

On the experimentation of the novel GCMR multicast routing in a large-scale testbed

Davide Careglio*, Dimitri Papadimitriou†, Fernando Agraz*, Sahel Sahhaf‡, Jordi Perelló*, Wouter Tavernier‡

*CCABA - Universitat Politècnica de Catalunya, Barcelona, 08034 Spain

†Alcatel-Lucent Bell Labs, Antwerp, Belgium

‡iMinds, Ghent, B-9050 Belgium

I. INTRODUCTION AND MOTIVATION

Originally defined in the 90s, multicast is nowadays (re)gaining interest given the increasing popularity of multimedia streaming/content traffic and the explosion of cloud services. In fact, multicast yields bandwidth savings complementing cached content distribution techniques and its potential benefits have been verified by studies several times since then (see e.g. [1]). By multicast routing, we refer to a distributed algorithm that, given a group identifier, allows any node to route multicast traffic to a group of destination nodes, usually called multicast group. To enable one-to-many traffic distribution, the multicast routing protocol configures the involved routers to build a (logical) delivery tree between the source and the multicast group, commonly referred to as the Multicast Distribution Tree (MDT). Nevertheless, the scaling problems faced in the 90s still remain mostly unaddressed and worst-case projections predict indeed that routing engines could have to process and maintain in the order of 1 million active routes within the next 5 years [2].

During last decade, some multicast routing schemes have been standardized but only the Sparse Mode (SM) and the Single Source Multicast (SSM) variant of Protocol Independent Multicast (PIM-SM [3] and PIM-SSM [4] respectively) have been deployed in the context of IPTV systems for routing multicast streams between VLANs, subnets or access networks [5]. However, the adoption of inter-domain multicast has failed, as it relies on an overlay routing executing on top of the unicast routing topology, which suffers from the same scaling limits. Additional issues are: i) the indirection added by the multicast address space, ii) the limits of shared trees between domains, iii) management and security complexity, iv) the limited number of applications making use of one-to-many connectivity via Internet multicast routing, and many others. A complete analysis of the deployment issues for the IP multicast routing architecture can be found in [6].

As part of the work conducted in the EULER FP7-project [7], we designed the Greedy Compact Multicast Routing (GCMR) scheme [8]. In this demonstration, we exhibit the successful operation of GCMR in the context of inter-domain routing over a large-scale network topology and compare its performance with the standard PIM protocol.

II. GCMR IN BRIEF

The main objective of GCMR is to minimize (i.e., to compact) the routing table size at each router by taking local (i.e., greedy) decisions at expenses of i) routing packets on paths

with relative small deviation compared to the optimal tree; ii) increasing the number of messages required to create the MDT. Instead of using unicast topology storage information to derive the multicast routing entries (as in the case of all multicast routing protocols such as PIM), in GCMR, the information needed to reach a given multicast source S is acquired by means of a two-stage search process. The algorithm search process is triggered whenever a node decides to join a source-specific multicast group G , denoted by (S,G) . In addition, it can be triggered whenever a failure occurs and part of the MDT needs to be restored. It is worth mentioning that the information exchanged in the search process between the nodes is limited to the (S,G) identifier, a sequence identifier, a branching cost metric and a couple of code/subcode fields to identify the type of message; neither topological nor confidential information needs to be disseminated. A detailed description of the GCMR algorithm can be found in [8], theoretical performance analysis and simulation comparison with other major multicast routing paradigms are documented in [9].

As GCMR does not rely on any unicast routing protocol, it can work together with any addressing scheme like IPv4, IPv6 or even geometric coordinates. In addition, GCMR can be implemented directly in any host, as its scope is not limited to routers, not requiring any host-router protocol like IGMP. Therefore, to the authors knowledge, it is the first name-independent, receiver-initiated, dynamic, distributed, end-to-end multicast routing algorithm. Finally, although GCMR has been designed to improve scalability of multicast routing in inter-domain environments, it can perform in other environments where routing scalability is also a main issue and only limited topology/routing information is available.

III. GCMR DEMONSTRATION

In this demonstration, we execute a prototype of the GCMR multicast scheme and evaluate its functionality and performance compared to the standard PIM on the iLab.t virtual wall (VW)¹ platform. iLab.t VW is a large-scale experimental Linux machine-based emulation testbed located in Ghent, Belgium. The prototypes of the GCMR and the PIM routing engines have been developed using the libraries of the Quagga² open source routing suite.

The iLab.t VW is a generic test environment, which assists researchers to validate and evaluate the performance of innovative network prototypes. Each of the 3 VW facilities

