A Cooperative Network Monitoring Overlay

Vasco Castro, Paulo Carvalho, and Solange Rito Lima

University of Minho, Department of Informatics, 4710-057 Braga, Portugal
e-mail: pmc@di.uminho.pt

Abstract. This paper proposes a flexible network monitoring overlay which resorts to cooperative interaction among measurement points to monitor the quality of network services. The proposed overlay model, which relies on the definition of representative measurement points, the avoidance of measurement redundancy and a simple measurement methodology as main design goals, is able to articulate intra- and inter-area measurements efficiently. The distributed nature of measurement control and data confers to the model the required autonomy, robustness and adaptiveness to accommodate network topology evolution, routing changes or nodes failure. In addition to these characteristics, the avoidance of explicit addressing and routing at the overlay level, and the low-overhead associated with the measurement process constitute a step forward for deploying large scale monitoring solutions. A JAVA prototype was also implemented to test the conceptual model design.

Key words: Network Monitoring; Quality of Service; Overlay Networks

1 Introduction

Monitoring of large networks raises multiple challenges regarding scalability, robustness and reliability of measurements. A monitoring model that captures the real network behaviour but that only works on small topologies is of limited applicability in today's networks. Therefore, in a monitoring system, it is necessary to find a compromise among all design goals contributing to a globally scalable and representative monitoring solution. It is known that monitoring systems where a single point is responsible for gathering and processing measurements obtained throughout the network suffer from severe scalability and robustness limitations. To address this problem, distributed solutions where monitoring data is collected and processed at each measurement point (MP) have been proposed. For instance, solutions based on active edge-to-edge measurements provide a straightforward way of measuring service quality, however, the potential interference of cross probing among boundary nodes on network behaviour needs to be carefully considered.

To reduce network overhead and improve spatial coverage, it is important to identify the most representative and critical network points in order to obtain an overall view of the network status involving only a subset of MPs. Resorting to composition of metrics between these MPs, i.e. through concatenation of partial metrics, the interference on network operation can be reduced, avoiding redundant measurements in overlapping links. The composition of metrics also allows observing trends, being more informative as a result of the underlying metric partitioning scheme.

In this context, this paper proposes a collaborative network monitoring overlay which resorts to the cooperation between representative MPs strategically located in the network to compute performance and quality metrics both intra-area and end-to-end. The aim is to pursue a flexible, scalable and accurate monitoring overlay solution that simplifies and systematises the cumulative computation of metrics by involving only a subset of network nodes.

This paper is organised as follows: related work is discussed in Section 2, the proposed monitoring model and its components are described in Section 3, the model prototype is presented in Section 4, the main key points and open issues of the solution are highlighted in Section 5 and the conclusions are summarised in Section 6.

2 Related Work

Active monitoring carried out on an edge-to-edge basis, i.e., between network boundaries, is particularly suitable for monitoring network performance and quality of service (QoS) [1]. This approach improves scalability as only edge nodes are involved in the monitoring process, removing the complexity of monitoring tasks from the network core. The use of synthetic traffic injected in the network for measurement purposes simplifies the estimation of metrics such as delay, loss, available bandwidth [2–4, 1]. Nevertheless, intrusive traffic may be significant in network domains involving a large number of boundary nodes. Active hop-by-hop monitoring aims to reduce the amount of synthetic traffic of active edge-to-edge measurements. Considering that in edge-to-edge probing, probes from distinct pair of edges may cross the same links, hop-to-hop monitoring strategies try to avoid repeating probes in those links. However, capturing network behaviour combining hop-by-hop measures is not an efficient and easy solution as it involves : (i) a high-degree of metrics' concatenation; (ii) monitoring agents in all network nodes; and (iii) additional traffic in the network for reporting metrics to management stations. To reduce the amount of data exchanged between management stations and MPs, several solutions have been pointed out, namely the use of flow aggregation [5], statistical summarisation [6] and network thresholds crossing alerts [7].

