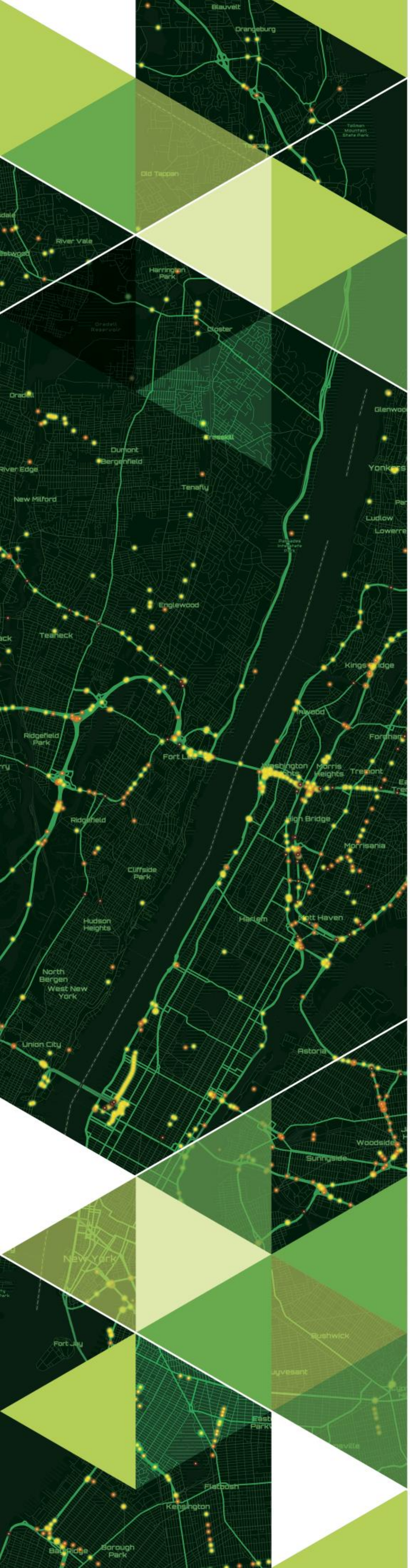


AN ESRI
TECHNICAL PAPER

June 2022

ArcGIS Enterprise Systems and Network Transport Time

380 New York Street
Redlands, California 92373-8100 USA
909 793 2853
info@esri.com
esri.com

esri THE SCIENCE OF WHERE®

Table of Contents

ArcGIS Enterprise Systems and Network Transport Time

Executive Summary

ArcGIS Enterprise technology is designed to operate in a variety of network conditions. Deployed systems work best when connections to databases and file shares, and between federated components, are on a local area network (LAN). This is because these connections are sensitive to per-bit network transport time (more so than bandwidth). Other types of connections, such as those over the HTTPS protocol from client applications, are more tolerant of longer per-bit transport times, such as those on wide area networks (WAN) and the internet.

Introduction

This paper discusses the performance and reliability considerations related to network transport time in ArcGIS Enterprise systems. For many years, all components were commonly placed in the same physical data center, with and low latency and high bandwidth connections. But, with the proliferation of cloud environments and hybrid cloud and on-premises environments, and the increasing use of public-facing systems and internet-based applications and devices, additional considerations come into play. Many organizations are pursuing more complex or creative networking configurations.

The goal of this paper is to assist with decision-making about where to place various components—such as databases, file shares, federated ArcGIS Server sites, Portal for ArcGIS, and client applications—that interact in an ArcGIS Enterprise system with respect to overall network architecture and different network pathways. This paper should help you understand network connectivity concepts and performance using standard approaches and consider the consequences of nontraditional architectural decisions such as

- Deploying ArcGIS Enterprise in a cloud data center and publishing web services that reference an on-premises relational database management system (RDBMS).
- Deploying Portal for ArcGIS in one data center and a federated ArcGIS Server site in another.
- Using on-premises desktop clients (ArcGIS Pro and ArcGIS Desktop) with an ArcGIS Enterprise deployment that is hosted in a cloud data center.
- Using a premium network service or pathway to converge one or more data centers (on-premises and cloud).

There are underlying principles that should inform your thinking. However, because implementation details and user perceptions vary, it is often not practical to make decisions based on these underlying principles alone. Typically, testing is required to make a realistic determination. Performance considerations are testable with relatively high degrees of confidence in the results. Although reliability considerations can be more challenging to assess with testing

Defining Terms and Concepts

Specificity and common definitions are important in attaining shared understanding. Foundational terms used throughout this paper are defined below.

