Measurement Based Networking approach applied to congestion control in the multi-domain Internet

N. Larrieu, P. Owezarski
LAAS–CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4
FRANCE
{*nlarrieu, owe*}*@laas.fr*

Abstract

How to provide quality of service (QoS), has been a major issue for the Internet for the past decade. Though many proposals have been put forward in the areas of differentiated and guaranteed services, none have met the needs of users and operators. Efforts have been stymied by the complexity of the Internet, its myriad systems of interconnection, and by the technological heterogeneity of these systems. They have also run up against poor general knowledge of traffic characteristics that are largely unknown. In particular, recent monitoring projects showed that Internet traffic exhibited huge variations, leading to non-stationary traffic, and thus making difficult to guarantee a stable QoS. This paper then proposes a new measurement based architecture (MBA) and its related mechanisms (as the measurement signaling protocol (MSP) for instance, aiming at signaling to network components network information in real time) suited for coping with actual non-stationary traffic, and with the actual split topology of the Internet for which each domain provides a particular QoS. The idea of our measurement based networking (MBN) approach relies on a real time analysis of traffic characteristics and QoS evolution, and on the design of mechanisms able to adapt their reactions accordingly. The benefits of MBN are illustrated on a case study: a new measurement based congestion control (MBCC) which aims at smoothing traffic (making it more stable and stationary) and optimizing the use of network resources. Some preliminary results, based on NS-2 simulations, show the perfect suitability of this new mechanism for improving traffic characteristics and multi-domain QoS in the Internet, given the complexity and variability of actual traffic.

Keywords

Traffic characterization, Measurement Based Networking, congestion control, QoS

1. Introduction

The Internet is becoming the universal communication network for all varieties of information, from the simple transfer of binary computer data to the transmission of voice, video or interactive information in real time. These new applications require new network services. What has been a network offering a single best effort service now has to evolve

into a multi-service network. The resulting question, of how to provide QoS, has been a major issue for the Internet for the past decade. Though many proposals have been put forward, such as IntServ, DiffServ, and others, until now none has been widely deployed.

The solutions that the Internet community has offered in the areas of differentiated and guaranteed services have not met the needs of users and operators (Internet Service Providers (ISP), carriers, etc.). Efforts have been stymied by the complexity of the Internet, its myriad systems of interconnection, and by the technological heterogeneity of these systems. They have also run up against poor general knowledge of traffic characteristics that are largely unknown, and that might deviate significantly from standard suppositions. Thus, recent advances in Internet traffic monitoring seem to provide important missing information. In particular, these studies showed that Internet traffic is not smooth, exhibiting large oscillations at all time scales. Traffic variability is responsible of instability issues of the Internet QoS, as well as a serious decrease of Internet performances [12]. In particular, it has been shown that the new Peer-to-Peer (P2P) applications, used most of the time to exchange large files as music or movies, are changing the characteristics of Internet traffic [9]. New Internet traffic is now characterized by long range oscillations creating long range dependence (LRD) in the traffic. LRD and oscillations are very damageable for the network QoS as they can provoke congestion, and provide very unstable services to users.

This paper then proposes to use monitoring techniques and to take advantage, in real time, of monitoring results for proposing a new Measurement Based Architecture (MBA) and its related mechanisms (for instance MSP – Measurement Signaling Protocol – for exchanging / signaling measurement data) suited to adapt to frequent changes in the traffic. Thus, section 2 starts by analyzing what are the characteristics of the Internet that makes QoS so difficult to enforce. Section 3 presents our measurement based approach (called MBN for Measurement Based Networking) and the related MBA principles (section 3.1). One of the key component of this architecture is MSP (Measurement Signaling Protocol) that aims at signaling to all concerned network components the traffic measurements and characteristics in real time (section 3.2). Examples of MBN oriented applications are really wide, but we chose to illustrate it by the design of a new measurement based congestion control mechanism for the Internet, called MBCC. Thus, section 4 presents traffic parameters that have to be measured and that will impact MBCC mechanisms. Moreover, the different proposed mechanisms of MBCC and MSP have been evaluated thanks to NS-2 simulations. Simulation results are presented in section 5. In particular, it is demonstrated that MBCC is able first to better optimize available bandwidth compared to TCP, thanks to its very accurate knowledge of traffic characteristics, and second to improve global traffic regularity even if some part of this traffic is not transmitted using MBCC. Finally, section 6 concludes this paper and presents some planned evolutions for the MBN approach.

2. QoS issues in the Internet

Guaranteeing QoS means providing the requested QoS under all circumstances, including the most difficult ones. Among the most difficult circumstances, Internet QoS is highly

sensitive to a wide variety of disruptions, often designated as unexpected traffic, be they induced by failures, by the Byzantine behaviors of some network elements, or more simply by the significant, though not abnormal, increase in traffic levels related for instance to the live diffusion of some popular event.