Inferring the traffic load of each topological link resorting only to measures of traffic entering and leaving the network, in addition to routing information, has been matter of study within the network tomography research area [8, 9]. Tomography concepts continue to deserve significant attention for estimating distinct aspects of network behaviour, including QoS and fault diagnosis [10–13]. In [14], network tomography is applied to the definition of a monitoring overlay, which resorts to a subset of the topology links (overlay links) to infer packet loss ratio in all network nodes.

Taking in consideration the mentioned strategies, this study proposes a network monitoring overlay solution which resorts to representative MPs to compute performance and quality metrics both intra- and inter-area. Performing network monitoring through representative and collaborative MPs allows to define a virtual monitoring topology based on a cumulative approach for multi-metric computation with reduced overhead.

3 A Cooperative Monitoring Overlay

The proposed model relies on a collaborative participation of representative MPs acting as peers, each one contributing with a disjoint measure component to the evaluation of a global measure. Achieving a measurement between any two points of the network in distinct administrative entities implies the cooperation between different areas regarding evaluating a final metric. Thus, end-to-end measurements are obtained through the aggregation of metrics calculated in each of the network areas involved.

Figure 1 illustrates the monitoring overlay network and the underlying physical topology. The overlay network consists of representative MPs and these are the only players taking part in the measurement process. Each MP in the overlay is expected to store the measurements to its neighbouring MPs. Thus, measurement data is distributed and stored throughout the overlay network. Based on a monitoring request, each MP in the measurement path provides the required measures for aggregation in order to calculate a set of metrics between any specified MPs. This distributed approach also has the advantage of avoiding the existence of a single point of failure. Distributing measurement data over several MPs also enables a rapid recovery of the measurement process by bringing alternative MPs in the process of rebuilding the measurement path in case of routing or network topology changes. Note that these changes do not necessarily imply a change in the overlay topology.

The proposed model allows measurements at two levels: Intra-area and Inter-area. Intra-area measurements are carried out on a regular time basis to ensure that MPs in the same

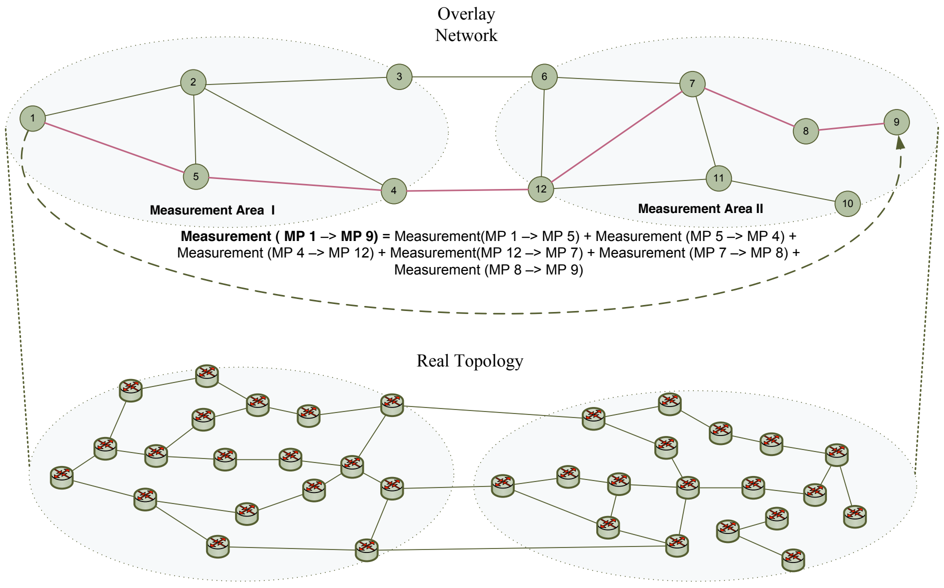

Overlay
Network
1
2
3
4
5
6
7
8
9
10
11
12
Measurement Area I
Measurement Area II
Measurement (MP 1 –> MP 9) = Measurement(MP 1 –> MP 5) + Measurement (MP 5 –> MP 4) +
Measurement (MP 4 –> MP 12) + Measurement(MP 12 –> MP 7) + Measurement (MP 7 –> MP 8) +
Measurement (MP 8 –> MP 9)
Real Topology

Fig. 1. Example of measurement between different administrative areas

area have a clear view of network status and quality of service. An MP may, at anytime, send or exchange measurement data between itself and any other MP within its area. Thus, by retrieving data from multiple MPs in the area and using composition of metrics, it is possible to calculate the value of a metric for a given measurement path. Thus, each MP stores information on the level of quality of service to its neighbouring MPs. Inter-area measurements are performed through the composition of the metrics resulting from intra-area measurements. Conversely to intra-area operation, this type of measurement does not need to be performed continuously, but on request. This process can be triggered, for example, by an application signalling process to assess the communication path before establishing an end-to-end session crossing different network areas.

