Network Working Group                                    Gabor Feher, BUTE
INTERNET-DRAFT                                      Istvan Cselenyi, TRAB
Expiration Date: May 2001                                Peter Vary, BUTE
                                                        Andras Korn, BUTE

November 2000

# Benchmarking Methodology for Routers Supporting Resource Reservation

<draft-feher-bmwg-benchres-method-00.txt>

## 1. Status of this Memo



## 2. Table of contents

## 3. Abstract

The purpose of this document is to define benchmarking methodology measuring performance metrics related to IP routers supporting resource reservation signaling. Beside the definition and discussion of these tests, this document also specifies formats for reporting the benchmarking results.

## 4. Introduction

The IntServ over DiffServ framework [3] outlines a heterogeneous Quality of Service (QoS) architecture for multi domain Internet services. Signaling based resource reservation (e.g. via RSVP [6]) is an integral part of that model. While this significantly lightens the load on most of the core routers, the performance of border routers that handle the QoS signaling is still crucial. Therefore network operators, who are planning to deploy this model, shall scrutinize the scalability limitations in reservation capable routers and the impact of signaling on the forwarding performance of the routers.

An objective way for quantifying the scalability constraints of QoS signaling is to perform measurements on routers that are capable of resource reservation. This document defines a specific set of tests that vendors or network operators can use to measure and report the signaling performance characteristics of router devices that support resource reservation protocols. The results of these tests will provide comparable data for different products supporting the decision process before purchase. Moreover, these measurements provide input characteristics for the dimensioning of a network in which resources are provisioned dynamically by signaling. Finally, these test are applicable for characterizing the impact of control plane signaling on the forwarding performance of routers.

This benchmarking methodology document is based on the knowledge gained by examination of (and experimentation with) several very different resource reservation protocols: RSVP [6], Boomerang [7], YESSIR [8], ST2+ [9], SDP [10], Ticket [11] and Load Control [12]. Nevertheless, this document aspires to compose terms that are valid in general and not restricted to these protocols.

## 5. Existing definitions





A previous document from the authors, "Benchmarking Terminology for Router Supporting Resource Reservation" [4] defines performance metrics and other terms that are used in this document. To understand the test methodologies defined here, that terminology document must be consulted first.

## 6. Methodology

### 6.1 Evaluating the Results

RFC2544 [4] describes considerations regarding the implementation and evaluation of benchmarking tests, which are certainly valid for this test suite also. Namely, the authors intended to create a system from commercially available measurement instruments and devices for the sake of easy implementation of the described tests. Simple test scripts and benchmarking utilities for Linux are publicly available from the Boomerang homepage [13].

During the benchmarking tests, care should be taken for selecting the proper set of tests for a specific router device, since not all of the tests apply to every type of Devices Under Tests (DUTs).

Finally, the selection of the relevant measurement results and their evaluation requires experience and it must be done with an understanding of generally accepted testing practices regarding repeatability, variance and statistical significance of small numbers of trials.

### 6.2 Test Set up

#### 6.2.1 Single Tester Device

The ideal way to perform the measurements is connecting a tester device (or, in short, tester) to both the incoming and outgoing network interfaces of the DUT. The tester sends signaling messages and data traffic to one or more incoming ports of the DUT, while the outgoing network ports of the tested device, where the processed signaling messages and the forwarded packets appear, are connected back to the tester. Thus the tester device is capable to measure performance metrics, such as the signaling message handling time, various traffic forwarding times and the signaling loss. This scenario can be seen in Figure 1 [4]. In this case the tester device is a signaling initiator and a signaling terminator at the same time, while additionally, it originates and terminates the data traffic also.





```
                            +------------+
                            |            |
               +------------|   tester   |<--------------+
               |            |            |               |
               |            +------------+               |
               |                                         |
               |            +------------+               |
               |            |            |               |
               +----------->|    DUT     |---------------+
                            |            |
                            +------------+
```

