



**FileCatalyst®**  
reinventing file transfer

# Accelerating File Transfers

## Accelerating File Transfers

By: John Tkaczewski & Matt Murphy

# CONTENTS

Contents.....	1
Introduction .....	3
Introduction to FileCatalyst.....	3
Understanding TCP.....	4
The Pitfalls of TCP When Transferring Large Data Sets .....	4
Latency and How It Affects TCP .....	4
Packet Loss and Its Effect on TCP .....	6
The Pitfalls of Common TCP-Based Protocols .....	7
Optimizing TCP.....	7
Available Solutions .....	8
FTP/SFTP/FTPS Server .....	8
Email.....	9
Cloud Services/Digital Delivery .....	9
Shipping Physical Media.....	10
Common Drawbacks of Physical Media: .....	10
FileCatalyst Acceleration Features.....	10
Congestion Control .....	11
No Congestion Control .....	11
RTT-Based .....	12
Loss-Based.....	12
Multi-Client File Transfers .....	13
Compression.....	13
Incremental Transfers .....	13
Progressive Transfers.....	14
FileCatalyst and the Competition .....	15
Aspera .....	16
Signiant.....	17
Open Source.....	18
FileCatalyst Security.....	19

Required Ports – without Reverse Proxy .....	19
Required Ports – with Reverse Proxy .....	21
Minimizing Port Usage .....	21
Encryption .....	22
SSL Cipher Restrictions .....	22
IP Filters .....	22
Login Security .....	23
HIPAA Security Compliances.....	23
Penetration Testing.....	23
File Transfer Acceleration Scenarios.....	25
Scenario 1 – Small to Medium Enterprise (SME).....	25
Profile – Performing Large File Transfers .....	25
Challenges – Bottlenecks Created By FTP .....	25
Solution – FileCatalyst Direct.....	25
Scenario 2 – Multi-National Enterprise.....	26
Profile – Globally Delivering Content .....	26
Challenges – Underutilization of the connection .....	26
Scenario 3 – Cloud .....	27
Profile – Migrating Data to the Cloud .....	27
Challenges – Slow Upload Speeds.....	27
Solution – FileCatalyst Direct and HotFolder .....	27
References.....	28



# INTRODUCTION

As file sizes continue to grow, managing and delivering large files is becoming an important consideration for organizations of all sizes. Companies have abandoned the commonly used FTP/TCP protocol as a delivery method in favor of alternative file transfer solutions which provide acceleration, reliability, management, and security.

Companies seeking file transfer acceleration are not limited to the high-tech sector. Organizations leveraging the benefits of acceleration are found in sectors such as media and entertainment, natural resources, supercomputing, legal, health, government, financial, manufacturing, and more.

Companies using TCP-based file transfer protocols to transfer large data sets may experience slow file transfers, or even failed and/or corrupt file transfers. This failure rate can be detrimental to organizations moving large data sets on a regular basis. This wastes valuable time, especially if these transfers take hours across an otherwise healthy network.

This white paper will address some of the issues organizations encounter when using TCP-based protocols. It will also outline some other common file sharing methods, and the issues inherent with each. It will then outline the FileCatalyst solutions, and how they overcome the issues surrounding slow file transfers.

Finally, this paper will present a number of scenarios that showcase the advantages of switching to an accelerated file delivery system, such as FileCatalyst, that includes reliability, security, automation, and tracking.

## INTRODUCTION TO FILECATALYST

FileCatalyst is a pioneer in managed file transfers and an Emmy® award-winning leader of accelerated file transfer solutions. FileCatalyst, founded in 2000, has more than one thousand customers in media & entertainment, energy & mining, gaming, printing, military, and government, including many Fortune 500 companies.

FileCatalyst is a software platform designed to accelerate and manage file transfers securely and reliably. FileCatalyst is immune to the effects of latency and packet loss impacting traditional file transfer methods like FTP, HTTP, or CIFS. Global organizations are now using FileCatalyst to address file transfer needs, including content distribution, file sharing, and offsite backups.

# UNDERSTANDING TCP

The Transmission Control Protocol (TCP), in conjunction with the Internet Protocol (IP), is the basic framework and set of rules that define the internet and how data is sent and received. TCP is also the framework used by all the common internet protocols, including FTP, SFTP, HTTP, SCP, CIFS, and SMTP.

TCP is a connection-oriented protocol; meaning that it establishes a connection between applications at each end. TCP sends and receives packets across a network between each endpoint.

TCP can break application data into packets that are easier to manage and send across a network. The packets are then numbered and sent in groups. The biggest advantages of TCP are stateful connections, guaranteed packet arrival, and built-in network congestion control.

Although TCP benefits from these advantages (and all the commonly-used internet protocols associated with it), there are some glaring disadvantages—especially with bulk data transfers over IP links where latency and/or packet loss are present.

## THE PITFALLS OF TCP WHEN TRANSFERRING LARGE DATA SETS

### LATENCY AND HOW IT AFFECTS TCP

To reliably transfer data across a network via TCP, the receiving party must send an acknowledgment (ACK) to the sending party confirming the packet was received. These ACKs must be sent in sequential order, and the sender cannot send another packet of data until it receives an acknowledgment that the previous packet was received. The time spent sending a packet and receiving the ACK is measured as Round-Trip Time (RTT). This is one of the reasons TCP can be slow: time is spent waiting for ACKs instead of transmitting data.

On local networks with computers sending and receiving data in close proximity to each other, ACKs spend less time in flight and do not slow down the data transmission. However, as the geographic distance increases, so does the RTT. The slower ACK reception causes exponential throughput degradation for bulk data transfers.

TCP responds to this by adjusting the acceptable amount of unacknowledged data allowed on the link. If the acceptable amount is surpassed, the transfer will stop and wait for an ACK. The optimal amount of unacknowledged data en route should equal

the end-to-end bandwidth, multiplied by the RTT. This sum is known as the bandwidth-delay product.