3.1 Model Operation

As mentioned before, measurement of multiple metrics can be carried out between any two MPs in the overlay network. This section presents a description of the phases involved in the measurement process, assuming that no optimisation tasks (e.g. caching or metrics' composition) are performed. The process, being sender-oriented, is rather simple and effective: an entity requiring measurement information issues a *Measurement Request* and, on success, will receive a *Measurement Report*.

Measurement Request - Initially, a monitoring entity sends a message to the initial MP indicating that it needs to obtain a set of metrics between a pair of MPs. For the topology in Figure 1, the measurement process takes place between MP1 and MP9. Upon receiving the request, MP1 sends a specific packet request for measurement purpose across the overlay network. Each MP in the overlay path will intercept this packet and attach measurement data between itself and the upstream MP, before sending it to the downstream MP. Figure 2 illustrates MP5 receiving the request and forwarding it after adding the measurement data between MP1 and MP5. This process is repeated until the destination is reached, i.e., each MP will successively attach its measurement data along the overlay, as shown in Figure 3. The final MP or the destination, upon receiving the packet measurement request, will add measurement data corresponding to the last segment of the path.

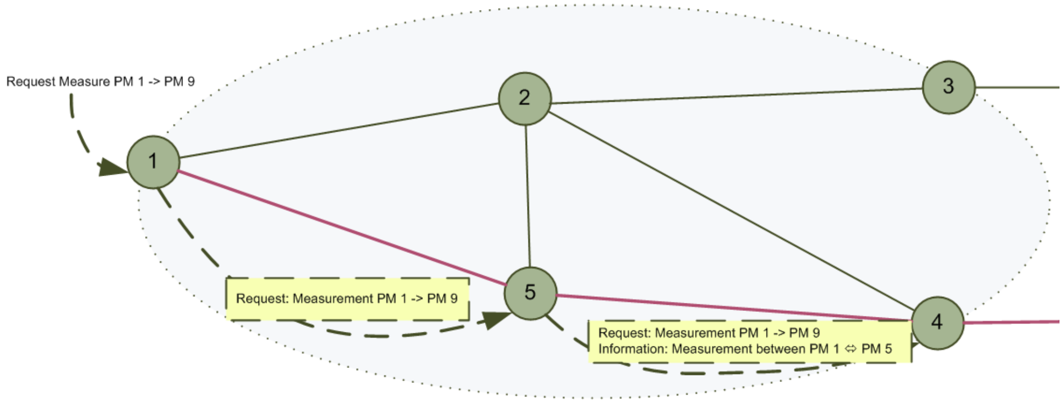

Request Measure PM 1 -> PM 9
1
2
3
5
4
Request: Measurement PM 1 -> PM 9
Request: Measurement PM 1 -> PM 9
Information: Measurement between PM 1 ⇔ PM 5