Response Time

Response time is the time elapsed from the beginning to the end of an operation. For example, the response time of drawing a map is measured from the time a user starts the pan or zoom operation to the time the map has completed redrawing. In contemporary applications, maps are complex objects often composed of multiple map services or layers drawn on parallel threads. This means that the end of the map draw event is ragged, that is, each service or layer is completed at different times.

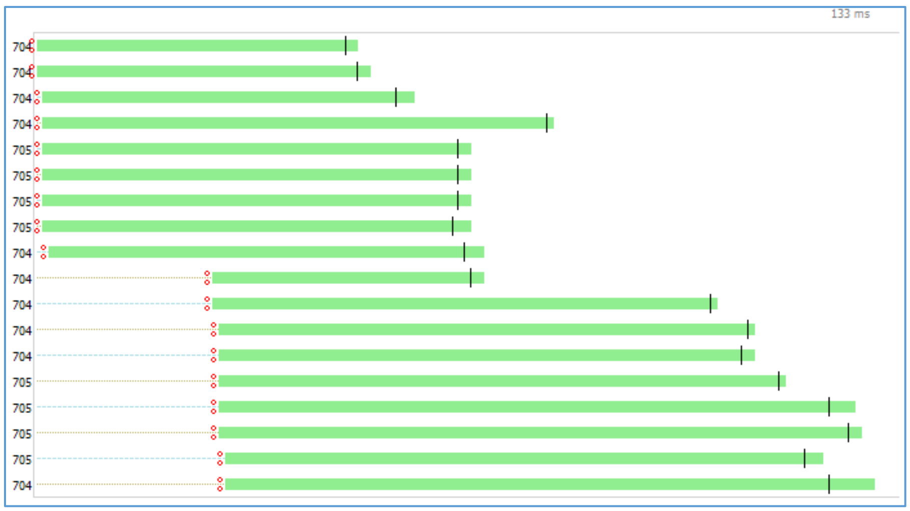


Figure 1: Many requests start and end at different times in a map draw operation.

It is often useful to measure the response time of an individual service request or layer draw operation. This makes response time easy to measure. However, this does not tell you whether response time is good or bad. It just gives you a number. Only your judgment can determine if what that number represents is acceptable.

Figure 2: The response time of one request is clear.

Performance and response time are related ideas. The terms are sometimes used synonymously. This paper uses response time to avoid the ambiguities associated with the term performance.

Reliability

Reliability is a measure of how often an operation successfully returns an appropriate value without error. For example, a complete service request returns its results without an HTTP error or failure or an error in the response body. Reliability is also a goal that relates to response time. For example, a user may say, "We want our users to see a reliable response time from the ArcGIS Enterprise environment." This indicates consistent, repeatable response time metrics such as "maps generally take from two to three seconds to draw."

Bandwidth

Bandwidth is the quantity of information that can be sent over a connection in a specific amount of time (such as one second). It is usually measured in megabits per second (Mbps) or gigabits per second (Gbps). Many people equate bandwidth with speed. This is not accurate. Bandwidth is a measure of capacity, not the speed at which things happen, which is dependent on many other factors.

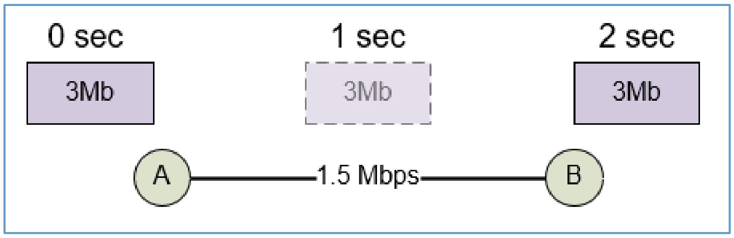


Figure 3: Bandwidth is computed capacity.

An analogy for bandwidth is a highway with multiple lanes of traffic. You may know that a road has three or four lanes, but do you know your likely driving speed down that road? You do not. You might guess that with four lanes, the road might be less congested than with three. However, it could be the opposite. Congestion, not bandwidth, can influence speed. However, as networks have been steadily improving over the decades, there are fewer situations where bandwidth is in short supply. This means that congestion is becoming less and less of a factor.

Per-Bit Transport Time

Per-Bit Transport Time is the time it takes one bit (the smallest piece of information) to travel over a connection.

