Journal of Network and Computer Applications 47 (2015) 107–130

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

A comprehensive survey on Carrier Ethernet Congestion Management mechanism

Hela Mliki*, Lamia Chaari*, Lotfi Kamoun

LETI Laboratory, Engineering School of Sfax (ENIS), University of Sfax, 3038 Sfax, Tunisia

ARTICLE INFO

Article history:
Received 18 October 2013
Received in revised form
5 August 2014
Accepted 5 September 2014
Available online 20 September 2014

Keywords:
Carrier Ethernet
Congestion control
Ethernet Congestion Management schemes
IEEE 802.1Qau standard

ABSTRACT

In the literature, congestion control schemes have been broadly studied. Nevertheless, this subject needs to be reviewed in the context of Carrier Ethernet. Congestion happens when the capacity cannot cover the resource demands. Nowadays, users are witnessing the networks saturation because of the expanding demands for bandwidth. The addressed issue is not only how to avoid congestion but also how to use all available capacity without overuse it or underuse it. To deal with congestion, link level algorithms drop or mark packets with increasing probability as buffer congestion increases. The most common solution is based on the transport layer algorithms to adjust the resources transmission rate by using these dropped or marked packets. Carrier Ethernet defined by layer two uses new proposals specified by the 802.1Qau standards committee in order to handle congestion problems. This paper surveys and studies different properties of Carrier Ethernet congestion control schemes. This paper also draws a parallel between the different schemes and point out the advantage and disadvantage of each one. Then, this paper presents a taxonomy of the Carrier Ethernet congestion control mechanism and correlate it with existing taxonomies.

1. Introduction

The original set of Ethernet LAN technology is popular due its capacity of affording the scaling and features needed for new generation carrier networks. The high bandwidth connectivity, the low-cost, and the ease of use are the most reasons behind the Ethernet network success. Originally designed as a technology for connecting computers in local area network (LAN), currently, Ethernet has also become the technology for metro (Raahemi et al., 2004), backbone (IEEE Computer Society, 2008; Nortel, 2007) and data center (O'Hanlon, 2006; Cisco System Inc, 2008) networks. To cope with the fast evolving requirements for the rapidly growing Internet and increasing demand on bandwidth, Ethernet is evolving and the IEEE 802.3 standard group study technologies to realize 10 Gigabit Ethernet passive optical networks (10G E-PON) and 100 Gigabit Ethernet. An overview of Ethernet technologies evolution was given by Lam and Way (2008).

Carrier Ethernet augments the original set of Ethernet LAN technologies with support for new capabilities required to deliver services. The additional capabilities enable end users to build Metro and Wide Area Network, and service providers to build network infrastructure or deliver Ethernet based MAN or WAN telecommunication services. It provides flexible bandwidth increments and the ability to add new services using one technology (MEF5, 2004; MEF12, 2005). Therefore, operators tend to switch toward packet based Ethernet/IP technologies across their access and core networks in order to cost effectively support the rapidly escalating bandwidth requirements.

Today, there are exponential increase demands for applications like social networking, cloud computing (Ethernet Technology Summit and Exhibition, 2012; Buyya et al., 2008) and streaming video necessitate more powerful data centers (Kachris and Tomkos, 2012; Buyya et al., 2008). This makes a hard challenge to the data centers networking requiring efficient interconnection design with reduced latency and high bandwidth. Therefore, data center networks often uses Fiber Channel for high reliability. Kachris and Tomkos (2012) survey next generation data center networks using optical interconnects. Then, provide a taxonomy and comparison between the proposed schemes considering their main features such as scalability and connectivity. However, technology requiring Fiber Channel is more expensive than the Ethernet one. Hence, the IEEE 802.1 standards committee studied the issue of using Ethernet as the infrastructure to enable the Data Center applications (O'Hanlon, 2006; Cisco System Inc, 2008).

Data Center traffic can be highly bursty, because when any single source needs full access to bandwidth to achieve the lowest latency in the absence of congestion, burst are injected without

* Corresponding authors.
E-mail addresses: mliki.hela@gmail.com (H. Mliki), Lamia.chaari@tunet.tn (L. Chaari), Lotfi.Kamoun@isecs.rnu.tn (L. Kamoun).

http://dx.doi.org/10.1016/j.jnca.2014.09.003
1084-8045/

traffic shaping or policing. This is different environment from the Internet-type of networks, where aggregation of many low-speed flows (like FTP, web-browsing) determines the traffic characteristics. Lossless networks, like IfiniBand (InfiniBand Trade Association, 2004), Fibre channel, RapidIO (RapidIO Trade Association, 2011) and PCIe (PCI Express Specification) provide rapid response time, prevent buffer from saturation and prevent from throughput collapse caused by congestion. Congestion is then an important issue in the design of Data Center networks, as in any computer network, in order to create a lossless network. On the other hand, network that occasionally may lose some packets are called best effort or loosely. The well known are ATM networks and TCP/IP over Ethernet. TCP is designed to allow and relies on packet loss to detect congestion. Such solution is simple, cheap, and exports the problem from the core network to the edge nodes. However, packet drops recovery causes increased latency (Gusat et al., 2009).

