A Network Cooperative Overlay System

M. Stiemerling[*+], S. Niccolini[+], S. Kiesel[+], J. Seedorf[+]
[+]*NEC Europe Ltd., NEC Laboratories Europe*
[*]*University of Göttingen, Institute of Computer Science*
stiemerling@cs.uni-goettingen.de
{niccolini|kiesel|seedorf}@nw.neclab.eu

1 Introduction

Today's Internet experiences a lot of different traffic, ranging from traditional client/server applications to the nowadays predominant peer-to-peer. Peer-to-peer applications are the largest contributor to the overall network traffic [7][8] and while they are the main reason for many subscribers to buy broadband Internet access, they impose severe challenges on traffic engineering and sometimes the cause for network congestion. Usually, peer-to-peer systems use more simultaneous connections than client server applications, thus competing more aggressive for network bandwidth. Furthermore, they may have an impact on network operator's business cases. However, there is not the single peer-to-peer system, but there is a plethora of heterogeneous systems. This increases the complexity of managing various systems for network operators. One possible way of decreasing the impact of peer-to-peer systems is to block them completely or to severely throttle their network throughput. However, this typically results in side effects, such as other applications are also affected, or bad customer feedback [9].

A way forward out of this situation is to give both parties, i.e., the peer-to-peer systems and the network operators, a means to cooperatively work together. We propose a *common handle* that consists of a generalized overlay system which has the inherit property of allowing network side feedback for performance control. This generalized overlay system is reusable by multiple applications concurrently and it shares a common control and management interface across all applications. This allows participating nodes to communicate between each other. Combining that control and management interface with the network side feedback allows both network and peer-to-peer applications to run in a cooperative way.

The remainder of this paper is structured in this way. Section 2 outlines the basics of the generalized overlay system, while Section 3 describes the idea of cooperation between network and peer-to-peer system. The paper concludes with the proposal for an enhanced host interface in Section 4 and the concluding remarks in Section 5.

2 Generalized Overlay System

The generalized overlay system described in this chapter is based on the Service-Aware Transport Overlay (SATO) system [1] developed in the European Commission's FP6 "Ambient Networks" Research project. The purpose of SATO is to provide a flexible and customisable transport service to the application layer by using overlay networks on top of the transport layer connectivity. Furthermore, it includes auxiliary services such as distributed lookup services (e.g., DHT).

2.1 General Concepts of SATO

Service-aware Transport Overlay

Service-aware Adaptive Transport Overlays (SATOs) enable the flexible configuration of virtual networks consisting of Overlay Nodes (ONodes) on top of the underlying network connectivity. The Overlay Network topologies are responding to the application needs and can follow point-to-point, point-to-multipoint and multipoint-to-multipoint paradigms. Many SATOs can be created and deployed simultaneously.

On Demand Overlay Set Up and Tear Down

Service-aware Adaptive Transport Overlays are designed for accommodating the requests of the application services. As a consequence, they are established on demand, based on particular requirements, and terminated when the service is not requested anymore, e.g., after the last user has disconnected.

Dynamic Inclusion of Network Elements

The SATO concept allows the transparent inclusion of data processing elements (SATO Ports, see Figure 1) in the end-to-end transport path (between a client and a server or Peer-to-Peer). These SATO Ports can be located on any device (e.g., desktop computer but also router or middleboxes) and provide any functions such as overlay routing, media adaptation, rate adaptation, metering, congestion control, etc.

Overlay Adaptation

SATOs can adapt as a consequence of ONodes joining or leaving the virtual network, due to ONodes with changing context (e.g., system load), or as a result of adaptation requirements out of the cooperation part (see Section 3). This introduces the notion of adaptive overlays that dynamically re-configure in order to optimise the service delivery but obeying network constraints.