Traffic disruptions more generally include all events that provoke a large change in network traffic characteristics, and that can badly impact the QoS provided by the network. In this context, it is important to be able to develop methods and methodologies for global monitoring of the network. These methods are essential for detecting and reacting to "disruptions". These are conclusions that have emerged from the French Metropolis* project, and from many other recent research projects across the globe, that have shown that Internet traffic is very far from being regular, and presents large variations in its throughput at all scales [11]. These projects have shown that Internet traffic exhibits characteristics such as self-similarity [13], (multi-)fractality [5], and long-range dependence [4], which is to say in all cases that traffic can vary significantly. These phenomena are due to several causes and in particular to congestion control mechanisms, especially the ones of TCP that is the dominant protocol in the Internet [11]. Thus, the notion of burstiness in TCP sources plus the LRD explain oscillations in the global traffic. Moreover, the increase of capacities in the Internet allows users to transmit larger and larger files (i.e. elephant flows) as music or movies for instance, it is clear that the scale of LRD is increasing, explaining why oscillations of Internet traffic, even with a coarse observation granularity, are so high [14]. Of course, oscillations are very damaging for the global use of network resources as the capacity freed by a flow after a loss for example cannot be immediately used by other flows: the higher the oscillations amplitude, the lower the global network performance [12]. Such high amplitude variations are notably due to the increase of the number of elephants in the network, and it is clear that they introduce oscillations with higher amplitudes than mice (short flows) [2]. Indeed, elephants, because of their long life in the network, have time to reach large values of the congestion control window, and thus, any loss event can provoke a huge reduction, followed by a huge increase of the sending rate.

Of course such significant variations have a bad impact on the stationary properties of the traffic, what forbids the enforcement of efficient mechanisms for guaranteeing a stable QoS during time for all users. The fact that traffic is not stationary is a well known feature, for instance on a daily basis, where traffic average is different during night and day, at lunch time compared to work hours, etc. However, it now appears that the huge variations of traffic due to users behaviors (exchange of large movies at high speed for instance) induce a finer grain huge variability in the traffic, we previously called in this paper "disruptions" or "unexpected traffic" that lead to non stationary behaviors. Our MBN proposal that is presented in section 3 deals with providing a new solution for networking able to take into account non-stationary traffic and huge variations on a link, as well as the large differences that appear on the traffic characteristics from one link to the other in the network.

Another issue with the current Internet related to the enforcement of mechanisms pro-

*More information is available at http://www.lip6.fr/metrologie/.

viding end-to-end QoS is due to its topology and administrative structure. It is illustrated on Figure 1. The Internet is generally defined as an interconnection of networks. That is of course true, but such definition does not include everything. In fact, the Internet can be more and more considered as a global worldwide network but split in several domains (also called AS for Autonomous Systems), administratively independent and independently designed and managed. Each network of each AS then does provide users with different services and QoS levels. In such a context, ensuring end-to-end QoS is a pain as the final QoS got by users will be the one of the worst network on the path from source to destination. In such framework, enforcing end-to-end QoS would require to set-up a global infrastructure and management procedure to avoid disparities between AS. But such hypothesis is unrealistic as carriers and ISP are competing, trying to be more attractive for customers than other competitors. Thus, finding a global agreement between all carriers and ISP defining how to handle Internet traffic is impossible. End-to-end QoS is then a multi-domain issue.

3. Measurement Based Networking Principles

3.1 Measurement Based Architecture design

Given such topological structure of the Internet, in addition of all the issues related to current traffic as un-stability, non-stationary nature, huge oscillating nature, correlation, dependence, and a huge versatility of traffic types during time, it is easy to understand that it is impossible to find a static solution optimal for all connections in the Internet. This statement leads us proposing MBN in order to react in real time and locally to some events in the network.

The first requirement of MBN is then to be aware of the network and traffic changes. It is then necessary to measure traffic and QoS parameters locally, as well as on long distances when the connection crosses several domains. Figure 1 depicts how measurement tools can be deployed in the Internet for this purpose. This figure more specifically depicts the case of a MBCC connection from a source to a destination crossing two Internet AS, and all the core and edge routers running MSP, i.e. having measurement capabilities and signaling to concerned network components these measurement results. Note that even if it is impossible to say that all links and all routers of the Internet will be monitored one day, we argue that taking into account the results of the measurement and monitoring tools effectively deployed in the Internet will be of great interest for improving Internet networking. MBN is designed that way: even if on some points measurement information is not present, the network continue to work with good performances and QoS. But performance and QoS can be much improved, and even become optimal, if measurement information is available.

