Open Box Protocol (OBP)

Paulo Loureiro[1], Saverio Mascolo[2], and Edmundo Monteiro[3]

[1] Polytechnic Institute of Leiria, Leiria, Portugal
loureiro.pjg@gmail.pt
[2] Politecnico di Bari, Bari, Italy
saverio.mascolo@gmail.com
[3] University of Coimbra, Coimbra, Portugal
edmundo@dei.uc.pt

Abstract. In this paper we propose a new explicit congestion control approach, Open Box Protocol (OBP). The OBP gives sources the capacity to look inside the network and to make their congestion control decisions, based, not only on packets loss or packets delay, but also based on another kind of information. For example, the most restricted interface capacity, the available bandwidth, the RTT variations or the presence of heterogeneous transmission means. In the paper we describe the OBP and discuss its evaluation results based on ns-2 simulations. The results show the OBP´s capacity to provide traffic sources with the required information, in static or dynamic network scenarios, and to make correct and quick congestion control decisions. Also, it is visible that the OBP avoids the full queue problem and tries to keep the queues near zero occupation.

1 Introduction

Congestion control algorithms have many objectives to address. They must maintain the performance independently of the flows size distribution; they must guarantee short flow completion times in presence of a mix of flows; they have to ensure all flows with a fair share of available bandwidth; they have to make an efficient use of high bandwidth-delay links; they must have the capacity to react to sudden changes in network paths (wireless links); algorithms must be stable, robust [1], [2] and deployable in the Internet. On the other hand, implicit congestion control algorithms [3], [4], [5], [6], the most used in the Internet [7], give traffic sources poor information about network state, typically information about packets loss [3] and / or round trip time (RTT) variations [5], [6], [8]. With these two pieces of information, congestion control algorithms have to control the congestion inside the network. The great number of proposals about congestion control algorithms shows that any algorithm that only uses these two pieces of information will have limitations in making a better use of network capacity.

Internet traffic is of dynamic nature and the characteristics of this traffic are changing. Beyond the traditional applications, Internet is used by a set of new applications, for example the multimedia applications VoIP and IPTV. This kind of applications has changed the characteristics of the internet traffic and therefore the

R. Perrott et al. (Eds.): HPCC 2007, LNCS 4782, pp. 496–507, 2007.

congestion control actions must be adjusted. Besides that, the wireless networks are in expansion. In this case the congestion control decisions may not be the best when using the detection of packets loss as the only criterion to identify the congestion situations. In wireless networks the corruption in packets is more frequent than in wire networks. Moreover, the current network capacity is larger than in the past. This new Internet has new congestion control challenges and the transport protocols have to control the congestion situations and efficiently use the network resources. It is important to bear in mind that short capacity networks still exist and need to coexist with high capacity networks.

Congestion control solutions based on router collaboration have good potential because routers are the place where the congestion normally occurs. Therefore, they can quickly detect the congestion situations. Using routers, the congestion symptoms are identified in a direct way because routers detect that the amount of packets that arrives at interfaces is bigger than the capacity of the interface to process all the packets. Router based congestion control mechanisms allow quick congestion detection and the actions taken to control the congestion can be more effective. The actions to control the congestion are taken with delay in reference to the instant when the congestion began. This factor is important because the congestion duration can be shortened, the packet loss reduced and the network resources better used.

Traditional TCP's congestion control and avoidance algorithms [14] are powerful but not enough to provide good service in a lot of network conditions since they handle the network as if it were a black box [9]. The goal of our work is to create a new congestion control model, which we call Open Box Protocol (OBP), using router collaboration to identify the network resources along the path and to provide this information to end systems. Additionally, the congestion control decisions are made by the end systems by using the information received from routers. Moreover, the model must have the capacity to efficiently use the network resources, avoid congestion, reacting well to sudden changes in network paths, being easily deployed in the Internet and coexisting with other congestion control protocols.

2 Background and Related Work

Over the past few years, several solutions have been proposed to give TCP better and more network feedback, beyond packets loss information and RTT variations. In addition, the research community has been specifying alternative solutions to the TCP architecture. As the OBP, some of these models are classified in category of “modification of the network infrastructure”. They are briefly explained as follows.

The ECN [9], [10] and the Quick-Start [11] are two solutions that use the router collaboration to address the congestion control problem. With the ECN, the router collaboration is done by detecting congestion situations and by informing the end systems about this situation.