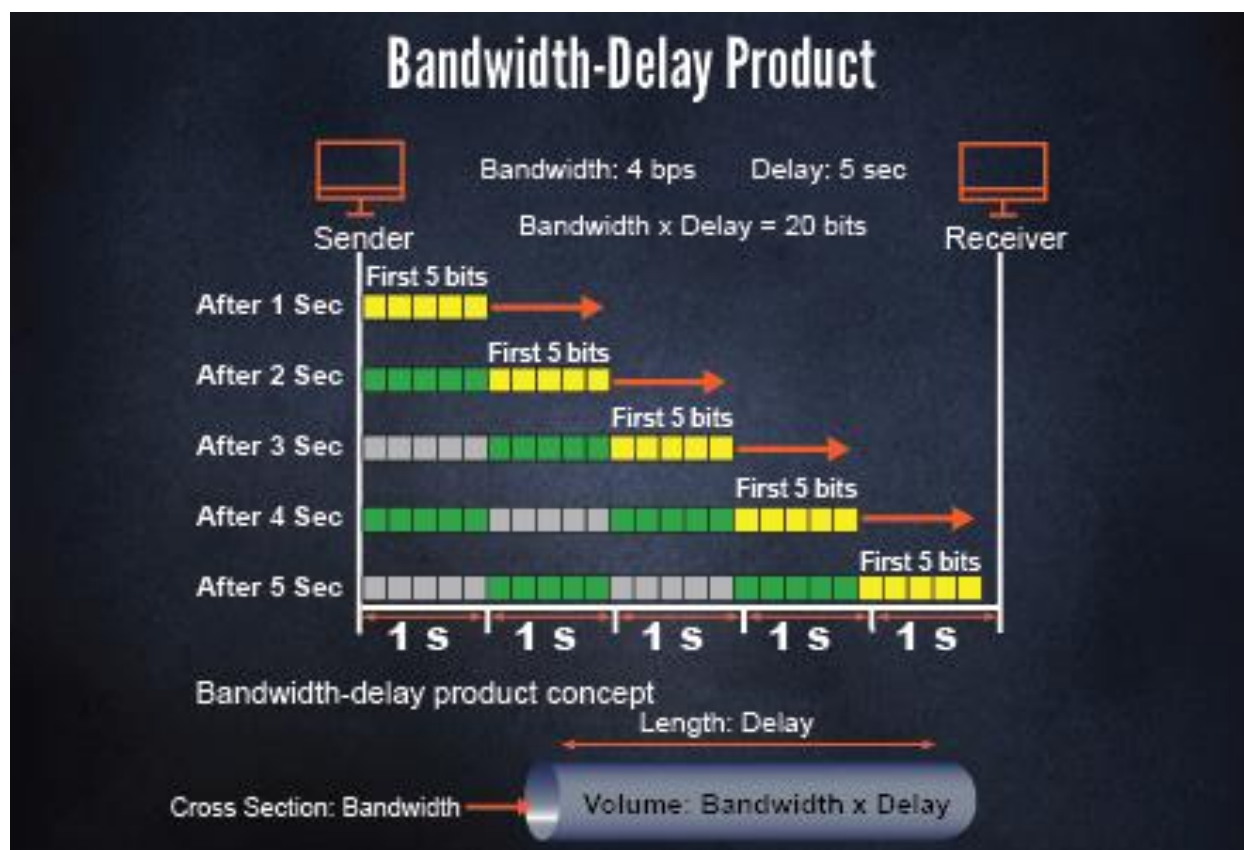


Image 1: Bandwidth Delay Product

TCP perpetually estimates this value and sets a "TCP window." When the bandwidth-delay product exceeds the TCP window, the result is "dead air," which creates even more wait time. Some satellite connections must deal with hundreds, or even thousands, of milliseconds of RTT<sup>1</sup>.

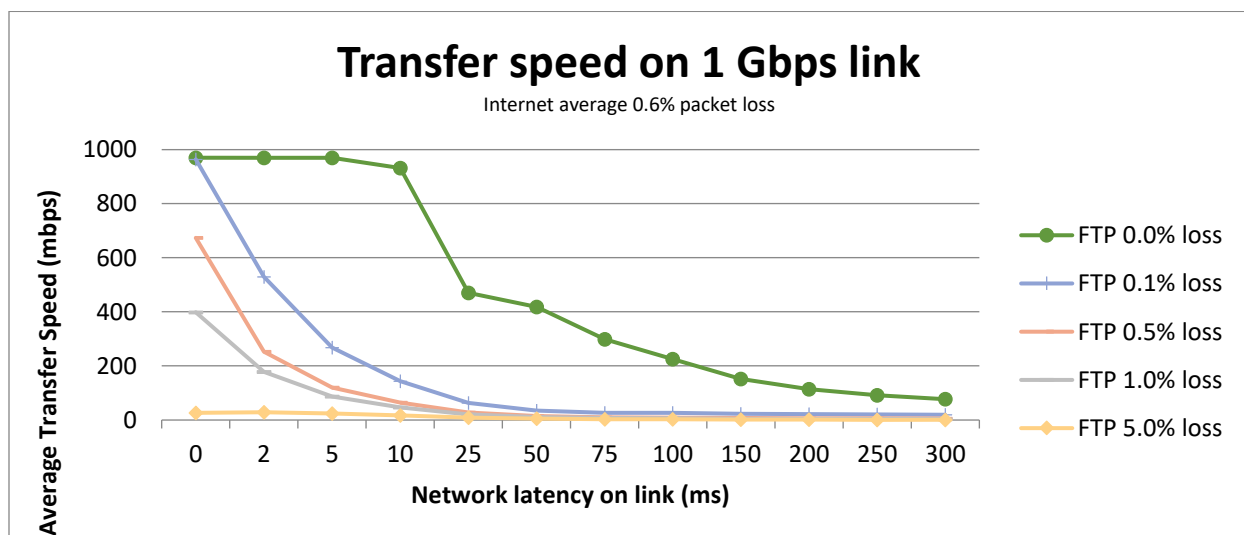


Image 2: The Effects of Latency on TCP

<sup>1</sup> V Jacobson, TCP Extensions for High Performance (May 1992)

## PACKET LOSS AND ITS EFFECT ON TCP

Network congestion typically causes buffer overflow on routers that are unable to handle a large amount of congestion placed on them. A router experiences buffer overflow when it does not have the capacity to accept all the incoming packets. This causes packet loss.

TCP cannot distinguish between packet loss caused by network congestion, and congestion caused by interference in wireless or satellite networks. Physical structures in the “route” of a wireless or satellite connection cause interference, and ultimately packet loss. TCP will cut the TCP window in half when packet loss is detected, which is too aggressive when inherent interference is present. The ideal solution should be able to react to congestion in a less aggressive manner.

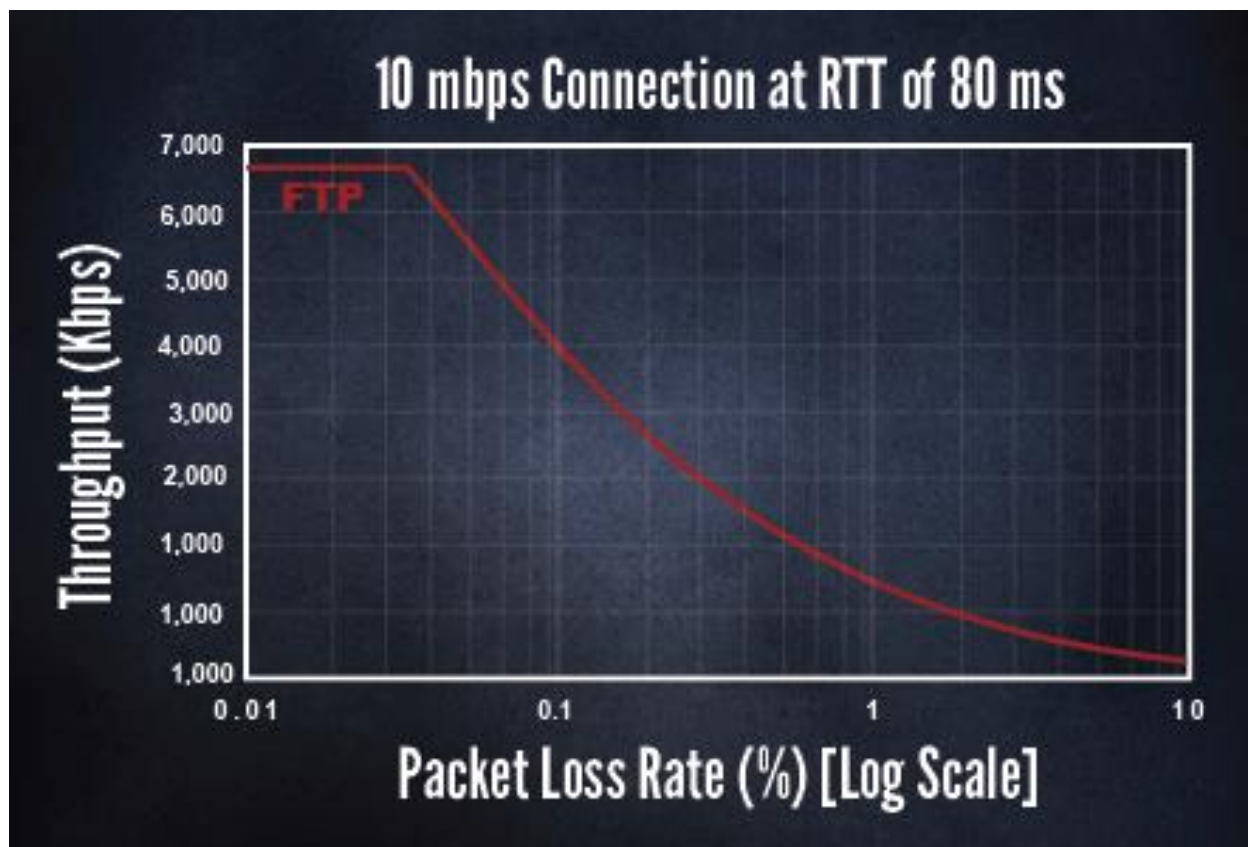


Image 3: The Effects of Packet Loss on TCP

## THE PITFALLS OF COMMON TCP-BASED PROTOCOLS

Speed constraints are not the only issues associated with TCP-based protocols. Some of the limitations and disadvantages include:

- **HTTP:** HTTP transfers have a size limit of approximately 2 GB. The file is placed on the computer's memory during a transfer. The larger the file, the more resource intensive the file transfer becomes.
- **FTP:** FTP does not use encryption by default when transferring files. Alternatively, FTP can be secured by using the SSL (Secure Sockets Layer) or SFTP protocols. Both SSL and SFTP are inherently secure.
- **Bandwidth Prioritization:** FTP (or any other TCP-based file transfer protocol) does not give users the ability to adjust bandwidth in order to speed up or slow down occurring file transfers.
- **Integrity Checking:** Many TCP-based protocols do not check the integrity of a file after it is transferred.
- **SMTP:** Size limits are commonly placed on SMTP transfers. This is not practical for sharing large files. If you are using your own individual mail server, however, the limits can be adjusted.
- **Blind Resuming:** When a file is paused and resumes, most TCP-based protocols will blindly append the file with no checks, sometimes resulting in a corrupt file.

