hold the full flow information between their ASes). When the BR comes alive again it must perform a consistency check with inner routers and ingress BRs of adjacent ASes to evaluate the state of the flows he has initiated (they are kept in stable storage). It then builds a BANDW message and sends it over the tree. Each core BR updates the amount of bandwidth of that AS with the value in the message. The main consequence of this kind of failure is to have unused resources reserved at the core BRs, and extra billing between ASes. A heartbeat packet exchange is used to detect adjacent BR failure. If an adjacent BR does not recover after a configurable threshold time, each adjacent BR will start the above mentioned trunk failure procedures for each active trunk to the failed BR.

Therefore, with these “heavier” recovery mechanisms all stale reservations can be cleared without storing any flow information at the core BRs.

3.4. Data Forwarding and Traffic Verification

The data forwarding part scales very well because each packet has the complete routing information and there is no per-flow state in the core BRs. The route in each packet is verified by the stub ASes that send the packet. Egress BRs of stub ASes can check if the stack of labels in the packet is consistent with the flow characteristics, can possibly even

be the ones to create the stack from a local identifier in the packet, or can delegate to the inner router depending on the scale of the number of flows. The unique period when this check is compulsory at the egress BRs is after receiving a message TRUNKFAILURE.

Policing in SIDSP is straightforward. The paths between egress BRs of a domain and ingress BRs of adjacent domains can have devices to police traffic based on class of service (virtual trunk identifiers). This measure scales well and it is the only necessary measure to perform billing between domains and check for the compliance of agreements between ASes.

IV. SIMULATIONS

We implemented the SIDSP and BGRP protocols on the ns-2 [11] simulator version 2.30. We tested their performance using the topology originally proposed in [7] to compare BGRP and RSVP. The topology models the progression of domains along an Internet path with a length of ten ASes, with tier-3 and tier-2 access networks at the beginning and at the end of the path, and backbone (tier-1) ASes in the middle of the topology. We used the same conditions as [7] with 100 source and sink stub AS BRs, where every source connects to every sink. Three different demand distributions were tested: *Flat topology*, with 5 sources and 5 sinks at every AS of the path; *Hierarchical topology*, with 11 sources and sinks at the four central (backbone) ASes in the path and 1 source and sink at the other ASes; and *Selected source topology*, with all sources equal to 1, and all sinks equal to 9 (models the download from centralized services). Fig. 5 shows the first four ASes in the path for the flat topology. We modeled each AS in the path with one input BR (nodes 0 to 9), one output BR (nodes 10 to 19), and one or two transit BRs (nodes 20 to 37) that inter-connect ASes on the path.

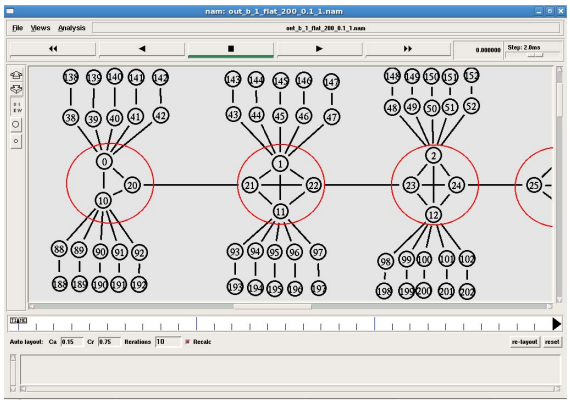

Figure 5. Flat topology simulated

[7] used a single node per AS and ignored the intra-AS paths. Each AS includes a varying number of external source stub AS BRs (nodes 38 to 87) connected to data sources (nodes 138-187), and external sink stub AS BRs (nodes 88 to 137) connected to data sinks (nodes 188 to 237).