Figure 1

6.2.2 Two Tester Devices

The benchmarking described in this document can be performed with two tester devices as well, separating the initiator and terminator functionality into two pieces of equipment. In this case the initiator tester device is the driver of the input network interfaces of the DUT, while the second one, the terminator tester device, is connected to the output network interfaces of the tested device measuring the performance metrics on signaling messages and traffic packets leaving the DUT. Figure 2 shows this scenario.

```
+---------+         +-------------+          +-----------+
|         |         |             |          |           |
| sender  |-------->|     DUT     |--------->| receiver  |
|         |         |             |          |           |
+---------+         +-------------+          +-----------+
```

Figure 2

The main benefit of the single tester device measurement setup is that the tester knows the exact time when a signaling message or a data packet enters to the DUT and when it leaves, thus it can calculate the time dependent performance metrics (e.g. signaling message handling time) easily. Using the two testers setup, the testers must be clock synchronized in order to measure performance metrics depending on time differences. Nevertheless, the scalability tests do not require the evaluation of performance metrics; therefore do not depend on the time synchronization.

The main benefit of the two tester scenario is that the load caused by the generation and the evaluation of test flows are shared between the two devices, unlike in the case of single tester setup, where all of the measurement tasks must be done at the same device.

During the benchmarking tests, if the clocks are properly synchronized in the two tester case, both test configurations are suitable to carry out the measurements.

Although the definition of the benchmarking methodologies, later in this document, uses the expressions of "initiator tester" and "terminator tester"; they do not have to be two physically separated





appliances, but in the case of single tester setup, both the initiator tester and the terminator tester refers to the single tester device.

However, the person who performs the tests can choose the tester setup at his or her will, the scenario configuration should always be described properly in the report of the benchmarking results.

6.2.3 Testing Unicast Resource Reservation Sessions

Testing unicast resource reservation sessions requires that the initial tester is connected to one of the networking interfaces of the DUT and the terminator tester is connected to a different networking interface on the tested device.

During the benchmarking tests, the initiator tester must use unicast addresses for data traffic flows and the resource reservation requests must refer to unicast resource reservation sessions. Both data packets and signaling messages transmitted by the DUT must be perceivable for the terminator tester.

### 6.2.4 Testing Multicast Resource Reservation Sessions

Testing multicast resource reservation sessions requires that the initial tester is connected to more than one networking interfaces of the DUT and the terminator tester is connected to more than one network interfaces of the tested device whose are different from the previous ones.

Furthermore, during the measurements, the data traffic flows, originated from the initiator tester, must be sent to multicast addresses and the tester device must request reservations referring to multicast resource reservation sessions. Of course, both data packets and signaling messages transmitted by the DUT must be perceivable for the terminator tester, just like in the case of unicast resource reservation sessions.

Since there are protocols supporting more than one resource reservation schemes for multicast reservations (e.g. RSVP SE/FF/WF); and in a view of the fact that the number incoming and outgoing networking port combinations of the DUT might be almost countless; the benchmarking tests, described here, do not require measuring all imaginable setup situation. Still, routers supporting multicast resource reservations must be tested against the performance metrics and scalability limits on at least one multicast scenario. Moreover, there is a suggested multicast test configuration that consists of a multicast group with four signaling end-points including one traffic originator and three traffic destinations.

The benchmarking test reports taken on DUTs supporting multicast resource reservation sessions always have to consist of the proper multicast scenario definition.





### 6.2.5 Signaling flow

This document often refers to signaling flows. A signaling flow is sequence of signaling messages.

In the case of measurements defined in this document there are two types of signaling flows: First, there is a signaling flow that is constructed from signaling primitives of the same type. Second, there is a signaling flow that is constructed in a special way: the signaling flow is consisted of signaling primitive pairs. Signaling primitive pairs are necessary in situations where one of the signaling primitive make changes in the states of the DUT. In this case, to avoid the effect of state changes, the pair of the signaling primitive restores the modified states in the DUT. A typical example for the second version of the signaling flows is an alternating reservation set-up and tear-down signaling message.

Moreover, the signaling messages should be equally spaced on the time scale when they are forming a signaling flow. This is mandatory in order to obtain measurements that might be repeated later. Since modern resource reservation protocols are designed to avoid message synchronization, thus, equally spaced signaling messages are not unrealistic in the real life.