## OPTIMIZING TCP

TCP makes use of two buffers referred to as "windows" to perform transfers: the congestion window on the sender machine and the receive window on the receiver machine. The congestion window is scaled up and down by the sender in reaction to packet loss. On clean links with little to no loss, the window can quickly scale to its maximum value. However, on lossy links, the window will quickly lower itself to reduce the re-transmission of redundant data. This is referred to as "congestion control." The size of the receive window determines how much data the receiving machine can accept at one time before sending an acknowledging receipt back to the sender. When a TCP connection is established, the window sizes are negotiated based on the settings on each machine. The lower value between the two machines will determine the size of the congestion window.

To optimize TCP performance, you must increase the value of the TCP windows<sup>1</sup>. The congestion window must be configured on the sender side, and the receive window must be configured on the receiver side. The congestion window should be tuned to maximize the in-flight data and reduce the "dead air" on your link. The amount of in-flight data needed to maximize the link is called the bandwidth-delay product (see *Image 1*). The receive window should be increased to match the size of the congestion



window on the sending machine.

Although tuning TCP can yield increased transfer rates, it can also create problems on networks containing packet loss. A single dropped packet will invalidate the TCP windows. When this occurs, the entire block is re-transmitted, substantially lowering the throughput of TCP transfers with large window sizes. Another possible issue is that tuning TCP for high speed transfers may reduce speeds for every day usage such as email and web browsing.

Changing the window sizes can be a complicated process. Increasing the TCP window size involves changing configurations on both the receiving and sending parties of the transfer. This is less than ideal for environments containing multiple endpoints. On Linux-based systems, administrators must manually edit the system config files. On Windows-based systems, the registry settings must be updated. In both cases, administrative privileges will be required. Once the window sizes are increased, the optimization may be marginal. At high speeds (greater than 1 Gbps) several concurrent flows will typically be required to achieve full link capacity. With high RTT (> 50ms), 10 or more streams may be required to achieve 1 Gbps. Using this method to reach 10 Gbps or higher may require 100 or more streams, creating a strain on CPU resources.

## AVAILABLE SOLUTIONS

Since TCP is the backbone of many existing transfer processes, most file transfer solutions are based on TCP. These options are suited for transfer scenarios that occur spontaneously and infrequently. Some TCP-based solutions include:

### FTP/SFTP/FTPS SERVER

Some organizations host their own FTP servers to provide their own file transfer service. This can be beneficial, but there several considerations when deploying your own server.

When sending files across the public internet, an organization's FTP server should have special security measures in place, including SSL protection for FTPS.

Even after the FTP server is deployed and configured (which takes skilled and experienced IT staff), the transfer is still performed via TCP. All the bottlenecks inherent to TCP still apply to the FTP server.

The FTP server may be able to perform file transfers, but it may lack many features included with other commercial MFT (Managed File Transfer) solutions. The drawbacks of an in-house FTP server include:

- Lack of tracking and reporting
- No email notifications for completed file transfers
- No MD5 checksum
- Unreliable resume and restart features
- No file delta capabilities
- Inability to transfer directly from the web browser

## EMAIL

In home and enterprise scenarios, email attachments can serve as an easy way to deliver files. Even though email is a common method, there are some limitations to consider, especially for enterprise applications:

- **File Size Limitations:** Email servers are configured to handle attachments of a certain size. Files that exceed the maximum size limit will be “bounced.”
- **Archival Storage Costs:** Emails sent within an organization are archived on a mail server. This means that every sent attachment is archived, thus taking up space on the server. These archived attachments may greatly increase IT storage costs.
- **Poor Network Utilization:** Emails rely on the Simple Mail Transfer Protocol (SMTP). This protocol is built on TCP, so all the inherent issues of TCP are present.

## CLOUD SERVICES/DIGITAL DELIVERY

Cloud services such as Dropbox, Hightail and others have recently become popular methods for transmitting large amounts of data. These services are relatively easy to use and scalable, but they also have drawbacks. With cloud-based transfers, the sender must upload the file(s) to the cloud, and then the receiver must download the file(s) from the cloud as a separate step after the initial upload is complete. Since cloud-based solutions use the internet, they also suffer from the limitations inherent to TCP.

Sending structured folders containing a large number of files can be difficult via cloud services because cloud services use basic HTTP upload tools to move files. It is difficult to move complex directory structures without first zipping the payload into a single archive. Most web-based HTTP upload tools used by cloud services have a limit of 2-5 GB per file.

Even with all the limitations of digital delivery methods, a digital delivery solution is still the ideal solution when file sizes are under 2 GB. However, for a cloud solution to be truly complete, it should overcome the issues of speed caused by packet loss and latency as well as work with files of any size.

The ideal solution should accelerate file transfers and maximize the already existing infrastructure. It should also be easy to implement and—most importantly—easy to use.

## SHIPPING PHYSICAL MEDIA

Rather than deploying a transfer acceleration technology, some companies use physical storage as a means of delivering data. Physical mediums include tape storage, hard drives, flash drives and DVD/Blu Ray disks.

Amazon Web Services (AWS) offers a service that physically ships storage called a “Snowball.” A large hard drive is shipped to the user’s location where they copy their data to the drive. Amazon then picks up the Snowball and ships it to the user’s desired location.

Shipping and archiving large data sets via physical storage can work on an occasional basis, but this is not an easily scalable method. The time it takes to copy and ship the data, along with the expense of shipping, can generate high costs and inefficiency.

## COMMON DRAWBACKS OF PHYSICAL MEDIA:

- **Preparation Time:** Storing and shipping data requires physical human interaction; from copying the data to printing out mailing labels and shipping. Where there is human interaction, there is also the potential for human error.
- **Cost:** Depending on the frequency and urgency, costs for using a courier service can easily add up.
- **Delivery Time:** Shipping physical media may take up to 5 business days or longer, depending on the destination.

## FILECATALYST ACCELERATION FEATURES

The User Datagram Protocol (UDP) can draw more performance from an IP network than TCP by omitting some of the features included with TCP. UDP is a “connectionless” protocol, meaning it does not depend on sequenced acknowledgments. Without acknowledgements, data transfers are faster than TCP on high latency networks, but have the potential to become unreliable when any form of packet loss is present on the link.

FileCatalyst’s core transport technology is based on the UDP protocol. However, FileCatalyst’s unique patented data transfer algorithm<sup>2</sup> (4) gains the performance benefits of UDP, while adding reliability exceeding that of TCP. FileCatalyst is not only

---

<sup>2</sup> Christopher Bailey, *Patent US9043486 - Data Transfer Method, System and Protocol*

more reliable than TCP, but also offers rate control features originally missing in traditional TCP based protocols. Furthermore, the FileCatalyst protocol can transmit data at precise rates, which is an important feature in many real-time applications.

Like TCP, FileCatalyst also breaks data into blocks. The major difference between FileCatalyst and TCP is that, with FileCatalyst, there is no delay while waiting to receive a block of data before commencing subsequent blocks of data. Transmission is initiated immediately, even if previous blocks have not yet been acknowledged. Regardless of network latency, data transmission remains constant with FileCatalyst, enabling transfers to occur at full line speed.