With the Quick-Start, the collaboration of routers is used to decide the value of the initial congestion window. The algorithms slow start, congestion avoidance, fast retransmission and fast recovery [14] are still used and therefore the problems associated to theses mechanisms remain.

Explicit Control Protocol (XCP) [12], [13] is designed to work well in networks with large bandwidth-delay products. The XCP generalizes the Explicit Congestion Notification proposal (ECN). Routers provide feedback, in terms of incremental window changes, to the sources in multiple round-trip times, which works well when all flows are long-lived.

Rate Control Protocol (RCP) [1] is designed to efficiently use high bandwidth-delay product networks, such as the long optical links; to be stable independently of link-capacity, round-trip times and number of active flows and to try to emulate processor sharing. Each router maintains a single fair-share rate per link. Each packet carries the rate of the bottleneck link. The RCP gives the same transmission rate to all flows. Another relevant point, at routers, the RCP uses information presents in packets, and received from end systems, to decide the fair-share rate. This means that routers cannot simultaneously receive RCP packets and packets from other transport protocols. The implementation of this model in Internet is conditioned by this factor.

The differences between the XCP, the RCP and the OBP are the kind of feedback that end systems receive and the entities that have to make decisions about congestion control. Opposite to the ECN and the Quick-Star, the OBP makes all congestion control decisions based on information received by routers. The OBP is computationally simpler than the XCP and the RCP, since routers do not have to make decisions about congestion control and only need to provide feedback information about the network state.

3 Open Box Protocol (OBP)

We will answer the following question: Is it possible for end systems to constantly see the network state between the source and the receiver? The answer is yes if we can represent the network path through a small set of variables and if we can continually put this information at sources. With this information, the sources can quickly make decisions about the efficient use of network capacity and quickly adapt to sudden network changes.

Any flow begins with an SYN packet that will be acked to give the initial values of the OBP protocol parameters. When this packet arrives at first router, the variables that represent the network state are updated. At second router the packet is evaluated and if any variable needs to be updated the router performs this exchange. Information about the network state will arrive at the end system inside the ACK packets. The end system, with this information, can make congestion control decisions for efficiently use the network resources and avoid the congestion.

3.1 State of Network Path

To represent the network path we only need the information that is used or important to make decisions about congestion control. It is not relevant to represent the network path at other levels. Fig.1 shows a network path with four routers. Each box represents the router output interface and has two kinds of information: the output interface capacity and the available bandwidth. In this example the *narrow link* is 45 Mbps (the most restricted interface capacity) and the available bandwidth is obtained in *tight link* (the most restricted available bandwidth).

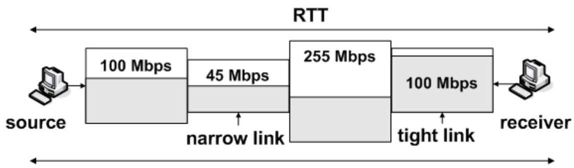

Fig. 1. Network state

To represent the network path from one end system to another one needs the following variables: narrow link (NL) - along the path, this is the link with the most limited capacity (Mbps); tight link (TL) - along the path, this is the link with the least available bandwidth (Mbps); round trip time (RTT) - time needed to send a data packet and to receive the associated ACK packet (seconds); heterogeneous path (HP) - having or not heterogeneous means along path (boolean), for example wireless links. With the exception of RTT, which is obtained by using information from TCP header, all the other variables are carried inside data packets in IP header.

3.2 Router Processing

The OBP considers that, when a new packet is inside the first router of the network path, and before the packet leaves, the router updates three variables: NL, TL and HP. When this packet is at the second router these variables are changed or not. The value of the NL is changed if the output interface capacity is less than the previous one. The value of the TL is changed if the available bandwidth is less than the previous one. The value of the HP is changed if this router has a wireless link. The available bandwidth at the TL is obtained through the following steps: at the output interface, and for short periods of time, all packets that get in are counted. Then, all of the packets that got in are divided by the period of time. This procedure gives us the used bandwidth.

3.3 Source and Destination Processing

Unlike other explicit congestion control protocols, the OBP congestion control decisions are made in sources. However, the OBP makes decisions using explicit information received from the network.