The signaling flow parameters are the type of the signaling primitive or pair of signaling primitives beside the period time of the signaling messages.

### 6.2.6 Signaling Message Verification

Although, the conformance testing of the resource reservation is beyond the scope of this document, defective signaling message processing can be expected in an overloaded router. Therefore, during the benchmarking tests, when signaling messages are processed in the DUT, the terminator device must validate the messages whether they fully conform to the message format of the resource reservation protocol specification and whether they are the expected signaling messages at the given situation. If any of the messages break the protocol specification then the benchmarking test report must indicate the situation of the failure.

Verifying data traffic packets are not required, since the signaling performance benchmarking of reservation capable routers should not deal with data traffic. For this purpose there are other benchmarking methodologies that verify data traffic during the measurements, like the one described in RFC 2544.

## 6.3 Scalability Tests

Scalability tests are defined to explore the scalability limits of a reservation capable router. This investigation focuses on the scalability limits related only to signaling message handling, examination of the data forwarding engine is not in the scope of this document.

During the scalability tests, no data traffic forwarding is allowed on the DUT.

### 6.3.1 Maximum Signaling Message Burst Size

Objective:
The maximum signaling burst size is the number of the signaling messages in a signaling burst that the DUT is able to handle without signaling loss.

Procedure:
1. Select a signaling primitive or a signaling primitive pair and form a signaling flow. The chosen signaling primitive or primitive pair should be the same during the whole test run. The signaling messages should follow each other back-to-back in the flow and after "n" number of messages the flow should be terminated. In the first test sequence the number "n" should be set to one.

Additionally, all the signaling messages in the signaling flow must be conform to the resource reservation protocol definition and must be parameterized in a way to avoid the signaling message processing errors in the DUT.

2. Send the signaling flow to the DUT and count the signaling messages received by the terminator tester.

3. When the number of sent signaling messages ("n") equals to the number of received messages, the number of messages forming the signaling flow ("n") should be increased by one; and the test sequence has to be repeated. However, if the receiver receives less signaling messages than the number of sent messages, it indicates that the DUT is over on its scalability limit. The measured scalability limit for the maximum signaling message burst size is the length of the signaling flow in the previous test sequence ("n"-1).

In order to avoid transient test failures, the whole test must be repeated at least 30 times and the report should indicate the median of the measured maximum signaling message burst size values as the output of the test. Among the test runs, the DUT should be reset to its initial state.

There are signaling primitives, such as signaling messages indicating errors, which are not suitable for this kind of scalability tests. However, each signaling primitive that is suitable for the test should be investigated.

Reporting format:
The report should indicate the type of the signaling primitive or signaling primitive pair and the determined maximum signaling message burst size.





Note:
In the case of routers supporting multicast resource reservation sessions, the signaling burst can be also formed by sending signaling messages to multiple networking interfaces of the DUT at the same time.

### 6.3.2 Maximum Signaling Load

Objective:
The maximum signaling load is the maximum number of signaling messages within a time unit that the DUT is able to handle without signaling loss.

Procedure:
1. Select a signaling primitive or a signaling primitive pair and form a signaling flow. The chosen signaling primitive or primitive pair should be the same during the whole test run. The period of the signaling flow should be adjusted that exactly "s" number of signaling messages come into view in one second. In the first test sequence the number "s" should be set to one.

Additionally, all the signaling messages in the signaling flow must be conform to the resource reservation protocol definition and must be parameterized in a way to avoid the signaling message processing errors in the DUT.

2. Send the signaling flow to the DUT for at least one minute, and count the signaling messages received by the terminator tester.

3. When the number of sent signaling messages ("s" times the duration of the signaling flow) equals to the number of received messages, the

signaling flow period should be decreased in a way that one more signaling message should fit into a one second interval of the signaling flow ("s" should be increased by one). But, if the receiver receives less signaling messages than the number of sent messages, it indicates that the DUT is over on its scalability limit. The measured scalability limit for the maximum signaling load is the number of signaling messages fitting into one second of the signaling flow in the previous test sequence ("s-1").

