ADVA™ Optical Networking

TECHNOLOGY WHITE PAPER

Ultra Low-Latency Financial Networking

Author:
Jim Theodoras,
ADVA Optical Networking

Financial networks are one of the most important applications in networking, yet seldom openly discussed. This is partly due to the nature of the business itself, where the slightest technical edge can lead to $Billions in profits. However, this secrecy can at times be self-defeating if it prevents advances that can occur through more open discussion of the technical hurdles to be solved. Slowly, the veil is lifting on the needs of financial networks, and the unique challenges faced.

This technology white paper takes a closer look at financial networks and their unique low-latency needs, examining all contributors to total latency between two geographically dispersed software applications (for example, information feed and matching engine).

Speed versus Latency

The needs of financial networks are often lumped into a single adjective, "fast". They need to be fast, or "I want the fastest network". But what is meant by "fast"? Too often, "fast" is misinterpreted to mean the largest capacity network, either the most 10Gs, 40G, or even earliest adoption of 100G. However, there are plenty of big, fast network pipes already out there that do not necessarily meet the needs of the financial industry. Closer examination reveals financial networks need low-latency, meaning the time between an instruction being issued and executed is as low as possible. For example, consider the analogy of being at a sporting event and wanting to take a still photo of a rapidly changing scene. Setting the shutter speed to 1/2000th of a second makes a camera fast, and will help capture the motion. But, if the lag between pressing the shutter button and the camera actually taking the picture is too long, the event will have passed before it can be captured. In this case, the speed was fast, but the latency was too high. Returning to the topic of financial networks, a 100G network may be fast, but depending upon the overall latency of the network, a trade command issued over the network may still not be the first to execute.

> *... financial networks need low-latency, meaning the time between an instruction being issued and executed is as low as possible.*

Sources of Network Latency by Network Layer

All modern networks are based upon the Open System Interconnection (OSI) Reference Model which consists of a 7 layer protocol stack, and today's financial applications reside at the upper layers of the protocol stack. Unfortunately, the physical transport of information between financial applications occurs at the bottom of the protocol stack, thus forcing all data to have to tunnel down all layers for transport, and then back up the stack again after transport. Each of these layer transitions adds latency. Here is a brief summary of various lag