New flows begin with a high transmission rate allowed by the network. The initial transmission rate depends on the available bandwidth at TL and the interface capacity at NL, and is calculated after the sources have received the SYN-ACK packet. This method assures short completion times for short flows. Every time the sources receive an ACK packet the transmission rate is tuned. This is done using the feedback information received from the network. These transmission rate adjustments are done to come near to zero the available bandwidth and the RTTs near the minimum. The RTT near minimum means that the queue occupation is near zero. The OBP model tries to efficiently use the network path capacity and tries avoiding congestion. This means that the available bandwidth must always be near zero;

The following equations show how the transmission rate is adjusted. The initial transmission rate $W_{(t0)}$ depends on the available bandwidth $AB_{(t0)}$, the capacity $CN_{(t0)}$ at NL and the constants α and β.

$$W_{(t0)} = \alpha * AB_{(t0)} + \beta * CN_{(t0)}$$

Every time a new ACK packet is received, the feedback information inside the packet is used to make adjustments in transmission rate. These adjustments are done based on feedback information and based on an equilibrium point. The equilibrium point is updated in multiples of RTT and is calculated based on the mean of the transmission rate during the previous period (this period is equal to an RTT average). The transmission rate is updated whenever an ACK packet is received and is always around the equilibrium point. Fig. 2 shows, for a flow, an example of the behavior of these two variables: the transmission rate and the equilibrium point. In this example, at the beginning, we can see that the transmission rate is always higher than the equilibrium point. This means that there is available bandwidth to be used. After the first RTTs the source finds the equilibrium point that enables filling the entire network path between the source and the destination.

The network warnings can include negative available bandwidth - this case corresponds to a transmission rate below the equilibrium point; or positive available bandwidth – in this case corresponds to a transmission rate above the equilibrium point. Although this image is obtained from a test with 100 flows, it is visible, for this flow, that the OBP quickly finds the maximum equilibrium point and stabilizes in this value. This situation corresponds to the efficient use of network resources.

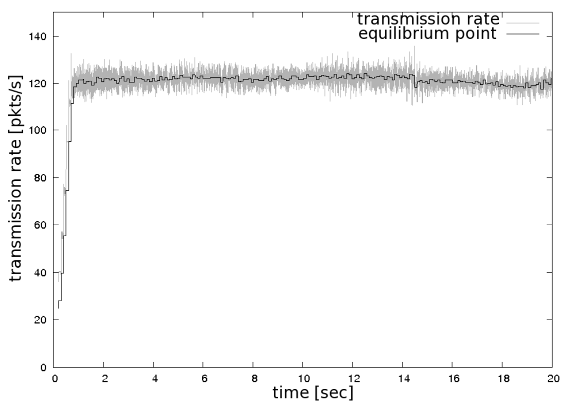

Fig. 2. Transmission rate vs Equilibrium point of a flow. This test had 100 flows, a bottleneck link equal to 1 Gbps and the RTT equal to 0.1 ms.

The transmission rate is updated every time an ACK packet is received. The transmission rate $W_{(t)}$ depends on the current equilibrium point $EP_{(k)}$, the available bandwidth $AB_{(t)}$, the capacity $CN_{(t)}$ at NL and a constant δ. Also, it is affected by the RTT if this value is different from the minimum RTT, affected by a constant μ.

$$W_{(t)} = EP_{(k)} + EP_{(k)} * [(\delta * AB_{(t)})/(AB_{(t)} + CN_{(t)})] + EP_{(k)} * \mu * [RTT_{min} - RTT]$$

This formula allows obtaining transmission rates around the equilibrium point. If the AB(t) is near zero the W(t) obtained is the EP(k). However, if the AB(t) received is negative or if the RTT is large, the value obtained of W(t) is less than the EP(k). In an extreme case, the W(t) obtained may be near zero, for example if the AB(t) has a high negative value or if the RTT is very high. This solution protects the network against congestion collapse, because it can instantly reduce the transmission rate to few packets [2]. This formula also enables a quick adaptation to sudden or transient events [2] as it admits changes in the transmission rate whenever ACK packet is received.

The equilibrium point is updated one time per RTT. When this occurs, the equilibrium point is updated with the mean of all transmission rates calculated whenever an ACK packet is received during the previous period.

$$EP(k) = \text{mean}(\sum W_{(t)}); \text{ during last RTT}$$

The formulas used by OBP assure that the increase in transmission rate is always decided by the feedback received from the routers. Also, the transmission rate can be updated every time an ACK packet is received. Opposite to this behavior the traditional congestion control algorithms allow the sources to increase the transmission rate without knowing if the network is near congestion.