Fig. 2. Example of MP5 handling a measurement request

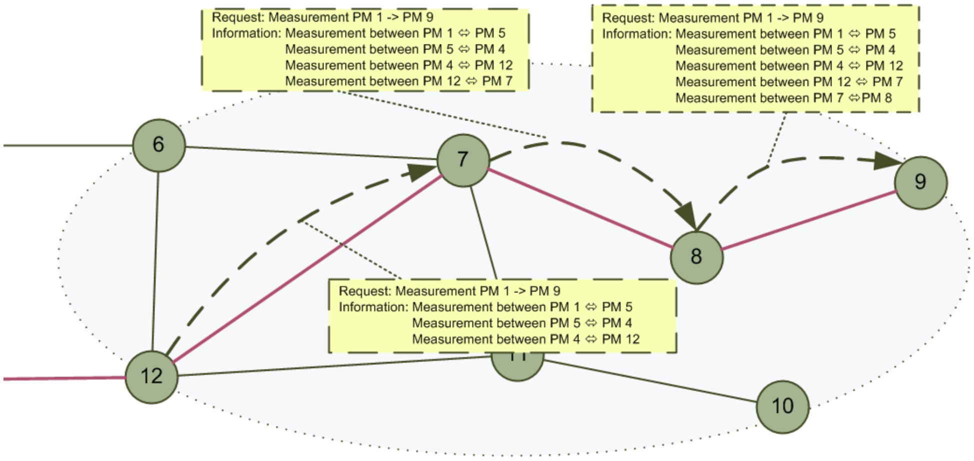

Request: Measurement PM 1 -> PM 9
Information: Measurement between PM 1 ⇔ PM 5
Measurement between PM 5 ⇔ PM 4
Measurement between PM 4 ⇔ PM 12
Measurement between PM 12 ⇔ PM 7
Request: Measurement PM 1 -> PM 9
Information: Measurement between PM 1 ⇔ PM 5
Measurement between PM 5 ⇔ PM 4
Measurement between PM 4 ⇔ PM 12
Measurement between PM 12 ⇔ PM 7
Measurement between PM 7 ⇔PM 8
6
7
9
8
Request: Measurement PM 1 -> PM 9
Information: Measurement between PM 1 ⇔ PM 5
Measurement between PM 5 ⇔ PM 4
Measurement between PM 4 ⇔ PM 12
12
10

Fig. 3. Measurement process across multiple MPs

Measurement Report

Once the measurement in the last MP is obtained, the resulting report message is sent back to the initial MP with the collected measurement data (see Figure 4). At this point, the initial MP is able to compose the required metrics in order to obtain the end-to-end (MP-to-MP) measurement view. This operation can assume distinct cumulative functions (additive, multiplicative, max-min, etc.) depending on the nature of the metric being evaluated.

In practice, this measurement operation can be considerably simplified as area border MPs (e.g. MP 12 in Figure 1) may already have up-to-date measurements from the remaining measurement path. This allows an immediate reply from that MP to the measurement requester, reducing measuring latency significantly. This process can be further improved through proper pro-active metrics dissemination among inter-area MPs.

One challenge of the present model is to identify the representative MPs. Although several works target this topic [15, 14, 16], this aspect requires further study. These issues will be revisited in Section 5.

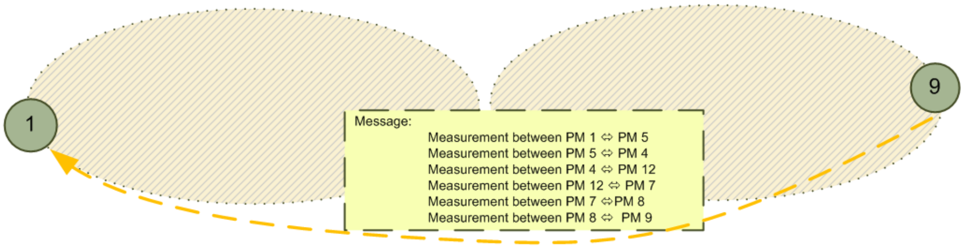

1
9
Message:
Measurement between PM 1 ⇔ PM 5
Measurement between PM 5 ⇔ PM 4
Measurement between PM 4 ⇔ PM 12
Measurement between PM 12 ⇔ PM 7
Measurement between PM 7 ⇔PM 8
Measurement between PM 8 ⇔ PM 9

Fig. 4. Example of a Measurement Report

4 The Implemented Prototype

4.1 Model Components

To test the conceptual model design goals, a model prototype was implementated in Java and MySQL for databases support. The prototype includes four main components: (i) the "Measure Requestor"; (ii) the "Packet Interceptor"; (iii) the "Measure Processor"; and (iv) the "Measure Receiver". Figure 5 illustrates the main interactions among these components.