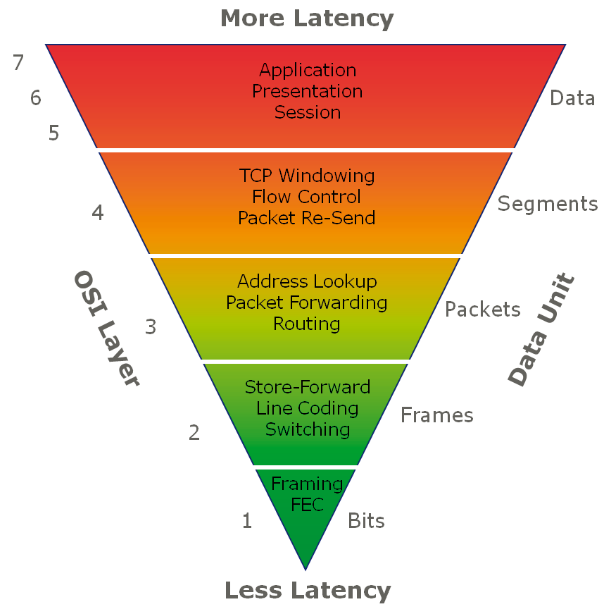

Figure 1: OSI reference model and latencies incurred at each network layer

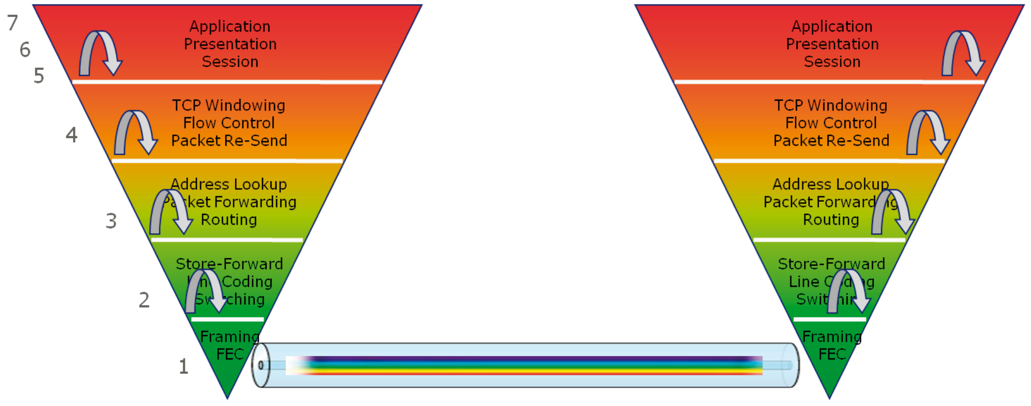

Figure 2: Data being transport must transition all protocol layers at least twice

inducing mechanisms that may be encountered as the information tunnels down the protocol stack.

Contrary to popular convention, not all network information is called “data”, as only information residing at Layers 5-7 is officially referred to as data. Layers 5-7 are where the application, presentation, and session layers formally reside, and may be distinct layers, or may be squeezed together into something quite different. Financial application software resides in these layers, and how they operate, access, and command the lower layers of the protocol stack is highly proprietary to each financial company. What can be said, however, is that this is where large investments into high performance computers (HPC) pay off, as latency at these layers is inversely proportional to computing power.

Information at Layer 4 is packaged into "segments". Layer 4 is referred to as the "transport" layer of the network, which can be confusing as no actual physical transport occurs at this layer. Rather, the transport of segments is guaranteed at this level through the setup of end-to-end connections, for example from a web browser to a remote server. Transmission Control Protocol (TCP) is one of the most common handshaking protocols, with file transfers, email, and now even the worldwide web based upon it. Unfortunately, despite the historical significance of this layer, it remains a substantial contributor to latency in financial networks. Layer 4 serves as the concierge for the applications that reside in layers above it. As an analogy, think of the concierge at the front of a high-rise hotel hailing cabs one at a time from their lineup on the street, as guests randomly descend from the floors above, each with a different destination, each patiently waiting their turn at a cab. A good concierge will ask where folks are headed, and group people with similar destinations into larger vehicles, like shuttles to the airport. TCP works in a similar manner, as it receives data from the financial applications being run, packages them into segments, loads them into a buffer, and then waits …

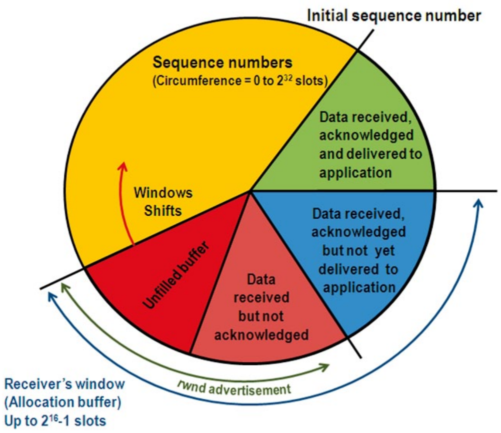
Initial sequence number
Sequence numbers
(Circumference = 0 to 2^{32} slots)
Data received, acknowledged and delivered to application
Windows Shifts
Data received, acknowledged but not yet delivered to application
Unfilled buffer
Data received but not acknowledged
Receiver's window (Allocation buffer) Up to 2^{16}-1 slots
rwnd advertisement

Figure 3: TCP sliding window flow control protocol (source: Wikipedia)

