

Seeing through transparency claims



Symmetric acceleration devices communicate in two ways across a WAN but as CRAIG STOFFER, VP marketing at Silver Peak Systems argues, only one has the answers to the real world problems faced when tuning WANs for symmetric communication

Many WAN acceleration solutions are “symmetric” in nature, requiring communication between devices on each end of a WAN link. In order for these solutions to work, there must be a reliable way for traffic to be encoded/compressed on one end of the WAN link and successfully directed to a device on the other end of the link where the traffic can be decoded prior to final delivery.

There are two primary ways in which symmetric acceleration devices communicate across a WAN. These are:

Header transparency: WAN acceleration appliances preserve the original source and destination IP addresses (src/dest) of the clients and servers when sending packets across the WAN. This information is used to route the accelerated traffic in the same manner as un-accelerated traffic. It is important to point out that while the headers remain unchanged in this scenario, payload information is fundamentally altered as compression, data reduction, and other acceleration techniques are applied to traffic.

Peering: During transit over the WAN, the original src IP address is replaced with the IP address of the near-end WAN acceleration appliance. The original dest IP address is replaced with the IP address of the far end acceleration device. Peering is most commonly implemented using Network Address Translation (NAT) or by encapsulation within tunneling protocols, like GRE.

By using intermediary IP addresses, peering establishes well-defined ingress and egress points for all accelerated traffic across the WAN. After acceleration takes place, the intermediary headers are replaced with the originals. As is the case with header transparency, when traffic is optimised and sent across the WAN using peering, payload information is inherently altered.

Proponents of header transparency argue that this method of communication is superior because it preserves existing policies across the WAN. By maintaining original headers, routers, monitoring devices, and other network elements within the WAN can theoretically enforce ACLs and quality of service policies on accelerated traffic just like they would on non-accelerated traffic. Naturally, this is attractive if significant time and effort has been spent creating these policies prior to implementing a WAN acceleration solution. Unfortunately, this ‘transparency’ argument is not as clear as one might

expect. In fact, a peek behind the curtain reveals that when peering is coupled with the right technologies, it actually provides added visibility with fewer side-effects than header transparency. Here’s why:

Issue: visibility into the payload

In both communication methods described above, payload information is obscured when traversing the WAN. This is a natural result of all compression and data reduction techniques. There is no way that an

acceleration device can compress traffic or eliminate repetitive bytes of data, for example, without altering the payload. In fact, that is the whole point of performing those acceleration techniques – to reduce the payload for faster transfer across the WAN.

The only way to provide this level of visibility is to collect statistics prior to optimisation and export them to an external device. That is why many WAN acceleration appliances (using both header transparency and peering) support Netflow and similar techniques.

Header transparency only provides visibility into headers. It does not enable downstream devices to garner useful information from the payload. Therefore, Header transparency must be deployed in conjunction with Netflow or a similar tool for this type of visibility.

Peering must also be deployed in conjunction with Netflow or a similar tool for visibility into the payload.

Issue: dealing with ephemeral ports

FTP, VoIP, MAPI, and many other “ephemeral” applications dynamically negotiate ports and embed this information within the data stream. This is in contrast to HTTP, CIFS, and other applications that use statically defined ports for communication. Because port information is required to enforce application-based ACLs and QoS in many environments, the only way a device within the WAN can properly

enforce these policies on ephemeral applications is to perform real-time inspection of the control stream. Policies based on static ports will not work in these environments because the port information is not known in advance of data transfer.

The only way to accurately enforce policies on applications using ephemeral ports is to perform deep packet inspection, whereby port information can be pulled out of the data stream in real-time. This can be done irrespective of the method of communication used across a WAN, but must be performed prior to payload optimisation. It is often recommended that deep packet inspection take place within the WAN acceleration device, where specific policies can be enforced prior to WAN traversal or data can be collected and exported to the necessary devices before the payload information is obscured. Any packet inspection that takes place downstream of the acceleration device will not yield meaningful results due to the effect that the optimisation process has on the data stream.

Header transparency provides visibility into the header only. It has no effect on the enforcement of policies in a WAN when ephemeral ports are employed, because the necessary information is embedded within the payload. As a result, WAN routers, firewalls, and other devices that rely on headers to make their decisions cannot be used to enforce application-specific policies in an accelerated WAN when header transparency is used. This is an important point that often gets lost when proponents of header transparency argue their case – this method of communication is only pertinent when enforcing policies on simple protocols (i.e. those that use static ports).

Peering: Like header transparency, peering in and of itself does not provide the necessary visibility into the payload to support policies on ephemeral ports. To achieve this, deep packet inspection is required prior to optimisation.

Issue: identifying accelerated vs non-accelerated traffic

It is often useful for network administrators to identify which traffic is being accelerated across the WAN and which traffic is not. This can be used for troubleshooting, and monitoring performance as well as resource planning.

Impact: header transparency because the original src/dest IP address is always used, it is impossible to identify what traffic is optimised and what traffic is not optimised across the WAN when using header transparency.

Impact: peering: when device peering is used, monitoring devices can easily identify accelerated traffic across the WAN by looking at the src/dest addresses used in the headers. Any

traffic that has the src/dest addresses of the acceleration devices can be easily identified as optimised traffic.

Issue: design complexity

In a large network, multiple paths might exist for a source and destination device to communicate across a WAN. Regardless of whether header transparency or peering is used, the best path across the WAN will be chosen in a connectionless fashion – as would be the case with any routed traffic.

Impact: header transparency does not specifically define an egress point on the far side of a WAN. If multiple paths

exist, it is possible that traffic can be accelerated on one end of a WAN link, but never ‘un-accelerated’ at the far end. This would result in the delivery of corrupt information to the destination device. This can be avoided with careful planning – i.e. ensuring that a WAN acceleration device exists on every possible path between a source and destination device. But, this can complicate network design and management, and can increase the overall expenditures spent on WAN acceleration.

Impact: peering: when peering is used, the WAN egress point is well defined. In other words, because the IP address of the far-end WAN acceleration device is used as an intermediary destination address, it is guaranteed that all traffic will ultimately make its way to this device. This ensures that the traffic is always decoded and decompressed prior to final delivery, which eliminates the chance of data corruption and simplifies network design.

Choosing the right approach

There are several ways to effectively communicate across a WAN but no single method by itself delivers the level of transparency required to preserve all existing policies across one. For a WAN acceleration solution to be truly transparent it must satisfy the following criteria:

- Work across all applications – not just those using a specific type of port or transport protocol.
- Co-exist with other devices in the network, such as IDS/IPS and firewalls.
- Be completely seamless to the client, server and application itself without any risk of data corruption.