Fig. 6 shows the number of reservation entries stored in each of the backbone BRs when all flows are setup for the three topologies. It shows the effect of reserving bandwidth per flow aggregation (BGRP) and per trunk (SIDSP) in the backbone. Fig. 7 shows the total number of bytes stored in the tables of the same BRs. For each AS the sum of the memory used in the right and left transit nodes is constant (e.g. nodes 21 and 22), but due to the different number of nodes at each side of the node, Fig. 7 shows an irregular pattern. Results confirm that the state in the backbone BRs is minimal compared to BGRP because no flow information is saved. That information is stored only on the stub BRs. The trade-off is the increase in the header size and thus on the bandwidth. But, as Fig. 8 shows, the larger the packet is the less relevant this overhead becomes.

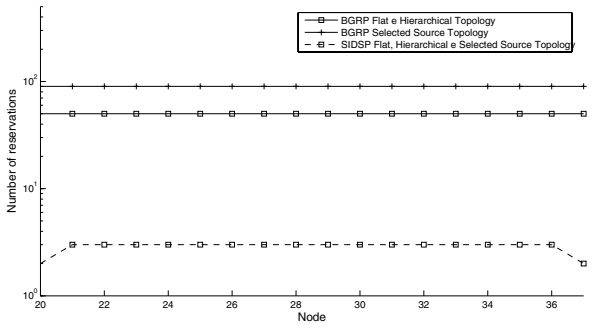

Figure 6. Number of reservations in core BRs

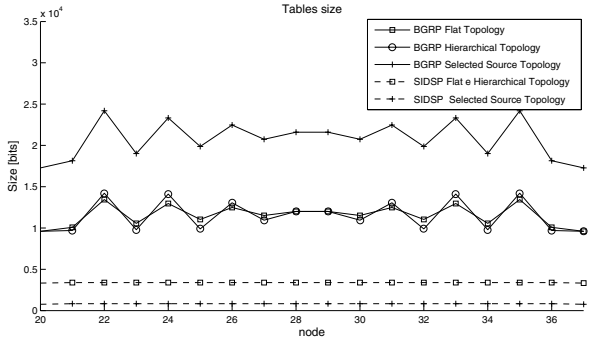

Figure 7. Size of the tables stored in core BRs

V. EVALUATION

Providing the Internet with QoS means that state information must exist somewhere. It is consensual for a long time that the information must be at the borders of the network due to scaling considerations. Our system does exactly that. One of the characteristics of packet networks (and the Internet) is their robustness in face of failures. Whether the healing mechanisms are performed at the core of the network or at the BRs of the stub ASes is irrelevant. Modern networks (e.g. MPLS) are getting simpler and

pushing the complexity to the borders. The same must happen with the Internet.

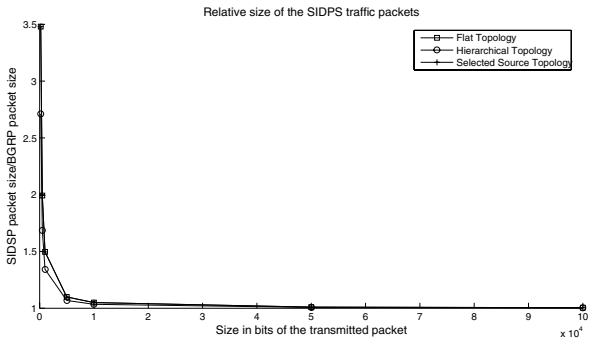

Figure 8. Packet size overhead

In our system trunks cannot traverse ASes. This is a major issue. If trunks could span several ASes then a management algorithm would have to be devised to perform several tasks: negotiation, recover, etc. Messages would have to be exchanged and processed and there would also be overlaps with the BGP protocol. By forcing trunks to be internal to ASes no knowledge of the overall network is necessary, neither is the definition of entities responsible to manage resources at these levels. Everything is contained inside the structure that already exists – the AS.

The BGP slow convergence and flow path definition problems can be reduced if the Internet's hierarchical structure is taken into account [12]. By extending BGP to take into account the hierarchical structure of bilateral QoS agreements connecting stub, tier-3, tier-2, and tier-1 ASes (and not just flat QoS agreements as in [13]), it would be possible to properly setup a flows' path, even during BGP instability periods. The hierarchy would reduce the overall problem, and define default routes resilient to instability.

