Congestion Control for Carrier Ethernet using Network Potentials

Dietmar Toelle, Rudi Knorr

Fraunhofer Institute for Communication Systems ESK

Hansastrasse 32, 80686 Munich, Germany

{toelle, knorr}@esk.fraunhofer.de

Abstract—**In this paper we introduce a congestion control method suitable for Carrier Ethernet / Metro Ethernet Networks. Ethernet is considered to replace ATM as provider access aggregation network technology due to its cost efficiency, but it still lacks many necessary features to provide Quality of Service (QoS) as required from carrier-grade networks. This congestion control method is based on network potentials, which is a new concept to make the network traffic load aware and to implicitly prevent overload situations. Network potentials are scalar measures for the ability of the network to carry traffic of QoS services.**

Index Terms—**Carrier-Grade Metro Ethernet, Congestion Control, Effective Bandwidth, Network Admission Control, QoS, Resilience, Triple Play**

I. INTRODUCTION

Ethernet was originally designed as a local area networking technology for computer interconnection. In recent years, the deployment of Ethernet as the link layer technology for first mile broadband access and even for Metropolitan Area Networks (MAN) has started. The main reason for its increasing popularity can be found in certain advantages over competing technologies: high bandwidth efficiency (low cost per port), a wide range of scalability (from 10 Mbit/s up to 10 Gbit/s and more) and low complexity in terms of operation, management and maintenance. Moreover, Ethernet as an unified transport technology in the LAN, Access and MAN can bring further cost savings because inefficient protocol translations and interworking mechanisms are not necessary.

However, for a high level of service availability, public networking makes high demands on underlying transport technologies in terms of Quality of Service provision and fault resilience. The essential question is how switched Ethernet can be deployed in the Access and MAN while meeting these requirements and keeping its own advantages.

The aspect of fault resilience of switched networks has been extensively studied in one of our previous papers [1]. Another important aspect is Quality of Service provisioning, especially for the emerging inelastic real-time services like Voice over IP (VoIP) or Video on Demand (VoD) as part of the Triple Play service scenario.

The major problem for QoS provisioning in Ethernet based Access Aggregation Networks (AANs) is the prevention of aggregation overflows. In these networks there is a strong aggregation towards a few service end points (cf. 1), e.g. IP Core Edge, or directly connected servers. If there is too much traffic aggregated, packets might be dropped due to congestion on certain links of the network and therefore QoS is reduced for at least some of the customers, if not all of them.
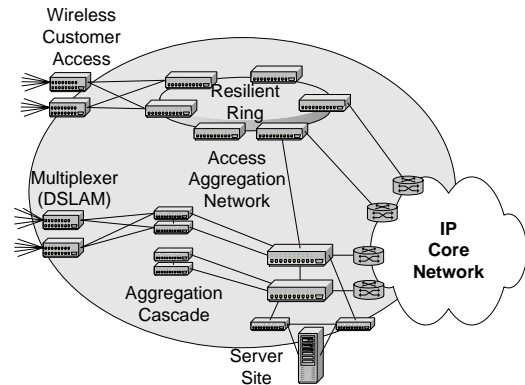

Fig. 1. Structure of an Ethernet Access Aggregation Network

Additionally to traffic overflows in the same service class, different service classes are also competing for the network resources and available resources might be reduced due to cross traffic from other service classes, especially if there are several service providers, which is the case for Triple Play service scenarios in AANs.

In order to have several service types requiring QoS at the same time there must be some form of resource request and reservation or access / admission control to control the overall network load by limiting the access to the services requiring QoS.

While the Internet Protocol (IP) supports various mechanisms for QoS provision (see [2] and the references therein), there are few for Ethernet. Ethernet itself supports only service separation and prioritization by tagging [3] and bridging [4]. Furthermore there is an admission control protocol based on an explicit resource reservation protocol, called subnet bandwidth management [5], which is quite complex and does not scale very well. Real-time services can be realized using protocols as described in [6] and the references herein, but again, their usage is limited to special network cases like in-house networks or industrial automation networks. Currently, admission control methods are either based on explicit resource reservation, which does not scale, or on management based network admission control schemes, which are more suitable for core networks [7].

