Computer Networks 76 (2015) 259–274

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Bounded latency spanning tree reconfiguration

Martino Fornasa [a,*], Michele Stecca [b], Massimo Maresca [a,b], Pierpaolo Baglietto [a]

[a] *Computer Platform Research Center, University of Genova, University of Padova, Italy*
[b] *International Computer Science Institute, Berkeley, CA, USA*

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form 5 November 2014
Accepted 7 November 2014
Available online 25 November 2014

Keywords:
Ethernet
Carrier ethernet
Wide area network
Metropolitan area network
Spanning tree
RSTP

ABSTRACT

One of the main obstacles to the adoption of Ethernet technology in carrier-grade metropolitan and wide-area networks is the large recovery latency, in case of failure, due to spanning tree reconfiguration. In this paper we present a technique called Bounded Latency Spanning Tree Reconfiguration (BLSTR), which guarantees worst case recovery latency in the case of single faults by adopting a time-bounded bridge port reconfiguration mechanism and by eliminating the bandwidth-consuming station discovery phase that follows reconfiguration. BLSTR does not replace the Rapid and Multiple Spanning Tree reconfiguration protocols, which remain in control of network reconfiguration, whereas it operates in parallel with them.

1. Introduction

While in the past Ethernet technology was prevalently adopted in the local domain and in enterprise networks, recently the bandwidth growth deriving from the diffusion of optical transmission has made it convenient to adopt Ethernet technology also in the metropolitan/wide-area domain and in carrier networks. The adoption of Ethernet technology in carrier networks is based on the IEEE 802.1Q-2011 standard [1], which includes: (i) the Virtual LAN (VLAN) concept to segregate the traffic related to different services in the user network, (ii) the Provider Bridge concept (formerly IEEE 802.1ad), based on stacked VLAN, to segregate the traffic related to different customers in service provider networks, and (iii) the Provider Backbone Bridge concept (formerly IEEE 802.1ah) which introduces a separate network associated to a private addressing space to interconnect different Provider Bridge networks. In Provider Bridge technology, which is of particular interest in this paper, the service provider bridges can be classified in two categories, namely, that of Provider Edge Bridges (PEB), connected to customer equipment, and that of Provider Bridges (PB), internal to the service provider network (see Fig. 1).

In general terms the Ethernet working model can be summarized by three distinctive features, namely, *spanning tree, address learning,* and *flood on unknown.* The *spanning tree* feature denotes the fact that the Ethernet frames are forwarded through an acyclic overlay topology, called *active topology*, which spans all the bridges, i.e., a spanning tree. Ethernet uses the Rapid Spanning Tree Protocol (RSTP) [2] to establish such an overlay topology and the Multiple Spanning Tree Protocol (MSTP) [1], an extension of RSTP, to establish more than one spanning tree instance on the same physical topology to improve robustness and link utilization. The *address learning* feature denotes the fact that the bridge forwarding tables are updated at the reception of each frame by associating the frame source MAC address to the frame arrival port, i.e., by learning the route to a station from the traffic generated by that

---

* Corresponding author at: via Gradenigo 6/A, 35131 Padova, Italy. Tel.: +39 328 2127759.

E-mail addresses: m.fornasa@cipi.unige.it (M. Fornasa), steccami@icsi.berkeley.edu (M. Stecca), m.maresca@cipi.unige.it (M. Maresca), p.baglietto@cipi.unige.it (P. Baglietto).

http://dx.doi.org/10.1016/j.comnet.2014.11.004
1389-1286/

station. The *flood on unknown* feature denotes the fact that when a bridge receives a frame directed to an unknown MAC address the bridge floods the frame on all its active ports. Both *address learning* and *flood on unknown* require the presence of a loop-free topology. In particular *address learning* requires the existence of single bidirectional paths between bridge pairs whereas *flood on unknown* is not compatible with the presence of cycles which would cause endless forwarding loops.

Being a distance-vector protocol [2], RSTP cannot provide an acceptable bounded reconfiguration time in case of failures and more in general in case of network modifications that worsen the network paths, because of the well known count-to-infinity phenomenon [3]. In particular it was shown that the RSTP convergence may end up lasting several seconds or even tens of seconds [4]. While such a large reconfiguration latency can be acceptable in enterprise networks, on the contrary, it is not compatible with carrier grade services, for which the worst acceptable reconfiguration latency is of an order of magnitude of the tens of milliseconds [5].