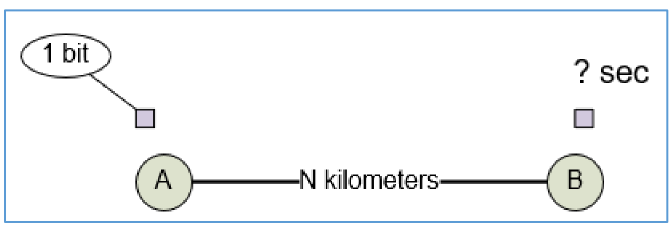


Figure 4: The per-bit network transport time is the time it takes one unit of data to travel over a connection.

While congestion is less of a factor influencing speed, per-bit transport times have become a greater factor. There are several reasons for this.

First, physical distance (over any given medium) is a deterministic factor for per-bit transport time. Longer physical distance translates to longer transport times. As networks have improved, organizations have spread out their IT infrastructure across even greater distances. So, while the maximum per-bit transport time is about the same, the typical distance of many network connections has increased. This makes per-bit network transport time more relevant.

Second, network sophistication and complexity can increase per-bit transport times. The more switches, routers, proxies, firewalls, intrusion detection systems, etc. there are between point A and point B, the more impedance there is to the speedy transport of the bit across the network. While networks have improved their capacities and reliabilities, they have also increased their complexities. Not every network component will add to transport time, but many will. Growing network complexity also makes per-bit transport time more relevant. Outcomes vary, but the increased complexity might increase reliability and decrease response time.

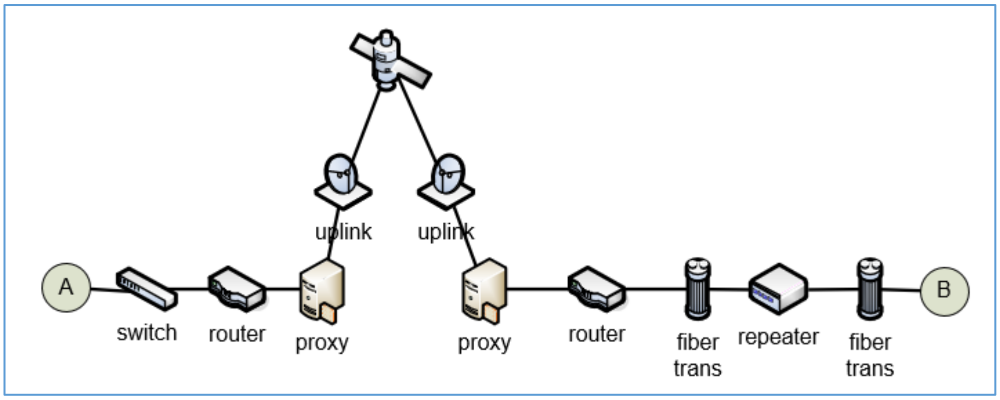


Figure 5: Adding more components can lead to network complexities

Chattiness

Network communications can be more or less chatty. *Chattiness* is an informal term—a more chatty communication has relatively more back-and-forth requests sent during the application or client usage, and a less chatty communication sends fewer but potentially larger or more complex requests. In human communication, texting or instant messaging is chatty—you make relatively short statements, even broken sentences, typically waiting for a reply before making subsequent statements. Sending an email is less chatty because you send a large amount of information at once and are likely to get a larger amount of information back.

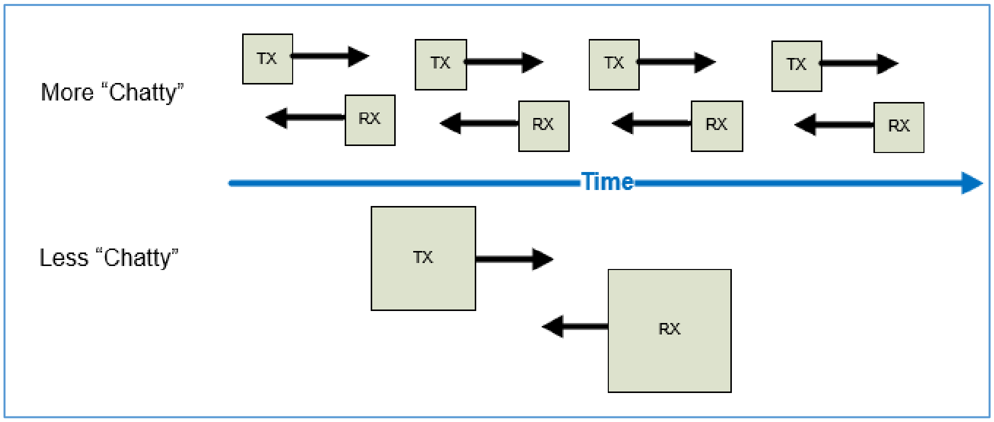


Figure 6: Chatty communications have more transmissions (TX) and receipts (RX) per logic unit of information