Thus, given the administrative topology of the Internet, we propose to use different measurement techniques. Then, intra-domain measurements as loss ratio, used and available bandwidth, number of flows, etc. can be made using passive equipments (as Netflow, SNMP based tools, DAG cards, etc.). This is possible and certainly very easy as the domain is administrated and managed by a single entity, and such measurement or

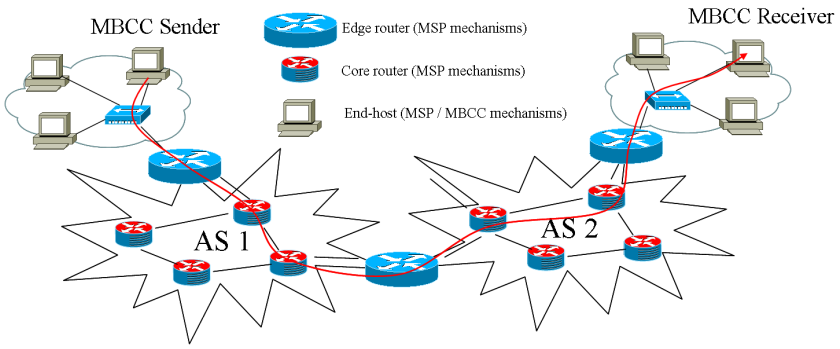


Figure 1: Network entities needed in the network to deploy MBN: case of the MBCC congestion control

at least management tools might be already present. In addition, passive measurement tools provide information on the traffic with a carrier point of view*. In fact, only delays measurement will be done in an active way, for easiness reasons.

On the other side, for inter-domain measurements it is impossible to use passive techniques as the other domains are not managed the same way, and their administrators may not use measurement techniques, or not necessarily the same techniques. In addition, even if they are performing measurements on their domain, in an open market where ISP have to compete with each other, they may not be willing exchanging such measurement information. So, in that case, it is required to address a measurement technique with a user point of view. Therefore, if you want to get information on other domain, the best solution consists in measuring what you need with active techniques, i.e. sending packet through the other domains, and measure what happens to these probe packets.

Then, all these measurements performed in real time and signaled to all concerned network components (traffic sources for instance), can give an accurate knowledge of network and traffic state, and allow them to perfectly adapt their sending rate (for instance) to available resources. Note that one important aspect of MBN deals with the design of a protocol for signaling measurement information. Such protocol has necessarily to work in intra-domain, but can also be extended for inter-domain signaling.

3.2 Measurement Signaling Protocol principles

The measurement signaling protocol is then a key component of the MBA architecture. The challenges for MSP are first to be efficient, i.e. to provide traffic information very rapidly, in order to make network components react according to very recent information on network state. But MSP must also be scalable. This means that MSP must generate a limited amount of traffic, not to impact the way the network is working. For scalability purpose, MSP components also have to store a limited amount of data: their link and

*It is generally said that passive measurement are carriers oriented measurements.

connection table has to be as small as possible, with a limited number of entries, in order to reduce the time required for searching any information in it.

Figure 1 depicts the way MSP is working illustrating MSP working on the MBCC case study – this part just gives the main general principles of MSP to reach its objectives). First of all, MSP is closely related to connections, i.e. the signaling path has to be the same as the one of the considered connection. We are then using the principle of RSVP [3] with a first packet that finds a path from the source to the destination, and then, a reverse packet that goes back to the source. The differences are that the back packet that is a reservation packet in RSVP is in MSP a signaling packet, transporting measurement information. As well, the signaling packets are send any times it is required, whereas in RSVP they are just sent once at the connection opening. In fact, MSP only uses the RSVP principle of being able to find a path and to go back along this path. It allows MSP to perfectly identify which are the concerned network components (routers) on the path, and to limit the number of sources and destinations for the measurement messages.

For addressing the scalability issue (that is one of the reasons why IntServ that uses RSVP is not used in the Internet), we chose:

- To consider only elephant flows. In fact, mice do not create troubles in the traffic, and most of issues that appear in the traffic are related to the sending of elephants. Then MSP routers just keep information on elephants passing through them. This is a way to limit the amount of entries in the connection table, as elephants just represent a very small proportion of the amount of flows;
- To send measurement information only when some disruptions arise in the traffic, in order to generate traffic only when the network conditions are changing*. It then should limit the amount of signaling traffic, and provide to senders or routers important information very rapidly (recall that measurement are achieved in all routers all along the path from source to destination, and potentially very close from the source).

We then expect to solve the scalability problem that has been previously encountered in the Internet while addressing the QoS management and enforcement issues. MSP performances will be accurately evaluated in section 5.2.

4. Measurement Based Congestion Control principles

Given the oscillatory and non-stationary traffic that causes so many issues for providing stable guaranteed QoS, this paper illustrates the MBN concept with the MBCC congestion control. MBCC goals deal with improving traffic characteristics and network performance by smoothing traffic (to limit oscillatory and non-stationary traffic effects), and optimizing the use of available bandwidth, thanks to the MBA measurement infrastructure and MSP. As well, MBCC also aims at providing more fairness between competing flows and continuing working with good performances even if measurements are missing.

