October 2019

ADAPTIVE BANDWIDTH USAGE IN SDWAN

Uni Chen

Alan Xiao-Rong Wang

Jingxia Xia

Jiaoming Li

ADAPTIVE BANDWIDTH USAGE IN SDWAN

AUTHORS:
Uni Chen
Alan Xiao-Rong Wang
Jingxia Xia
Jiaoming Li

ABSTRACT

Presented herein are solutions for adaptive bandwidth usage in wide area networks (WANs) such as software-defined wide-area networks (SDWANs or SD-WANs). The solutions presented herein include inbound link bandwidth detection, and usage/overrun publication along with a percentage of traffic to be moved. The solutions presented herein also adjust outbound traffic based on a peer site's overrun alarm and percentage information. The solutions presented herein may be employed for a Multiple Producer sites, Single Consumer site (MPSC) use case and a load-balance coexists with preferred policy use case in SDWAN. The solutions presented herein also allow for fast convergence.

DETAILED DESCRIPTION

The solutions presented herein may be employed, for example, for multiple producers, single consumer scenarios and scenarios where load-balance is to co-exist with preferred policy. Generally, the scenario where load-balance co-exists with preferred policy may be more typical in scenarios involving software-defined wide-area network (SDWAN or SD-WAN).

Multiple producers, single consumer may be a typical customer scenario/requirement. In such a scenario, there may be large Multiprotocol Label Switching (MPLS) bandwidth on Data Centers (e.g., 1G) but small MPLS bandwidth on branches (e.g., 100M or less), and the branches' internet links may be 1G. A customer may desire to utilize a MPLS link as much as possible, and most traffic may use the MPLS link at first. When the traffic increases and finally exceeds remote branches' MPLS bandwidth, some traffic may be moved to the internet link (1G). Multiple datacenters may send traffic to the same branch at the same time; however, the datacenters are not aware of when to move the traffic and how much traffic should be moved from the MPLS link to the internet link.

For example, as shown in Figure 1, below, site 1's MPLS link ingress bandwidth is 100M. If all of the Data Centers (DC1, DC2, and DC3) send 100M of traffic in total to site 1’s MPLS link, then site 1's MPLS link is almost overrun. However, all of the Data Centers would be unaware that this is the case.