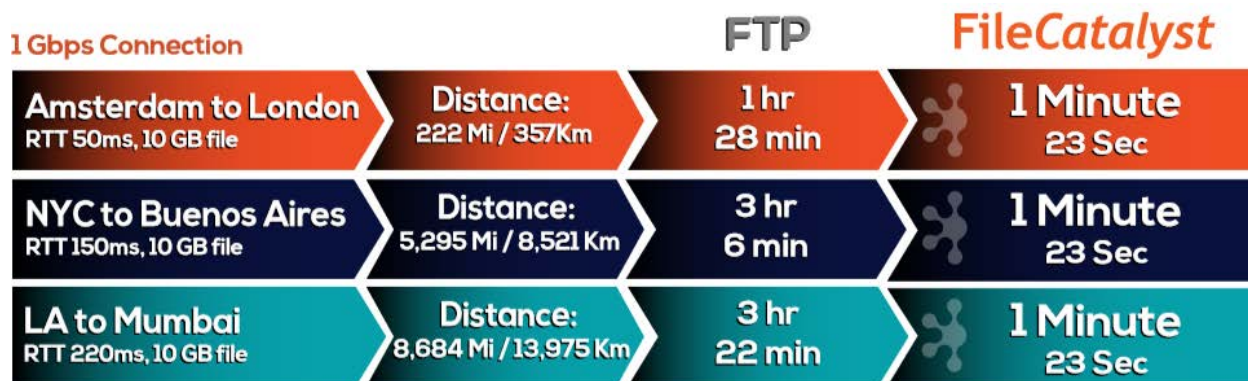


Image 4: The Benefits of Transferring a 10 GB File on a 1 Gbps Connection Using FileCatalyst

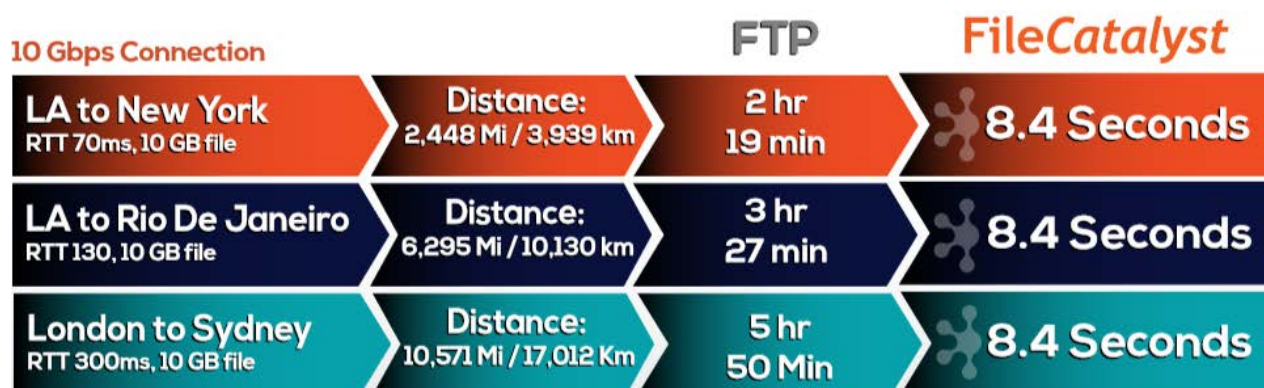


Image 5: The Benefits of Transferring a 10 GB File on a 10 Gbps Connection Using FileCatalyst

## CONGESTION CONTROL

Congestion control allows FileCatalyst to adapt to changing network conditions, ensuring that the transmission remains optimal and avoids congestion collapse. FileCatalyst provides three modes of congestion control: no congestion control, RTT-based, and loss-based, which is the default setting.

## NO CONGESTION CONTROL

This setting allows users to send data as fast as possible with only minimal background traffic checks. This option is excellent for dedicated links or links configured via



Software-Defined Networking (SDN) which is specific for file delivery tasks.

## RTT-BASED

“RTT-based” congestion control establishes a baseline average RTT before the data starts to flow. Once the transmission begins, RTT is continuously monitored. While the RTT stays within a certain range of the baseline RTT, the speed of the transfer will increase. Once the RTT begins to go above a certain threshold, the speed is decreased. How much the RTT can spike above the baseline average is controlled by the congestion control aggression setting provided by FileCatalyst.

This type of congestion control is suited for wireless or satellite links where there is packet loss from sources other than congestion. TCP will slow down, for example, when a packet is lost due to interference. When FileCatalyst is using the RTT-based congestion control, it ignores individual packet losses and focuses only on RTT.

## LOSS-BASED

There are circumstances in which RTT-based congestion control may not work properly. For example, when a router's queue is very small the RTT may never spike when congestion is present. If the RTT remains low, FileCatalyst will continue to increase the transmission rate, even when congestion is present. For these scenarios, the only way to detect congestion is by monitoring packet loss. As outlined previously, packet loss may come from sources other than congestion, so this is best used on land-based networks where packet loss is due to real congestion.

Loss-based congestion control reacts to packet loss by slowing down, similar to TCP, but far less aggressively. TCP can be quite aggressive in its congestion avoidance, which may under-utilize your link; FileCatalyst's loss-based congestion control algorithm was designed to maximize link utilization while still avoiding congestion. Like RTT-based congestion control, the loss-based congestion control mode aggression can be tuned.

## MULTI-CLIENT FILE TRANSFERS

The FileCatalyst protocol can transfer multiple files from a single data source concurrently, letting an organization better optimize their computing and network resources during a file transfer. FileCatalyst's Multi-Client feature can transfer multiple growing files at once, as well as autoarchive smaller files into a single file, thereby greatly increasing throughput on data sets containing many small files.

## COMPRESSION

Data compression is a general term referring to a technology that can encode large files in order to reduce their size. Not all data can be compressed, but compression can significantly shrink the size of data that can be compressed—resulting in less information to transfer. This method naturally results in a faster transfer.

Data compression applies an algorithm to the data which stores repetitive bits of information as a “shorthand.” Once the shorthand is sent, the receiving end uses a decoder tool that restores these pieces back to their original state. After the decoding process is complete, an identical copy of the original file is formed on the receiving end.

## INCREMENTAL TRANSFERS

In some scenarios, similar files may already exist on both sides of the transfer, but changes may have been made to the file at the source location. When a difference is detected, an algorithm calculates the differences between the source and destination and stores them as discrete files. These small delta files containing the changes are then transferred to the destination.

Once the delta files are sent to the destination, the changes are applied to the destination file as a “patch.” The patch updates the destination file, resulting in an identical copy of the revised file. The benefit is quite clear: sending a 4 MB delta instead of a 2 TB creates an incredible difference in transfer time.

## PROGRESSIVE TRANSFERS

FileCatalyst can send files as they are being written on the disk. This is very beneficial when the process of creating the final file takes significant time.

Progressive transfers allow FileCatalyst to transfer a file as it is being built by another application. This is especially beneficial for live video production workflows, when the encoding process for a large video file may take several hours to complete. Without FileCatalyst, the transfer could not start until the encoding process is complete.

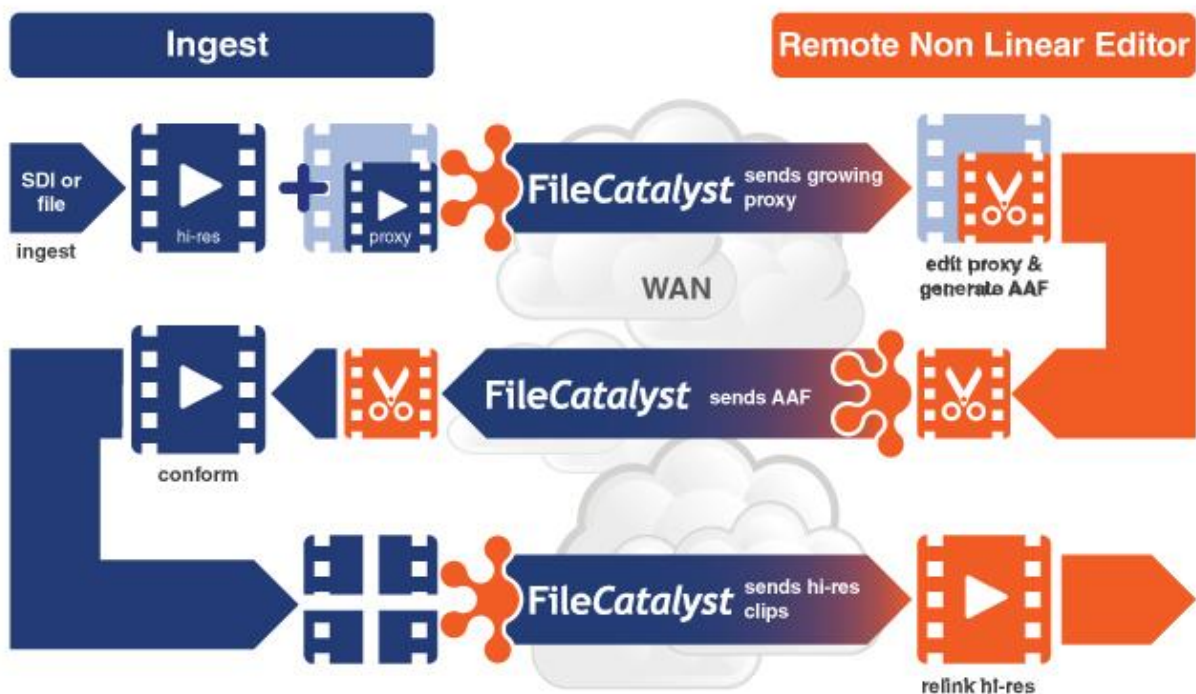


Image 6: A Live Video Production Workflow Employing FileCatalyst's Progressive File Transfer Features.

