A Novel Distributed Spanning Tree Protocol for Provider Provisioned VPLS Networks

Madhusanka Liyanage*, Mika Ylianttila†, Andrei Gurtov‡

*Centre for Wireless Communication (CWC), University of Oulu, Finland.
madhusanka@ee.oulu.fi
† Centre for Internet Excellence (CIE), University of Oulu, Finland.
mika.ylianttila@cie.fi
‡ Helsinki Institute for Information Technology HIIT, Aalto University, Finland and ITMO University, Russia
gurtov@hiit.fi

Abstract—Spanning Tree Protocol (STP) is a widely used protocol to maintain a loop free Layer 2 (L2) switching network. On the other hand, Virtual Private LAN Service (VPLS) is a L2 Virtual Private Network (VPN) service which is becoming very popular among many industrial enterprises. In a VPLS network, VPN connections through the provider network are invisible to L2 network devices and protocols. It causes to several issues while utilizing STP in a VPLS enabled Ethernet network.

In this paper, we propose a novel Distributed STP (DSTP) to maintain a loop free Ethernet network over a VPLS network. DSTP proposes to run a modified STP instance in each remote network segment and evades the transportation of STP messages through the provider network. In addition, we propose two Redundancy Identification Mechanisms (RIMs) to mitigate the impact of invisible loops in the provider network. Simulation results verify that DSTP is capable of maintaining a loop free Ethernet network over a VPLS network. Furthermore, DSTP significantly reduces the convergence time of the spanning tree and STP overhead over the provider network.

I. INTRODUCTION

The redundancy is a necessary requirement in a switched network. It protects the communication from the loss of connectivity due to the failures of network devices or links. However, this redundancy provision may result in physical layer loops. These physical layer loops will cause serious problems in switched networks such as broadcast storms, multiple frame transmissions and MAC (Media Access Control) address database instability. Hence, the Spanning Tree Protocol (STP) is introduced to create a loop free logical topology [1]. STP is a powerful tool and it is commonly used in every type of Layer 2 (L2) networks.

On the other hand, many industrial enterprises are still relying on legacy SCADA (Supervisory Control and Data Acquisition) and process control devices. These devices are incapable to support higher layer protocols and still supporting L2 protocols only. They often expect flat network topologies or single broadcast domains. However, higher layer communication networks such as IP/MPLS (Multiprotocol Label Switching) based networks provide high speed connectivity, higher reliability, scalability and security than lower layer networks. Virtual Private LAN Service (VPLS) is one of the most promising layer 2 provider provisioned VPN service to transport legacy Ethernet traffic on top of the well-advanced IP/MPLS based provider networks. A VPLS provides multipoint-to-multipoint Ethernet connectivity and extends the Ethernet broadcast domain to geographically distributed locations. VPLS has gained the immense popularity among many industrial enterprises since it offers a successful solution to many problems such as high-speed connectivity, any-to-any forwarding at L2 networks. Moreover, emerging technologies of Internet of Things and Smart Spaces [2] make VPLS a promising mechanism for network communication between many heterogeneous participants in a localized high-dynamic environment.

In a VPLS network, connections through the provider network are invisible to the switching network devices and L2 protocols. However, these transparent links in the provider network cause many negative effects on L2 protocols. For instance, STP suffers many issues such as broadcast storms due to loops in the provider network, spanning tree instability, security breaches through the provider network, extra cost to transport STP BPDPs and high spanning tree convergence time. Hence, it is infeasible to implement the existing STP in a VPLS enabled Ethernet network.

Our Contribution

In this paper, we propose a novel Distributed STP (DSTP) to maintain a loop free Ethernet network over a VPLS network. DSTP proposes to run a modified STP instance in each remote segment of the VPLS network. Thus, it eliminates the requirement to transport STP BPDPs through the provider network. In addition, we propose two Redundancy Identification Mechanisms (RIMs) called Customer Associated RIM (CARIM) and Provider Associated RIM (PARIM) to mitigate the impact of invisible loops in the provider network. Thus, DSTP is capable of establishing and maintaining a loop free Ethernet network over a VPLS network by solving above incompatibility issues. DSTP successfully transmits broadcast frames over the VPLS network without causing any broadcast storms. Furthermore, DSTP significantly reduces the convergence time of the spanning tree and STP overhead over the provider network. Also, DSTP achieves the scalability by significantly reducing the number of STP messages transmitted through the
provider network. We conduct several simulations to verify these features and illustrate the performance advantages of DSTP.