Nevertheless, Ethernet networks, since it is appeared on 1973, run without congestion control in the data link layer. In order to eliminate long latencies caused by packet retransmission and reordering, Ethernet uses flow control at link level defined in the IEEE 802.3x that allows us to pause an Ethernet link. This technique is used in case of lack of sufficient resource buffer to receive the transmitting packet in the downstream switch. This can prevent from congestion collapse and reduce loss rate in the network, but also it may cause congestion spreading (Pfister and Norton, 1985) or tree saturation effect. Therefore a congestion control mechanism is a critical component in the network. In order to avoid drops due to congestion in the network, it is important for Ethernet used in data center or metro networks to support congestion management. Consequently, IEEE 802.1 standards committee is developing new proposals for congestion signaling in Carrier Ethernet networks.

In an effort to keep up with demand and maximize QoS, IEEE802.1Qau (IEEE, 2010) is deploying congestion control mechanism for Carrier Ethernet. The standard specified congestion management mechanisms enabling switches to notify congestion information to edge stations with the functionalities of adjusting their transmission rate to reduce frame loss.

Depending on network topology each packet usually routed over a certain number of intermediate nodes (switches). According to IEEE802.1 Qau, the intermediate switches typically have a queue that grows when congestion occur, then frames are dropped when the queue length exceeds a defined threshold. In addition, flows can be managed by both sender and intermediate nodes. Furthermore, congestion control measurement can be proceeded by intermediate nodes or/and by the receiver (destination). Congestion information, called also feedback or congestion notification, can be reflected back to the reaction point by the intermediate node (congestion point) or by the destination node. These entities are then called reflection points. Depending on the location of the reflection point, three congestion control varieties can be defined as follows:

- Reflection point resides at the intermediate node (switch): It consists of Backward notification Congestion Management (BCM), the advantage of this mechanism is to deal with a sudden increases in load faster. Ethernet Congestion Management (ECM) (Lu et al., 2007) and Quantized Congestion Notification (QCN) (Alizadeh et al., 2008) are such mechanisms.
- Reflection point resides at the destination node: It consists of Forward notification Congestion Management (FCM). This mechanism outperforms BCM in fairness and stability. Forward Explicit Congestion Notification (FECN) (Jiang et al., 2007) is such mechanisms.
- Reflection point resides at the intermediate node (switch) and the destination node: It consists of a combination between the BCM and the FCM mechanism. Consequently, we call it the Hybrid Notification Congestion Management mechanisms. Hybrid Notification Congestion Management mechanisms are: 3 point QCN (3ptQCN), Extended Ethernet Congestion Management (E2CM) (Gusat et al., 2007) and Enhanced Forward Explicit Congestion Notification (E-FECN) (So-In et al., 2008).

These Ethernet congestion control mechanisms are the main focus of the present study in this paper.

The congestion control mechanisms should be able to provide some performance features.

- *Fairness*: is based on Max–Min fairness (Bonald et al., 2006) which considers fairness in sharing a scarce queue's space and gives each connection an equal part. The Max–Min fairness approach provides fairness by equally sharing the bottleneck. Ethernet congestion control proposed mechanisms allow equal throughput rate to all flows involved in the bottleneck path. Therefore, mechanisms' fairness results in sharing only the resources between the source and the most congested intermediate node.
- *Prevent feedback implosion*: A general drawback mentioned with respect to feedback messages is the additional traffic in the reverse direction on what it may be a congested path. Nevertheless, the need of congestion state information for senders, in order to be responsive, has to be taking into account. Therefore some proposed mechanism sends feedback message within probability value (like FECN, E2CM). Others try to reduce the feedback generation by getting rid of positive feedback from the congestion point (like QCN).
- *Stability*: Switch queue occupancy processes should not fluctuate, causing underutilization of the resource or overutilization leading to frame drops. However, queue size should be controlled to be maintained at a target threshold length. This is crucial when trying to control a shallow queue.
- *Responsiveness*: Ethernet link throughput can vary with time due to flow fluctuation or the appearance of bottlenecks or the arrival of new sources, etc. the algorithm needs to adjust source rates according to these variations.
- *Simple to implement*: The algorithm should be simple in order to be implemented in hardware. Complicated rates computing, parameters tuning and control loop gains should be omitted.

The main problem to be solved is to achieve lossless flow transport in Carrier Ethernet. In order to resolve this problem Ethernet protocol should be enhanced to support the new Carrier Ethernet services. Congestion control algorithms are a key component of data transport in Carrier Ethernet. Efficient congestion control schemes for Carrier Ethernet should be implemented to avoid congestion and to get lossless transport traffic. In fact, the sources need to be notified to adjust their sending rate according the status of the network in order to manage congestion and avoid traffic loss. In this context, this paper surveys, classifies and compares the existing schemes for congestion control implemented for Carrier Ethernet.

The contributions of this paper are as follows: (a) the paper presents a survey of existing closed loop flow control techniques used in the literature and identifies the advantage and disadvantage of these techniques. (b) The paper studies the Carrier Ethernet technologies and services. (c) The paper elaborates a taxonomy of the Carrier Ethernet congestion control mechanism and correlate it with existing taxonomies in literature. (d) The paper presents a comparative study of different approaches of congestion control for Carrier Ethernet in satisfying the performance features in MAN, WAN and data center networks context.