Transactions on Engineering and Computing Sciences - Vol. 11, No. 1
Publication Date: February 25, 2023
DOI:10.14738/tecs.111.13967.
Secara, I. A., & Hordiiuk, D (2023). The Importance of High-Bandwidth Low-Latency Network Systems in the Modern Age. Transactions on Engineering and Computing Sciences, 11(1). 37-44.

SOCIETY FOR SCIENCE AND EDUCATION

The Importance of High-Bandwidth Low-Latency Network Systems in the Modern Age

Ion-Alexandru Secara
CTO at PostureHealth Inc.

Dariia Hordiiuk
CTO at AiSport.

ABSTRACT

This paper investigates the importance of high-bandwidth and low-latency network systems in the modern age. It introduces the concepts of networks systems, bandwidth and latency. Then the paper analyzes factors affecting bandwidth and latency in network systems. The factors analyzed include the distance over which the network connection spans, the router and wireless protocols used, as well as the type of cabling used in wired connections. Finally, the paper uses the aforementioned analysis to investigate the choices and solutions that financial institutions use in order to optimize the speed and capacity of their algorithms, when performing high-frequency trading. Choices analyzed include the proximity of the algorithms to the exchange servers, as well as the type of cabling used in wired connections to facilitate fast data transfers.

Keywords: Network Systems, Network Bandwidth, Network Latency, Networks used in High-Frequency Trading.

INTRODUCTION TO NETWORK SYSTEMS

The last 20 years have brought a sharp increase in popularity of computing machines. As people's lives become more and more dependent on computers, the need for better communication, connectivity and interoperability increases, hence the increased demand for reliable and fast network connections.

A network system is an assembly of independent computers that can communicate with each other via a shared medium, for the purpose of transmitting data and sharing resources [1]. Local area networks (LANs) span over a relatively small area, such as a single building or a campus. LANs can connect hundreds of computers and can be used by thousands of users with various responsibilities, privileges and access rights. The development of various network and communication protocols has made it possible for various entities, including organizations, institutions and business, to develop and use LANs according to their own individual requirements. Wide area networks (WANs) are comprised of multiple LANs located in different geographical locations, which can extend over multiple cities, countries and continents. One of the most widely used WANs is the Internet [1].

The COVID-19 pandemic has played a significant role in the development and innovation of network systems. As a large percentage of the world's population transitioned to working from

home, the importance of reliable and fast network connections became crucial. As the usage of social media, multimedia applications, including high-definition video, real-time collaboration and virtual reality has been increasing, network requirements, including bandwidth and latency, have become ever more important.

FEATURES OF HIGH-BANDWIDTH LOW-LATENCY NETWORK SYSTEMS

Network Bandwidth

Network bandwidth is a measurement of the highest capacity of a wired or wireless network to transmit a certain amount of data from one point to another, over a network connection, in a certain amount of time. Generally, network bandwidth is measured in the number of bits per second (bps), megabits per second (Mbps) or gigabits per second (Gbps) [2]. Network bandwidth is commonly mistaken with network speed. However, those two terms refer to two separate aspects of a network's performance. While the bandwidth measures the highest capacity at which a network can transmit data, the speed refers to the actual rate at which data can be transferred [3]. For instance, if a network's bandwidth if 1 Gbps, then the network's speed can never be higher than 1Gbps. The higher the bandwidth of a network, the more traffic and data the network can handle, while decreasing the chances of data congestion. Data congestion can lead to packet loss [3], which introduces the next feature analyzed by this paper - network latency.