The rest of the paper is organized as follows. Brief descriptions of STP, VPLS and STP implementation issues are presented in Section II. A prior study of this topic is summarized in Section III. The proposed DSTP and redundancy identification mechanisms are described in Section IV. We discuss our simulation model and the numerical results in Section V. Section VI concludes the paper.

II. BACKGROUND

A. Spanning Tree Protocol (STP)

Spanning Tree Protocol (STP) is a widely used network protocol which ensures a loop-free network topology for a L2 Ethernet network. STP is standardized as IEEE 802.1D [3]. Basically, STP prevents bridge loops. As a result, it prevents the broadcast radiation that results from these bridge loops. STP calculates a spanning tree which interconnects all the bridges in the network. Then, STP disables all the links which are not a part of the spanning tree.

STP is a collective operation of all bridges in the network. Thus, each bridge should have a sound knowledge about the network. STP defines special data frames called BPDUs to exchange their STP related information among the bridges. Initially, each bridge transmits configuration BPDUs to share the information with others. All bridges collaboratively select one bridge as the root bridge which is the root of the spanning tree. Then, other bridges select root ports and designated ports to establish the spanning tree. Then, each bridge blocks any other active ports which are neither a root port nor a designated port.

STP defines 5 port states for each port: namely blocking, listening, learning, forwarding and disabled. However, the port transition in STP is inefficient for networks which need faster convergence time. Usually, STP takes 30 to 50 seconds to respond to a topology change. Hence, IEEE introduces the Rapid Spanning Tree Protocol (RSTP) as 802.1w in 2004 [4]. It introduces new convergence behaviors and bridge port roles to STP. However, the basic features of STP remain the same for the RSTP as well. RSTP is just speeding up the operation of STP by introducing new port convergence procedures.

We only consider the traditional STP under this article. However, the changes which are proposed under DSTP can be directly applicable to RSTP as well.

B. Virtual Private LAN Service (VPLS)

VPLS is a provider provisioned L2 VPN service which provides Ethernet based multipoint to multipoint communication on top of the provider network. A VPLS network mainly consists of three main segments, namely Customer Edge Equipments (CEs), Provider Edge Equipments (PEs) and a provider backbone network. Figure 1 illustrates a simple VPLS network.

CE devices are the interfacing devices between the provider and customer networks. PE devices contain all the VPN intelligence. The VPLS architecture builds a full mesh of Pseudo-Wires (PWs) between these PEs. These PWs are overlaid on top of the provider network. The provider network can be operated based on several network protocols, such as IPv4, IPv6, MPLS.

C. Implementation Issues of STP in a VPLS network

The traditional STP faces many issues while using it in a VPLS enabled network. Here, we briefly discuss six critical issues among them.

First, STP cannot eliminate the loops which are generated through the provider network. In a VPLS network, a remote network segment can be connected to multiple PEs in order to achieve the network redundancy and support load balancing (E.g. Customer site2 in Figure 2). Since the VPLS builds a full mesh of PWs, there is a PW between every pair of PEs, including PEs belonging to the same network segment. These PWs are built on top of the provider network and they are invisible to STP in the customer network. As a result, broadcast frames can be looped back to the same network segment through these PWs.

Second, PWs in the provider network jeopardizes the root path calculation process of STP. During the root path calculation process, each bridge selects the root path based on the accumulative cost of every link from root bridge to the particular bridge. However, the link cost of PWs are invisible to bridges. It may lead to invalid root path calculation and indirectly affects the performance of STP.

Third, STP is vulnerable to several security breaches. The most common attack is the Denial-of-Service (DoS) attack. Here, the attacker sends a massive amount of raw configuration BPDUs and/or topology change notification (TCN) BPDUs to jeopardize the operation of STP [5] [6]. Furthermore, the root role claiming and dual-home attacks also badly influence the operation of STP. Many of these attacks are initiated by eavesdropping ongoing BPDUs in the switched network. In a non secure VPLS, an attacker can easily eavesdrop STP BPDUs at the non-secure public network.

Fourth, STP protocol needs to transport a large amount of BPDUs over the provider network during the root bridge election process. Thereafter, the root bridge broadcasts HELLO BPDUs in every 2 seconds. Furthermore, every topology...
change in the network exchanges a lot of TCN BPDU's. This increases the overhead of STP and consumes the expensive bandwidth of the provider network. Ultimately, the customer has to pay an extra fee to transmit these STP BPDU messages.

Fifth, the extensive propagation and queuing delays in the provider network increase the spanning tree convergence time and forwarding table instabilities.