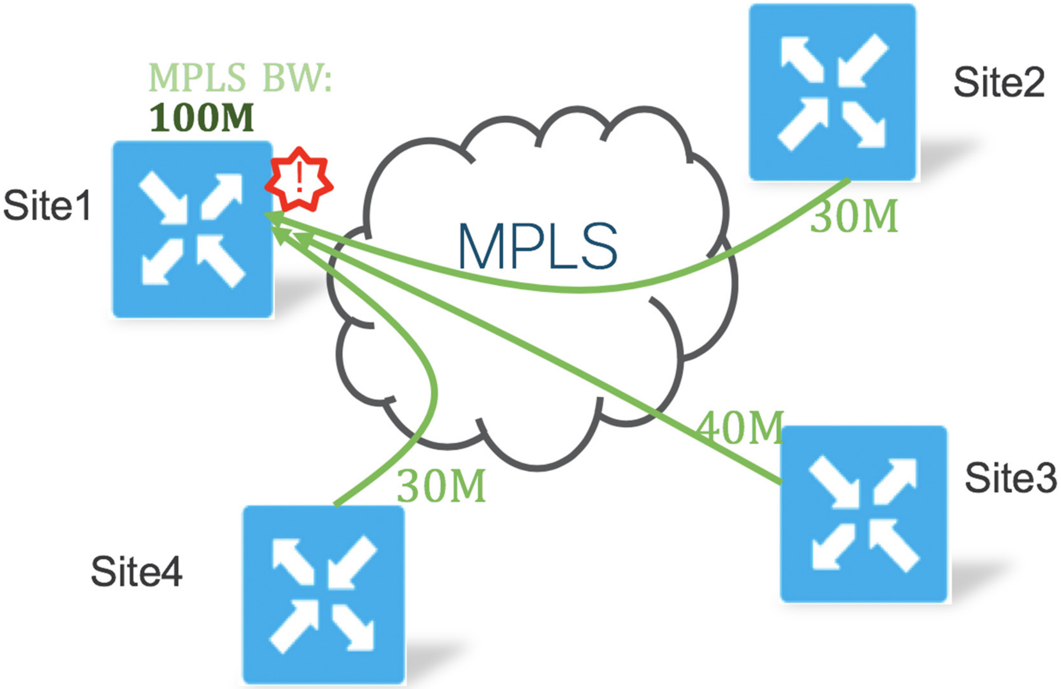

Figure 1

As mentioned above, a typical use case may be load-balance co-exists with preferred policy. For example, remote MPLS bandwidth may be 100M and the internet link may be 1G. MPLS may be customer configured as the preferred link for critical traffic; normally such traffic is about 10M~50M. Sometimes such traffic may reach 80M or even more. By default, all other traffic will be load-balanced between MPLS and the internet link. However, a problem is that load-balanced traffic does not “care” about preferred traffic. In other words, it will not move traffic from the MPLS link to the internet link when there is increasing preferred traffic on the MPLS link.

Brief workflow:

- Define two thresholds: a "BW threshold" and a "Healthy threshold.”
    - BW threshold: when the ingress bandwidth usage percentage exceeds this threshold, start sending an alarm to all remote sites (datacenters).

  - Healthy threshold: a healthy target, bandwidth usage percentage. The healthy threshold may be a value that is slightly smaller than a value of the "BW threshold." The purpose is to avoid flapping around the "BW threshold," and in case any calculation differences between local and remote sites (datacenters).
  - The thresholds are configurable.
- As shown in Figure 2, below, site 1 send an "alarm" to all other sites after it detects its ingress is overrun.
- All other sites are to take action: remove some traffic from MPLS link to other links.

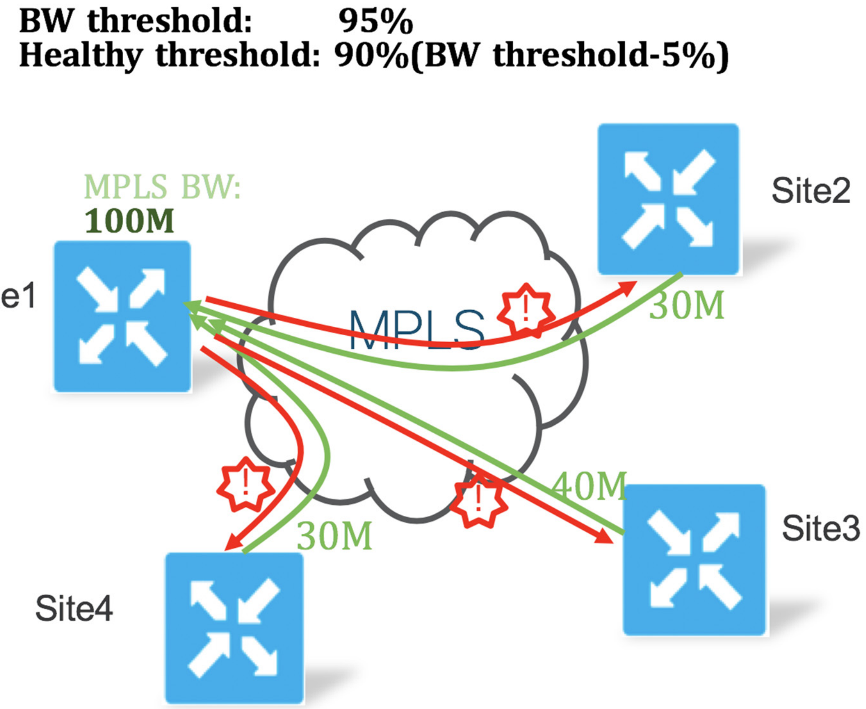

Figure 2

- In the example shown in Figure 3, below, after site1 has detected that its ingress traffic bandwidth has crossed the "BW threshold", site 1 sends an *alarm* to all other sites (datacenters) along with a percentage (***X***) of traffic that the other sites should move from this link.
  - In this example, ***X*** = (100M/100M)*100% - 90% = 10%

- Each of the other sites that received this *alarm* and the percentage (***X***) will move ***X*** (e.g., 10%) of traffic from this link (MPLS) to other links (internet).

  DC1 : 10% * 30M= 3M

  DC2 : 10% * 40M = 4M

  DC3 : 10% * 30M = 3M

- In this example, all the datacenters, in total, moved (3M + 4M + 3M = 10M) 10M of traffic from the MPLS link, and site1's MPLS link becomes healthy again (90M, 90% = (100M - 10M) / 100M)).

