

The Global Data Transport Problem: Summary of the Digital Fountain Solution

Global corporations and media companies need to share information among their regional offices, clients, customers and partners. The amount of information that is considered “mission critical” is growing all the time and for this information, accelerated and predictable delivery is critical. Unfortunately, over long-distance WANs, the performance of traditional TCP-based data transfer technologies is poor, no matter how much bandwidth is available. Solving this problem and accelerating global transfers is where Digital Fountain comes into play.

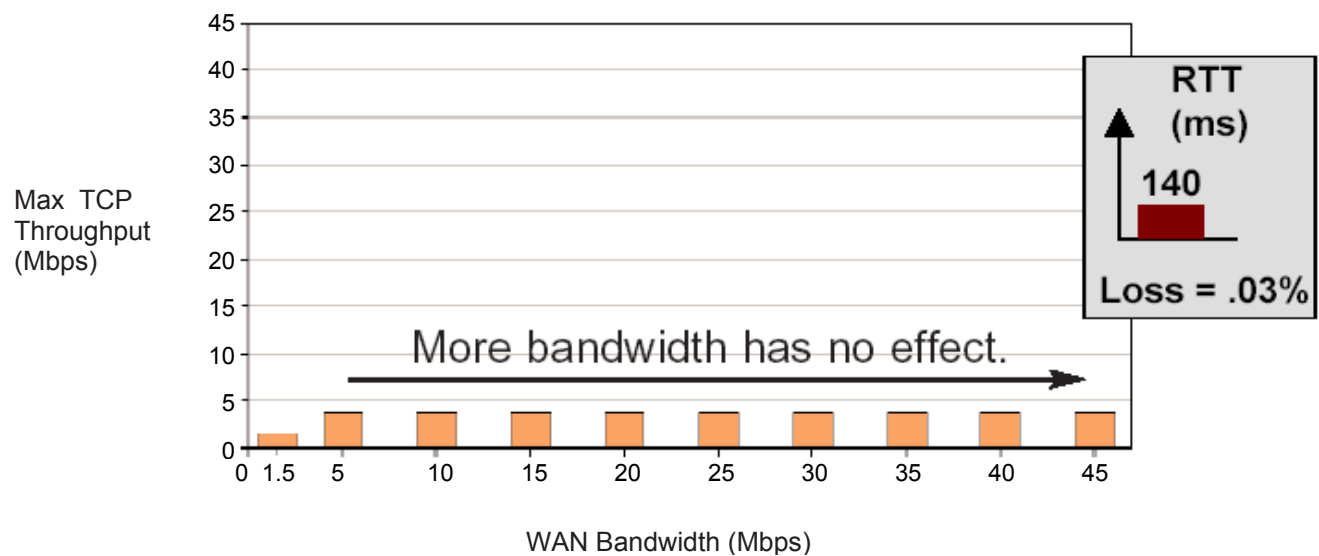
LOS ANGELES TO ...	DATA SIZE (2 GB) (RTT, LOSS)	TRANSFER TIME	
		TCP/ FTP / HTTP	DIGITAL FOUNTAIN
Hong Kong	(360, 1%)	17 hours	28 minutes
London	(140, 1%)	7 hours	28 minutes
New York	(70, 1%)	3 hours	28 minutes

The problem with TCP

The performance problems with TCP and related applications are nonintuitive. One would think that adding more bandwidth would always provide greater performance. But, over global WANs, this is not the case. And, as corporations increasingly have more bandwidth than a T1, the problems with TCP are becoming more and more apparent.

The factors that limit TCP throughput are packet loss and Round Trip Time (RTT). RTT is the time it takes a packet to travel from the sender to the receiver and back. RTT is a function of distance (and/or router hops). The greater the distance, the longer the RTT. As loss or RTT increases, the lower the maximum achievable TCP performance, no matter how much bandwidth is available.

FIGURE 1: Maximum Throughput for TCP-based Applications for given WAN Bandwidth (Los Angeles to London)



With 140 ms of RTT and just 0.03% loss (10X better than typical carrier SLA), maximum achievable throughput with TCP is approximately 4 Mbps regardless of available bandwidth.

To understand why this is, we need to look at how TCP works. TCP is a reliable transport that uses a send-acknowledge process. To achieve reliability (the assurance that all data is completely and accurately transferred), data is sent sequentially and acknowledged as it is received. If any data is not acknowledged, it is considered lost and is re-sent.

Data loss is frequently due to network congestion and so TCP also incorporates congestion-avoidance features. The congestion-avoidance algorithm attempts to detect congestion by constantly measuring RTT and packet loss. As RTT increases, TCP assumes this means that the network is congested and that the routers between the source and destination are busy, causing longer queues and hence delays. In response to long RTT, TCP reduces its transmission rate. The response to packet loss is similar.

Using RTT as a proxy for congestion is reasonable on a LAN since long RTT usually means there is congestion. But, in the case of a long-distance WAN connection, long RTT is ambiguous. It could indicate that there's congestion, but it could also merely reflect the fact that it takes longer to send and acknowledge a packet when distance extends to thousands of miles. Unfortunately, TCP cannot discriminate between these two cases and so long RTT causes TCP to lower the transmission rate in both cases.

The result of above is that over a long distance WAN, you can add more and more bandwidth and get ZERO increase in throughput!