At destination end systems, and at transport level, the ACK has three new variables, NL, TL and HP. These variables are filled with information received from data packet. Information present in ACK packet arrives at the source.

3.4 Deployable in Internet

To be deployable in internet any new congestion control approach should consider the overhead in terms of packet header size, the added complexity at end systems or routers and the interaction with other transport protocols [16]. In terms of packet size overhead this model has three new variables: NL, TL and HP. The NL and the TL are represented in units of KBps and use 3 bytes each. The HP is a boolean (1 bit).

In terms of complexity, the OBP implementation puts the load processing in the sources side. Moreover, we can have the network being used by flows that are controlled by different congestion control algorithms, among which the OBP.

4 Evaluation

To evaluate the OBP model, we have created simulations on the ns-2 simulator [15] (version 2.30) with the OBP implementation. Fig. 3 shows the network topology used in the simulations.

To evaluate the performance of OBP we did tests with long flows. We also did with flows created by Poison distribution and the flows size was defined by Pareto distribution. We equally tested the OBP and compared it with the TCP Reno, the XCP, the RCP and the TCP Reno with Quick-Star.

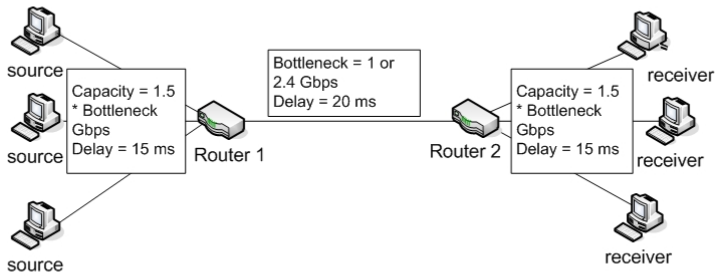

Fig. 3. Network topology

4.1 Behavior with Long Flows

In these two tests we generated 100 flows at instant 0 seconds. The bottleneck link was equal to 1 Gbps or 2.4 Gbps. The RTT was 100 ms, and the size was 1000 bytes per packet. The transmission rate used the following configurations: α = 0.0, β = 0.002, μ = 1.0, δ = 1.0. The objective of these tests was to verify if the available bandwidth, at bottleneck link, was near zero; which meant that sources generated enough traffic to fill the path, and if the RTT had no oscillations, which meant that the routers' queues were near zero occupation.

Convergence and stability. We can see in Fig. 4 that, in few RTTs, the sources generated traffic to fill the path and the feedback received about the available bandwidth was near zero. The available bandwidth can be less than zero. This means that, at those instants the amount of traffic generated is greater than the capacity of output interface. In this case, packets were momentarily stored in the routers' queues.

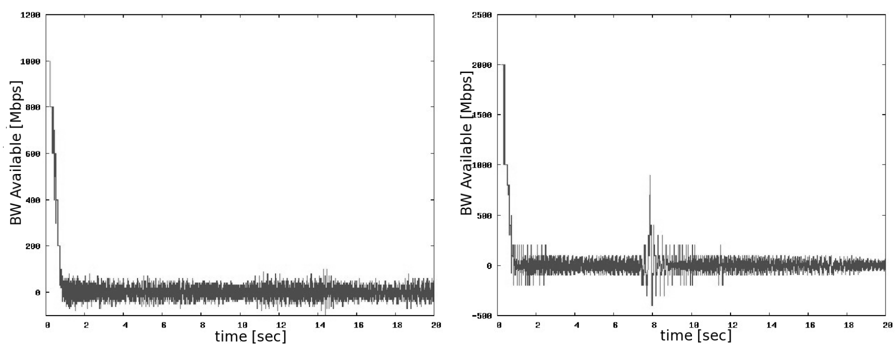

Fig. 4. Available bandwidth sent to sources by the bottleneck router, with the bottleneck interface equal to 1 Gbps and 2.4 Gbps

The state of routers' queues. The OBP also uses the time that a packet spends inside the routers' queues to make congestion control decisions. This situation allows maintaining the queues' occupation near zero, or with tendency to near zero. Fig. 5 shows that the RTT is always near 0.1 ms. By these results we can conclude that the OBP has capacity to generate traffic that tends to use the path capacity, without filling the routers' queues.