Per-bit network transport time has more impact on response time in chatty communications than with less chatty communications. Chatty communications are slower when per-bit transport times are longer. This may prompt the question, Why is any communication chatty? The answer is that in suitable network conditions, chattiness can be optimal for response time.

Concepts Summary

Bandwidth has grown substantially in contemporary networks. In most cases, it is no longer the primary factor when explaining response time and reliability issues. Per-bit network transport time often has greater relevance than bandwidth.

Reliability tends to be better than it was in the past. The key to reliability is redundant paths and self-healing network protocols. On longer network links, the increased reliability has sometimes come at the cost of decreased response time because it takes longer to heal a disrupted path.

The relative degree of connection chattiness is an important consideration when laying out your system components across network segments.

Network Communications in ArcGIS Enterprise

This section will discuss the relationships between components in ArcGIS Enterprise and the concepts just described.

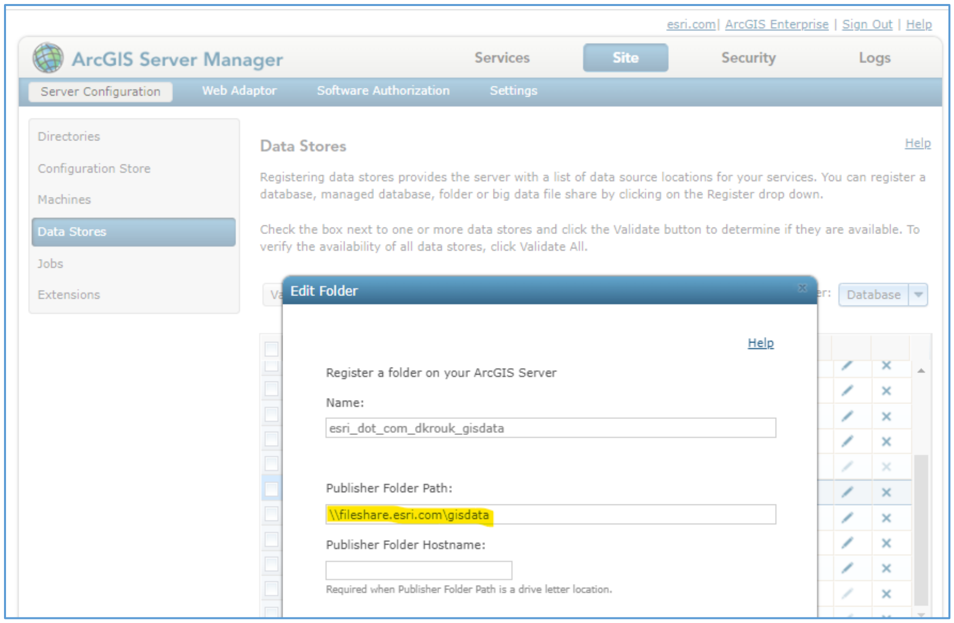
Software Design

Esri designs its software to function in varying network conditions with different characteristics. However, the IT industry patterns supported by the software, and employed in a system design, mean that some usage patterns are more effective than others.