Network Latency

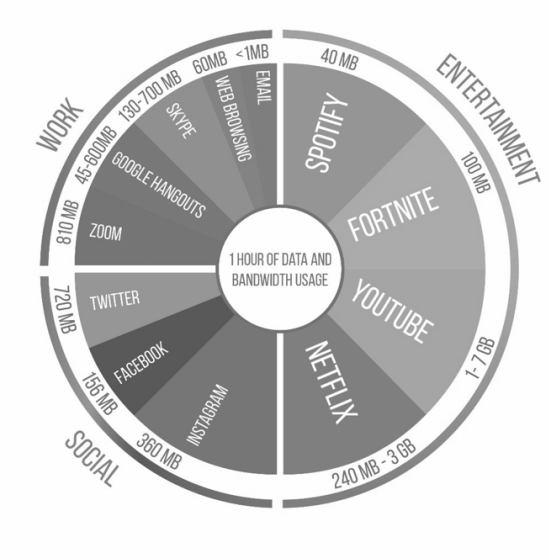
Network latency is used to describe any type of delays that occurs in the communication of data over a network [4]. Network connections that experience reduced delays are called low-latency networks, while network connections that experience increased delays are called high-latency networks. High latency reduces the efficiency of data transmission over a network and can significantly decrease the bandwidth of a network. A common way of measuring a network's latency is by running the 'ping' command on any public websites available on the internet. This command will measure the Round-Trip Time, referred to as RTT, or the time necessary for a package to be delivered to the respective website and then transmitted back to the sender [5]. Latency is generally measured in milliseconds.

Factors Affecting Bandwidth and Latency in Network Systems

Sharing of Network Bandwidth:

As mentioned above, bandwidth represents the maximum highest capacity of a wired or wireless network to transmit a certain amount of data from one point to another, over a network connection. Since it depends directly on the hardware (modem, router, wired cabling etc.) and software (limit set by network administrator, applications not being able to handle more traffic) capabilities of the network, computing machines and applications that are running, bandwidth is not unlimited [3]. If a single user with a single computer is connected to a network, then the bandwidth of that network is shared by the different services and applications of that user's computer. If multiple users or multiple devices are connected to the same network, then the bandwidth of that network is shared between all the users and devices connected to the same network [2]. Depending on the applications and services that are running on the user's computer or the activities of the other users and devices connected to the same network, the experience of each user can be greatly impacted, because the bandwidth of

the network has to be shared. For example, streaming 4K videos will use a significant amount of the bandwidth available on the network and it is usually referred to as a network hog [3]. A network bandwidth hog is service, network endpoint or application that consumes significant amounts of bandwidth in a network, reducing the bandwidth available to other users, services or applications. In comparison to streaming 4K videos, having a VoIP (Voice over Internet Protocol) call, consumes significantly less network bandwidth [2]. To better showcase this, below you can see an illustration that depicts the amount of bandwidth that is used by various activities, on average.



1. Illustration depicting the average amount of bandwidth that different pieces of software use, over the span of one hour [6]

As shown in Illustration 1 above, Netflix uses the most bandwidth out of the services that were analyzed. Streaming one hour of Standard Definition content takes up to 1GB. Streaming one hour of High-Definition content takes 3GB and streaming one hour of Ultra High-Definition Content can go up to 7GB. Meanwhile, on the other end of the spectrum, using email or streaming music on Spotify uses the least amount of bandwidth [6].

Hardware and Software Limitation on Network Bandwidth:

Apart from impact on network bandwidth resources that other applications and users have, network bandwidth can also be affected by the type of network connection, wireless or wired, and the hardware used. When it comes to wireless connections, the network bandwidth depends on the routers, and the wireless protocols that are being used. Currently, the standard wireless protocol is the 802.11ax. This protocol is also referred to as Wi-Fi 6, with a bandwidth of 10 Gbps, while using less power and being more reliable in crowded environments. 802.11ax was introduced in 2019 and was meant to replace 802.11ac, which is referred to as Wi-Fi 5 [7]. Compared to 802.11ax, 802.11ac has a bandwidth of 1300 Mbps on a 5 GHz network band and 450 Mbps on a 2.4 GHz network band. Given that 802.11ax was introduced 5 years after 802.11ac, the bandwidth increase is significant [8].