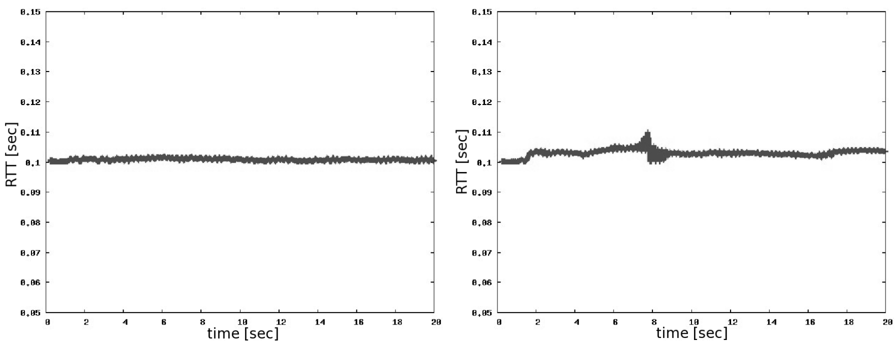

Fig. 5. Packets' RTT received at sources. The bottleneck link is 1 Gbps and 2.4 Gbps.

4.2 OBP Behavior with Variation of Traffic Characteristics

In this section our goal is to find out the OBP performance in presence of traffic with dynamic characteristics. The simulations used the Pareto distribution to define the flows' size, with the mean of 125 packets (bottleneck link 1 Gbps) and 300 packets (bottleneck link 2.4 Gbps), and the shape was 1.8. The arrival of flows was defined by the Poisson distribution with the lambda equal to 950 packets. The duration of the tests was 20 seconds and the packets had 1000 bytes.

The results of these tests are analyzed through the available, the average completion time per flow and the RTT. The average completion time represents the difference between two instants, SYN packet departure and arrival of the last data packet.

Convergence and stability. Fig. 6 shows the available bandwidth. Although these tests generated 950 new flows per second and the flows' size varied between 56 and 16340 packets (1 Gbps) or between 134 and 39216 packets (2.4 Gbps), the results show the capacity of OBP model to manage the transmission rate and to fill the network path without congesting the bottleneck link. The Poisson distribution is used to define the instant of flows creation, so there are some variations in the available bandwidth along the time. The stability of the OBP is equally confirmed through these results and the available bandwidth tends to zero.

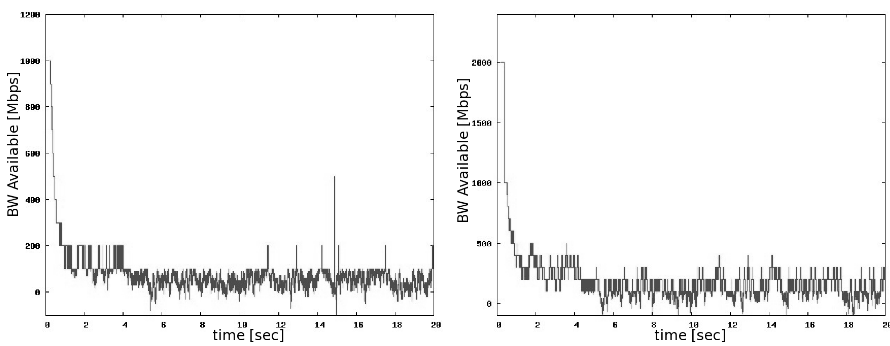

Fig. 6. Available bandwidth. The bottleneck is 1 Gbps and 2.4 Gbps.

The state of routers' queues. The RTT calculated at the sources is always near the minimum RTT. The OBP tries to use the available bandwidth, without filling the routers' queues. This is visible in Fig. 7, which shows the RTT that includes the propagation delay (0.1 ms) and the extra time spent inside routers' queues. This extra time is near zero.

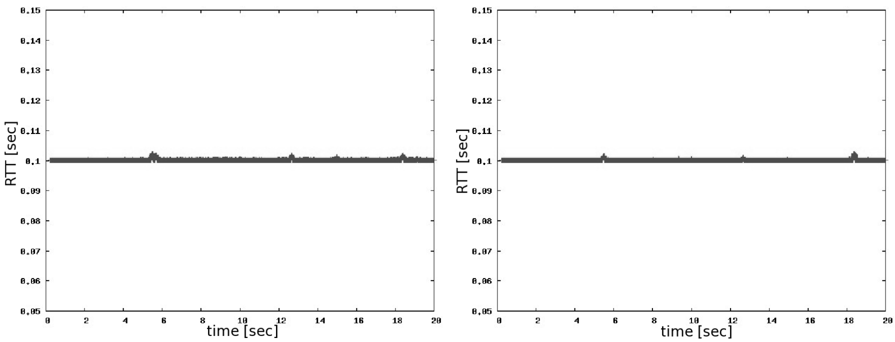

Fig. 7. Packets' RTT. The bottleneck link is 1 Gbps and 2.4 Gbps.