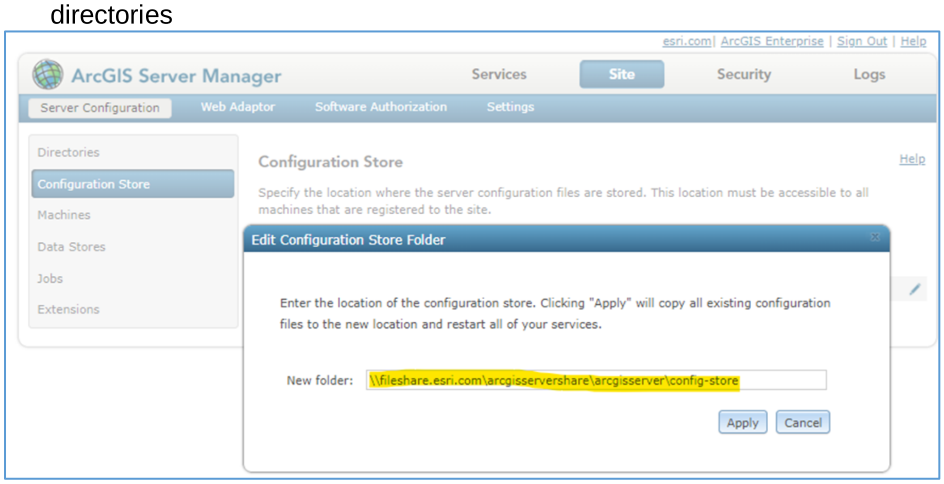
Connections to Databases and Files

In ArcGIS Enterprise, several components connect directly to databases (RDBMSs) and files (files on a file share being most relevant in this case). Those components include the following:

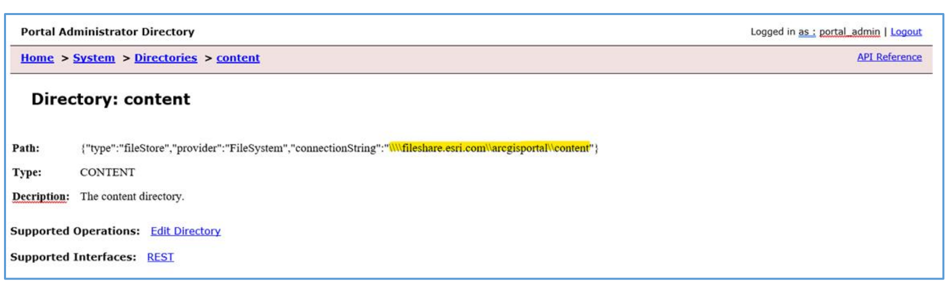
A. ArcGIS Server connections to databases and files to support map (and other) services



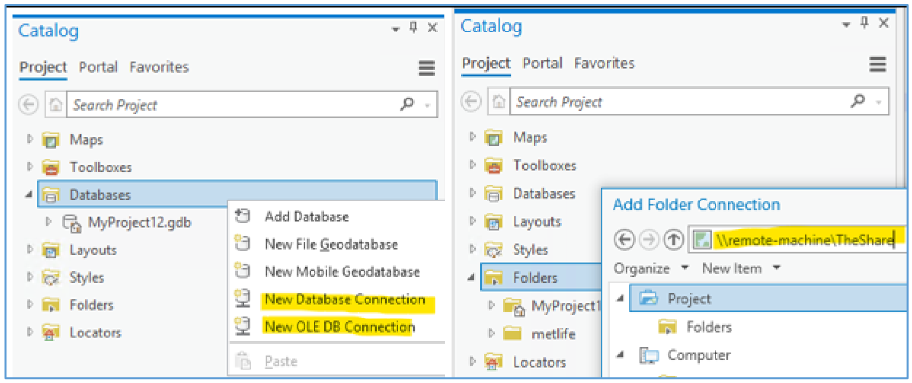
B. A multiple machine ArcGIS Server site's configuration store and server directories



C. The content directory in a highly available ArcGIS Enterprise (a two-machine Portal site)



D. ArcGIS Pro direct database connections and file share (mapped drive or UNC path) connections to feature classes



The underlying database and file share protocols that support these connections are relatively chatty. And this is not unique to Esri® software. As a result, the best response time is over network paths with short per-bit network transport times.

When it comes to reliability, the impacts are more varied. For example, ArcGIS Server will attempt to create a new connection to an RDBMS if the network connection is broken. Requests will fail until a new connection is established. Typically, ArcGIS Pro will ask you to refresh or reestablish an RDBMS connection that is broken. The ArcGIS Server and Portal for ArcGIS connections to their config-store, server directories, and content directory will typically lead to an internal failure when the connection is broken. However, once the network link is healed, the same operation will succeed as the Esri software will eventually reestablish a connection. Note that broken connections will disrupt the requests and functionality that rely on the information stored in those locations. For server systems, reduced network reliability expresses itself as variable system reliability and, to some degree, decreased response time.

Connections to HTTPS Endpoints

Most clients (web browsers, native mobile apps, ArcGIS Pro, etc.) connect to ArcGIS Enterprise through HTTPS. When you use services, web maps, web apps, etc., you are using HTTPS.

HTTPS connections tend to be relatively less chatty than database and file connections. As a result, most HTTPS connections are more tolerant of longer per-bit network transport times.