The progressive transfer feature, combined with concurrent and/or multi-client transfers, allows FileCatalyst to transfer several growing files at once. This feature also allows for auto-discovery of new growing files in a predefined directory.

# FILECATALYST AND THE COMPETITION

FileCatalyst uses a patented proprietary User Datagram Protocol (UDP) file transfer protocol which includes congestion control, bandwidth throttling, incremental transfers, security features, and real-time management. FileCatalyst is well known in the media and entertainment space as an alternative to Aspera, Signiant, and a few open source projects.

FileCatalyst has been used for some of the largest sporting events in the world to support television broadcast workflows. FileCatalyst has been awarded two technology Emmy awards (2015 and 2016) for their work in the industry. Under many circumstances, FileCatalyst's underlying technology is faster, more flexible, and more reliable than competing solutions.

Independent research from Anhalt University in Germany shows that FileCatalyst is the fastest and most reliable protocol for networks containing latency and packet loss<sup>3</sup>. The study tested multiple commercial vendors including ExpeDat (Data Expedition), Tixel (TIXStream), Catapult, Velocity, and FileCatalyst. During the tests, FileCatalyst Direct was the only solution able to deliver full line speed.

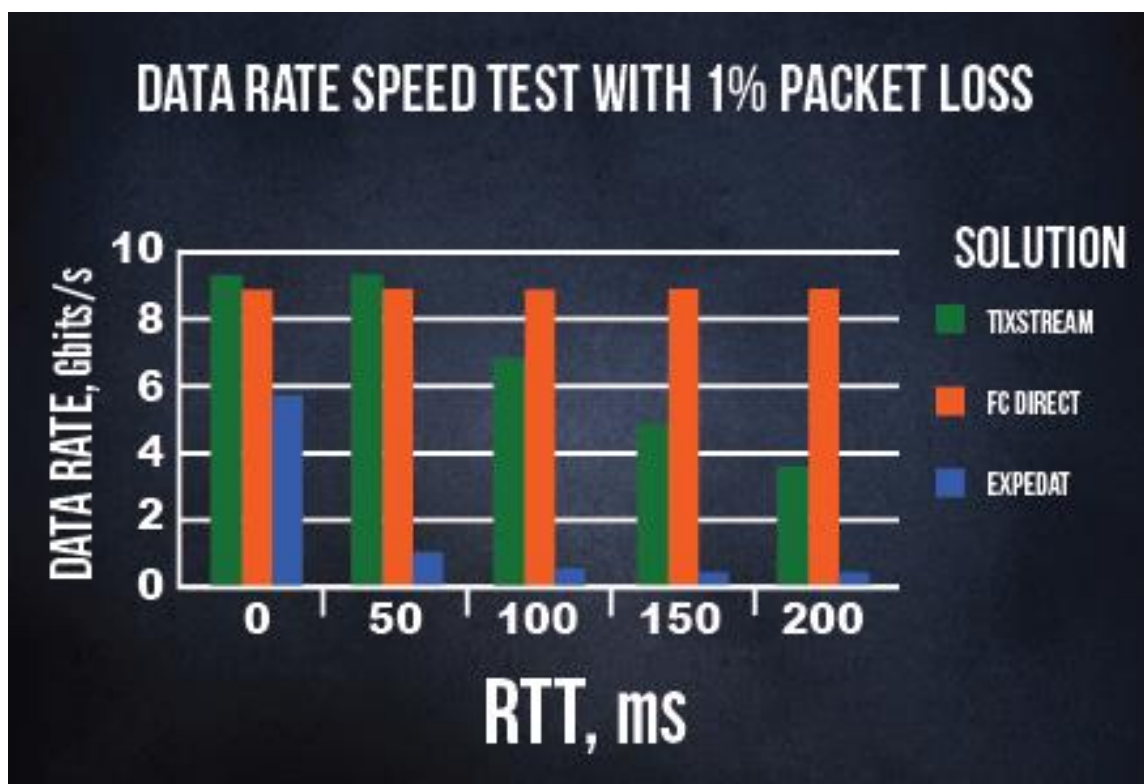


Image 7: Data Rate Speed Test Between TIXStream, FileCatalyst Direct, and ExpeDat

<sup>3</sup> Dmitry Kachan, *Comparison of Contemporary Solutions for High Speed Data Transport on WAN 10 Gbit/s Connections* (2013)



## ASPERA

Aspera chose not to participate in the independent study from Anhalt University in Germany mentioned above. However, the company's documentation indicates that the maximum achievable baseline speed is 1 Gbps<sup>4</sup>.