However, wired network connections have different considerations to take into account. Wired networks use cables to establish a connection. The material that the cable is made out of can greatly impact the network bandwidth available to the connected devices. Cables that are made out of copper have a bandwidth of approximately 40 Gbps [9]. In the meantime, cables that are made out of optical fiber have a current bandwidth of 100 Gbps [10], but recent experiments carried out by Japanese scientists have achieved speeds measured in the hundreds of Tbps (terabits per second), using a four-core optical fiber [11]. The main difference in bandwidth between copper cables and optical fiber cables is due to the nature through which the data signal is being sent through. In optical fiber cables, the data signal is sent as pulses of light, compared to electric signals in copper cables. The pulses of light travel much faster than electric signals, but they also travel further than electric signals [10].

When traveling over extended distances, cables made out of copper and optical fiber suffer from attenuation, meaning that the signal that is being transmitted, gets weaker as the distance travelled gets higher. However, as mentioned above, signals transmitted through optical fiber cables can travel much further than signals transmitted through copper cables [10]. Given the smaller effect of attenuation and the increased speed of signal transmission, optical fiber cables are widely used to connect different countries and even continents. For example, one of the longest undersea optical fiber cables, the SEA-ME-WE-3, has a length close to 39 thousand kilometers and is used to connect 33 countries, on 3 different continents [12]. The SEA-ME-WE-3 uses 2 optical fiber cable pairings, each carrying 48 wavelengths of 10Gbps, allowing it to transmit data from Japan to the UK in only 0.35 seconds [10]. Infrastructure like the SEA-ME-WE-3 enable the world to be connected in real-time.

Hardware and Software Limitations on Network Latency:

Similar to network bandwidth, network latency is also affected by the type of network connection, wireless or wired, and the hardware used, as well as the proximity of the device to the router.

When it comes to wireless networks, the average latency measured varies between 2ms to 20ms, but it can be considerably higher, depending on the router, proximity to the router and the wireless protocol used. On the other hand, wired connections generally have a latency of 0.1ms to 0.3 ms, providing a significant decrease in latency compared to wireless connections [13]. This comes to no surprise, since the types of signals used to transfer data, travel faster through cables, than through air, and they also offer more stable connections, which lead to a smaller amount of packet loss [13]. This becomes increasingly important especially when being connected to the internet, since the signals have to be transmitted through multiple routers, as they travel over long distances across the world.

Another important factor affecting network latency, as in the case of network bandwidth, is represented by the distance between the sender and the recipient. As mentioned above, the amount of time that it takes for a request or piece of data to be transmitted from the sender to the destination and back to the sender is known as the RTT. If the sender is located in San Francisco and the receiver is located in Sacramento, approximately 90 miles apart, data would generally be transmitted within 10 milliseconds. However, if the sender is located in San

Francisco and the receiver is located in Chicago, approximately 2,200 miles apart, the data would take longer to be transmitted; up to 50 milliseconds [14].
As the world becomes more and more interconnected, the need for increased reliability and high performance is at an all-time high.

DESIGN OF HIGH-BANDWIDTH LOW-LATENCY NETWORK SYSTEMS

The architecture and implementation of high-bandwidth low-latency networks are developed to tackle the factors presented previously in this paper. While most applications and services, require stable network speeds and latency to operate at an acceptable level, there are certain domains that have developed networks that can support high-bandwidth and low-latency operations.

For instance, let's consider the network design that financial institutions (investment banks, hedge funds etc.) use to perform high-frequency trading, referred to as HFT, where even a nanosecond can make a significant difference in the amount of money that is won or lost on any trade. To outline the importance of HFT for financial institutions, the global HFT market size is forecasted to reach USD 27.01 billion by 2030, from USD 9.02 billion in 2017 [15].