In previous work [9] on traffic characteristics analysis, the oscillating nature of internet traffic has been analyzed. In particular, it has been shown that very high oscillations on long ranges are due to TCP that is not suited for transmitting large files on high speed

*Of course, a periodic sending of measurement information is also included in MSP for coping with very slow variations on traffic characteristics (i.e. traffic without disruption), but thanks to the concept of reacting only on traffic disruption, the period can be very large, thus inducing a quite limited signaling traffic.

networks. Such oscillations are a characteristic of actual current traffic that contains more and more long flows (elephants). [9] also showed that such kinds of oscillations are responsible of LRD in the traffic, LRD being able to characterize the dependence in the traffic on both short and long terms.

Therefore, the main objective is to bring more stability to elephant flows. To increase elephant flows regularity (i.e. to suppress observable oscillating behaviors at all scales), the new TCP Friendly Rate Control (TFRC) congestion control mechanism seems to be able to provide a great contribution. TFRC has been designed to provide a service suited for stream oriented applications requiring smooth throughputs. TFRC, then, tries as much as possible to avoid brutal throughput variations that occur with TCP because of loss recovery. By associating such a congestion control mechanism to elephants, i.e. to the main part of the traffic, we have been expecting to be able to control traffic oscillations, and then to increase global QoS and performance of the network. The benefits of using TFRC instead of TCP have been demonstrated in [6].

However, if TFRC is able to cope with TCP oscillations, it is not able to perfectly adapt to more brutal disruptions in the traffic due to failures on some links inducing a re-balancing of the related traffic, or to legitimate flash crowd. The measurement based approach is proposed as a solution to cope with such disruptions. As well, we expect MBCC to be an optimal solution, TFRC performing, in average, a little bit less efficiently than TCP SACK (Selective ACKnowledgement) [6]. However, in order to take advantage of all the TFRC benefits, MBCC has been designed as an extension of TFRC to which some capabilities for using measurement results coming from monitoring equipments in the network have been added. Making this choice also helps to fulfill one of the requirements that is to continue providing good results (better than the ones of the current Internet) even if monitoring information are missing, as it has been demonstrated in [6].

The main principle of MBCC then consists in using the TFRC algorithm for computing the nominal sending rate of each connection, and to correct it thanks to the knowledge of the available and consumed bandwidth in the network. Then, if some bandwidth is available, sources should generate more traffic than the throughput corresponding to a TCP flow without creating any loss or congestion in the network. The network congestion level should be significantly decreased by having "proactive" sources able to adapt in real time their sending rate according to available resources. As well, such mechanism will help to improve fairness between flows, as the correction on the sending rate will not depend on the RTT value, but on the real bandwidth available, equally shared between competing flows.

As in [6], it is sufficient to use MBCC for elephants For a normal period (when monitoring information is correctly received and when there is no congestion, i.e. there is some available bandwidth), each elephant flow can get an additional fraction of available resources. This fraction is computed by dividing the total available bandwidth by the average number of elephant flows in the network at this time. It makes sense to divide the available bandwidth by the average number of active flows (N) crossing this link, as it has been demonstrated that elephant flow arrivals nearly follow a Poisson process [2]. Indeed, in a Poisson process, as mean equals variance, the average number is significant because the process values will never be far from this average.

At last, for a congested period, MBCC senders have to reduce their sending rate for resolving congestion, and this trying to be as fair as possible. Thus, MBCC sources send the minimum value between the possible TFRC throughput and the effective throughput got by the flow at this time in the bottleneck of the network (this information being given by monitoring tools met all along the path).

So, the equations of this algorithm can be summed up as follows:

- For a period without congestion ($p = 0$):
  $X_{MBCC} = X_{TFRC} + aBW_{flow}$;
- For a congested period ($p \neq 0$):
  $X_{MBCC} = min(X_{TFRC}; cBW_{flow})$;

Where:

- aBW_{flow} is the Available BandWidth per flow in the bottleneck link(s) of the path. It is computed with the ratio $\frac{\text{total available bandwidth}}{N}$, this information is provided by MSP routers met on the path;
- cBW_{flow} is the Consumed BandWidth per flow in the bottleneck link(s) on the path, this information is provided by MBCC receiver with other end-to-end information such as RTT and loss ratio for instance;

5. Experimental validation of MBN approach applied to congestion control

In this section we present experiment results which validate MSP and MBCC mechanisms. In particular, section 5.2 quantifies optimal parameter values for MSP and MBCC in order to find the good trade off between low load for signaling information, good response time and accurate reaction for MSP and MBCC agents. Based on these otpimal values section 5.3 studies in details MBCC advantages (by considering specific statistical parameters detailed in section 5.1) for network stability and resources utilization compared to traditional congestion control mechanisms such as the TCP's ones.