Sixth, it is highly possible to use different types of vendor switches and different type of port configurations in different remote network segments. As a result, STP BPDU's can be discarded at the provider network facing interfaces. It may effect the proper operation of STP.

III. RELATED WORK

STP is based on an algorithm that was invented by Radia Perlman [1]. Later, several STP versions were standardized as IEEE 802.1D [3], 802.1w [4] and 802.1s [7]. These protocol versions were designed to work only on traditional Ethernet networks. These STP specifications did not consider special network scenarios such as VPLS. Thus, all STP versions face same set of issues in VPLS enabled Ethernet networks.

Furthermore, Cisco proposed two proprietary versions of STP as Per-VLAN (Virtual LAN) Spanning Tree (PVST) and Per-VLAN Spanning Tree Plus (PVST+). When multiple Virtual LANs are used in an Ethernet switched environments, PVST and PVST+ create a separate spanning tree for each VLAN [8]. Later, Cisco’s proprietary version of RSTP to create a spanning tree for each VLAN is proposed as Rapid Per-VLAN Spanning Tree (RPVST) [8]. However, these STP versions are suffering the same set of issues in VPLS enabled Ethernet networks since these versions did not consider the impact of VPLS networks.

On the other hand, IETF standardizes two frameworks to develop a VPLS by using Border Gateway Protocol (BGP) [9] and Label Distribution Protocol (LDP) [10]. Thereafter, several versions of VPLS architecture are proposed [11] [12] [13] [14] [15] [16]. However, none of these VPLS architectures paid close attention on how to use L2 customer protocols in a VPLS network. Hence, these architectures fail to solve above implementation issues of STP. Several Host Identity Protocol (HIP) based VPLS architectures were proposed to establish a secure VPLS network [12] [16] [17]. These secure architectures are able to protect STP from attacks which are initiated from the provider network. However, they still experience other above stated issues.

Many of these architectures were proposed to evade the transmission of L2 protocol PDUs over the provider network [9] [10] [12]. It will solve some of the implementation issues of STP such as security and convergence issues. However, this is not a stable solution due to two reasons. First, broadcast mechanism will not be efficient and it will result in long packet distribution delays. Second, it will not prevent the impact of loops over the provider network such as broadcast storms, multiple frame transmissions and MAC address database instability. Therefore, none of STP versions or VPLS architectures provides a complete solution for all the implementation issues of STP in a VPLS network.

IV. PROPOSED DISTRIBUTED SPANNING TREE PROTOCOL

We propose a novel Distributed Spanning Tree Protocol (DSPT). The proposed solution seeks to offer improved STP to solve above implementation issues and enhance the scalability of STP to span over provider provision networks such as VPLS. Basically, DSTP slices the spanning tree over remote customer sites. It proposes to run a customized version of STP on each remote customer site.

However, the existing STP versions cannot be used as local STP instances since they are not capable of identifying loops over the provider network. Hence, we add a Redundancy Identification Mechanism (RIM) to DSTP and customize the STP accordingly. RIM identifies the possible loops over the provider network and eliminates them. Two RIM implementations are proposed, namely Provider Associated RIM (PARIM) and Customer Associated RIM (CARIM).

A. Provider Associate Redundancy Identification Mechanism (PARIM)

The first step of PARIM is to elect a Designated PE (DPE) for each network segment. It is a collective operation of all PEs. Initially, every PE broadcasts a Network Advertisement Packet (NAP) to all other PEs through the provider network. This NAP contains the Network Segment Identifier (NSI) of the remote network segment where the PE is connected, the priority of the PE, the link cost value, sender ID and DPE ID. The provider assigns a unique NSI for each remote network segment and a priority value for each PE. The PE with the lowest priority value has the highest rank. The link cost value represents the bandwidth of the link that connects the PE to the provider network. The link cost is calculated according to the procedure proposed in IEEE 802.1D specification [3]. The sender ID can be the IP address of the PE or the host identity. In the beginning, each PE sets DPE ID to its own ID.

Similar to the root bridge election process in STP [3], PEs exchange NAPs until they select one DPE for each network site. If there are more than one PEs represent a network segment, the sequence of events to determine the best DPE is the lowest priority value, the lowest link cost value, the lowest sender ID. Once DPEs are elected, all other PEs are set at the broadcast blocking state. In the broadcast blocking state, PEs do not forward any broadcast frame and drop all receiving broadcast frames. Only DPEs are allowed to flood broadcast frames.

Furthermore, DPE sends periodic HELLO packets via the provider network to other PEs in the same network segment to notify the aliveness of the DPE. An addition of a new PE or a removal of a DPE will trigger a new DPE selection process.