¹<http://www.iminds.be/en/develop-test/ilab-t/virtual-wall>

²<http://www.nongnu.org/quagga/>

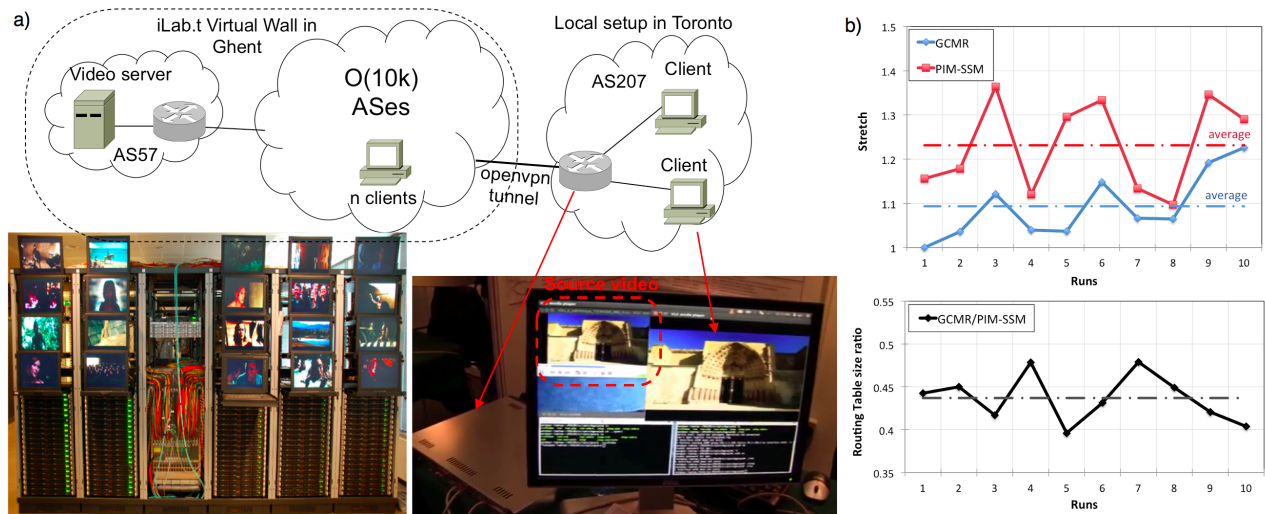


Fig. 1. a) Setup of the demonstration and b) examples of results obtained with the platform comparing stretch and routing table size in ten different runs.

of iLab.t consists of 100 server blades interconnected by a non-blocking 1.5 Tb/s VLAN Ethernet switch. On the iLab.t VW, virtualization technique using OpenVZ Linux containers allows for multiple virtual nodes to run on one machine enabling large-scale experiments (20-50 times the number of physical machines in the iLab.t).

As shown in Fig. 1a), the demo will consist of two parts, the iLab.t VW and the local setup at the conference booth. A large-scale topology mimic a realistic portion of the Internet will be setup in the iLab.t VW. The aim is to reach an experimental facility that can emulate $O(10k)$ routers each one running a Debian 6 Linux distribution, the Quagga routing suite and the considered multicast routing protocols (GCMR and PIM-SSM). Several hosts will be also setup in the iLab.T VW. One of them will be a video server while the others clients. An openvpn tunnel will be setup to interconnect the iLab.t VW with the local setup. The local setup will consists of a Linux router and two clients.

The execution of the demo will consist of two parts. In the first part, the PIM protocol will run in the routers. To setup the paths of the MDT, PIM also needs the multiprotocol BGP in the routers and the IGMP in the hosts. Once the video server starts the transmission, several clients will join the multicast group and several statistics will be collected. Once n clients joined, a failure in a link is caused and the recovery time measured. In the second part, the same scenario is repeated but using the GCMR protocol alone as it does not require unicast routing or IGMP. At the end, a comparison in terms of routing quality (i.e. stretch), routing table size, communication costs, and recovery time will be shown. Remind that the stretch metric measures the distance between the multicast tree obtained with the protocol and the optimum one. Expected results will show that, using more than one order of magnitude smaller routing tables, GCMR obtains better stretch and very quick recovery time than PIM at the expense of higher communication cost. In particular, it is worth mentioning that to setup a new branching path in case of a failure, PIM needs to wait first a stable state of BGP routing which may tend to explore all alternatives (problem known as path exploration [10]). On the contrary, in GCMR, the recovery

time comprises only the time for the failure-detecting node to initiate a search and receive answers from its neighbors that point to the least cost branching path.

A preliminary version of this demo consisting of 207 routers has been successfully presented during the Hands-on-FIRE! event co-located in the 2013 FIA Week in Dublin, Ireland. Fig. 1b) illustrates a couple of results obtained with this configuration, which compare the stretch and the routing table size of PIM and GCMR protocols in ten different runs. As observed, GCMR presents in all runs lower stretch than PIM (i.e. better quality as it is closer to the optimal 1-stretch) and requires around 44% less bits than PIM.

ACKNOWLEDGMENT

This work has been partially funded by the EULER FP7-258307 and DOMINO (TEC2010-18522) projects .

REFERENCES

- [1] S. Ratnasamy, A. Ermolinskiy, S. Shenker, "Revisiting IP multicast", in *Proc. Sigcomm '06*, Pisa, Italy, pp. 15.26, Sep. 11-15, 2006.
- [2] T. Li (ed.), "Design Goals for Scalable Internet Routing", *IETF RFC 6227*, May 2011
- [3] B. Fenner et.al., "Protocol Independent Multicast - Sparse Mode (PIM-SM)", *IETF RFC 4601*, August 2006.
- [4] H. Holbrook, B. Cain, "Source-Specific Multicast for IP", *IETF RFC 4607*, August 2006
- [5] L. Zheng, J. Zhang, R. Parekh, "Survey Report on Protocol Independent Multicast - Sparse Mode (PIM-SM) Implementations and Deployments", *IETF RFC 7063*, Dec. 2013.
- [6] C. Diot, et al., "Deployment Issues for the IP Multicast Service and Architecture" *IEEE Network*, vol. 4, no. 1, pp. 78-88, Jan. 2000.
- [7] EULER FP7-258307 European project, *Experimental UpdateLess Evolutionary Routing*, <http://www.euler-fire-project.eu>.
- [8] P. Pedroso, D. Papadimitriou, D. Careglio, "Dynamic compact multicast routing on power-law graphs", in *Proc. Globecom 2011*, Houston, TX, USA, Dec. 5-9, 2011.
- [9] D. Papadimitriou, D. Careglio, P. Demeester, "Performance analysis of multicast routing algorithms", in *Proc. ICNC '14*, Honolulu, HA, USA, Feb. 3-8, 2014.
- [10] R. Oliveira, B. Zhang, Dan Pei, Lixia Zhang, "Quantifying Path Exploration in the Internet", *IEEE/ACM Trans. Net.*, vol. 17, no. 2, Apr. 2009.