There are exceptions to this principle. For example, between an Enterprise Portal and a federated ArcGIS Server, there are back-channel communications. While those connections employ HTTPS, the nature of the communication is moderately chatty. In this case, the consequence of slight delays is easily expressed in the response time and failed communication caused by network problems is disruptive to the system. An unreliable network may lead to functional failures and decreased system response time.

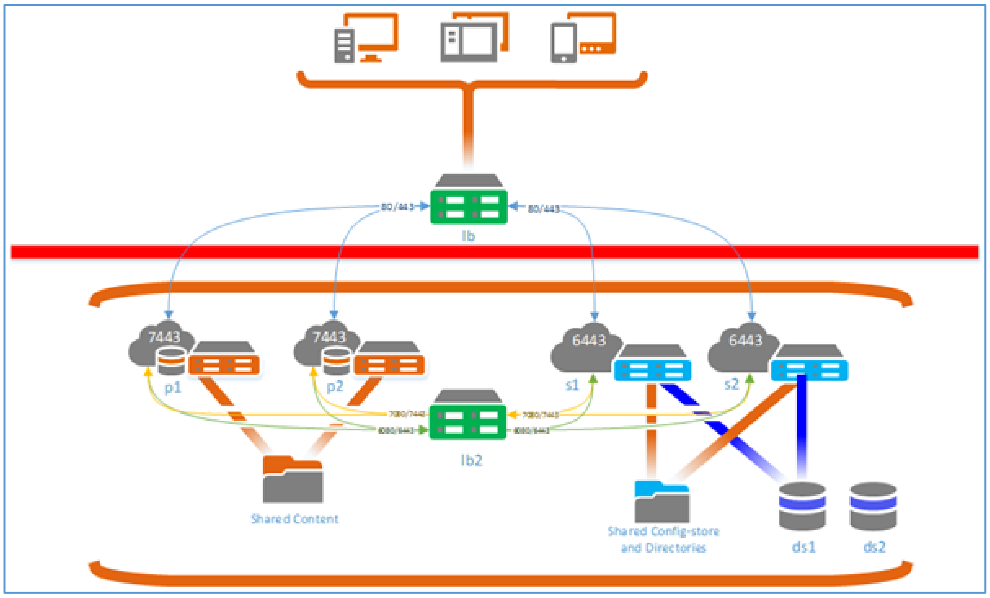


Figure 7: Back-channel communications are shown as yellow and green lines.

Some user-facing cases over HTTPS are moderately chatty or request-dense. This makes them relatively more sensitive to per-bit network transport times when it comes to response time. Some operations in the ArcGIS Utility Network fall into this category. The same can apply to a web map, depending on its complexity and use. Most client applications will report an error or experience a functional problem when a request fails to return a response due to network disruption. The situation is often healed by reattempting the operation in the application or reloading the browser-based application.

So, it tends to be true that HTTPS connections are more friendly to longer per-bit network transport times. This means response time is less noticeably impacted by longer per-bit network transport times than in the case of database and file connections. But, for some use cases, such as the back channel in federation and particularly dense user communications, there can be a noticeable response time degradation. And, while the software will sometimes be able to create a replacement connection for one disrupted by the network, there are both system stability and response time consequences to this kind of problem.

Connections to Streaming Desktops or Applications

Streaming desktops or applications (e.g., VDI, virtual desktops, RDP, Citrix, AWS Workspaces, Azure WVD) are technologies designed to allow clients to function well over longer transport times than they otherwise would. They accomplish this by running the client software remotely (closer to the servers) and streaming the display information back to the user. This kind of technology makes it practical to use ArcGIS Pro in one location while consuming databases or file shares in another location without suffering the full response time impact of the network distance. It does this by keeping the chatty communications off the long-distance network connections. These streaming technologies are not produced by Esri. However, they are well-established and effective solutions with Esri products like ArcGIS Pro.

Network Characteristics

There are two kinds of networks with respect to per-bit transport time and reliability: LANs and WANs.

Local Area Networks

LANs typically have a per-bit network transport time of about 1 millisecond. Occasionally, per-bit network transport times on LANs can be as high as 5–10 milliseconds.

Contemporary LANs benefit from multiple pathways and network protocols that allow the rapid healing of broken pathways. Reliability is high because communications are quickly rerouted over other pathways. Client applications (e.g., Esri software) are often unaware of the problems on the network.

Wide Area Networks

WANs are all networks other than LANs. These include the internet (an enormous public WAN), an organization's private network that connects multiple buildings and locations, and commercial and premium network services that provide optimized pathways between data centers (whether those data centers are on-premises or on the cloud).