Yes, the Achilles heel to TCP is that it waits for the acknowledgement from the previous transmission before sending the next segment. Actually, this would be horribly inefficient, so TCP uses what is called a "sliding window" flow protocol where it simultaneously fills a new segment, while sending an unacknowledged segment, while receiving acknowledgement from the previous segment sent. A good analogy for this would be the rental car kiosk at the airport, where a steady stream of shuttle busses stop on a regular basis, and travelers board the busses on a best case basis, sometimes boarding the bus early and having to wait to leave, sometimes boarding the bus just before it leaves (minimum latency), and sometimes missing the bus and having to wait for the next bus. To make matters worse (and more complicated), the size of the bus and the distance between kiosk and terminal directly impact the rate at which you can transfer the passengers. With TCP, the size of the bus is called the TCP receive window, and the distance is equivalent to the sum of all time delays between the TCP sender and receiver. These two parameters actually limit the amount of information that can be transmitted between two points in a financial network, and is often referred to as the TCP bottleneck.

The TCP bottleneck is typically alleviated through a variety of techniques, such as turning off segment re-send, modifying window sizes, using a modified version of TCP called SCTP, etc. More recently, with the emergence of multi-core microprocessors, each core can be assigned separate TCP sessions, and then segments are carefully spread across all the sessions in a technique called "TCP striping".

Information at Layer 3 is packaged into "packets". Layer 3 contains the network layer, where logical (Internet Protocol or IP) addressing and switching of packets occurs. Once a flow of packets (session) is midstream, additional latency is limited only to bus delays, which tend to be minor in the grand scheme of things. Most of the latency penalty at Layer 3 occurs when a packet first arrives, as its IP address must be looked up, and then the packet forwarded to appropriate port. Address lookup and packet forwarding consume the most power, money, and latency in the IP routers that form the internet as we know it today.

Bit rate (Gbit/s)	Time (ns)	
	64 byte	512 byte
0.1	5120	40960
1	512	4096
10	51.2	409.6
40	12.8	102.4
100	5	41

Table 1: Minimum time to transmit a packet

Information at Layer 2 is packaged into "frames". Layer 2 contains the data link layer where physical (Media Access Control or MAC) addressing and switching occurs. Assuming a Layer 2 switch is under-subscribed, meaning the incoming frames sum up to less than the switching bandwidth available, then the only sources of latency will be framing-up to any line code present (e.g. 8b/10b) and buffering. The latency of buffering will depend upon the type used, either store-and-forward or cut-through. With the former, the latency incurred will be a function of the size of the frame being buffered, as the entire frame must be buffered before being sent on its merry way. The longer the frame, the longer you must wait before it can be sent. This is one of the reasons higher data rates are thought of as faster – because any store and forward buffering in the link

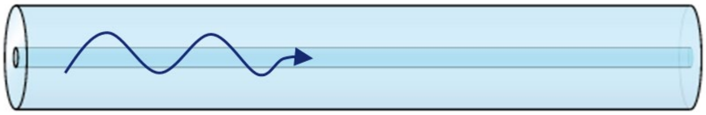

Figure 4: Propagation delay in glass fiber ~ 1ms per 100km round trip

will happen quicker. Cut-through buffering offers much lower latency that is not dependent on frame size, as a frame is sent out as it is received. The problem with cut-through buffering is that by the time the error code is checked at the end of a frame, it has already been sent.

There is a common misconception that there is nothing that can be done to minimize Layer 1 transport delays, as information is already traveling at the speed of light in a glass fiber.

Information at Layer 1 is simply raw "bits". Layer 1, the lowest layer, is where the actual physical transport of information occurs. Layer 1 contains one of the largest contributors to latency, propagation delay – or actual time-of-flight of the information through the transport media. There is a common misconception that there is nothing that can be done to minimize Layer 1 transport delays, as information is already traveling at the speed of light in a glass fiber. In fact, there is a lot that can be done to minimize transport delay, and often more latency can be shaved here, for less money, than all the other layers combined. Let's look at some of these techniques and strategies.