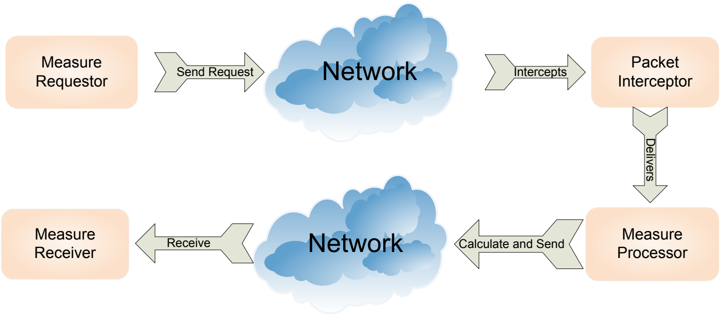


Fig. 5. Interaction of model components

Measure Requestor - This component is responsible for initiating the measurement process between two MPs. In the developed prototype, this is a command line application that receives as parameters, the source and destination MPs, and the set of metrics to measure.

Packet Interceptor - This component is responsible for capturing measurement packets. These packets are differentiated in the network through the use of `router alert` option within IPv4 header, avoiding packet processing at upper protocol layers. In a Linux router, this can be accomplished resorting to `iptables` and proper rules to verify the option `router alert` (it requires the extension `xtables-addons`), intercepting, in this way, the measurement packets. Captured packets are taken from kernel to user space (through `libnetfilter_queue`) for processing at MPs. The use of `router alert` option avoids the use of explicit MP addressing, allowing for a more flexible overlay topology definition.

Measure Processor - This component is responsible for processing and concatenating measurement data, playing a relevant role in the model prototype due to its functionality. Once a packet request is intercepted at an MP, this component detects the new request, validates it and appends the required metrics to the measurement packet. This process involves identifying the latter upstream MP before adding its measurement contribution. Then, the component builds an IP packet setting the `router alert` option, updates the data payload accordingly and sends the packet to the downstream MP. Once the last MP is reached, the "Measure Processor" opens a TCP connection to the initial MP for sending the aggregate measurement outcome. Figure 6 depicts the modules within "Measure Processor" and how they interact to provide this component functionality.

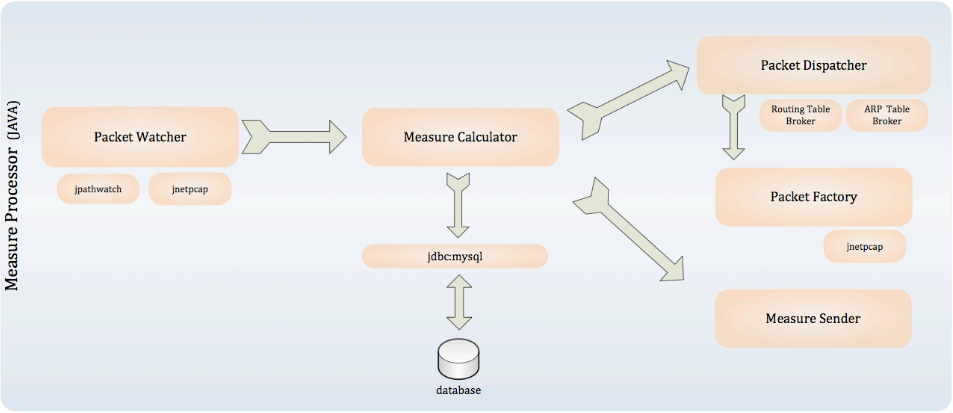


Fig. 6. The component Measure Processor

Measure Receiver - When the measurement process starts, a measurement packet request is issued and, simultaneously, the request is stored in a database, remaining in listening mode on an UDP port. Upon receiving the corresponding measurement result, this component updates the database for the corresponding request. As mentioned, the measurement reply aggregates all the metrics collected along the overlay measurement path.

4.2 Model Primitives

In the proposed model, the measurement primitives are structured in XML (Extensible Markup Language). Although XML structuring tends to be verbose, characteristics such as its universal format, self-descriptive nature, simplicity and extensibility, and the numerous available APIs for manipulating it, are a clear advantage.