2.2 Elements of SATO

The core of the SATO system is the SATO overlay nodes. These nodes are overlay peers participating in the peer-to-peer routing and possibly application processing (see Section 2.1). Each overlay peer hosts the overlay manager and multiple types of SATOPorts. The overlay manager is in charge of controlling the overlay, providing a distributed lookup service for users and media relays, plus for communicating the network side feedback. The multiple types of SATOPorts on each overlay peers can be instantiated as several running SATOPorts, where each SATOPort instance can be assigned to a different SATO overlay. Figure 1 shows the overlay peer building blocks.

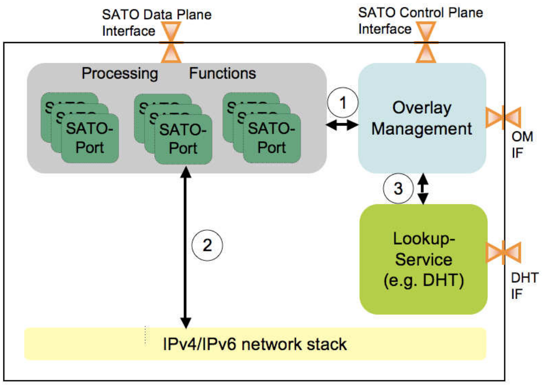

Figure 1 SATO ONode Architecture

Figure 1 shows the overlay node architecture. The *Overlay Management* (OM) is the central contact point on each SON. It is in charge of operating the SATO system, i.e., taking requests from applications to setup, reconfigure or tear down a SATO. Applications can request a SATO service through the SATO control plane interface. The OM interface (OM IF) is used by the OM to communicate with other OMs on other nodes. The OM of the SATO system is aware of the service requested by the application and about the network conditions in which the service is operated (the so-called service logic).

Each SON is part of the SATO *lookup service*, which is typically a distributed hash table (DHT), but not limited to, e.g., a centralised lookup service approach could also be chosen for a deployment of SATO. The *lookup service* and the OM communicate with each other using the interface 3. This *lookup service* is used to store information required for operating the SATO system and information required for running the applications. The OM is primarily using the *lookup service* to store information about available processing functions, such as packet relays, and their location in the network, i.e., the IP addresses.

The *processing functions* block that hosts the SATO-Ports provides the real power of the SATO system. SATO-Ports have one external interface to the application, which is the SATO data plane interface in Figure 1. The interface 2, called SATO network interface, is connecting the SATO-Ports to the network stack of the SON and thus to other SATO-Ports. The control of the SATO-Port by the OM is provided by interface 1.

3 Cooperation

The second key element of the network cooperative overlay system is the network side feedback mechanism. This network side feedback mechanism is a distributed monitoring suite providing information collected by participating overlay peers about the state of the network. The participating overlay peers are using both active and passive measurement techniques to provide peers with information about the network conditions. To this we refer as *overlay-internal data*.

The rationale of this approach is to make the overlay layer and the underlay cooperating with each other and thus achieving the common objective of the most effective and "best" usage of network resources. This is either achieved by triggering reconfiguration in the SATO level or by providing information about the SATO overlays towards the network, so that the network is aware of the status of the overlay system. Figure 2 shows the schematic approach taken by the cooperative overlay system.

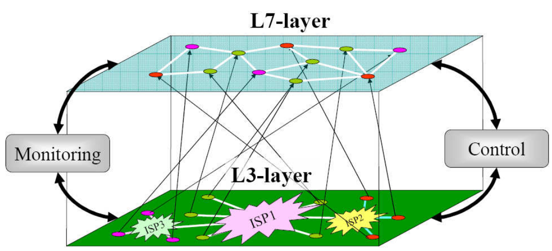

Figure 2: Schematic representation of the cooperative approach

Indeed, in the context of P2P file sharing, a recent study [5] proposes ISPs to run an "oracle" service for peer-to-peer applications, for influencing and make more efficient the overlay neighbour selection (when a peer supplies the oracle with a list of potential neighbours, the oracle ranks them according to certain performance criteria, e.g. proximity to the user, available bandwidth, etc., practically yielding the ISP a simple handle to cooperatively manage peer-to-peer traffic). This oracle is called *external data*.