In this paper we address the problem of designing an efficient network admission control mechanism for congestion control in Ethernet based MANs. As a basis we introduce the concept of network potentials and their usage to implement an efficient network admission control method.

Throughout the paper we use the following terminology: The expressions MAN and AAN are used interchangeably. Customers requesting a service are referred to as clients and the corresponding service end points are referred to as servers. Note that in this context the IP Core Edge router is also considered to be a server. In this paper we focus on Ethernet stand-alone solutions due to their simplicity. We want to establish an alternative to SDH/Sonet based solutions like RPR and POS.

In the remaining part of this paper we will provide a general overview of the concept of network potentials in section II and the related resource usage. In section III we outline the distribution of network potentials. Section IV describes a first simulation model and preliminary results and section V concludes this paper and indicates our future work.

II. NETWORK POTENTIALS

This section is intended to give an overview of network potentials. Network potentials are simply scalar measures of the ability of the network to transport traffic with the appropriate level of quality. As already pointed out, the problem is the inability of the network to recognize and avoid aggregation overflows on its own. That is why we use two different types, vertical and horizontal network potentials.

A) VERTICAL NETWORK POTENTIALS

A Vertical Network Potential (VNP) is assigned to each link and each direction of a network regardless which technology is used. In the aforementioned Ethernet MAN case, each active link is assigned two VNPs representing the maximum usable transmission bandwidth of the link in either direction.

The potential of a link might be reduced due to certain configurable parameters. For Ethernet links, the VNPs might be reduced due to limitations of the transmission rate using management settings, e.g. shaping mechanisms such as token buckets, or rate shapers. Conversely, the VNPs might be increased due to aggregation of links. A common aggregation protocol for Ethernet is the link aggregation control protocol (LACP, IEEE 802.3ad).

Furthermore, networks also may consist of wireless links or shared wired media like Metro Ethernet PON [8]. In these cases it is far more difficult to assign the VNPs. For Ethernet PON, the uplink VNP corresponds to the transmissions rate of the link divided by the number of timeslots.

Determining the VNPs of various transmission technologies are separate problems, which are not in the focus of this document.

B) HORIZONTAL NETWORK POTENTIALS

In order to actively avoid aggregation overflows and to make transparent the actual load situation throughout the network, horizontal network potentials (HNPs) are introduced. HNPs in general represent the maximum end-to-end transmission rates for specific traffic relations, i.e. the minimum of all VNPs that the end-to-end flows traverse. A traffic relation herein represents an end-to-end flow between two service end points, e.g. the flow between a video client and a video server. By assigning a HNP to a traffic relation the HNP is not only related to the traffic relation itself but also to

its traffic class. In other words there might be more than one HNP for the same traffic relation indicating several services in parallel, e.g. different quality levels of the same service or different services from the same server.

However, due to the fact that there are potentially thousands of end-to-end traffic relations, it is not feasible to use individual HNPs per traffic relation. This would be equivalent to the explicit resource reservation mechanism of subnet bandwidth management. It is rather recommended to form the HNPs based on the network links. This significantly brings down the number of HNPs to be maintained within the network. Further explanations how to setup the HNPs and their usage incident thereto will follow in section III.

C) NETWORK POTENTIAL OCCUPANCY AND TRAFFIC

The above mentioned network potentials only indicate the ability to carry a certain amount of traffic from one end of a specific traffic relation to the other, i.e. the unused network resources between these two points. If the HNP is nonzero and larger than a required minimum amount to realize a service, a service customer can access this free network resource and send the traffic across it to get the requested service. Each requested service produces a certain amount of traffic, denoted required bandwidth (RB), which, for instance, can be characterized by its effective bandwidth [9], its traffic envelope [10], or its peak bandwidth. Another way to describe the amount of traffic is by means of wavelet methods [11]. A traffic flow can be described by:

$$X(t) = \sum_k U_{J_0,k} * \phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty} \sum_k A_{j,k} * U_{j,k} * \psi_{j,k}(t) \qquad (1)$$

where ϕ are the scaling functions, ψ the wavelet functions, U the scaling coefficients, and A*U is a special construction for the wavelet coefficients. The coarsest time scale is denoted by J_0. The resulting time series $X(t)$ can be used to calculate the required bandwidth of this time series.