A measurement message, following a simple format, comprises two parts or nodes: "Measure Request" ("mr") and "Measure Response" ("mrp"), as illustrated in Figure 7. The first part, generated by the starting MP, defines a header specifying the initial request for measurement. Thus, the node "mr" is composed of the following sub-nodes:

(i) hs (host source) - network address identifying the source MP;
(ii) hd (host destination) - network address identifying the destination MP;
(iii) id (identification) - key identifier associated with the request for measurement;
(iv) ms (measures) - set of metrics.

The second part of the message, consisting of node "mrp", allows MPs in the measurement path for appending measurement data after intersecting the measurement request. Each MP provides information regarding the upstream MP, the current MP, a timestamp, and the values for the metrics defined in node "ms". The structure of node "mpr" is as shown in Figure 8.

For the sake of clarity, a simple example of an XML measurement request for packet loss and delay between MPs 192.168.99.100 and 192.168.117.101 is provided in Figure 9.

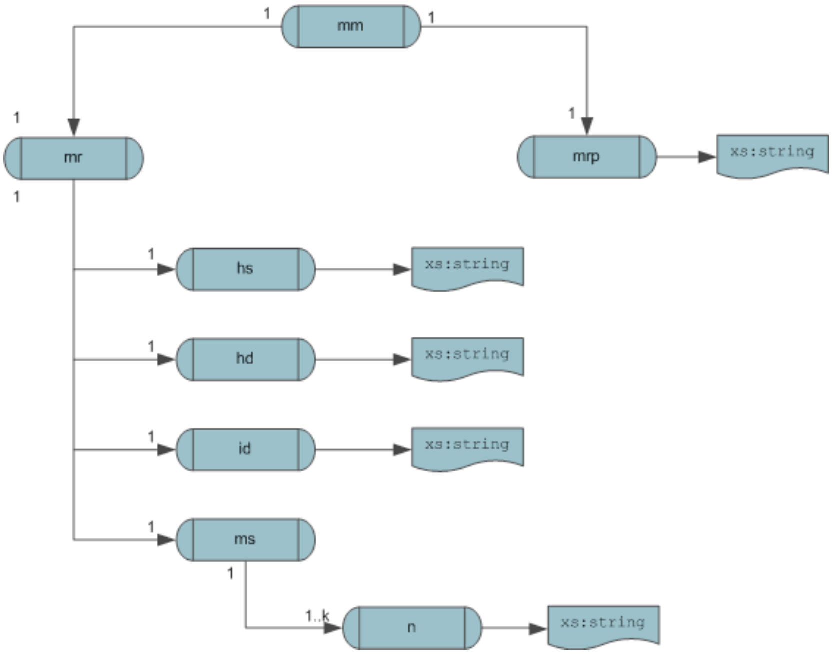
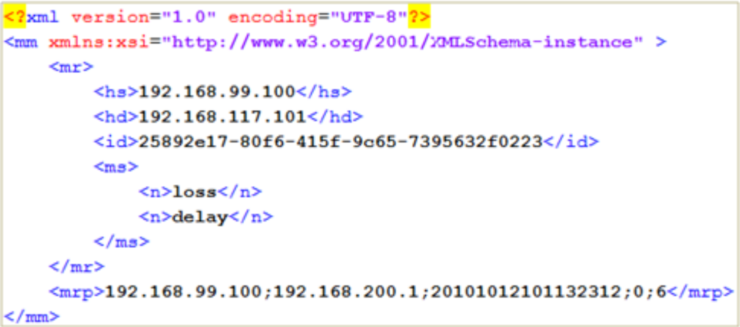


Fig. 7. Measurement message in XML

```
node1;node2;timestamp;metric1;metric2;...metric(k)|
node2;node3; timestamp;metric1;metric2;...metric(k)|
...|...|node(n-1);node(n);timestamp;metric1;metric2;...metric(k)
```

Fig. 8. Structure of node "mrp"



```
<?xml version="1.0" encoding="UTF-8"?>
<mm xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
    <mr>
        <hs>192.168.99.100</hs>
        <hd>192.168.117.101</hd>
        <id>25892e17-80f6-415f-9c65-7395632f0223</id>
        <ms>
            <n>loss</n>
            <n>delay</n>
        </ms>
    </mr>
    <mrp>192.168.99.100;192.168.200.1;20101012101132312;0;6</mrp>
</mm>
```