One of the key attributes in the design of high-bandwidth low-latency networks is the distributed nature of their architecture. High-frequency trading is almost exclusively run by algorithms. Those algorithms receive real-time high-speed streams of data, which means that financial institutions need the algorithms to be blazingly fast and able to process significant quantities of data. The workload for each transaction is split into multiple processes that are then executed on separate dedicated servers, which can include hundreds of machines [16]. As Arista mentions in its white paper, machines that are used in algorithmic trading can receive market data from other machines, analyze trading decisions on other machines and execute on trading decisions on another set of machines [16]. This means that the bandwidth and latency of the network connecting all those machines becomes one of the most important factors in high-frequency trading.

As previously analyzed in this paper, the distance between the sender and the receiver can make a significant difference in the time it takes to transmit data - the network latency. This is especially true for financial institutions who engage in high-frequency trading operations. Therefore, one of the first considerations for financial institutions is to place their servers and computing machines as close as possible to the servers of the exchange [17]. This reduces the time that the data signals need to traverse in order to be processed by the algorithms of the financial institution. According to Don Henderson, chief technology officer at NYSE Technologies, there are three types of connections used for HFT: remote connections, proximity and co-location [18]. While co-location and proximity connections have the smallest distances to the exchange servers, allowing for efficient arbitrage strategies, remote connections have also started using high-speed wired connections in order to enable fast trading decisions. According to Don Henderson, "There are longer term investors, who have multiple 10 Gbps pipes connected to our infrastructure. They may use it for market data but also to send trades to us." [18]

Placing servers close to the exchange server is one of the key factors in reducing latency. However, given that the data being transferred does not require significant network bandwidth resources, why do financial institutions choose to invest in high-bandwidth network connections?

When the amount of output bandwidth is overloaded by packets over a very brief period of time, usually microseconds, microbursts occur. Microbursts cause data buffers to fill with packets of data, which leads to new packets of data being delayed or even dropped [19]. Looking at the Better Alternative Trading System, a stock exchange referred to as BATS, the average bandwidth resources used were measured to 86 Mbs, over the period of one second. However, when this average bandwidth resources used are measured over a period of one millisecond, the amount of bandwidth resources used can reach 382 Mbs [19]. Therefore, according to Don Henderson, investors generally use 10 Gbps wired connections [18]. And while 10 Gbps is high given the highest bandwidth measurement came to 382 Mbs, as financial institutions look to extend their number of trading systems, strategies and exchanges that they are connected to, the amount of bandwidth resources required can increase even further.

Therefore, in order to create high-bandwidth low-latency network connection to the servers of the exchange, financial institutions use proximity and advanced wired connections to increase the efficiency of their trading algorithms. Most of those wired connections are made out of optical fiber. As highlighted in a report by Spread Networks from 2019, a newly installed optical fiber cable that connected the Chicago exchange to the New York exchanged, decreased the latency experienced by 3 milliseconds [20]. This is extremely important for financial institutions, and it is shown by the fact that 200 companies paid a combined $2.8 million dollars to use the newly installed optical fiber cable [20].

CONCLUSION

In conclusion, as the world becomes more dependent on computers, the need for better communication, connectivity and interoperability increases. Therefore, the bandwidth and latency capabilities of network systems become key factors when designing and implementing network systems. Despite the hardware, software and distances over which connections span over, network systems, especially wired connections using optical fiber cables, continue to keep the world inter-connected. For instance, the design of network systems is of the utmost importance for financial institutions who use algorithms to perform high-frequency trading.

In future work, the paper should analyze advancements in wireless and wired network connections that can greatly increase network bandwidth and reduce network latency.

Bibliography

[1] D. R. Winkelman, "Chapter 1: What is a Network?" [Online]. Available: https://fcit.usf.edu/network/chap1/chap1.htm.

[2] A. Froehlich and K. Ferguson, "bandwidth (network bandwidth)," TechTarget, May 2021. [Online]. Available: https://www.techtarget.com/searchnetworking/definition/bandwidth.