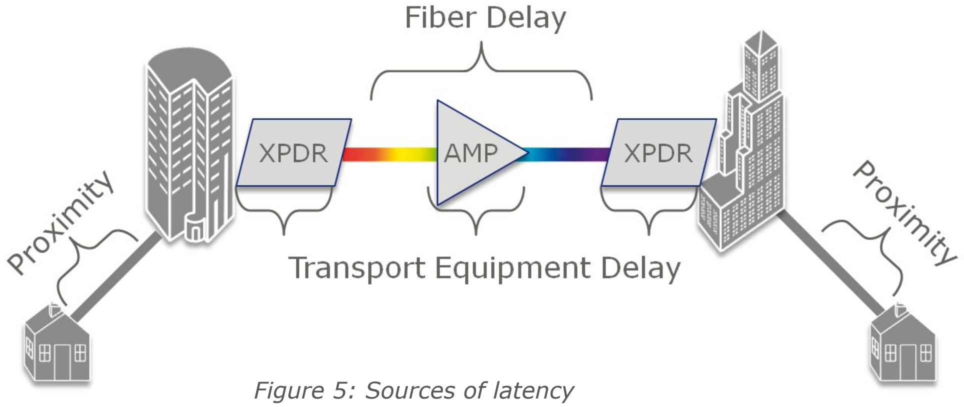

Figure 5: Sources of latency

Sources of Transport Delay

The three major sources of delay in getting between two points in a financial district are fiber optic cable, proximity, and transport delays.

Fiber optic delay is simply the time the data spends in the fiber optic cable. The shorter the cable, the lower the delay. While cables are envisioned to run directly between two locations, the truth is much less ideal. Financial districts are complex rats' nests of fiber cables crisscrossing a city, crossing roads, going up and down manholes, and along available easements. Shortening the network's fiber length often means finding a dark fiber provider that has pieced together a shorter route.

Proximity delay is how close you can get to the available fiber access points. It is a basic geometric problem. There can only be so much real estate nearby a fiber junction, and locating your equipment in buildings closest to this point lowers the delay, often at a co-location provider's site.

Transport equipment delay is the amount of latency that is added by any and all optical gear that the data encounters as it is transported along the fiber. In

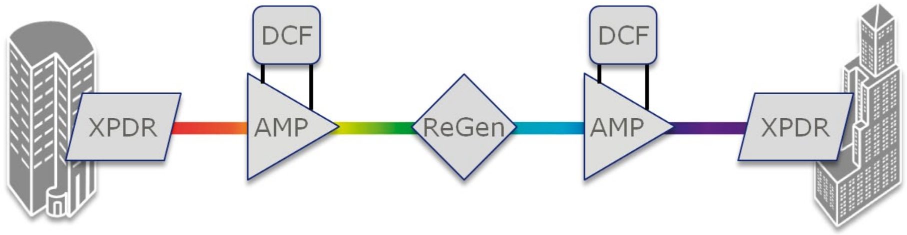

Figure 6: Transport elements

the past, when trade execution times were in the seconds, the transport gear's delay was considered in-consequential. Today, as other sources of latency have been substantially reduced, optical transport equipment delay is a real contributor to path latency. Through careful consideration of all the equipment in the information path, and the functions they provide, ADVA Optical Networking has developed proprietary technology to substantially reduce latency.

... as other sources of latency have been substantially reduced, optical transport equipment delay is a real contributor to path latency.

Optical Transport Functions

The fiber optic link between two financial sites is anything but bare cable. There are several optical transport functions that are needed for one's data to ride on a fiber:

Color conversion – The majority of optical transport links today are Wavelength Division Multiplexed (WDM), which means everyone is assigned a color channel. In order to transport one's data over a WDM network, it must first be converted from grey to a color. This function is called "transponding". If done incorrectly, milliseconds can be incurred here. ADVA Optical Networking has a wide range of technologies that enable our transponder latencies to be best-in-class, from high nanoseconds to low microseconds, depending upon functions needed.