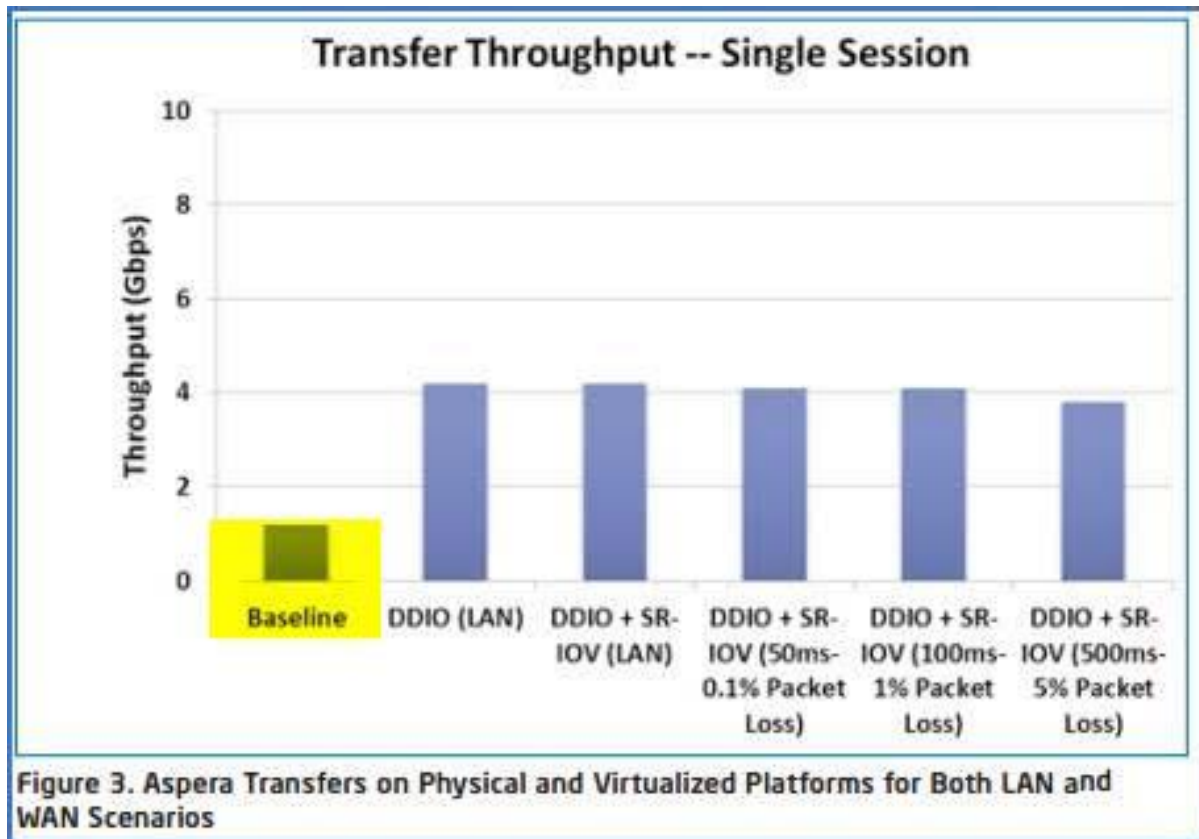


Image 8: Aspera's Published Maximum Throughput Speed

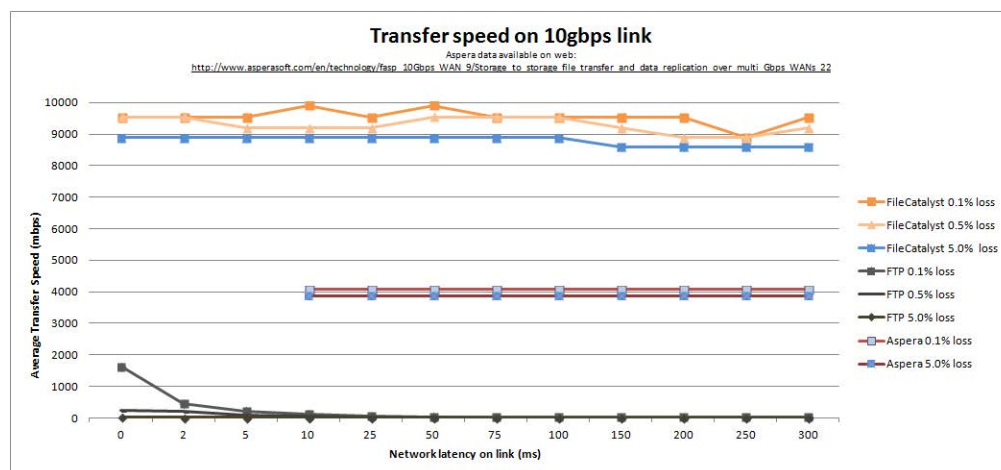


Image 9: FileCatalyst Transfer Speeds for a Single File Transfer

<sup>4</sup> IBM Corporation, *IBM Technologies for Ultra-High-Speed Data Transfer in Life Sciences*

For Aspera to achieve 4 Gbps, the Intel Distributed Discrete Input/Output (DDIO) instruction set had to be enabled, and encryption had to be disabled. To achieve 10 Gbps, Aspera required a minimum of three concurrent file transfers, as well as disabled encryption. FileCatalyst's 10 Gbps implementation does not require concurrent transfers, specific hardware or instruction sets to achieve multi-gigabit speeds.

Aspera's deployment model is client/server-based, and requires a license on the server and at each location with client software installed. Conversely, FileCatalyst's license is based on the server and the maximum number of allowable concurrent connections. For example, if the FileCatalyst Server is licensed for 10 concurrent connections, an enterprise may deploy the client software to as many locations as required, but only 10 instances can run at a given time. New locations are added at no additional cost for the customer.

## SIGNIANT

Signiant is another file transfer acceleration software solution. Signiant may not be able to reach speeds over 2 Gbps, as no documentation, demo or content referencing higher speeds could be found at the time of this publication. Furthermore, Signiant's technology doesn't efficiently handle large and complex directory structures (2-5 million files). Signiant is also unable to maintain file attributes such as file owner, group, permissions, creation dates, or last modified date during transfers.

Signiant's Media Shuttle does not offer the ability to automate and schedule file transfers, or to synchronize folders. This functionality missing from Media Shuttle is available via FileCatalyst HotFolder. Signiant's primary focus is on Software as a Service (SaaS) based offerings, which does not allow for perpetual, un-metered, on-premise installations. SaaS-based offerings also require a continual connection to a central server on the public internet. This deployment model is incompatible with private closed networks.

Signiant's billing is based only on a per-usage model plus a monthly base price. This is a good model for enterprises that require file transfer acceleration sporadically and infrequently. Per-usage models also have their downsides, as prices can increase as the amount of transferred data increases. This is not an ideal solution for enterprises that rely on file transfer acceleration to routinely send large amounts of data, as it can become costly.

FileCatalyst not only includes a per-usage model for scenarios which require temporary acceleration for short term projects, but also includes a perpetual on-premise license – which is not offered by Signiant. A perpetual on-premise license provides enterprises with the predictability and scalability that per-usage models lack.

## OPEN SOURCE

There are several open source projects that provide an accelerated file transfer solution via UDP. Some solutions are more mature than others, and they all use different technologies to solve the same problem.

Some commercial solutions that claim to use UDP acceleration have simply integrated an open source project into their core file transfer technology. These solutions inherit the strengths, but also the weaknesses, of the open source project they leverage. FileCatalyst has developed and patented a UDP-based protocol that does not use any code from any open source UDP technology.

One common problem with these open source solutions is their lack of a Graphical User Interface (GUI). Some provide only a bare-bones sender/receiver Application Programming Interface (API), meaning that the end user must compile from the source. Other solutions may only come with a Command Line Interface (CLI).

Another common problem with open source solutions is the lack of checkpoint restarting and automatic MD5 checksum verification. This doesn't allow file transfers to resume automatically if the link temporarily fails, which can potentially lead to corrupt file transfers. A lack of firewall traversal support through a reverse/forward proxy is also common with open source solutions. While this is not an issue for internal transfers, most organizations send files over the WAN, which will almost certainly have at least one firewall somewhere on the route.

Most open source solutions do not fare well in network conditions where packet loss or high latency is present. Finally, the congestion control options included in open source UDP-based solutions do not adapt to ever-changing network conditions.

# FILECATALYST SECURITY

## REQUIRED PORTS – WITHOUT REVERSE PROXY

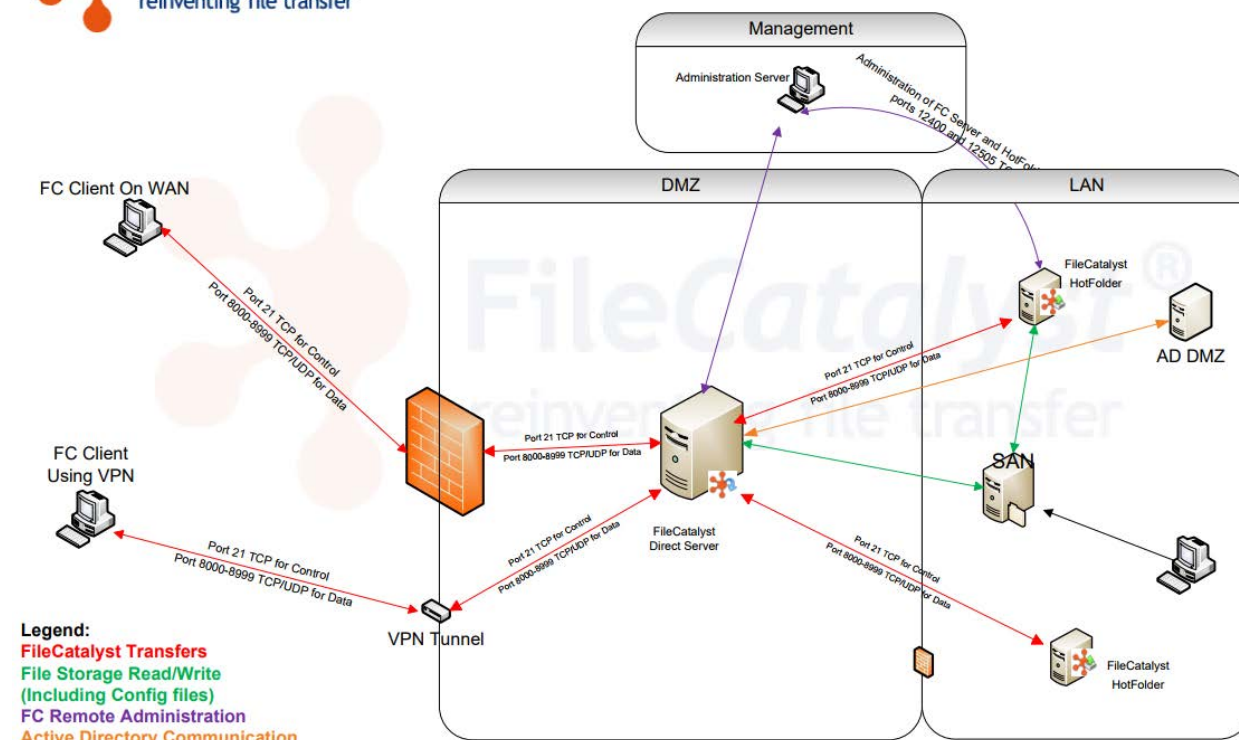


Image 10: FileCatalyst DMZ Deployment without Reverse Proxy

FileCatalyst Server requires one TCP port (default 21) to be open to inbound traffic in order to establish a control connection. The client, or connecting system, uses a short-lived port that's allocated automatically from a range predefined by the IP stack software. This is usually called an "ephemeral port," and is used as the source port to establish the control connection. Ephemeral ports typically range from port 1024 to 4999. The control connection is used to perform authentications via a username and password, share information regarding available files/directories on the server, and to negotiate protocols/ports for data transfers.