Digital Fountain's Solution

At its core, Digital Fountain's patented Meta-Content is a reliable data transport that dramatically simplifies the processes required to completely and perfectly deliver data over any digital network. This technology is essentially immune to network impairments such as packet loss and delay, making it fast and predictable. With Meta-Content technology, organizations gain complete control over throughput up to the available bandwidth, irrespective of RTT and loss, allowing them to realize dramatically increased transfer speeds and predictable delivery times.

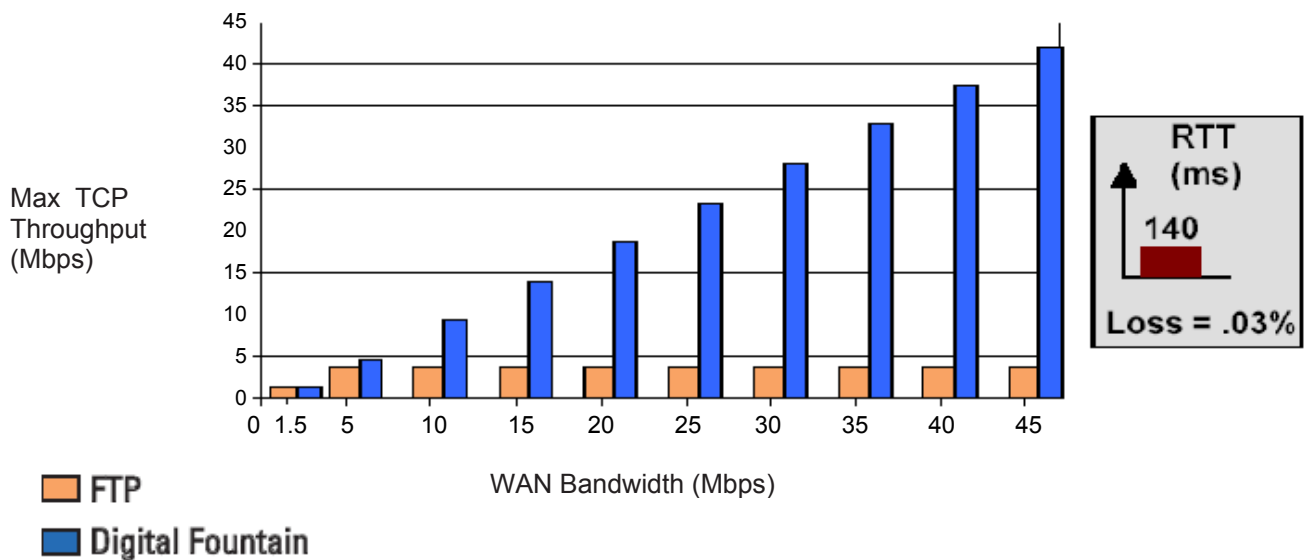
With Meta-Content, rather than sending the original data, a precise mathematical "recipe" for the original data is sent. The mathematical recipe is a series of values of different equations. A never-repeating stream of as many as required Meta-Content packets can be generated from the original data.

The benefit of Meta-Content is that to receive a complete and perfect copy of the original data, a receiver needs only to receive any combination of Meta-Content packets equal in aggregate length to the length of the original data. But, the key is that it doesn't matter which Meta-Content packets are received. Any combination of sufficiently many Meta-Content packets is equally as good as any other combination to regenerate an exact copy of the original data.

For example, assume the original data is 100 packets in length. To receive the data with TCP, all 100 packets need to be received in order. If any are lost, they have to be resent. With Meta-Content, 100 packets need to be received, but any 100 packets will do. Since the Transporter Fountain can generate a never-repeating stream of as many as required Meta-Content packets from the original data, Meta-Content packets never ever need to be resent. If one is lost, the next one that comes along is just as good. One could receive the first five packets, miss the next three, and then receive the next 95 and be done. One could receive every other packet, but when any 100 are received, delivery will be complete. Meta-Content packets are never repeated and so receivers never receive the same one twice. Once the receiver has a sufficient number of packets, it sends a "complete" message back to the Transporter Fountain to signal that it can stop generating and sending Meta-Content.

Here's an example of how Digital Fountain provides accelerated transfers over the WAN. Whereas with TCP, transmission rate is automatically determined by the RTT and packet loss, with Digital Fountain Transporter Fountain, transmission rate can be directly controlled regardless of distance, RTT, loss, congestion, etc. In essence, Digital Fountain takes control of rate from the independent machinations of TCP and gives it to the administrator. It is this control that allows for such greater speeds. In the example below, the orange bars represent maximum TCP throughput, while the blue bars represent Digital Fountain maximum throughput.

FIGURE 2: Maximum Throughput for given WAN Bandwidth (Los Angeles to London)



With 140 ms of RTT and 0.03% of loss, TCP performance is capped to about 4 Mbps independent of available bandwidth; Digital Fountain performance is capped only by available bandwidth.

The Mathematical Foundation of Meta-Content

The foundation of Meta-Content is analogous to a basic algebraic technique called simultaneous linear equations. That concept says that if you had, for example, three independent equations and three unknowns, you could simultaneously solve for the three unknowns. And, if you had four independent equations and three unknowns, you could choose any three of the four equations and simultaneously solve for the unknowns. Essentially, Meta-Content technology generates a never repeating stream of linear equations from the original data. To solve for the original data, all one needs to do is have enough equations, but it doesn't matter which ones.

Meta-Content is also very fast. Even on a low-end home PC, original data can be regenerated from Meta-Content at 10's of Mbps. With faster hardware or by keeping the process in RAM, there's really no limit to how fast the regeneration process can go. In any practical scenario, Meta-Content generation is faster than the WAN and so computation time is not a factor.