Fig. 9. Example of a Measurement Request between neighbouring MPs

4.3 Testing the prototype

As proof-of-concept of the present model, a virtualised network topology (using VMware) was considered for testing the proposed solution. Figure 10 illustrates a simple network monitoring overlay including two distinct monitoring areas and three representative MPs (MP1, MP2 and MP3). As expected only these nodes detect measurement requests and act accordingly. The virtual machines 2 and 4 run IMUNES (Integrated Network Topology Emulator / Simulator) to emulate common IPv4 backbones, running OSPF as routing protocol. This virtualised testbed allowed to carry out preliminary tests to validate the full life-cycle of an application measurement request, from its occurrence to the final response, reporting the corresponding measurement data in XML.

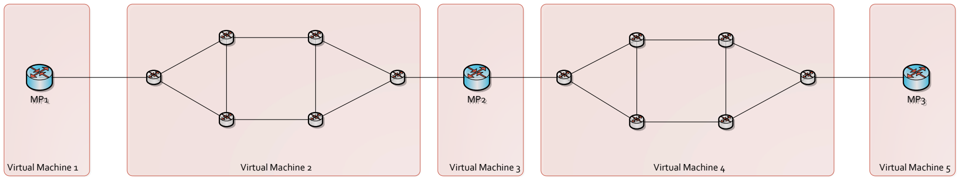


Fig. 10. Virtualised network for test purposes

5 Model Key Points and Open Issues

This section highlights the proposed model key points regarding its design and functionality and discusses open issues that may contribute positively to ongoing and future developments.

5.1 Key points

The present model proposal for a cooperative network monitoring overlay, taking advantage of decentralising the control and data plane, exhibits several key properties, namely:

Autonomy - Each MP is responsible for maintaining its own measurements, providing them on request. Therefore, its location does not need to be pre-determined, conferring a high-degree of decentralisation to the model. The decentralisation inherent to the proposed model allows for a high-degree of autonomy as all MPs only rely on themselves upon receiving a measurement request. All information required to satisfy a measurement request is contained in each of the corresponding MPs. The autonomy degree can be improved, if each MP is aware of representative MPs in the same measurement area. This would allow to take more advantage of metrics' composition, providing also a better response in case of MP failure, for instance, through auto-configuration.

Robustness to failure - As mentioned above, measurement data are not centralised on a single network point being disseminated throughout MPs in the overlay, thereby ensuring that if an MP fails: (i) it does not represent the loss of all measurement information, only monitoring between that MP and its neighbouring MPs is affected; (ii) there is no need for reconfiguring the overlay network as the inclusion or exclusion of MPs is transparent to the network entities that wish to obtain an MP-to-MP (or end-to-end) measurement.

Adaptability - Topology changes do not require the reconfiguration of the entire overlay network, or intervention in all MPs. In fact, upon a topology change, the only need is to reconfigure neighbouring adjacencies so that the existing MPs take into account the new MPs.

Scalability - Attending to the nature of the model, expanding the overlay topology does not imply a direct increase in monitoring traffic. Topology growth only leads to large payloads of measurement request packets, as consequence of an eventual increase in the number of MPs, i.e. for a monitoring request traversing a longer measurement path.

Low overhead - The solution resorts to special-purpose probing packets requiring low processing from the network equipment, therefore, the interference of monitoring with the normal network operation is minimised. The overhead of reporting measurements to a central management or monitoring entity is also avoided, as measurements occur on demand. For large networks, fragmentation of measurement requests may however occur, as discussed in the following section.

End-to-end capability - The implemented prototype demonstrated that it is possible to build up an end-to-end or any other MP-to-MP combination based on local measurements.

5.2 Open issues

Location of representative MPs - In the proposed model, as in real network operation, there is clear added-value for having MPs on (or near to) area border routers. This results from their strategic location both from technical and administrative perspectives. However, the selection of representative MPs inside a measurement area requires a deep analysis of aspects such as the centrality of MPs, (overlapping) routing paths and aggregate traffic behaviour in order to devise a suitable set of representative MPs. The challenge lays on finding the minimal set of MPs able to provide the most representative and accurate monitoring view of the area. An equivalent study can also be carried out in the inter-area context.