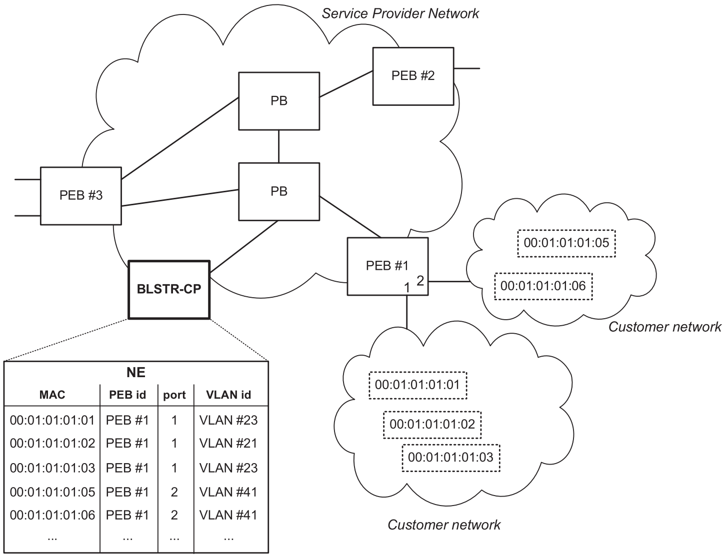
Several approaches to reduce Ethernet reconfiguration latency were proposed in the past. A first approach is to devise special techniques for specific physical topologies of large diffusion, such as ring [6,7]. A second approach is to exploit different MSTP instances to perform rapid rerouting of traffic after a fault [8,9]. A third approach is to abandon the Ethernet working model and to replace the spanning tree approach with a link state protocol [10,11]. A detailed discussion is provided in Section 8.

The Bounded Latency Spanning Tree Reconfiguration (BLSTR) technique [12] proposed in this paper guarantees bounded latency of spanning tree reconfiguration after a bridge failure or after a link failure. BLSTR does not replace the spanning tree reconfiguration protocol (RSTP/MSTP), whereas it operates in parallel with it. Specifically, BLSTR maintains a copy of the bridge configurations and of the bridge forwarding tables deriving from all the possible single resource faults, quickly propagates fault notifications at their occurrence, deactivates frame forwarding on failure-affected bridges for a limited amount of time that linearly depends on bridge time synchronization accuracy, activates the appropriate configurations and forwarding tables and activates forwarding again. As a consequence BLSTR not only provides time-bounded reconfiguration but it also eliminates the effect of the bandwidth-consuming flooding needed to fill out the forwarding tables after reconfiguration.

BLSTR follows the direction proposed in [13,14] for the routing domain, according to which routing decisions are taken in a centralized way and then distributed to the network nodes. In the same way as [15] proposes to relay fault information on dedicated packets to spread fault information on a link-state network, BLSTR is based on distributed active fault notification and centralized alternative configuration computation. However, the spanning-tree distinctive features (tree overlay topology, distance-vector approach, root bridge concept, address learning, absence of time-to-live field in frame header) require a dedicated approach.

BLSTR exhibits the following characteristics:

- It is fully compatible with RSTP/MSTP, and as such it can be included in current generation Ethernet bridges as an additional software component.
- Its time critical operation is fully distributed, i.e., each bridge reconfigures itself in case of network failures, and as such it exhibits the same robustness as RSTP/MSTP.
- It guarantees bounded reconfiguration latency in the order of magnitude of tens of milliseconds on wide-area networks after a bridge failure or after a link failure.



NE

MAC	PEB id	port	VLAN id
00:01:01:01:01	PEB #1	1	VLAN #23
00:01:01:01:02	PEB #1	1	VLAN #21
00:01:01:01:03	PEB #1	1	VLAN #23
00:01:01:01:05	PEB #1	2	VLAN #41
00:01:01:01:06	PEB #1	2	VLAN #41
…	…	…	…

Fig. 1. Network Endpoint table (NE) for a sample Provider Bridges network (PEB: Provider Edge Bridge, PB: Provider Bridge).