5.1 Simulation principles

These new congestion control (MBCC) and signaling (MSP) mechanisms have been implemented and evaluated using NS-2. It has then been needed to develop a set of tools for monitoring available and consumed bandwidth in the simulated network and to exchange the measurement results between routers and traffic sources.

Simulation topology

The topology used is described on figure 2. In these simulations, we have created several bottleneck links to simulate a multi-domain topology. Elephant flows, using either MBCC or TCP SACK, and cross traffic using TCP New Reno are transmitted in order to make them compete in bottlenecks. The goal is then to study the impact of smooth MBCC flows on Internet traffic.

These core links (the ones of AS 1 and AS 3) represent the most "congested" links on the considered path, i.e. the ones that will mainly influence the MBCC sending rate. This difference should induce important congestion periods where adaptability skills of MBCC can be analyzed, and its performance level compared with others congestion control mechanisms especially the ones of TCP SACK [7]. Every simulation is based on real

traffic traces collected on the Renater* network. These traces are replayed in NS-2 with a special methodology detailed in [10] whose goal is to make simulation realistic, i.e. replay in simulations traffic samples such as reproducing all characteristics of real traffic.

In simulation, short and long flows are differentiated. Mice do not induce any transfer problem in the network. Then they will be transmitted using TCP and more precisely TCP New Reno that is the most frequent version of TCP in the Internet. At the opposite, elephant flows create in the network long range oscillations which induce congestions. Thus, simulations compare the case in which elephants are transmitted using our new MBCC congestion control mechanism, and the one in which they are transmitted using TCP SACK*.

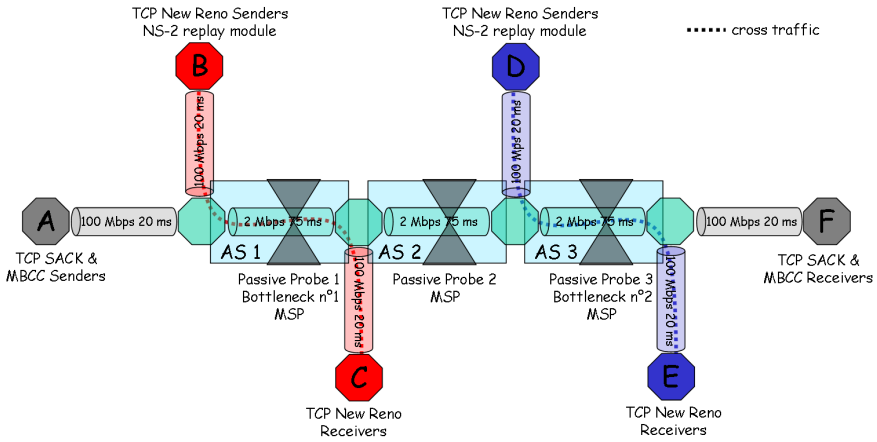


Figure 2: Network topology used in NS-2 simulations

Analysis parameters

For evaluating MBCC, several parameters have been evaluated in simulations:

- signaling traffic impact by computing the percentage between average throughput needed for signaling traffic and global average throughput on the network;
- throughput evolution by traffic type (TCP or MBCC) to study the traffic variability: computing of average throughput (A), standard deviation (σ) and stability coefficient (SC $= \frac{A}{\sigma}$);
- loss process evolution in order to evaluate MBCC adaptation capabilities compared to TCP;
- traffic oscillation range by computing the Hurst factor*.

*Renater is the French National Network for Education and Research.

*TCP SACK serves as the TCP reference as it is the best performing version of TCP. Thus, we did not consider others TCP's versions such as TCP Vegas given that [8] already demonstrated TCP Vegas was less performing than TFRC.

*Hurst factor, noted H, is a quantification of traffic LRD and represents also a good evaluation of traffic oscillation ranges. To compute it we use a wavelet based analysis of the traffic detailed in [1]. Note that the use of LRD for quantifying traffic oscillation has already been validated in [9].

5.2 Evaluation of MSP optimal configuration

Several simulations have been achieved, each simulation consisting of two different scenarios: in scenario 1, elephant flows (traffic from node A to node F) are transmitted using MBCC while in scenario 2, elephants are transmitted using TCP SACK. In these two scenarios, the cross traffic (traffics from nodes B to C and D to E), which is a normal Internet traffic consisting of both mice and elephants, is sent using TCP, and more precisely TCP New Reno.