In order to avoid transient test failures, the whole test must be repeated at least 30 times and the report should indicate the median of the measured maximum signaling load values as the output of the test. Among the test runs, the DUT should be reset to its initial state.

In the case of this test, there are also signaling primitives which are not suitable for this kind of scalability tests. However, each signaling primitive that is suitable for the test should be investigated just like in the case of the maximum signaling burst size test.





Reporting format:
The report should indicate the type of the signaling primitive or signaling primitive pair and the determined maximum signaling load value.

### 6.3.3 Maximum Session Load

Objective:
The maximum session load is the maximum number of resource reservation sessions that can exist simultaneously in a reservation capable router.

Procedure:
1. Set up "n" number of reservation session in the reservation capable router by sending the appropriate signaling messages to the DUT. In the first test sequence the number "n" should be set to one.

2. In the case of soft-state protocols wait for a specified amount of time ("T") while still maintaining the established reservations with refresh signaling messages. Hard-state protocols can skip this step. Time "T" must be at least as long as the protocol specifies as reservation time out. This waiting is necessary to assure that DUT is

able to refresh the reservations.

3. Check whether all the "n" number of reservations exist in the DUT. When all of them stayed alive, then repeat the test sequence by increasing the number of reservations by one ("n"+1). However, when any of the reservations was dropped by the DUT, then the test sequence cancels and the determined maximum session load is the number of resource reservation sessions set up successfully in the previous test sequence ("n"-1).

In order to avoid transient test failures, the whole test must be repeated at least 5 times and the report should indicate the median of the measured maximum signaling load values as the output of the test. Among the test runs, the DUT should be reset to its initial state.

Reporting format:
The report should indicate determined maximum session load value.

Note:
When the number of reserved sessions grows over a number that counts to a very high value in the given technology conditions, then the test can be canceled and the report can state that the resource reservation protocol implementation performs the maximum number of reservation sessions over that limit (e.g. "Over 10.000 sessions").

Checking the active resource reservation sessions in a reservation capable router might be difficult if the router does not support any interface to monitor its interior states. Lack of such support other methods should be used. One ultimate, but slow method is to send overrated data traffic across all of the resource reservation





sessions and whether the DUT drops the right amount of data traffic, then it means that all the reservation sessions are alive.

### 6.4 Benchmarking Tests

Benchmarking tests are defined to measure the QoS signaling related performance metrics on the resource reservation capable router device.

During the tests the DUT must not bump into its scalability limits. This means that the router must not drop any signaling messages or data packets. In the case of signaling message or data traffic loss, the test must be stopped, and the parameters of the test must be re-adjusted to prevent the DUT to leave its steady state operating range.

During all of the benchmarking tests described here, the initiator tester loads the DUT by sending signaling flows and traffic flows to the terminator device across the DUT. Moreover, the signaling end-points must also assure that the DUT maintains a certain number of resource reservation sessions during the test lifetime.

Every the performance metric is measured under different router load conditions, where this load is a combination of independent load types:

a. Signaling load
b. Session load
c. Premium traffic load
d. Best-effort traffic load

The initiator tester device generates the signaling load on the DUT by sending a signaling flow to the terminator tester. This signaling flow is constructed from a specific signaling primitive or a signaling primitive pair and has the appropriate period parameter.

The session load is generated by the signaling end-point reserving resource reservation sessions in the DUT via signaling. During the test, in the case of soft-state protocols, the initiator tester device must maintain the reservation sessions with refresh signaling messages periodically, when the resource reservation protocol defines it. These reservation sessions should not need to be loaded with data traffic.

The initiator tester device generates the premium traffic load by sending a data traffic flow, which refers to an existing resource reservation session, to the terminator tester across the DUT. The traffic must consist of equally spaced and equally sized data packets. To generate traffic load, it is recommended to use UDP packets, however any other transfer protocol can be used. The premium traffic must be reported by its traffic parameters: data packet size in octets, the calculated bandwidth of the stream in kbps unit and





the transfer protocol type. The data packet size should include both the payload and the header of the IP packet.