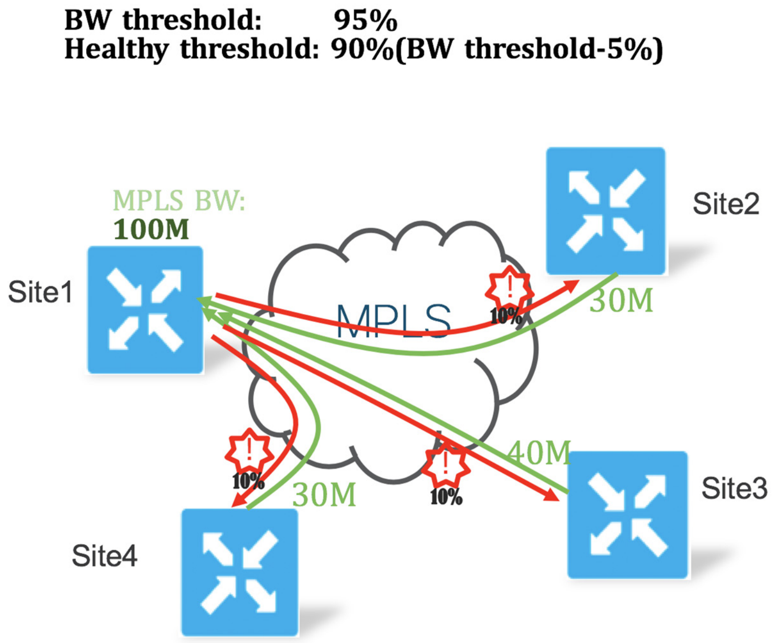

Figure 3

Additional Description:

- How to implement the "alarm":
  - The "alarm" message may be carried in an Overlay Management Protocol (OMP) message in SDWAN.

  - The "alarm" message may be in Virtual Private Network’s (VPN's) packet header (e.g., Network Service Header (NSH), Generic Routing Encapsulation(GRE) header, etc.).
  - The "alarm" could be in any message (e.g., BFD, TCP payload, solution specific message, etc).
- How to move traffic after the "alarm" is received:
  - SDWAN could adjust each link's "weight" to achieve this.
  - VPNs could also adjust "weight".
  - Any controller could move traffic based on this value.

Another example implementation:

- Some datacenters are unable to move all expected traffic due to no other available link and/or some other reason.
    - (e.g., in the above example, DC1 cannot move traffic to other links after it received the alarm because there is not another available link.)
- If some datacenters are unable to move all expected traffic, then the branch may send another alarm to all datacenters after the branch detected at a later time that its ingress traffic bandwidth still crossed the "BW threshold.” This would enable other datacenters to move some more traffic from the link.
    - (e.g., DC2 and DC3 could move some more traffic after they received the alarm again from the same site over the same link.)
- In order for the system to enable fast convergence while some datacenters cannot move traffic, a coefficient Y may be defined.
    - When a datacenter received the alarm from a remote site, the datacenter could move (X*Y*local traffic) to other link(s);
    - The default value and the minimum value for Y is 1, and the maximum value for Y is 1/X.
    - Y is per link per remote site.

  - Y will be set to the default value if no continuous alarm is received for the same link from the same site.
  - Y = 1/2X, Y may be set to 1/2X if the datacenter received the second continuous alarm from the same link of the same site.
  - Y = 1/X, Y may be set to 1/X if the datacenter received the third continuous alarm for the same link from the same site.
  - This means each datacenter could make itself to be healthy after receiving at most three alarms from the same branch if there is an available alternative link since it moved all traffic from the unhealthy link.
    - It also means that all load-balance traffic is moved away from the link, leaving the link for critical traffic.
  - Each implementation could define its own algorithm of coefficient Y based on its reality and requirement, or could make coefficient Y configurable.

With this implementation, the second typical scenario problem "load-balance co-exists with preferred policy" described above could be resolved, as all load-balanced traffic could be moved from the MPLS link to other links when there is more and more critical traffic (which prefer MPLS).

The solutions presented herein may be more light-weight compared to other proposed solutions since the alarm may only be triggered after ingress traffic crosses the threshold as opposed to sending messages periodically.