Each simulation lasts 300 seconds. 100 elephants and 2000 mice have been replayed. One of the main goals of these experiments was to study the impact of MSP signaling traffic on network congestion and MBCC behaviors. Thus, we try to find out the best values for the different MSP parameters:

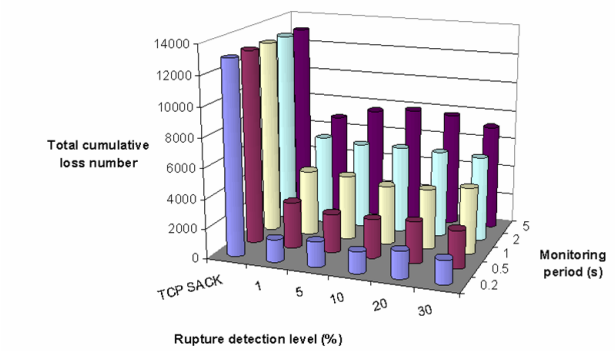
- First parameter is related to the measurement system granularity. In fact instantaneous throughput is computed as the average on a short period (P). This parameter has a strong impact as the coarser the granularity, the smoother the traffic appears. As a consequence, this granularity has a strong impact on the MSP traffic generated. Several periods P from 0.2 to 5 seconds have been tested;
- Second parameter is related to the disruption detection threshold on monitored links, i.e. what is the minimum variation between two consecutive measurements for which we can consider that the network conditions have changed and then need to be signaled to traffic sources to adapt to these new conditions? This threshold is expressed in terms of percentage of the total link capacity;
- Finally, the Time Out (TO) values correspond to the periodic behavior of MSP in case where no disruption arises in order to be able to take into account slow evolutions of traffic throughput. As we already said, it is useless to consider small values for TO as only slow evolution trends are concerned by this mechanism. In any case, it can be much larger than P. We then empirically selected (P, TO) couples respecting this principles: the couples are then: ($P = 0.2s$ and $TO = 2s$) or ($P = 0.5s$ and $TO = 4s$) or ($P = 1s$ and $TO = 5s$) or ($P = 2s$ and $TO = 8s$) or ($P = 5s$ and $TO = 10s$).

We then run several simulations using several traces to find out the optimal ($Period_{optimal}$, $Threshold_{optimal}$) pair. Results are depicted on Figure 3 representing the total cumulative number of losses, the overhead traffic due to MSP (in percentage of the total traffic) and the stability coefficient.

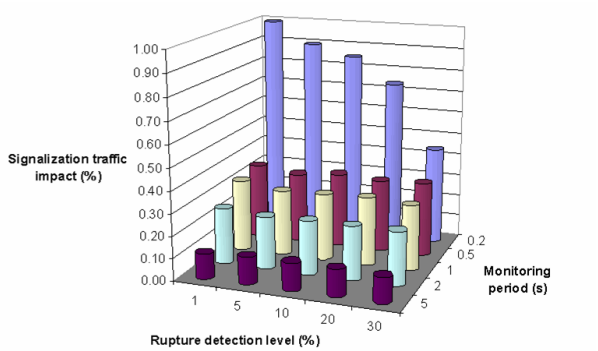
Optimal parameters inference

First, we are going to infer the optimal period value. One of the main goals of MBCC is to optimize the best network resources utilization by generating as less losses as possible. Thus, it is mandatory to take into account this criterion for the optimal pair ($Period$, $Threshold$) selection. In figure 3(a), only results with $Period \leq 1s$ are acceptable* (whatever the threshold value): $losses_{MBCC} \leq \frac{losses_{TCPSACK}}{3}$. Another main goal of MBCC is to transfer data with a smooth throughput to avoid oscillating behaviors which induce a bad network resource utilization. Thus, in figure 3(c), only results with $Period \geq 1s$ are acceptable (whatever the threshold value): $SC(\text{MBCC}) \geq SC(\text{TCP SACK})$. Considering results about signaling traffic impact, all results are quite good, and the ratio of MSP traffic is rather low (less than 1 % in the worst case if $Period = 0.2s$). This then

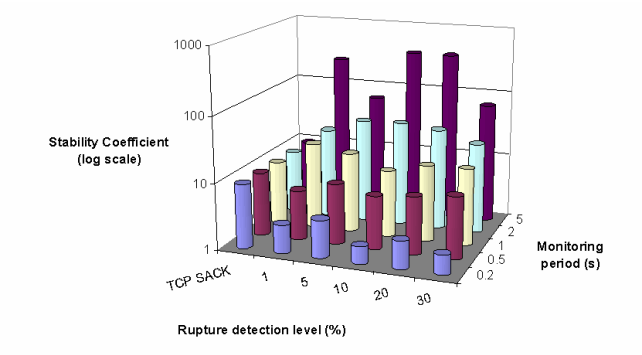
*For $Period \geq 2s$ loss levels between MBCC and TCP SACK are too close.



(a) Network congestion



(b) Overhead of signaling traffic / total traffic



(c) Traffic stability

Figure 3: Studied parameters evolution related to MSP working values

does not add any requirement. Thus, by crosschecking the three precedent parameters (congestion, stability and signaling traffic), only results got with a $Period = 1s$ meet all the choice's criteria.