Average completion time. The average completion time is a good metric for elastic flows and for short flows because it gives the necessary time to transfer all data of the flow. Fig. 8 shows that there are not great variations in completion time of flows with identical size. The exceptions are related to the instants of flows' creation and, also, to the network load at those instants. The behavior of the OBP is good with the bottleneck link of 1 Gbps likewise 2.4 Gbps. These results are again analyzed in the next section where they will be compared with the results of other transport protocols.

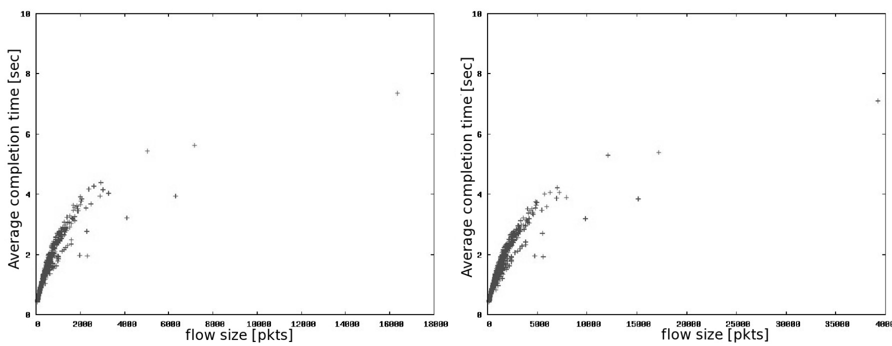

Fig. 8. Average completion time per flow size. The bottleneck link is 1 Gbps and 2.4 Gbps.

4.3 Comparative Results Among the OBP, the TCP Reno, the TCP Reno with Quick-Start, the XCP and the RCP

In this section our goal is to compare the OBP performance with other transport protocols. In these tests we compare the OBP with the TCP Reno, the XCP, the RCP and the TCP Reno with Quick-Start with the request rate equal to 100 KBps. These tests were done using the Pareto distribution to define the flows' size, with mean of 1250 packets and shape 1.8. The bottleneck link was 1 Gbps and the queue' size was 2000 packets. The arrival of flows was by Poisson distribution with lambda 95. The duration of tests was 20 seconds and packets had 1000 bytes. The results of these tests are analyzed through the average completion time per flow, the average throughput and the RTT of each packet.

Completion time and Throughput. From Fig. 9 and Fig. 10, the average completion time for the OBP is more stable than for the TCP Reno, the TCP Reno with Quick-Start and the XCP. This is visible by the smallest variations between flows with near sizes. The longer flows are also concluded more quickly by the OBP than by the TCP Reno, the TCP Reno with Quick-Start and the XCP.

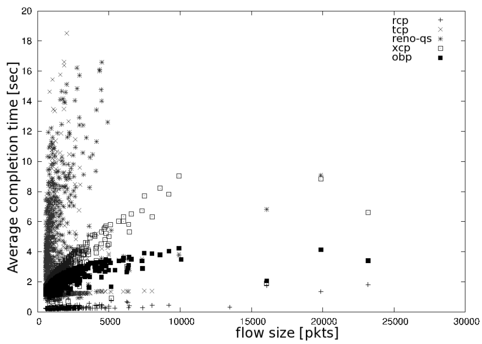

Fig. 9. Average completion time per flow

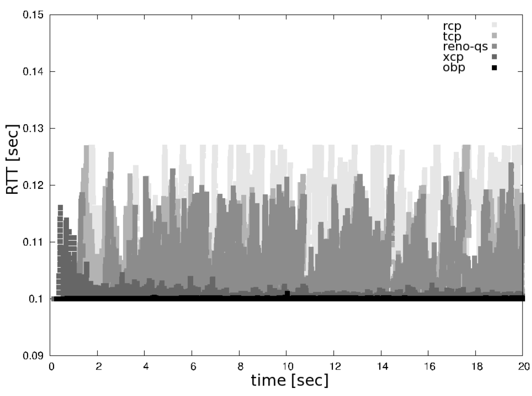

Fig. 10. Packets' RTT

The results of the RCP have to be analyzed together with the packets' RTT results. The RCP is very aggressive and defines high transmission rates. This is visible in Fig. 10 where the packets' RTT is always higher than the propagation delay value. The consequence of this aggressiveness is the packets loss, showed by Fig. 11, where the RCP' throughput is less than the OBP or the XCP.