Establishing data connections, depending on your settings, requires specific TCP and/or UDP ports to be open for inbound/outbound traffic on the server side of the firewall, as well as inbound/outbound ports on the client side. FileCatalyst uses TCP for data connections when using "FTP mode" for transfers, or when the client requires a directory listing from the server. FileCatalyst uses UDP to transfer data when in "UDP mode". FileCatalyst Server defines the range of ports used for data transfers. The



default port range is 8000 to 8999. However, this range may be customized to an organization's needs.

When TCP connections are established, all connections are outbound from the FileCatalyst client to the FileCatalyst Server, regardless of the direction of the transfer. As with control connections, the connecting system uses an ephemeral port as its source port and will connect to a port in the range as defined by the FileCatalyst Server.

When UDP mode is used for uploads, the source and destination ports are in the same range (default 8000-8999). This means that the client side must allow outgoing UDP—and the server must allow incoming UDP—on all ports in the defined range.

When UDP is used to download, the source port range is defined by the server and the destination port range is defined by the client. The default client-side port is 0, meaning that the client will pick any open port. FileCatalyst uses a firewall hole punching mechanism (similar to Skype) that performs a NAT traversal. If the default values fail, or if others are required, the value may be set to a specific incoming port. Firewalls and NAT devices may be set accordingly.

Note that even when these ports are open on the firewall, FileCatalyst does not listen for connections unless it has been pre-negotiated by an existing FileCatalyst client. This mitigates the risk of 3rd parties attempting to establish un-authorized communications with a FileCatalyst server or client.

## REQUIRED PORTS – WITH REVERSE PROXY



**Legend:**  
FileCatalyst Transfers  
File Storage Read/Write  
(Including Config files)  
FC Remote Administration  
Active Directory Communication

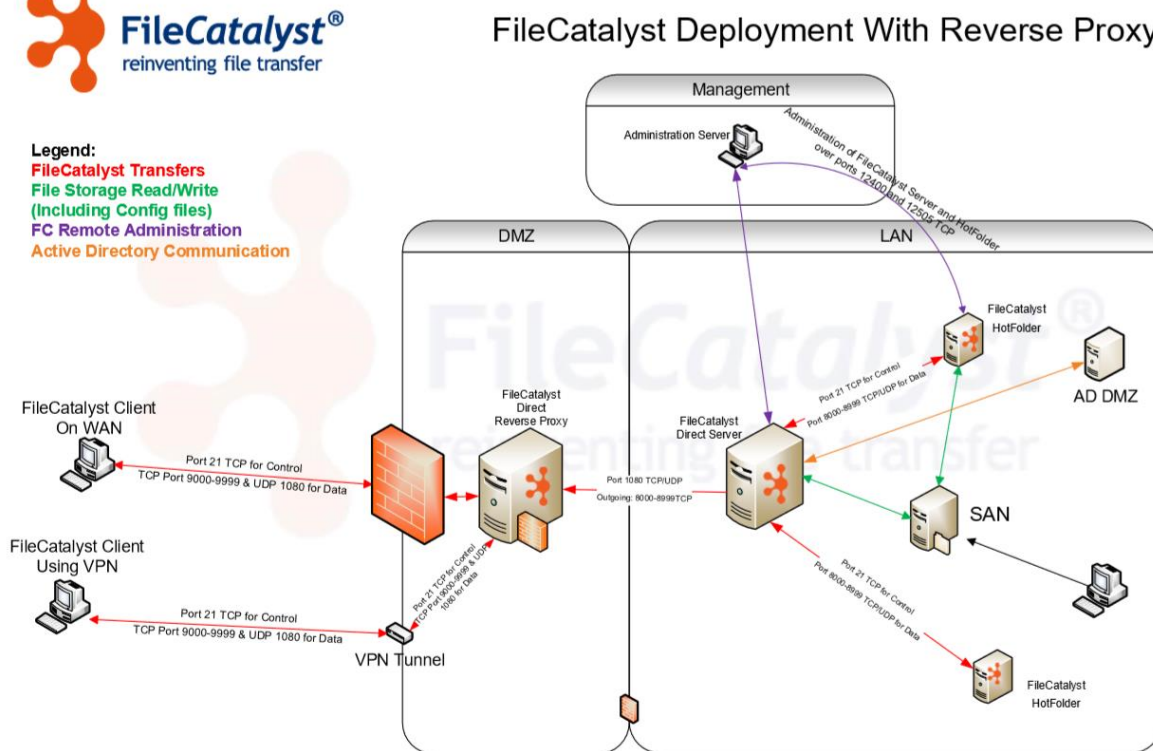


Image 11: FileCatalyst DMZ Deployment with Reverse Proxy

In the scenario above, the Reverse Proxy server is deployed in the Demilitarized Zone (DMZ) and facilitates the connectivity between the FileCatalyst Direct Server, located on the secure network, and the unsecured public internet. In this scenario, the port connectivity between the internet and the DMZ remains essentially the same as when not using a Reverse Proxy. However, the connectivity between the Direct Server on a secure network and the Reverse Proxy on the DMZ requires only a single configurable outgoing port from the secure network to the DMZ. The maximum achievable file transfer speed using a Reverse Proxy is 1 Gbps.

## MINIMIZING PORT USAGE

Depending on the volume of concurrent client connections, it may be desirable to fine tune the number of ports in the data port range. One is required for every concurrent transfer taking place on the server. Thus, a range of 8000-8999 could potentially support up to 1000 concurrent transfers. If the anticipated volume of concurrent transfers is known, the port range can be adjusted.

FTP transfers on the Windows Operating System (OS) may require additional ports to support higher transfer volumes. The Windows OS does not release a closed TCP socket for up to 3 minutes but rather places it in a CLOSE\_WAIT state. The

consequence is that FTP transfers made up of several smaller files will quickly exhaust ports. When using FTP for transfers with a Windows-based FileCatalyst server, minimizing the data port range should be done with caution. Linux and other UNIX based operating systems do not suffer from this limitation.

## ENCRYPTION

By default, the FileCatalyst TCP control connection and TCP/UDP data connections do not use encryption. All control data is text-based and sent in clear text, and should often be secured. FileCatalyst Server provides mechanisms that secure both the control and data connections. The TCP control connection and TCP data connections (FTP mode) can be secured using SSL (Secure Sockets Layer). When SSL is enabled, you must set the connection mode to “FTPS/Implicit” in order to connect with a 3rd party FTP client.

By default, FileCatalyst generates self-signed certificates for SSL communications. Valid Certificate Authority (CA) certificates should be employed to prevent a MiM (Man in the Middle) attack. FileCatalyst’s client and administration software provide options for strict validation of the domain to which it is connecting. If enabled, FileCatalyst will not connect if there is a domain mismatch.

When transferring in UDP mode, you must enable the Advanced Encryption Standard (AES) option for the data connection to encrypt data. When enabled, FileCatalyst uses AES encryption to ensure that intercepted data is useless to everyone except the intended recipient. Since AES exchanges a shared encryption key, you must also enable SSL to ensure the encryption key is not intercepted. In addition, FileCatalyst rotates the AES key for every new file in transit as an added measure of security.

By default, FileCatalyst uses 128-bit AES. However, a stronger encryption may be enabled if your country permits it.

## SSL CIPHER RESTRICTIONS

FileCatalyst Server allows the selection of specific SSL ciphers, which are considered appropriate for encrypted communication. The entire Java SSL/TLS set is utilized by default. The ciphers used can be modified (i.e.: enforce a minimum 128-bit encryption cipher).

## IP FILTERS

FileCatalyst provides an IP filter feature that allows administrators to permit or deny specific IP addresses. Furthermore, administrators can permit or deny entire ranges of IP addresses. Connection attempts from IP addresses outside of the defined rules will

be dropped by FileCatalyst Server instantly.