In a second time we are going to infer the optimal threshold value. In this case, only experiments based on traffic stability can bring information to choose a specific threshold value. For the two others parameters (signaling and losses), results are really too close for giving us some usable information. Thus, we take into account the threshold where SC is maximum, this is the case for $Threshold = 1\%$.

To conclude, the optimal values pair is ($Period_{optimal} = 1s$, $Threshold_{optimal} = 1\%$). They will be used in the next section to quantify MBCC advantages compared to TCP SACK.

5.3 MBCC contribution to global traffic regularity in a multi-domain configuration

This second experiment aims at comparing the impact of MBCC and TCP SACK on network performance, traffic regularity and resource optimization. Thus, this section should highlight MBCC capabilities to improve traffic regularity for MBCC flows, and should also show how MBCC can improve traffic profile in term of stability for flows which do not use MBCC but that are competing with others MBCC flows in Internet bottlenecks. For that, the topology used is the same as in the previous experiment (cf. figure 2). Scenarios are the same except that the number of flows is increased (by a factor of 10) and simulations last 1800 s in order to introduce longer elephant transfers, as it is the case in the real Internet. This experiment can also help to evaluate the scalability of the proposed mechanisms given that number of elephant flows is increased in an important way. Moreover, MSP routers are configured with the optimal values pair defined previously ($Period = 1s, Threshold = 1\%$), and parameters evaluated in simulations are the same as in the previous section. The results are detailed in the next paragraph and table 1.

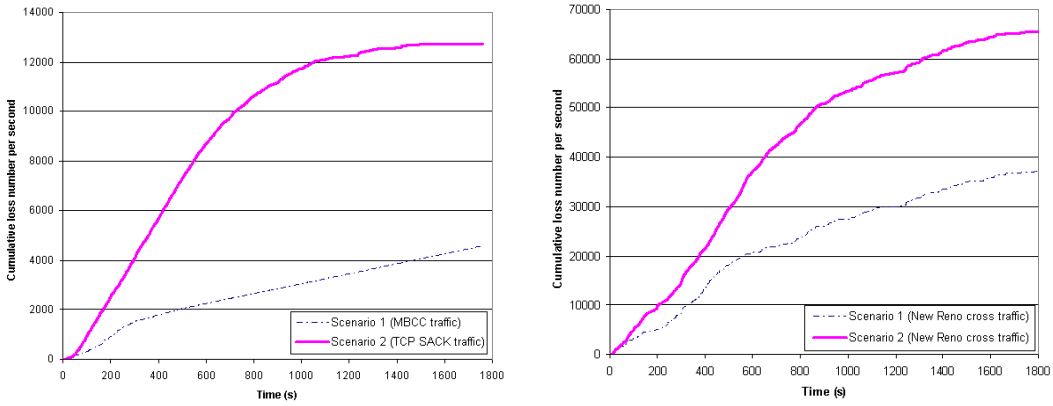
First, traffic throughput was computed. Table 1 shows result values for scenarios 1 and 2. This experiment then exhibited that MBCC performs better than TCP SACK, as throughput and resource usage are higher, and the traffic is also much smoother. Moreover, it seems that another interesting information deals with cross traffic in bottlenecks 1 and 2 when there is MBCC elephant traffic in the network (scenario 1). We can see that in the case where TCP SACK is used for transmitting elephants from A to F (scenario 2), cross traffic average throughput is lower and exhibits more variability than when MBCC is used in the network (cf. for instance $SC(\text{TCP New Reno}_{\text{Scenario 2}}) < SC(\text{TCP New Reno}_{\text{Scenario 1}})$).

Table 1 Traffic variabily analysis

	Scenario 1			Scenario 2		
	MBCC	TCP New Reno Bottleneck n°1	TCP New Reno Bottleneck n°2	TCP SACK	TCP New Reno Bottleneck n°1	TCP New Reno Bottleneck n°2
Average Throughput (B /s)	109434.9	111822.5	111572.5	109420.2	101651.1	101357.3
Throughput Standard Deviation (σ) (B /s)	31840.4	57127.7	60299.4	44704.3	83943.6	84291.9
Stability Coefficient (SC)	3.437	1.957	1.850	2.448	1.211	1.202
Hurst parameter (H)	0.51	0.74		0.92	0.83	

This result is confirmed with the loss process analysis. Indeed, figure 4 depicts a more important loss level in the network when using TCP SACK (cf. curves of scenario 2 on figures 4(a) and 4(b)) than when using MBCC (cf. curves of scenario 1 on figures 4(a) and 4(b)) and this for both elephant and global cross traffic. Indeed, traffic variability in scenario 2 is more important, then congestion occurred more easily in the network and the loss number is higher.