The initiator tester device generates the best-effort traffic load by sending a data traffic flow, which does not refer to any resource reservation sessions, to the terminator tester across the DUT. The traffic must consist of equally spaced and equally sized data packets and must be reported by its traffic parameters as it is described in the case of the premium traffic load description.

These four load types have influence on each other from their nature, but during the tests these cross-effects must be minimized. The signaling load must establish as few temporary resource reservations in the DUT as possible. For this reason, when a new resource reservation session is set up in the DUT as a side effect of a signaling message in the signaling flow, the signaling end-points must arrange to restore the number of reservations in the router as soon as possible. Furthermore, signaling messages are realized as data packets in the real word, however during the measurements they are not treated as premium or best-effort traffic.

### 6.4.1 Performing the Benchmarking Measurements

The test methodology is the same for all performance metrics. Moreover, it is also easier and less time-consuming to perform the measurements for all performance metrics at the same time in a test cycles.

The goal is to take measurements on a DUT running a resource reservation protocol implementation under different loaded conditions. The load on the DUT is always the combination of the four load components mentioned before.

Procedure:
The procedure is to load the router with each load component at a desired level and take measurements on all of the performance metrics. Once, the measurements are complete, repeat the test with a different load distribution.

During the test sequences, in order to avoid transient flow behavior influencing the measurements, the measurements should begin after a delay of at least "T" and after the setup of the common load on the DUT. The value of "T" depends on the parameters of the load components and the resource reservation protocol implementations, but, as a rule of thumb, it should be enough for at least 10 packets from the traffic flows and 10 signaling messages from the signaling flow to pass through the DUT and at least one refresh period to expire in the case of soft-state protocols.

During the measurement of the performance metrics in a practical load setup, not just one, but 100 measurement result sets should be collected. The output of the test sequence is the median of the performance metrics measured.

In order to avoid transient test run failures, that may cause invalid results for the entire test, the test run must be repeated at least 10 times and the report should indicate the median of the measured values. Moreover, after each test run the DUT should be reset to its initial state.

To complete the benchmarking tests all applicable signaling primitives should be included in at least one signaling flow that is used for benchmarking purposes.

At first sight, this procedure may look easy to carry out, but in fact there are lots of difficulties to overcome. The following guidelines may help in reducing the complexity of creating a conforming measurement setup.

1. It is reasonable to select different load levels for each load component (load levels) and then measure the performance metrics with all combinations of these individual load levels. Thus, the measurements results can be thought of as a four-dimension table, where each dimension is a load component.

2. The number of different load combinations depends on the number of different load levels within a load component. Working with many different load levels is highly unfeasible and therefore not suggested. Instead, there are proposed levels and parameters for each load component.

The data traffic parameters for the traffic load components have to be selected from generally used traffic parameters. It is recommended to choose a packet size of: 54, 64, 128, 256, 1024, 1518, 2048 and 4472 bytes (these are the same values that are used in RFC 2544 that introduces methodology for benchmarking network interconnect devices). Additionally, the size of the packets should always remain below the MTU of the network segment. The packet rate is recommended to be one of 1, 10, 100 or 1000 packets/s. Since the number of combinations for these traffic parameters is still large, the highly recommended values are 64, 128 and 1024 bytes for the packet size and 10 and 1000 packets/s packet rate. These values adequately represent a wide range of traffic types common in today's Internet. Thus, there are 6 different load levels for the traffic load generation.

The number of session load levels should be at least 4 and the actual value of the session load should be equally distributed between 1 and the maximum session load value.

The number of signaling load levels should be at least 4 as well, and the actual value of the signaling load should be equally distributed between 1 and the maximum signaling load value.

3. The load component levels should be extended by the situation, when there is no outcome of the particular load component. This means that there is no traffic flow in the case of traffic load components;





or there is no signaling flow in the case of the signaling load component; or there are no resource reservation sessions in the case of the session load component.

Including these levels, the suggested number of test are: 5 (signaling load) * 5 (session load) * 7 (premium traffic load) * 7 (best-effort traffic load).