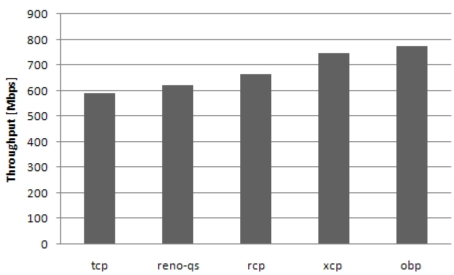

Fig. 11. Average throughput reached in 20 second

5 Conclusions and Future Work

In this paper we present Open Box Protocol (OBP). This solution enables the efficient use of the network capacity because the sources make congestion control decisions based on information about the network state. The OBP gives sources capacities to look inside the network and to make their congestion control decisions, based on the most restricted interface capacity, the available bandwidth, the RTT variations and the heterogeneous transmission means.

We have shown through analysis and experimental evaluation that the OBP has capacity to put information about the network state in the sources. Also, it is visible that the OBP can efficiently use the network bandwidth, keeping the routers' queues near zero occupation. We have equally shown that OBP can have better performance than others congestion control solutions. Moreover, the OBP implementation puts the load processing on the sources side, in opposition to other congestion control approaches, which make congestion control decisions for all flows by the same routers, as the case of the XCP and the RCP. The OBP can be used in networks where there are flows that use other congestion control protocols, because the OBP only needs to receive from routers congestion control information.

As part of our future work, we plan to test the OBP in wireless networks as in networks shared by flows that use different congestion control approaches.

References

1. Dukkipati, N., Kobayashi, M., Zhang-Shen, R., McKeown, N.: sa: Processor Sharing Flows in the Internet. In: IWQoS, Passau, Germany(2005)
2. Floyd, S.: Metrics for the Evaluation of Congestion Control Mechanisms. Internet draft - IETF, Expires (August 2007)
3. Jacobson, V.: Congestion Avoidance and Control. In: ACM SIGCOMM (1988)
4. Floyd, S., Henderson, T.: New Reno Modification to TCP's Fast Recovery. RFC 2582 (1999)
5. Jin, C., Wei, D.X., Low, S.H.: FAST TCP: Motivation, Architecture, Algorithms, Performance. Hong Kong. In: Proceedings of IEEE Infocom (2004)
6. Floyd, S.: HighSpeed TCP for Large Congestion Windows. RFC 3649 (December 2003)

7. Medina, A., Allman, M., Floyd, S.: Measuring the Evolution of Transport Protocols in the Internet. Computer Communications Review (April 2005)
8. Mascolo, S., Casetti, C., Gerla, M., Sanadidi, M., Wang, R.: TCP Westwood: End-to-End Bandwidth Estimation for Efficient Transport over Wired and Wireless Networks. In: Proceedings of ACM Mobicom, Rome, Italy (2001)
9. Salim, J.H., Ahmed, U.: Performance Evaluation of Explicit Congestion Notification (ECN) in IP Networks. RFC 2884 (2000)
10. Ramakrishnan, K., Floyd, S., Black, D.: The Addition of Explicit Congestion Notification (ECN) to IP. RFC 3168 (2001)
11. Floyd, S., Allman, M., Jain, A., Sarolahti, P.: Quick-Start for TCP and IP. RFC 4782 (2007)
12. Falk, A., Pryadkin, Y., Katabi, D.: Specification for the Explicit Control Protocol (XCP). Internet-Draft, Expires (May 9, 2007)
13. Katabi, D., Handley, M., Rohrs, C.: Internet Congestion Control for High Bandwidth-Delay Product Networks. In: Proceedings of ACM Sigcomm, Pittsburgh (August 2002)
14. Stevens, W.: TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms. RFC 2001 (1997)
15. The Network Simulator, http://www.isi.edu/nsnam/ns/
16. Floyd, S., Allman, M.: Specifying New Congestion Control Algorithms. Internet-Draft (December 2006)