An often overlooked feature of color conversion is "muxponding", or the aggregation of lower speed traffic into a higher speed signal. For example, today most information feeds are 1Gbit/s, yet most transport links are 10Gbit/s. So, frequently 10 different infor-mation feeds are squeezed into a single transport color. The standard way of doing this is ODU en-capsulation, which can add double digit microseconds to a link, on each end. ADVA Optical Networking has proprietary encapsulation techniques, in addition to standard offerings, allow best-in-class muxponding.

Optical amplification – A signal traveling down an optical fiber gets smaller and smaller with each kilometer of distance. Optical amplifiers are used to boost the signal as it weakens. Traditionally, a type of optical amplifier called EDFA (Erbium Doped Fiber Amplifier) is used, and the delay through EDFA's was considered negligible. However, in today's trading environment where every nanosecond counts, the hundreds of nanoseconds incurred at each amplifier along a path can no longer be ignored. Some high gain, dual stage EDFA's can have delays in the low microseconds. ADVA Optical Networking is one of the few optical transport vendors that still builds their own amplifiers today, allowing us to offer exclusive low-latency amplifier designs with delays in the nanoseconds.

In addition to EDFA amplifiers, which use an external spool of Erbium doped fiber to provide amplification, ADVA Optical Networking also has a long history of best-in-class Raman amplifiers that pump the actual fiber in the ground to provide gain. Since no additional fiber is needed for gain medium, Raman amplifiers tend to be faster than EDFA's. When both EDFA's and Raman optical amplifiers are carefully used in tandem, maximum signal amplification, with minimal latency can be achieved.

Dispersion compensation – Just as rain in the sky, or glass prisms can spread light into a rainbow of colors, so do fiber optic cables. This smearing of an optical data signal into multiple colors is predominately an issue at 10Gbits/s, and can cause a signal's assigned color to bleed into neighboring channels. This prism effect is called Chromatic Dispersion (CD). Tradition-ally, long spools of a special type of fiber called Dispersion Compensating Fiber (DCF) have been used to reverse the smearing of colors. However, as these spools can add up to hundreds of additional kilometers of fiber in a path, they are poorly suited to low-latency applications.

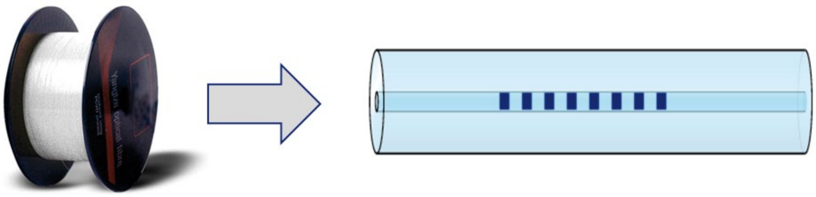

Figure 7: Spool of DCF versus FBG

ADVA Optical Networking has developed alternative dispersion compensation techniques that remove the need for large spools of DCF – Fiber Bragg Gratings (FBG's). FBG's can be thought of as a reverse prism written into a piece of fiber optic cable less than a meter long. They essentially have negligible delay, and greatly reduce the end-to-end latency of a path.

Electrical regeneration - Finally, even with the help of optical amplification and dispersion compensation, on longer links a signal will degrade to the point where it needs to be regenerated. How the signal gets regenerated greatly determines the additional path delay incurred. Traditional regeneration techniques can add hundreds of microseconds of unnecessary delay. ADVA Optical Networking has developed a large family of low-latency regenerators that use proprietary technology to reduce latencies into the nanoseconds.

ADVA Optical Networking has developed a special family of low-latency transponders, at practically every conceivable data rate, that have all been optimized for financial networks.

Why ADVA Optical Networking is ahead of the Pack

ADVA Optical Networking has been a trusted partner to the financial community for over a decade. Through close relationships in the industry, we have developed unique technologies to solve our customers' problems. Because of these efforts, we now offer an unmatched product family for financial networking, including, but not limited to the following:

1. **Ultra-low latency wavelength conversion**
   ADVA Optical Networking has developed a special family of low-latency transponders and muxponders, at practically every conceivable data rate, that have all been optimized for financial networks.