WANs can have different per-bit network transport times. On the internet, per-bit network transport times could be in the 150–500 millisecond range. For private or premium network services, per-bit network transport times could be less than 10s milliseconds to less than 150 milliseconds.

Reliability also varies with WANs. The foundation of reliability is redundancy. If a WAN link has redundancies in its gateway (access from the LAN) and pathways (routes on the network), it has the elements that will allow it to heal broken pathways or gateway outages. Experience will vary by specific WANs and WAN technologies. WANs are slower in using redundancies to heal broken pathways than LANs. So, slow-healing network pathways can lead to a slower response time. Some WANs will also be less effective in healing their disruptions. This expresses itself in reduced reliability compared to LANs.

System Design

Keeping these patterns of software design and network characteristics in mind, you can design and deploy systems that are aligned with your values. You may be tolerant of response time risk in exchange for the benefits of using a given network pathway. The same applies to reliability risk; you may be more (or less tolerant) of it given your objectives and priorities.

Least Risk Design Pattern

Placing the following connections on LAN links will provide the least risk design (relative to network characteristics):

1. Database connections
2. File share connections

3. Back-channel federation communications between the Enterprise Portal and federated servers
4. Exceptionally request-dense HTTPS user workflows

Low Risk Design Pattern

For user client connections, such as those from ArcGIS Pro to databases or file shares, you can use the streaming desktop or application pattern to increase your options for #1 and #2 above. These technologies allow ArcGIS Pro to run in the same data center as the databases and file shares while being used in another location (where the user is). So, if all you wish to accommodate over a WAN is your user client connections to databases and file shares, the streaming pattern gives you more design flexibility. However, if you also want to place your ArcGIS Server connections in your databases and file shares over a WAN, desktop or application streaming will not address that.

Streaming solutions are effective in reducing the impact of per-bit network transport times on applications like ArcGIS Pro. But their effectiveness in reducing the impact of network reliability is less clear. Different streaming technologies and different network conditions will lead to different outcomes.

Other Design Patterns

You can opt to place any or all the connections mentioned above on WAN links. Although none of those efforts are expected to provide better response time or reliability, they may effectively provide equivalent response time and reliability. Or, they may simply be acceptable, given your organization's other priorities.

To minimize your risks for these designs

1. Use LAN links for as many of the chatty communications as possible.
2. Use WAN links with the lowest-available per-bit network transport times.
3. Thoroughly test your use cases in preproduction environments before committing to your design ideas.

Exploring Specific Cases

This final section will address some common system design questions raised by Esri customers.

Can I Put ArcGIS Enterprise in a Public Could Data Center and Serve Data from On-Premises Databases?

Esri does not recommend this in most cases because it places chatty communications over WAN network links. However, a small group of Esri customers adopts system designs like this. And, if your WAN link has a very low per-bit network transport time and high reliability, your use cases may not be strongly impacted by this design choice.

Thorough testing of services could establish whether the response time impacts are acceptable. Logs should be reviewed following the response time testing to look for evidence of problems in making or maintaining connections to the databases from ArcGIS Server. Reviewing logs is useful because reliability varies, and users may not fully recognize a reliability problem as such in their testing. Log review can also be used to quantify response time results allowing for objective comparisons between the design alternatives.

Can I Put Portal for ArcGIS in One Data Center and a Federated ArcGIS Server Site in Another?

Esri does not recommend this. The smooth functioning of ArcGIS Enterprise is quite sensitive to the reliability and response time of the back channel federation communication between the Enterprise Portal and its federated Server sites.

While it is possible to test for acceptance, the variety of use cases suitable for testing is quite broad and could change as new features are introduced in future releases. Therefore, it is not practical to use log data to objectively determine the response time and reliability differences between design alternatives in a durable manner.

Can I Use ArcGIS Pro On-Premises with ArcGIS Enterprise in a Cloud Data Center?

Some customers report that their ArcGIS Pro workflows that use HTTPS in a distant data center (on the cloud or otherwise) are well functioning in their ArcGIS Enterprise implementation. The common characteristics of HTTPS communications from clients to servers provide a good theoretical basis for this reported satisfaction. Because HTTPS tends to be less chatty, you can expect it to be more tolerant of longer per-bit network transport times.

