



Silver Peak

| white paper

WAN DEDUPLICATION: GETTING MORE BY SENDING LESS

LOWER WANBANDWIDTH COSTS
AND IMPROVE TRANSFERTIMES BY
ELIMINATING REPETITIVE DATA



WAN DEDUPLICATION: GETTING MORE BY SENDING LESS

*LOWER WAN BANDWIDTH COSTS AND IMPROVE
TRANSFER TIMES BY ELIMINATING REPETITIVE DATA*

Data deduplication is one of the most significant technologies to impact the storage community in recent years. By identifying redundant data segments and storing only a single instance of information, this technology dramatically reduces storage space and allows more data to be protected over time.

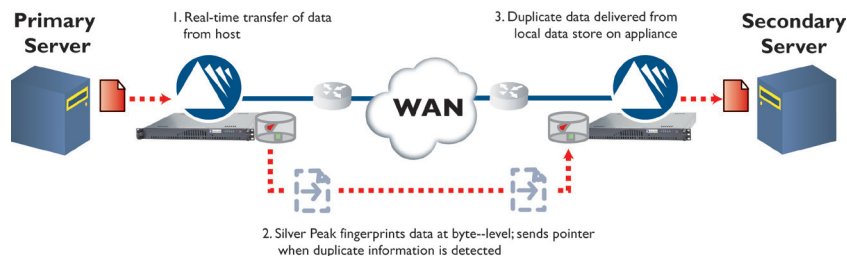
But that is just the beginning. In recent years, data deduplication has also made its ways into the ranks of the networking community. More specifically, it has become an important tool for optimizing application performance across the WAN. By eliminating the transfer of repetitive IP traffic, deduplication significantly improves WAN utilization and accelerates data transfers between geographically disperse locations. This saves bandwidth costs and helps to overcome many obstacles when communicating across a WAN.

Because WAN deduplication works on all IP traffic, it plays a key role in a variety of IT initiatives, including server centralization, virtualization, and application delivery. In addition, it is essential to improving the performance and reliability of data replication, backup, and recovery across the WAN. In this respect, WAN deduplication is actually a nice complement to storage deduplication, resulting in even higher cost savings and better Recovery Point and Time Objectives (RPO/RTOs) across the enterprise.

DUPLICATING THE SUCCESS OF DEDUPLICATION

WAN deduplication (also called “disk based data reduction”) works by deploying acceleration appliances in each enterprise location (i.e. on both ends of a WAN connection). The WAN acceleration appliances inspect IP traffic in real-time and store a local instance of information in an application independent data store at the appropriate enterprise location. The local instance is transparently populated based on day-to-day usage, containing a subset of the enterprises working data set that is most relevant to each location. Each instance of information is stored only once per location, enabling an appropriately sized WAN acceleration appliance to hold weeks or months worth of data.

WAN acceleration appliances equipped with deduplication technology examine outbound packets to see if a match exists in the local instance at the destination location. If a match exists, then the repetitive information is not sent across the WAN and instructions are sent to deliver the data locally. If the data has been modified, only the delta is transmitted across the WAN, maximizing bandwidth utilization and application performance.





Deduplication overcomes WAN challenges that often plague common business continuity processes, including backup, replication, and disaster recovery. More specifically, this technology delivers the following benefits:

- Improve data transfer times. By delivering repetitive information from local data stores (as opposed to re-sending it across the WAN), WAN transfers are handled at LAN-like speeds. More advanced solutions perform data reduction on both TCP and UDP traffic, delivering significant performance improvements across a wide range of traffic types.
- Maximize WAN efficiency. Deduplication can reduce as much as 99% of WAN traffic by eliminating the transfer of duplicate information. When performed at the byte level, repetitive patterns can be detected and eliminated even when the backup/replication solution is performing similar functions at the block level.
- Increase geographic distances. By reducing the impact of latency, enterprises can extend the distances between data centers and disaster recovery locations, increasing operational flexibility.

COMPLEMENTARY SOLUTIONS

WAN deduplication is quite complementary to storage deduplication. While the latter focuses on improving storage capacity, the former focuses on delivering the best possible performance across the WAN. When deduplication is enabled in the WAN, enterprises typically see a 10–20x performance improvement above and beyond what is achieved with deduplication in the storage medium alone. This can be attributed to several factors.

For one, WAN acceleration devices typically provide greater accuracy than storage devices when searching for repetitive patterns. This is because individual bytes of data are examined as opposed to blocks, which enables more repetitive patterns to be discovered — even within the same replication stream. In addition, when data deduplication is performed at the network layer, it works across all IP traffic (regardless of the application). Therefore, data sent via email, file, or web transfer will immediately register as a “hit” when it is sent across the WAN as part of a backup or replication process. In other words, the application itself may not consider the data repetitive, so data deduplication may not work from a storage standpoint. However, it is duplicate data from a WAN perspective, so a WAN acceleration appliance will treat it as such.

WAN deduplication also works in a bi-directional fashion. In other words, when data is sent from point A to point B, both locations are aware of the information and can deliver the information locally using references, regardless of which direction the traffic is flowing. This can dramatically improve the speed upon which an enterprise can recover data.

For example, if information was recently sent across the WAN in one direction as part of normal operations (that is, replication/backup or simply via email or FTP), then it can be immediately detected when resent in the opposite direction as part of the recovery process. Rather than retransmit an entire data set across the WAN in that scenario, this information can be delivered from local data stores for greater efficiency and performance.

WAN acceleration appliances also incorporate other optimization techniques into the mix that complement data protection processes. For example, payload and header compression are often used in conjunction with data deduplication to further reduce the amount of WAN bandwidth required for backup and replication. By providing compression within the WAN acceleration appliance, this functionality can be offloaded from the host replication server, ensuring better scalability and performance.

In addition, significant performance improvements can be provided with compression even when non-repetitive information is sent across the WAN. WAN acceleration can also be used to reduce the impact of both packet loss and jitter that occurs when router links are oversubscribed and drop or re-order packets, and they deliver specific enhancements to overcome latency that is inherent to different traffic types, such as TCP.

Lastly, Quality of Service (QoS) techniques can be used to prioritize traffic and allocate necessary bandwidth for business-critical functions, such as data replication.



BENEFITS OF A COMBINED SOLUTION

The most efficient way to save on capacity and improve performance when transferring and storing data is to eliminate redundant information. When data deduplication is used by a WAN acceleration appliance in conjunction with other WAN optimization techniques, the following benefits can be achieved:

- Meet and exceed Recovery Time Objectives (RTO)
- Improve Recovery Point Objectives (RPO)
- Increase geographic distances between data centers
- Avoid costly WAN bandwidth upgrades
- Avoid database synchronization issues that arise when backup and replication tasks are not completed within allocated windows
- Manage WAN capacity to better handle peak loading, emergency contingencies, and business growth

Data deduplication is a proven technique that improves the performance, reliability, and efficiency of data backup and recovery. By utilizing this technology in both the storage medium and across the WAN, enterprises can improve their data protection processes even more. In addition, deduplication can be combined with other WAN optimization techniques to enable a variety of other strategic IT initiatives, including server centralization, virtualization, and application delivery.

Deduplication is one of the rare technologies that have had a significant impact on both the networking and storage communities. While these groups do not always see eye to eye, on this one point they agree—when it comes to backing up and recovering data, less can actually be more.