Metrics composition and dissemination - The process of metrics' dissemination and composition deserves further development. In particular, combining a pro-active approach of disseminating metrics inside a measurement area with the possibility of avoiding overlapping measurement paths, the measurement latency and overhead may be considerably reduced.

Fragmentation - IP fragmentation of a measurement request may occur when the number of MPs in a measurement path increases. This problem can be avoided if fragmentation is handled within the measurement layer. This can be easily achieved fragmenting a measurement request, e.g. per metric under evaluation, if required. Alternatively, the use of measurement payload compression may also remove the need for fragmentation.

6 Conclusions

This paper has presented innovative research work regarding the definition of a network monitoring overlay which resorts to a cooperative interaction among representative MPs to monitor the quality of network services. In the proposed model, measurement overhead and redundancy are reduced through the composition of metrics from non-overlapping measurement paths, both intra- and inter-area. This aspect along with the ability of accommodating network topology and routing changes aim to contribute to a scalable and flexible end-to-end monitoring solution. A JAVA prototype has been implemented to test the conceptual design goals of the model. Future work will be focused on tackling the open issues identified above and performing large scale monitoring tests.

References

1. Habib, A., Khan, M., Bhargava, B.: Edge-to-edge measurement-based distributed network monitoring. Computer Networks **44** (2004) 211–233
2. Blefari-Melazzi, N., Femminella, M.: Measuring the edge-to-edge available bandwidth in a DiffServ domain. Int. J. Netw. Manag. **18** (2008) 409–426
3. Duffield, N., Lo Presti, F., Paxson, V., Towsley, D.: Inferring link loss using striped unicast probes. In: INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Volume 2. (2001) 915 –923 vol.2
4. Jain, M., Dovrolis, C.: End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput. IEEE/ACM Trans. Netw. **11** (2003) 537–549
5. Lin, Y.J., Chan, M.C.: A scalable monitoring approach based on aggregation and refinement. IEEE JSAC **20** (2002)
6. Asgari, A.H., Egan, R., Trimintzios, P., Pavlou, G.: Scalable monitoring support for resource management and service assurance. IEEE Network **18**(6) (2004) 6–18
7. Wuhib, F., Stadler, R., Clemm, A.: Decentralized service-level monitoring using network threshold crossing alerts. Communications Magazine, IEEE **44**(10) (2006) 70 –76
8. Vardi, Y.: Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data. Journal of the American Statistical Association **91**(433) (1996) 365–377
9. Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S., Diot, C.: Traffic Matrix Estimation: Existing Techniques and New Directions. In: ACM SIGCOMM. (2002)
10. Gu, Y., Jiang, G., Singh, V., Zhang, Y.: Optimal probing for unicast network delay tomography. In: INFOCOM'10, Piscataway, NJ, USA, IEEE Press (2010) 1244–1252
11. Burch, H., Chase, C.: Monitoring link delays with one measurement host. SIGMETRICS Perform. Eval. Rev. **33** (2005) 10–17
12. Arya, V., Duffield, N., Veitch, D.: Temporal Delay Tomography. In: INFOCOM. (2008) 276–280
13. Huang, Y., Feamster, N., Teixeira, R.: Practical issues with using network tomography for fault diagnosis. SIGCOMM Comput. Commun. Rev. **38** (2008) 53–58
14. Chen, Y., Bindel, D., Katz, Y.H.: Tomography-based Overlay Network Monitoring. In: in ACM SIGCOMM Internet Measurement Conference (IMC), ACM Press (2003) 216–231
15. Ratnasamy, S., Handley, M., Karp, R.M., Shenker, S.: Topologically-Aware Overlay Construction and Server Selection. In: INFOCOM. (2002)
16. Ni, J., 0002, H.X., Tatikonda, S., Yang, Y.R.: Network Routing Topology Inference from End-to-End Measurements. In: INFOCOM. (2008) 36–40