The introduction of high capacity SATO ports in the P2P system that are owned and operated by network operators can play a role in favour of the transport overlay. The application has information on the overlay status, while the owner of the transport infrastructure has detailed information on the network status. When a subset of peers is operated by the owner of the network, these two information sets can be merged and exploited.

Based on this integrated view, the peer operated by the owner of the network infrastructure can take actions to optimize connections to other peers, i.e., the peer-to-peer network's topology.

4 Host Interface

To be acceptable for a wide-range of peer-to-peer application designers, it is necessary to define a better host interface, i.e., between application and network stack, for a generalized overlay system, compared to today's socket interface. There are multiple attempts to define an enhanced socket interface. The Overlay Sockets [4] aims at simplifying the standard TCP/IP socket interface for overlay usage and adding overlay specific elements, such as overlay identifiers. The Ambient Service Interface (ASI, [1]) describes a new interface that allows applications to specify limited requirements for their service.

None of these considers the full needs of a cooperative and generalized overlay system. This system requires better local feedback mechanisms of the networking stack to notify the application of local network changes (e.g., interface down, change of IP address, etc) and changed network conditions (e.g., congestion, uplink down, etc). This, in combination with the cooperation mechanism in Section 3, allows the generalized overlay system and also other applications to better react to changing network conditions, including but not limited to network congestion.

5 Conclusion

Our proposal combines the SATO idea of a generalized overlay system with the network side feedback mechanism to achieve a network operator-friendly cooperative overlay system. The network cooperative overlay system combines two different views on the network by leveraging the overlay-*internal data* obtained by the overlay system and the *external data* provided by an operator. The goal of this combined approach is to lead both providers and users to a win-win situation, where users get better performance while at the same time reducing network resource consumption.

The overlay system described in this paper has been documented for Peer-to-Peer SIP usage, as one possible example, in [2] [6]. As next step, we will extend it for the specific requirements of peer-to-peer video distribution in the framework of the NAPA-WINE project, while working towards a generalized communication overlay system.

ACKNOWLEDGEMENT: This work was partially supported by NAPA-WINE, a research project supported by the European Commission under its 7th Framework Program (contract no. 214412). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NAPA-WINE project or the European Commission.

6 References

[1] Ambient Networks Deliverable: "System Design of SATO & ASI", website http://www.ambient-networks.org/Files/deliverables/D12-F.1_PU.pdf, December 2007.

[2] M. Stiemerling and M. Brunner, A Peer-to-Peer SIP System based on Service-Aware Transport Overlays, VoIP Themenheft, Praxis in der Kommunikationstechnik, October 2007.

[3] http://www.ambient-networks.org/, December 2007.

[4] J. Liebeherr, J. Wang, and G. Zhang , Programming Overlay Networks with Overlay Sockets, 5th COST 264 Workshop on Networked Group Communications (NGC 2003), LNCS 2816, pages, 242-253, September 2003.

[5] V. Aggarwal, A. Feldmann, C. Scheideler, "Can ISPS and P2P users cooperate for improved performance?" ACM SIGCOMM Computer Communication Review, Vol.37, No.3, July 2007.

[6] M. Stiemerling and M. Brunner, Peer-to-Peer SIP implementation report, Internet draft (work in progress) draft-stiemerling-p2psip-impl-02.txt, November 2007

[7] IPOQUE's Peer-to-Peer study 2006 ,http://www.ipoque.com/userfiles/file/p2p_survey_2006.pdf, May 2008

[8] IPOQUE's Peer-to-Peer study 2007, http://www.ipoque.com/userfiles/file/p2p_study_2007_abstract_de.pdf, May 2008

[9] Comcast blocks some traffic ,http://www.msnbc.msn.com/id/21376597/, May 2008