Figure 2 illustrates a block diagram of a VPLS network which uses the proposed DSTP with PARIM. A DSTP instance runs on each customer network site separately. Since only one PE is used to interconnect the customer network site1, PE1 is the obvious selection as DPE for the site1. Without the loss
of generality, we assume that PE2 has the higher priority than PE3. Hence, PE2 is selected as DPE for the site2. Then, PE3 enters to the broadcast blocking state.

The main advantage of PARIM is that the customer does not need to do any STP modification at the customer network. The provider associates to implement RIM and it is invincible for the customer. However, PARIM has some disadvantages. First, it increases the overhead of STP on the provider network and adds an extra cost for the customer. Second, PARIM increases the complexity of the VPLS architecture.

B. Customer Associated Redundancy Identification Mechanism (CARIM)

Similar to PARIM, the first step of CARIM is to elect a Designated CE (DCE) for each network segment. It is a collective operation of all CEs in a network segment. Every CE broadcasts the Network Advertisement Frames (NAF) through the customer network. A NAF contains the priority of the CE, the link cost value, a MaxHop field, sender ID and DCE ID. The link cost value represents the bandwidth of the link between the CE and the PE. The sender ID can be the MAC address of the CE or the host identity. In the beginning, each CE sets DCE ID to its own ID. MaxHop is used to avoid broadcast storms due to NAFs. The default value of MaxHop is 7 since IEEE recommendation is to consider the maximum diameter of a network as seven bridges [3]. However, it can be changed according to the customer’s requirements.

Similar to the root bridge election process in STP [3], CEs exchange NAPs until they select one DCE for each network site. If there are more than one CEs in the network segment, the sequence of events to determine the best DCE is the lowest priority value, the lowest link cost, the lowest bridge ID. The bridge ID and the link cost are calculated according to the procedure proposed in IEEE 802.1D specification [3]. Furthermore, these NAF are generated only by CEs. When any other bridge receives a NAF, it floods the NAF by reducing MaxHop by one.

Since CEs are also L2 devices, CEs are participating in the root bridge selection process of STP. However, only DCEs are participating in this process after the election of DCEs.

C. Operation of the Distributed Spanning Tree Protocol

The first step of the DSTP is to run the RIM procedure. The customer has the flexibility to choose either PARIM or CARIM. RIM elects the DPEs or DCEs and sets any other PEs or CEs into the broadcast blocking state.

The second step of DSTP is to elect the root bridge. The root bridge election is local to each site. None of the CE will forward STP BPDUs to PEs and all STP BPDUs are dropped at CEs. The local root bridge election procedure is similar to the root bridge election procedure of the traditional STP [3]. However, only DCEs are participating in the local root bridge election procedure in DSTP with CARIM.

Third step of DSTP is to calculate the path costs and decide the root, designated and block ports for each bridge. As the STP instance is local to each section, convergence time is
comparably lower in DSTP than the traditional STP. TCN BPDUs are also local to each site. Hence, the convergence and the forwarding table updates are much faster in DSTP than the traditional STP.

When a local device transmits a broadcast frame (e.g. Address Resolution Protocol (ARP) request frame), it propagates through the spanning tree and eventually reaches to the edge devices (CEs, PEs) of the customer network site. Only designate devices (DCE or DPE) transmit the received broadcast frame to all the other customer network sites through the provider network. All other edge devices drop the received broadcast frame.

When an edge device (CEs, PEs) receives a broadcast frame from any other remote customer network site, only designated boundary devices (DCE, DPE) broadcast it to the local customer network site. All other boundary devices drop the received broadcast frames.

V. EXPERIMENT RESULTS

We simulate the proposed DSTP on the OMNET++ network simulator [18] to analyze its performance. DSTP with PARIM, DSTP with CARIM and traditional STP [3] versions are implemented in the simulation model. Figure 4 illustrates the network model which is used for simulations.

![Fig. 4: The simulation model](image)

Various experiments are performed to measure the convergence time of the spanning tree, performance of broadcast mechanism, additional STP overhead over the provider network and scalability. The simulation model contains eight User Equipments (UEs). The links within customer network sites have the propagation delay of 1 ms and the propagation delay between PEs is set as 5 ms. Each experiment is repeated for 10 times and only averages are recorded.

A. The convergence time of the spanning tree

In the first experiment, the convergence time of the spanning tree is analyzed. We measure the time required to elect the root bridge by changing the propagation delay between PEs. Figure 5 illustrates the average time required to select the root bridge against the propagation delay between PEs. The root bridge selection is the first and the most time consuming step of the spanning tree establishment procedure.