It is quite important to use a description of the network traffic which takes into account QoS parameters such as delay or loss constraints, because the network potentials do not take them into account at all. Norros' formula [12], for instance, yields the required bandwidth for a traffic flow taking into account the traffic characteristics through mean, variance and burstiness on the one hand, and QoS parameters like maximum loss or delay through loss rate and queuing delay given a certain buffer length on the other hand. This is a good means to care for QoS constraints by assigning different effective bandwidth values to a traffic flow.

Once all the parameters are known for the network and the services, the required bandwidth per traffic flow (TF) of a customer's service can be calculated. The required bandwidth of all traffic flows and all service classes is summed and this sum (Σ RB) is used as an indicator for the resource usage and therefore reduces the corresponding HNP.

The maximum horizontal potential for a dedicated link can be determined by reducing the vertical potential of this link by the sum of all traffic flows (TF) occupying resources on this link. Equation (2) yields the maximum HNP for link L for N traffic flows per service class and K service classes.

$$\max\left(HNP_L\right) = VNP_L - \sum_{k=1}^{K}\sum_{n=1}^{N} RB\left(TF_{k,n}\right) \qquad (2)$$

The horizontal potential for a specific service S can be determined using equation (3).

$$HNP_{S1,L} = \min\left(HNP_{S1,L-1}, \max\left(HNP_L\right)\right) \qquad (3)$$

Summarizing this section, we introduced network potentials to represent the unused traffic resources throughout the network. The VNPs form a potential field of link bandwidth values. The HNPs form a mesh of available end-to-end network resources. If network resources are used, the potentials are reduced such that they represent the remaining free resources. The next section describes the interaction of potentials and their usage.

III. POTENTIAL DISTRIBUTION

As pointed out in section II B, the HNPs represent the available end-to-end network resources for a distinct traffic relation and traffic class. While the VNPs can be determined by a single network node on its own, the HNPs must be set up first and maintained during operation. The HNPs are initialized by the corresponding servers.

Assuming that there is a limited number M of services requiring QoS and therefore some kind of resource reservation or network admission control (NAC), the HNPs can be used to implicitly perform resource reservation of individual services by applying resource reservation for their complementary ressource usage. The resource reservation is not performed directly but rather the indicator of free resources becomes zeros and access to the network resources can be restricted on the corresponding HNP being zero.

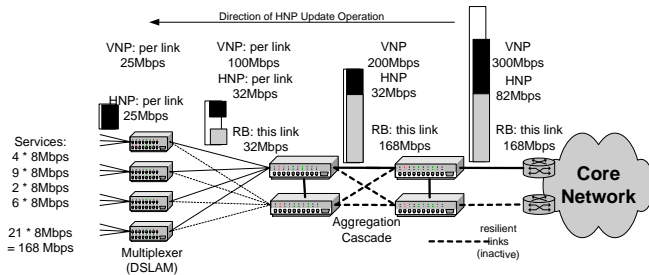The following simple example shows the application of HNPs.

Fig. 2. The very right black bar represents the initial HNP of the server, which is lower than the VNP (white bar). The HNP is transmitted to every active link containing a VNP. If the HNP is higher than the VNP the HNP is reduced to match the VNP and the new value is propagated instead of the original. The VNP of the link connecting the CPE to the DSLAM is significantly lower. Thus, the HNP that is seen by the CPE is as high as the VNP of the CPE link. Note that all links at an aggregation node are offered the same potential, i.e. the potential is not distributed among the links. A CPE can request as many services as there is enough HNP. The bars shaded in gray represent the RB for the active services traversing the corresponding link.

Consider a service requiring QoS, e.g. VoIP or Video on Demand (VoD), being realized by a dedicated network resource, e.g. a server or a share of bandwidth of a network interconnection point. This network resource is also represented by an initial HNP and a VNP. The VNP is the VNP of the link by which the resource is attached to the network. The initial HNP is determined by:

$$HNP_{init} = \min\left(VNP, N * RB_{Service}\right) \qquad (4)$$

where $RB_{Service}$ is the required bandwidth per traffic flow of that service, and N is the number of allowed concurrent users requesting this service. In figure 2 the principle of HNP distribution and potential reduction of the initial HNP due to lower VNPs is depicted. The potential distribution could be based on modified BPDUs, the well defined bridge packet data units commonly used for spanning tree protocols.