## LOGIN SECURITY

FileCatalyst Server provides mechanisms that block brute force password attacks by automatically blocking the offending IP address and/or disabling the compromised user account. This feature may be enabled or disabled, and the administrator may set the number of failed attempts that trigger the block.

## HIPAA SECURITY COMPLIANCES

FileCatalyst can ensure HIPAA compliance from a technical standpoint by ensuring the following:

1. **Access Control** - The ability to access the system using user accounts. Restrictions based on IP and authentication against a directory can also help ensure access control.
2. **Audit Controls** - Every login is audited, and every event is logged in the system. Full reports on file transfers are available via FileCatalyst Central.
3. **Integrity** - FileCatalyst can ensure that data isn't compromised by performing MD5 Checksums on the files once complete. If the checksum does not match, the file is deleted on the destination side and re-sent.
4. **Authentication** - FileCatalyst can authenticate against a local database or a directory service. When authenticating against a directory service, no passwords are stored locally.
5. **Transmission Security** - FileCatalyst can use SSL and AES to ensure transmission security.
6. **Reverse Proxy** - The ability to configure the FileCatalyst Server on a secure corporate network using a Reverse Proxy, deployed on the unsecured DMZ network, is a common deployment strategy for security-conscious organizations.

## PENETRATION TESTING

In general, the use of Java Virtual Machine (JVM) eliminates several security problems common with Native C Applications such as Memory Corruption, Privilege Escalation, and Injection<sup>5</sup>. External parties have conducted penetration testing on FileCatalyst products, of which the results are proprietary and cannot be shared. However, all identified issues have been addressed. Some of the public domain issues that have been identified and resolved include:

---

<sup>5</sup> David Svoboda, Is Java More Secure than C? (2015)



On the core transport:

1. POODLE
2. HeartBleed

On the web based security:

3. Injection flaws
4. Cross-Site Scripting (XSS)
5. Cross-Site Request Forgery (CSRF)

For the core transport, FileCatalyst uses patched SSL libraries to eliminate security issues. FileCatalyst recommends always maintaining the latest version of Java to ensure that newly discovered vulnerabilities are patched.

Concerning web security, the FileCatalyst web development team regularly tests FileCatalyst's web based products against The Open Web Application Security Project (OWASP) Top 10 requirements.

# FILE TRANSFER ACCELERATION SCENARIOS

## SCENARIO 1 – SMALL TO MEDIUM ENTERPRISE (SME)

### PROFILE – PERFORMING LARGE FILE TRANSFERS

A company headquartered on the Eastern Coast of the USA regularly sends and receives large files to and from India and Australia. They currently rely on FTP as their main transfer method, but they also ship physical media. They have recently noticed that they aren't able to take full advantage of their connection. They currently use the following connection:

- 150 Mbps Bandwidth
  - Cost: \$800/month
  - Annual: \$9,600/year

### CHALLENGES – BOTTLENECKS CREATED BY FTP

Since all their current file transfers utilize FTP, they can only realize about 25% of their 150 Mbps connection. Not only is their connection underutilized, but they also must deal with the inherent issues of FTP including “dead air” and packet loss. They even lost a contract due to a failed transfer.

Since they can only leverage 25% of their potential bandwidth, they are wasting 75% of their \$9,600 annual costs. This waste equals \$7,680 annually.

### SOLUTION – FILECATALYST DIRECT

The company sought out FileCatalyst, and after the team discussed their unique workflows and needs, FileCatalyst was able to provide an accelerated file transfer solution software at a one-time price of \$9,000.

After deploying FileCatalyst, the company accelerated their file transfers by 80%. Not only did they accelerate their connection speed and maximize their productivity, but they were also able to realize a return on their investment in less than a year and a half.

## SCENARIO 2 – MULTI-NATIONAL ENTERPRISE

### PROFILE – GLOBALLY DELIVERING CONTENT

A large enterprise organization distributes and publishes news content. They are headquartered in New York, with 4 branches: Los Angeles, Berlin, Moscow, and London. Their editors are in Moscow, and most of the production and content creation happens between Los Angeles and New York.

### CHALLENGES – UNDERUTILIZATION OF THE CONNECTION

Their 200 Mbps link is quite fast theoretically. The average latency of pan-Pacific transfers is 200ms, with an average packet loss rate of 1%. Under these conditions, FTP will only transfer at a top speed of 490Kbps. If they can only get speeds of 490Kbps, they are only using a mere 1% of the connection they are paying for.

### SOLUTION – FILECATALYST SERVER AND HOTFOLDER

The company sought out various solution providers and decided to evaluate FileCatalyst. They installed FileCatalyst Server across all 4 of their locations, as well as FileCatalyst HotFolder at every location. Between every location and team project, a total of 42 nodes were set up to send and receive files.

Once the FileCatalyst Server and FileCatalyst HotFolder were installed and configured, they noticed that they could leverage FileCatalyst Central to manage and monitor every node from a web browser. Administrators can monitor and manage all 42 nodes, monitor transfers and manage alerts, regardless of which office they are located in.

They also realized that they could leverage FileCatalyst's ability to send large files through download links via email, allowing them to easily distribute assets to their coworkers between offices.

With FileCatalyst's UDP-based acceleration solution, the transfer speeds now reach near link speed. Production transfers that formerly took 100 minutes to deliver now take one minute.

## SCENARIO 3 – CLOUD

### PROFILE – MIGRATING DATA TO THE CLOUD

A research and development firm headquartered in San Diego assessed options for migrating their large archive into the cloud for easier remote collaboration.

They have Amazon Web Services (AWS) storage in both Europe and Asia, and want to migrate their data sets to these locations. Not only do they want to migrate their archived backups to the cloud, but they also want to perform backups on a weekly basis.

### CHALLENGES – SLOW UPLOAD SPEEDS

They initially began the cloud migration without acceleration, relying on the basic TCP/FTP method to transfer their archive. The initial command line tool provided by AWS had very slow throughput when transferring from San Diego to Europe and Asia. The process was deemed an unacceptably slow and ineffective process.

They considered using Amazon Snowball, by which a large drive is shipped to the HQ to have the data copied onto a physical drive and shipped by Amazon to their offices and uploaded to the cloud.

### SOLUTION – FILECATALYST DIRECT AND HOTFOLDER

The company chose to evaluate FileCatalyst. The installation process included an instance of FileCatalyst Server in Amazon's Elastic Compute Cloud (EC2) hosting in EC2 on a Virtual Machine (VM) in Europe and Asia. The servers were then connected to their S3 storage in the respective regions. This immediately yielded a speed increase of five times.

Not only was the speed increased, but they were also able to leverage FileCatalyst HotFolder to sync files to the server automatically. Now all the employees with HotFolder installed always have access to the latest data in their cloud storage.

# REFERENCES

1. Jacobson, V, et al. "TCP Extensions for High Performance." *Emerging Issues Task Force*, Emerging Issues Task Force, May 1992, [www.ietf.org/rfc/rfc1323.txt](http://www.ietf.org/rfc/rfc1323.txt).
2. Bailey, Christopher P. "Patent US9043486 - Data Transfer Method, System and Protocol." *Google Patents*, Google, 26 May 2015, [www.google.com/patents/US9043486](http://www.google.com/patents/US9043486).
3. Kachan, Dmitry, et al. "Comparison of Contemporary Solutions for High Speed Data Transport on WAN 10 Gbit/s Connections." *ThinkMind(TM) Digital Library*, 24 Mar. 2013, [www.thinkmind.org/index.php?view=article&articleid=icns\\_2013\\_2\\_40\\_10167](http://www.thinkmind.org/index.php?view=article&articleid=icns_2013_2_40_10167).
4. IBM Corporation, IBM. "IBM Technologies for Ultra-High-Speed Data Transfer in Life Sciences." *Big Data Technologies for Ultra-High-Speed Data Transfer in Life Sciences*, IBM, 3 Jan. 2018, [www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZZW03369USEN](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZZW03369USEN).
5. Svoboda, David. "Is Java More Secure than C?" *SEI Insights*, SEI Blog, 5 Oct. 2015, [insights.sei.cmu.edu/sei\\_blog/2015/10/is-java-more-secure-than-c.html](http://insights.sei.cmu.edu/sei_blog/2015/10/is-java-more-secure-than-c.html).