Finally, the last line of table 1 shows that MBCC impacts in a very positive way traffic LRD. In fact, thanks to MBCC, the LRD is much reduced in the elephant traffic (cf. scenario 1 where $H = 0.51$) compared to the reference TCP SACK traffic where LRD is very high ($H = 0.92$ in scenario 2). Consequently, there are less oscillations (cf. related stability coefficient values of table 1). Moreover, the analysis of cross traffic LRD shows



(a) Elephant traffic

(b) Cross traffic

Figure 4: Congestion level estimation

that global traffic when elephants are transmitted using MBCC (cf. $H = 0.74$ in scenario 1) is less dependent on long ranges than using TCP SACK (cf. $H = 0.83$ in scenario 2), this feature inducing more stability on traffic profile and then less congestion in the network.

6. Conclusion and future work

In this paper, we proposed a new approach and then new mechanisms for improving Internet QoS with the ultimate goal to be able to enforce a stable QoS. As a consequence, this paper proposes a new measurement based architecture and related signaling mechanisms allowing the deployment in the network of a new congestion control mechanism called MBCC. Its principle is to adapt to the huge variations of traffic when disruptions arise. The main strength of MBCC deals with using monitoring and measurement tools which start to be widely deployed and that should be generalized in a short future. The principle of MBCC consists in using in real time the information on traffic characteristic evolutions that monitoring tools can provide, in order to perfectly adapt the reaction to the actual network and traffic conditions. In fact, MBCC changes the classical principle of congestion control: even if the connection keeps an end-to-end principle, a hop by hop control of this connection is added, where hops are made between two network equipments that are monitored. Given the current issues with Internet traffic, MBCC has been designed for being able to smooth traffic (what is a major requirement for being able to provide stable and guaranteed services), limit the number of losses, optimize the use of resources, and provide fairness. The experiment results proved that MBCC reaches its objectives. These benefits in terms of traffic regularity or loss level, will be directly felt by users as a QoS improvement.

As a final conclusion, it is clear that the results got with MBCC demonstrate the benefits of our measurement based networking approach applied to congestion control. In this paper we have shown with MBA, MSP and MBCC mechanisms that MBN is a right approach for addressing the issue of improving multi-domain end-to-end QoS. But we do believe that MBN can be a universal solution for managing the Internet and its traffic. In particular, MBN has been designed in order to be able to provide a suited solution that can adapt to any kinds of networks, any traffic nature and conditions, etc. In particular, MBN should have applications in many domains as traffic management, traffic engineering, security management or QoS optimization.

References

[1] P. Abry and D Veitch. Wavelet analysis of long range dependent traffic. In *Trans. Info. Theory, Vol.44, No.1*, pages 2–15, January 1998.

[2] N. Ben Azzouna and F. Guillemin. Analysis of adsl traffic on an ip backbone link. In *Proceedings of Globecom 2003*, December 2003.

[3] R. Braden and L. Zhang. Resource reservation protocol (rsvp) – version 1 message processing rules. *RFC*, (2209), September 1997.

[4] A. Erramilli, 0. Narayan, and W. Willinger. Experimental queuing analysis with long range dependent packet traffic. In *IEEE/ACM Transactions on Networking, Vol. 4, No. 2*, pages 209–223, 1996.

[5] A. Feldmann, A. Gilbert, and W. Willinger. Data networks as cascades: Investigating the multifractal nature of internet wan traffic. In *Proceedings of ACM SIGCOMM'98*, 1998.

[6] N. Larrieu and P. Owezarski. Tfrc contribution to internet qos improvement. In *Proceedings of the fourth COST 263 international workshop on Quality of Future Internet Services (QoFIS'2003)*, October 2003.

[7] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanov. Tcp selective acknowledgement options. *RFC*, (2018), October 1996.

[8] P. Owezarski and N. Larrieu. Coherent charging of differentiated services in the internet depending on congestion control aggressiveness. In *Computer Communication Journal, Vol. 26, issue 13*, August 2003.

[9] P. Owezarski and N. Larrieu. Internet traffic characterization – an analysis of traffic oscillations. In *Proceedings of the 7th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC'2004)*, July 2004.

[10] P. Owezarski and N. Larrieu. A trace based method for realistic simulations. In *Proceedings of the IEEE International Conference on Communications (ICC'2004)*, June 2004.

[11] K. Park, G. Kim, and M. Crovella. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *IEEE ICNP*, 1996.

[12] K. Park, G. Kim, and M. Crovella. On the effect of traffic self-similarity on network performance. In *SPIE International Conference on Performance and Control of Network Systems*, November 1997.

[13] K. Park and W. Willinger. *Self-similar network traffic: an overview*. In Self-similar network traffic and performance evaluation, J.Wiley & Sons, 2000.

[14] W. Willinger, M. Taqqu, R. Sherman, , and D. Wilson. Self-similarity through highvariability: statistical analysis of ethernet lan traffic at the source level. In *ACM Sigcomm'95, pages 100–113*, 1995.