Permanent connection between different customer sites can also be implemented. The associated bandwidth of this permanent connection is allowed for by reducing the HNPs on the appropriate links. If this reserved bandwidth is unused, the bandwidth can be used for low priority best effort services which do not carry HNPs.

Another important feature for which the potential distribution could be used is the advertisement of services. Consider a possible network scenario with different network and service providers. The network provider only caters for the connectivity and the service provider offers various services and content. Using horizontal network potential distribution, the service providers can advertise their services and make dedicated service discovery protocols obsolete. The service provider, however, is responsible for the correct mapping of higher layer QoS requirements to Ethernet QoS features and the correct setting of tags / labels for the HNPs.

If there are multiple services, there will be multiple HNPs seen by the customers. The HNPs are distributed up to the links connecting the customers to the provider network. There are several options for how customers can access the various services. If there are complex first provider nodes like IP DSLAM or even distributed BRAS, these nodes can be used to perform the network admission control on their own. On requesting a service by submitting a special service request, the BRAS can decide whether or not the service can be established by checking the corresponding potential and relaying the request to the server as depicted in figure 3.
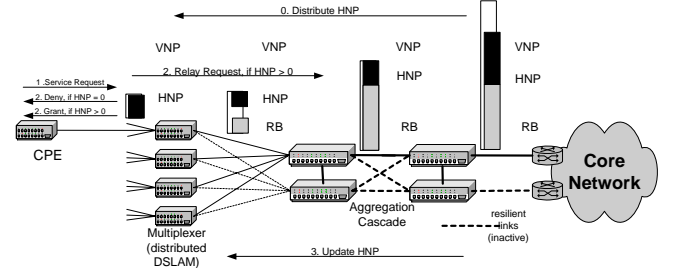

Fig. 3. The initial HNP of the server is first of all distributed throughout the network (step 0). Each DSLAM or distributed BRAS maintains the potentials and adjusts them to match the potentially lower VNPs of the customer links. If the BRAS supports the service request protocol it can directly answer incoming service requests (step 1) by granting the service and relaying the request or denying it (step 2). On a positive acknowledgement the HNP is updated appropriately (step 3). After the termination of the service the potential is updated again. If the BRAS does not support the protocol, the request must be relayed to the server, the BRAS just maintains the potentials. The principle is the same but the reaction time on a request will increase.

IV. SIMULATION AND PRELIMINARY RESULTS

In order to verify our model we set up a basic simulation (cf. 4) as a first step. We are using modified BPDUs to distribute the potentials. There are currently two services available within the network in addition to the common best effort service, which does not carry a potential. A single service is offered by each of the two servers respectively. Each server is the root for its service tree and its spanning tree respectively. Included in the simulation is the service request / grant mechanism.

The traffic model is set up as follows: By now and for simplicity and comparability we set the datarate of each service to constant rate at multiples of 64kbps. The VoIP services are set to 128kbps (factor 2), the video services to 2560 kbps (factor 40). Today, a service request would result in reserved bandwidth along the path, e.g. by RSVP-like mechanisms or the application of the modified Kaufman-Roberts-Algorithm [7] [13].
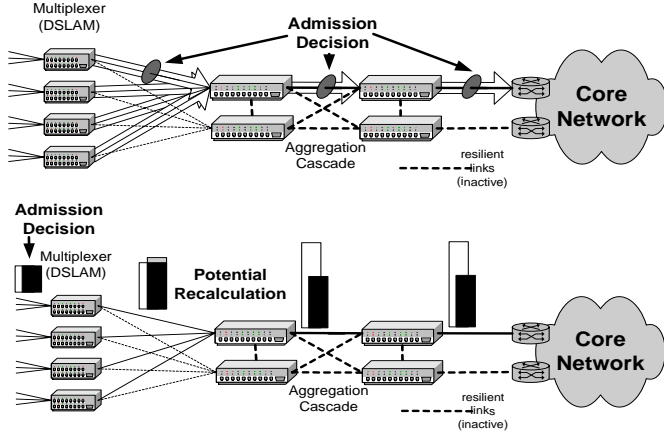

Fig. 4. Simulation scenario of an Ethernet MAN. Our method is compared with the standard method of requesting and reserving dedicated bandwidth using network budgets [7]. The top figure illustrates the simulated network using the link based bugdet network admission control. The bottom figure show our potential based network admission control.