[3] SolarWinds, "What is Network Bandwidth?" Geekbuilt., [Online]. Available: https://www.solarwinds.com/resources/it-glossary/network-bandwidth.

[4] IR Team, "Network Latency - Common Causes and Best Solutions," Integrated Research Ltd., [Online]. Available: https://www.ir.com/guides/what-is-network-latency.

[5] A. Günther and C. Hoene, "Measuring Round Trip Times to Determine the Distance Between WLAN Nodes," International Conference on Research in Networking, pp. 768-779, 2004.

[6] N. Teran, "1 Hour of Data and Bandwidth Usage," Show Me Cables an INFINITE brand, 11 May 2020. [Online]. Available: https://www.showmecables.com/blog/post/at-home-data-usage.

[7] B. Mitchell, "802.11 Standards Explained: 802.11ax, 802.11ac, 802.11b/g/n, 802.11a," Lifewire Tech For Humans, 16 November 2021. [Online]. Available: https://www.lifewire.com/wireless-standards-802-11a-802-11b-g-n-and-802-11ac-816553.

[8] Intel, "Different Wi-Fi Protocols and Data Rates," Intel, [Online]. Available: https://www.intel.com/content/www/us/en/support/articles/000005725/wireless/legacy-intel-wireless-products.html.

[9] i.e. Smart systems, "Copper vs. Fiber Optic Cabling," i.e. Smart systems, 26 May 2021. [Online]. Available: https://iesmartsystems.com/copper-vs-fiber/.

[10] R. Noori, "How Does Fiber-Optic Cable Bandwidth Work?" The Network Installers, 11 May 2022. [Online]. Available: https://thenetworkinstallers.com/blog/fiber-optic-cable-bandwidth/.

[11] M. Tyson, "Researchers Shatter Bandwidth Record Using Existing 4-Core Fiber Cables," Tom's Hardware, 3 June 2022. [Online]. Available: https://www.tomshardware.com/news/nict-researchers-shatter-bandwidth-record.

[12] Submarine Networks, "SEA-ME-WE 3," Submarine Networks, [Online]. Available: https://www.submarinenetworks.com/systems/asia-europe-africa/smw3.

[13] S. Moon, "Wi-Fi vs. Ethernet for Gaming – 7 Things You Should Know," BinaryTides, 13 May 2021. [Online]. Available: https://www.binarytides.com/wifi-vs-ethernet-for-gaming/.

[14] Perry, "What Is Internet Latency?" netBlazr, [Online]. Available: https://www.netblazr.com/what-is-internet-latency/.

[15] DATAINTELO, "High-frequency Trading Market Research Report," DATAINTELO, 2021. [Online]. Available: https://dataintelo.com/report/high-frequency-trading-market/.

[16] Arista, "Architecting Low Latency Cloud Networks," Arista, 2016. [Online]. Available: https://www.arista.com/assets/data/pdf/CloudNetworkLatency.pdf.

[17] The Week Staff, "Wall Street's secret advantage: High-speed trading," The Week, 11 January 2015. [Online]. Available: https://theweek.com/articles/493238/wall-streets-secret-advantage-highspeed-trading.

[18] MarketsMediaGroup, "Can We Get Closer: What's Next in Co-Location?," MarketsMediaGroup, 6 June 2012. [Online]. Available: https://www.marketsmedia.com/closer-whats-co-location/.

[19] Arista, "Low Latency High Frequency Financial Trading Networks," Arista, [Online]. Available: https://www.arista.com/en/solutions/low-latency-high-frequency-financial-trading-networks.

[20] R. Gillette, "High Frequency Trading and the Use of Algorithms," Thinking Through Writing, 5 April 2019. [Online]. Available: https://www.snoqap.com/posts/2019/4/5/high-frequency-trading-and-the-use-of-algorithms.

[21] S. Murmson, "What Can Affect Bandwidth?" Chron., [Online]. Available: https://smallbusiness.chron.com/can-affect-bandwidth-68780.html.