The limits of customer satisfaction with this pattern appear to be driven by the following factors:

- The specific per-bit network transport times
- The reliability of the specific network links
- The nature of the specific user workflows

For system design cases where you believe that user workflows will be relatively consistent over time, this design alternative can be investigated through testing. Logs of HTTPS activity would allow for objective measures and comparisons of the response time of design alternatives. However, they will not likely offer great insight into whether network reliability is a problem. Reports from users would likely be the primary source of information and evaluation for that aspect. That is not a very systematic source of information.

Can I Use a Premium/Commercial Network Service to Converge One or More Data Centers?

Services such as AWS Direct Connect or Azure ExpressRoute provide optimized, private network links to remote data centers. Each commercial product offers specific features, and product availability or characteristics may vary by area. However, in principle, a private network with a premium service level would be expected to offer better per-bit network transport times and possibly reliability than the internet. It is unlikely, however, that the experience over one of these links would be as good as the experience on a LAN.

A premium service is better than not having one. But the question of whether it is good enough depends on which connections you want to place on the link, your tolerance related to response time and reliability, and your organization's tolerance for risk and it's competing priorities.

Measuring Transport Time

This article emphasizes the significance of network transport time. There are different ways to measure it.

Ping

```
C:\>ping machin1.esri.com

Pinging machine1.esri.com [10.10.51.22] with 32 bytes of data:
Reply from 10.10.51.22: bytes=32 time=23ms TTL=124
Reply from 10.10.51.22: bytes=32 time=23ms TTL=124
Reply from 10.10.51.22: bytes=32 time=23ms TTL=124
Reply from 10.10.51.22: bytes=32 time=23ms TTL=124

Ping statistics for 10.10.51.22:
  Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
  Minimum = 23ms, Maximum = 23ms, Average = 23ms
```

In this example, the network transport time is 23 milliseconds.

Capture Network Packets

In this example, the SYN–SYN/ACK–ACK TCP handshake transpires over 25 milliseconds.

Source	src_prt	Destination	dst_prt	TCP_timedelta	Stream	Info
10.212.12.62	1381	10.10.51.22	443	0.000000000	30	1381 → 443 [SYN] Seq=0 Win=65280 Len=0 MSS=1360 WS=256 SACK_PERM=1
10.212.12.62	1382	10.10.51.22	443	0.000000000	31	1382 → 443 [SYN] Seq=0 Win=65280 Len=0 MSS=1360 WS=256 SACK_PERM=1
10.10.51.22	443	10.212.12.62	1381	0.025470000	30	443 → 1381 [SYN, ACK] Seq=0 Ack=1 Win=8192 Len=0 MSS=1460 WS=256 SACK_PERM=1

Figure 8: TCP session initiation in a packet capture is a useful metric related to per-bit network transport time

Either method provides a specific, repeatable observation that is reasonably accurate and related to the measurement target.

Conclusion

Esri has not established formal support statements for various network services and characteristics. However, the IT patterns used in its software design can inform your organization's system design choices. Database connections and file connections like short per-bit network transport times and high reliability. HTTPS connections are more accommodating of variable transport times and reliability, with a few exceptions. The back channel of federation is not particularly compatible with longer per-bit transport times or lower reliability, for example. Streaming desktop and application solutions make rich client applications like ArcGIS Pro more compatible with longer network distances.

While there are only a few principles to bear in mind, actual network characteristics, use cases, and user tolerances vary. Adopting designs other than the least risk design pattern described in this paper should be accompanied by thorough testing, analysis, and acceptance by the organization.

About the Author

Daniel Krouk works as a team lead and consultant in Esri's Professional Services Division. He and his team support advanced ArcGIS Enterprise system deployments and troubleshooting efforts, particularly at the intersection of enterprise IT systems and Esri software.

Esri, the global market leader in geographic information system (GIS) software, offers the most powerful mapping and spatial analytics technology available.

Since 1969, Esri has helped customers unlock the full potential of data to improve operational and business results. Today, Esri software is deployed in more than 350,000 organizations including the world's largest cities, most national governments, 75 percent of Fortune 500 companies, and more than 7,000 colleges and universities. Esri engineers the most advanced solutions for digital transformation, the Internet of Things (IoT), and location analytics to inform the most authoritative maps in the world.

Visit us at esri.com.

Contact Esri

380 New York Street
Redlands, California 92373-8100 USA

1 800 447 9778
T 909 793 2853
F 909 793 5953
info@esri.com
esri.com

Offices worldwide
esri.com/locations

For more information, visit **esri.com/URL**.