2. **Ultra-low latency inline amplification**
   ADVA Optical Networking builds its own amplifiers, allowing it to offer a special family of low-latency amplifiers. Whether EDFA or Raman, ADVA has optimized optical amplification for latency, thus obviating the need for regeneration or expensive coherent detection techniques.

3. **Zero-latency dispersion compensation**
   As already mentioned, long spools of dispersion compensating fiber (DCF) can be replaced with Fiber Bragg Gratings (FBGs). However, doing so degrades the optical signal, and ADVA Optical Networking has developed technologies that compensate for the impact of FBGs in transport networks.

4. **Multi-protocol support**
   Financial networks are comprised of more than simply Ethernet. High performance computing (HPC) clusters are typically connected with InfiniBand. Storage Area Networks (SANs) are typically interconnected with Fibre Channel. Only ADVA Optical Networking allows all three to be simultaneously transported over the same low-latency network, either on separate color channels, or multiplexed into a single 10GE carrier signal.

5. **In-service latency measurements**
   While Exchanges are scrambling to offer more frequent updates to trading latencies, ADVA Optical Networking offers real-time, in-service latency measurements today. Knowing latencies and their statistics brings a valuable new tool to tuning trading algorithms for maximum impact.

Conclusion

The field of low-latency networking for financial players continues to evolve and innovate at a relentless pace. Just a little over a year ago, trades were won or lost by milliseconds. By the end of the year, that was shaved to microseconds. Today 250ns can make the difference between winning and losing a trade. This new business reality requires increased focus on every source of latency in a path between two points. It is no longer sufficient to simply upgrade routers, or buy the latest multi-core processor based computer blades, as this only helps at the higher protocol layers. All sources of latency must be minimized at all network layers. Optimizing the physical transport layer and the optical equipment in the path offers the greatest latency reduction for the least cost/effort.

About ADVA Optical Networking

ADVA Optical Networking is a global provider of intelligent telecommunications infrastructure solutions. With software-automated Optical+Ethernet transmission technology, the Company builds the foundation for high-speed, next-generation networks. The Company's FSP product family adds scalability and intelligence to customers' networks while removing complexity and cost. Thanks to reliable performance for more than 15 years, the Company has become a trusted partner for more than 250 carriers and 10,000 enterprises across the globe.

Products

FSP 3000

ADVA Optical Networking's scalable optical transport solution is a modular WDM system specifically designed to maximize the bandwidth and service flexibility of access, metro and core networks. The unique optical layer design supports WDM-PON, CWDM and DWDM technology, including 100Gbit/s line speeds with colorless, directionless and contentionless ROADMs. RAYcontrol™, our integrated, industry-leading multi-layer GMPLS control plane, guarantees operational simplicity, even in complex meshed-network topologies. Thanks to OTN, Ethernet and low-latency aggregation, the FSP 3000 represents a highly versatile and cost-effective solution for packet optical transport.

FSP 150

ADVA Optical Networking's family of intelligent Ethernet access products provides devices for Carrier Ethernet service demarcation, extension and aggregation. It supports delivery of intelligent Ethernet services both in-region and out-of-region. Incorporating an MEF-certified UNI and the latest OAM and advanced Etherjack™ demarcation capabilities, the FSP 150 products enable delivery of SLA-based services with full end-to-end assurance. Its comprehensive Syncjack™ technology for timing distribution, monitoring and timing service assurance opens new revenue opportunities from the delivery of synchronization services.

ADVA Optical Networking
North America, Inc.
5755 Peachtree Industrial Blvd.
Norcross, Georgia 30092
USA

ADVA Optical Networking SE
Campus Martinsried
Fraunhoferstrasse 9 a
82152 Martinsried / Munich
Germany

ADVA Optical Networking
Singapore Pte. Ltd.
25 International Business Park
#05-106 German Centre
Singapore 609916

For more information visit us at www.advaoptical.com

ADVA™
Optical Networking

Version 07 / 2012