![Fig. 5: The average time required to select the root bridge](image)

We observe a linear increment of the root bridge selection delay with the propagation delay between PEs for both STP and DSTP with PARIM. However, DSTP with PARIM significantly reduces the root bridge selection delay of STP. On the other hand, proposed DSTP with CARIM requires the smallest amount of time to select root bridges and the root bridge selection delay is independent of the propagation delay between PEs.

Thus, the proposed DSTP solves the impact of large propagation delays at VPN tunnels and allows to interconnect customer sites regardless of the distance between them.

B. The performance of broadcast mechanism

In the second experiment, the performance of broadcast mechanism is analyzed. Here, UE7 broadcasts an ARP request frame. We measure the number of broadcast frames in the network.

Figure 6 illustrates the total number of broadcast frames generated against the simulation time. Simulations results are reported only for first 100 ms.

According to experiment results, DSTP with PARIM and DSTP with CARIM stop the generation of broadcast frames after 13 ms. After that, no more broadcast frames are available in the network. Also, they generate only one extra broadcast frame which is discarded at PE3 or CE3. However, the traditional STP generates the same broadcast frame repeatedly due to the invisible loop over the provider network. This will cause a broadcast storm.

Thus, we can conclude that the proposed DSTP successfully delivers broadcast frames over the VPLS and prevents the generation of broadcast storms.
In the third experiment, the additional STP overhead on the provider network is analyzed. We measure the number of STP related messages transmitted over the provider network and Figure 7 illustrates the total number of STP messages transmitted over the provider network against the simulation time. The simulations results are reported only for first 25 s. According to experiment results, STP overhead on the provider network is very high for the traditional STP. It is linearly increased with the simulation time due to periodic Hello BPDU messages. DSTP with PARIM also has some overhead due to DPE selection messages and periodic Hello packets. The overhead is linearly increasing with the simulation time due to periodic Hello packets. However, DSTP with CARIM transmits the periodic Hello BPDU only for PEs which are connected to the same customer network site. Therefore, the linear increment in DSTP with PARIM is much slower than the traditional STP. On the other hand, DSTP with CARIM has no additional overhead of STP messages since it delivers the STP BPDU only within customer network sites.

D. Scalability

In the fourth experiment, we study the scalability of each architecture by increasing the number of PEs in the network. Here, we assume that each new PE connects a new customer network site to the VPLS network. Figure 8 illustrates the total number of STP messages transmitted through the provider network during the root bridge selection phase.
in DSTP with PARIM than the traditional STP scenario. On the other hand, DSTP with CARIM does not exchange any periodic STP message via the provider network.

Thus, we can conclude that traditional STP is not suitable for implementing in a large scale network with a large number of PEs. DSTP offers the scalability by significantly reducing the number of STP messages transmitted through the provider network in a large scale network. Ultimately, it reduces the additional overhead and STP related cost of the customer.

VI. CONCLUSION

In this paper, we proposed a novel Distributed STP (DSTP) to maintain a loop free Ethernet network over a VPLS network. DSTP proposes to run a modified STP instance in each remote segment of the VPLS network. Furthermore, we proposed two Redundancy Identification Mechanisms (RIMs) called Customer Associated RIM (CARIM) and Provider Associated RIM (PARIM) to prevent functional issues due to invisible loops in the provider network.

We conducted several simulations to verify these features and illustrate the performance advantages of DSTP. DSTP successfully transmitted broadcast frames over the VPLS network without starting any broadcast storms. Furthermore, DSTP significantly reduced the convergence time of the spanning tree and STP overhead over the provider network by outperforming existing STP versions. DSTP also increased the scalability by significantly reducing the number of STP messages transmitted through the provider network in a large scale network. Ultimately, it reduced the additional overhead on the provider network and STP related cost of the customer. Thus, DSTP is capable of establishing and maintaining a loop free Ethernet network over a VPLS network by solving existing incompatibility issues of STP.

In future, we are focusing on studying the performance and compatibility issues in other layer 2 protocols such as ARP (Address Resolution Protocol), RARP (Reverse ARP) while using them in a VPLS based Ethernet network. We try to identify correlations between these issues and explore common solutions. The ultimate goal is to develop a common platform to implement all layer 2 protocols in a VPLS enabled Ethernet environment.

ACKNOWLEDGMENT

This work has been performed in the framework of the CELTIC CP2012 project SIGMONA. The research is supported by TEKES Finland and grant # 14-07-00252 of the Russian Foundation for Basic Research.

REFERENCES