This basic simulation model was intended to verify the potential distribution and service request / grant mechanism. An event based simulation model has been chosen to verify the model. In a next step we plan to extend the simulation model to include the verification of the QoS features. The services might be VBR services, so burstiness might be critical to network traffic and the QoS. Therefore we will include a representation of the services based on wavelets and include the queuing behavior of the network. In order to achieve this, we will implement a two-layered simulation. The simulation will be comprised of an event based top level layer for potential distribution and service request / grant / termination and an underlying time based queuing simulation for QoS verification.

Comparing network budget based network admission control with potential based NAC, our method reduces the complexity significantly. The following table shows a comparison between the different NAC methods. The budget based methods calculate the ressource usage using the Kaufman-Roberts-Algorithm (KRA) and assign the physical resources using the simplex algorithm in polynominal time.

Scenario as depicted in figure 4	Link budget based NAC	Potential based NAC
Physical resource assignment	Simplex Algorithm	HNPs filtered by VNPs
Admission decisions per service request	L	1
Number of values req. for calculation	(1+1) x N	(M+1) x N
Type of operations required	KRA, Simplex	Summation
Size of Assigment Matrix	m x N (N > D)	M x D (number of DSLAMs)

Tab.1. Comparison between LBNAC and PotNAC: L means average length of connection in hops, M means number of concurrent service types, m is the number of concurrent services, and N means total number of links.

Our method does not require complex operations, so the NAC can be performed by the network nodes themselves.

V. FUTURE WORK AND CONCLUSIONS

We introduced a novel network admission control mechanism for Ethernet based MANs. The introduced concept of network potentials can be used to prevent overload situations in networks by denying access to the network, if the potentials, i.e. free network resources, are depleted. This method is very efficient, so it can be performed by the nodes themselves. The next steps will include the incorporation of special resilience and load sharing mechanisms as well as service discovery and traffic engineering. The spanning tree protocol could be modified to distribute the potentials and decide which port should be blocked, based on the network potentials and thus perform traffic engineering.

REFERENCES

[1] T. Langguth, D.Toelle, et al, "Improving Resilience in Ethernet-based Access Networks," IEE ISSLS 2004, Edinburgh, March 2004.
[2] G. Armitage, "Quality of Service in IP Networks," Macmillan Technical Publishing, April 2000.
[3] "Virtual Bridged Local Area Networks," IEEE Standard 802.1Q-2003.
[4] "Media Access Control (MAC) Bridges," IEEE Standard 802.1D-2004.
[5] R.Yavatkar, et al, "Subnet Bandwidth Manager," RFC 2814, May 2000.
[6] R. Steffen, R.Knorr, "Quality of Service MAC Protocol for Shared Media In-house Networks," IEEE CCNC 2004, Las Vegas, January 2004.
[7] M. Menth, "Efficient Admission Control and Routing for Resilient Communication Networks," Dissertation, Uni Wuerzburg, August 2004.
[8] "Ethernet Passive Optical Network (EPON) Tutorial," June 2004, Available from: http://www.efmalliance.org.
[9] F. Kelly, "Notes on effective bandwidth," Stochastic Networks: Theory and Application, Eds.: F.P.Kelly, S.Zachary, I:Ziedins, Clarendon Press, Oxford, 1996, pp. 141-168.
[10] S. Bodamer, J. Charzinski, "Evaluation of Effective Bandwidth Schemes for Self-Similar Traffic," 13th ITC Specialist Seminar on IP Measurement, Monterey, September 2000.
[11] R. Riedi, M. Crouse, V. Ribeiro, R. Baraniuk, "A Multifractal Wavelet Model with Application to Network Traffic," IEEE Transactions on Information Theory, Vol. 45, No. 3, April 1999.
[12] I.Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," IEEE Journal on Selected Areas in Communications, Vol. 13, No.6, Aug. 1995, pp.953-962.
[13] J. Roberts, U. Mocci, J. Virtamo Editors, "Broadband Network Teletraffic, Performance evaluation and design of broadband multi services networks," Action COST 242, Springer- Verlag, July 1996.