Decoupling resource management from call control establishment is also an interesting feature that simplifies the overall problem. The resource management is turned into a local feature inside an AS and can be managed depending on the current circumstances. Differentiated services [14] are flexible enough to use the reserved but unused bandwidth of the virtual trunks by lower classes of service taking the most profit of the network.

VI. CONCLUSIONS AND FURTHER WORK

The system presented is very simple and can scale well in the current Internet. There is a clear distinction of tasks between the intra-domain and the inter-domain parts and the interface is clear. The hard-state approach is the main mechanism for state information reduction. It pushes the Internet into a more connection-oriented philosophy but guaranteeing QoS has always this effect. The paper discussed also the effect of failures on the state maintenance. In our opinion they are not dramatic and can be feasible in the real Internet.

The use of virtual trunks to interconnect ASes assuming the hierarchical structure that the Internet has become can be the beginning of a more stable part of the network suited for real-time services. It might be not so dependent on the Inter Domain Routing Protocol and consequently will stay more static and rigid but also more reliable to transient problems. In terms of resource usage there is no real waste because virtual trunks can always be used by other classes of traffic when there is availability.

Subjects of further work include the study of the requirements for the Border Gateway Protocol to be better adapted to the almost hierarchical structure of the Internet and to take into account QoS requirements, the definition of an intra-domain signaling protocol to work with SIDSP, and the interaction with end hosts.

REFERENCES

[1] D. Vali, S. Paskalis, L. Merakos, and A. Kaloxylos, "A survey of internet QoS signaling," *IEEE Communications Surveys & Tutorials*, vol. 6, no. 4, pp. 32-43, 4th Q. 2004.

[2] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, "Characterizing the Internet Hierarchy from Multiple Vantage Points," in *Proc. IEEE INFOCOM'02*, vol.2 pp. 618-627, Jun. 2002.

[3] N. Kushman, S. Kandula, and D. Katabi, "Can you hear me now?! It must be BGP," *ACM SIGCOMM Computer Communication Review* vol. 37 no. 2, pp. 75-84, Apr. 2007.

[4] M. Yannuzzi, X. Masip-Bruin, O. Bonaventure, "Open Issues in Interdomain Routing: A Survey," *IEEE Network* vol.19 no.6, pp. 49-56, Nov.-Dec. 2005.

[5] R. Braden, et al., "Resource Reservation Protocol (RSVP)," *IETF RFC 2205*, Sep. 1997.

[6] F. Baker, et al., "Aggregation of RSVP for IPv4 and IPv6 reservations," *IETF RFC 3175*, Sep. 2001.

[7] P. Pan, E. Hahne, and H. Schulzrinne, "BGRP: A Tree-Based Aggregation Protocol for Inter-Domain Reservations," *J. Commun. and Networks*, v.2, n.2, pp. 157-167, Jun. 2000.

[8] R. Sofia, R. Guerin, and P. Veiga, "SICAP: A Shared-Segment Inter-Domain Control Aggregation Protocol," *Tech. Report, ESE, Univ. of Pennsylvania*, Oct. 2002.

[9] R. Bless, "Dynamic Aggregation of Reservations for Internet Services," *Proc. 10th Int. Conf. on Telecommunication. Systems – Modeling and Analysis (ICTSM'10)*, vol.1, pp.26-38, Oct. 2002.

[10] E. Rosen, et al, "Multiprotocol Label Switching Architecture," *IETF RFC 3031*, Jan. 2001.

[11] Network simulator (version 2.30). Retrieved from http://www.isi.edu/nsnam/ns/

[12] L. Subramanian, M. Caesar, C. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica, "HLP: A Next Generation Inter-domain Routing Protocol," *Proc. ACM SIGCOMM'05*, pp. 13-24, Aug. 2005.

[13] D. Griffin, et al., "Interdomain Routing through QoS-Class Planes," *IEEE Communications Magazine*, vol.45, no.2, pp. 88-95, Feb. 2007.

[14] S. Blake, et al. "An Architecture for Differentiated Services," *IETF RFC 2475*, Dec. 1998.