Reporting format:
As the whole report description requires a four-dimension table, which is hard to visualize for a human being, therefore the results are extracted into ordinary two-dimensional tables. Each table has two fixed load component quantities and the other two load component levels are the row and column for the table. Naturally, these load component levels must be described properly. Following the suggested load levels, 25 different tables should be prepared to describe the benchmarking results.

On set of such tables describe the benchmarking results when a specified signaling primitives compose the signaling flow used to generate the signaling load. There should be one set of tables for each signaling primitive or signaling primitive pair.

Note:
Of course in the case of multicast resource reservation sessions, the combination number of the different multicast scenarios multiplies the number benchmarking tests also.

## 7. Acknowledgement

The authors would like to thank the following individuals for their help in forming this document: Joakim Bergkvist and Norbert Vegh from Telia Research AB, Sweden, Balazs Szabo, Gabor Kovacs from High Speed

Networks Laboratory of BUTE.

## 8. References

[1] S. Bradner, "Benchmarking Terminology for Network Interconnection Devices", RFC 1242, July 1991

[2] R. Mandeville, "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998

[3] Y. Bernet, et. al., "A Framework For Integrated Services Operation Over Diffserv Networks", Internet Draft, May 2000, <draft-ietf-issll-diffserv-rsvp-05.txt>

[4] G. Feher, I. Cselenyi, A. Korn, P. Vary, "Benchmarking Terminology for Routers Supporting Resource Reservation", Internet Draft, November 2000, <draft-feher-benchres-method-01.txt>





[5] S. Bradner, J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, March 1999

[6] B. Braden, Ed., et. al., "Resource Reservation Protocol (RSVP) - Version 1 Functional Specification", RFC 2205, September 1997.

[7] J. Bergkvist, I. Cselenyi, "Boomerang Protocol Specification", Internet Draft, June 1999, <draft-bergkvist-boomerang-spec-00.txt>

[8] P. Pan, H. Schulzrinne, "YESSIR: A Simple Reservation Mechanism for the Internet", Computer Communication Review, on-line version, volume 29, number 2, April 1999

[9] L. Delgrossi, L. Berger, "Internet Stream Protocol Version 2 (ST2) Protocol Specification - Version ST2+", RFC 1819, August 1995

[10] P. White, J. Crowcroft, "A Case for Dynamic Sender-Initiated Reservation in the Internet", Journal on High Speed Networks, Special Issue on QoS Routing and Signaling, Vol 7 No 2, 1998

[11] A. Eriksson, C. Gehrmann, "Robust and Secure Light-weight

Resource Reservation for Unicast IP Traffic", International WS on QoS'98, IWQoS'98, May 18-20, 1998

[12] L. Westberg, Z. R. Turanyi, D. Partain, Load Control of Real-Time Traffic, A Two-bit Resource Allocation Scheme, Internet Draft, April 2000, <draft-westberg-loadcntr-03.txt>

[13] Boomerang Team, "Boomerang homepage - Benchmarking Tools", http://boomerang.ttt.bme.hu

## 9. Authors' Addresses:

Gabor Feher
Budapest University of Technology and Economics (BUTE)
Department of Telecommunications and Telematics
Pazmany Peter Setany 1/D, H-1117, Budapest,
Phone: +36 1 463-3110
Email: feher@ttt-atm.ttt.bme.hu

Istvan Cselenyi
Telia Research AB
Vitsandsgatan 9B
SE 12386, Farsta
SWEDEN,
Phone: +46 8 713-8173
Email: istvan.i.cselenyi@telia.se

Andras Korn
Budapest University of Technology and Economics (BUTE)
Institute of Mathematics, Department of Analysis





Egry Jozsef u. 2, H-1111 Budapest, Hungary
Phone: +36 1 463-2475
Email: korn@math.bme.hu

Peter Vary
Budapest University of Technology and Economics (BUTE)
Department of Telecommunications and Telematics
Pazmany Peter Setany 1/D, H-1117, Budapest, Hungary
Phone: +36 1 463-3110
Email: vary@ttt-atm.ttt.bme.hu