Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

# Latency – Time for lawyers to get up to speed?

*Paul Hinton, Edwin Baker, Chris Hill*

*Kemp Little LLP, London, UK*

*Keywords:*
Latency
Jitter
Packet delay variation
Algorithmic trading
High-frequency trading
Cloud computing
EaaS
Latency legal issues
Latency contracts
Latency service levels
Innovation
Financial services IT
Systems integration

ABSTRACT

While latency issues have always existed in IT systems, it is only in recent years that latency has become a significant focal point for both IT suppliers and customers. This paper will provide an overview as to how latency arises in IT systems, why latency is becoming important across a variety of business sectors and then explores some of the prominent legal issues inherent in contracting for low-latency solutions.



## 1. Introduction

Latency is the delay experienced in an IT system between an input being sent and an output being received. Anyone who has experienced delays on satellite phone calls or video conferencing calls has experienced the frustrations of high latency. Although latency issues have always existed in every IT system, it is perhaps only in recent years that latency has become a significant focal point for both IT suppliers and customers.

At present, latency is particularly relevant within industries where speed of access and use can offer a distinct commercial advantage, for example when trading in financial products or when gambling or gaming online. In most countries, the Internet is reliably available and offers sufficient bandwidth at reasonable prices for access to digital commerce services. Access and bandwidth will continue to improve, software has reached a point where it can offer almost any functionality required and the devices which enable access are becoming increasingly sophisticated, user friendly and cheap. Consequently, as the value of goods purchased via the Internet increases and methods and sophistication of online browsing and purchasing continue to improve, latency issues will become increasingly relevant in other sectors. Consumers in the future may be empowered and able to purchase most commodities online, accessing vast global markets and ensuring the best price and deal terms possible, by entering offers and enabling all sellers to bid against each other for the sale (we have already seen aggregated online consumer purchase models of this nature, e.g. Groupon). Whatever the winners may be in terms of future digital commerce models, it is likely that user experience will be the key differentiator between services and, as such, reducing latency and ensuring consistent fast provision of service is likely to be a key focus for both consumers and suppliers.

## 2. What is latency?

'Latency' is the term used to describe the time it takes for a piece of data (a "data packet") to cross a data connection from sender to receiver. Latency is present in any trans-

0267-3649/$ – see front matter © 2012 Kemp Little LLP. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.clsr.2012.03.011

mission of data, whether the distance involved is a few centimetres (e.g. between individual components of a computer), a few metres (e.g. between computers located in the same room) or thousands of miles (e.g. over the Internet). Latency is usually measured by the time it takes for a data packet to be sent to its destination and a response to be received (known as a "round-trip").

"Jitter" or "Packet Delay Variation" (PDV)[1] is the term used to describe how latency varies over a given period of time and between data packets sent across the same data connection. If latency of successive data packets has only minor variations, the jitter is low and conversely if there are major variations for successive data packets over a given period, the jitter is high.

The total latency experienced in a round-trip and the variables influencing jitter are caused by a number of factors which can be grouped into 3 broad categories:

- propagation latency;
- transmission latency; and
- processing latency.

### 2.1. *Propagation latency*

This is the time it takes for a signal to travel from one end of a communication link to the other. There are two factors which determine propagation latency: distance and speed. A copper wire or a fibre optic cable carries a signal at roughly 67% the speed of light. This provides a propagation latency of around 0.005 milliseconds[2] ("ms") per kilometre, meaning that, for example, a straight line communication link between London and Tokyo of around 9600 km would have a propagation latency of approximately 58 ms one-way and a 96 ms round-trip. Unlike fibre optic or copper cables, the use of wireless methods of transmission such as radio waves provides the benefit of greater speed i.e. full speed of light, but this is subject to the impacts of propagation which in the case of radio waves can be caused by physical phenomena such as reflection, refraction, diffraction, absorption and polarization. In addition, when using wireless transmission to connect remotely to the Internet via satellite, the distances involved are usually far greater than compared with cabled infrastructure: for example, a geosynchronous satellite is around 36,000 km above the earth's surface which, assuming no atmospheric propagation, equates to 125 ms for a single trip or 250 ms for a round-trip (i.e. an eighth or a quarter of a second respectively). Therefore connections to the Internet via wireless transmission (as opposed to via cable) will mean data packets travel faster but in practice may suffer a higher degree of propagation latency and considerable jitter. Propagation latency can be minimised by keeping the total distance between the communicating devices as short as possible. This can be achieved by ensuring the two ends of the connection link are within the same geographic region, city or, in the case of 'co-location', in the same building.

The commercial demand to reduce propagation latency is evidenced by a recent announcement of a new submarine cable system between London and New York costing an estimated US$300m.[3] Previously this route was seen as being saturated (the last cable was installed around 2003), however the demand for fast high bandwidth connectivity is now seen to be increasing sufficiently to justify this significant investment.

### 2.2. *Transmission latency*

Transmission latency is the delay experienced in transmitting quantities of data across a communication link. Every communication link has a "speed" which measures the amount of data per second the link is capable of transferring (e.g. 10 Mbit/s). The smaller the capacity of the communication link, the longer it takes for data to be sent across it, and the higher the transmission latency. Conversely, the greater the capacity of the communication link, the quicker the data is sent, and the lower the transmission latency.

### 2.3. *Processing latency*

Processing latency is the delay caused by IT hardware and software interacting with the data packet from its origin all the way to its destination. Each system or stage of the software/hardware which in some way interacts with a data packet being transmitted causes a certain amount of delay. It is this total cumulative delay that determines the amount of processing latency.

| | Sources of processing latency |
|---|---|
| Application latency | Operating system, software, |
| Hardware latency | CPU, memory, data storage |
| Network latency | Protocol overheads, firewall, encryption, security check |
| Interface latency | Buffering, routing, framing, packetisation, serialisation, fragmentation |

Processing latency can be reduced through improvements and advances in hardware and also through the improvement and optimisation of software and networking protocols. In recent years we have seen suppliers continually competing to bring out ever faster products to assist customers in reducing processing latency.

[1] While Packet Delay Variation (PDV) is the most technically accurate description of variations in latency, "jitter" is commonly used in the telecommunications field to have the same meaning. Jitter does however have additional and different meanings outside of this context. For clarity, references to "jitter" throughout this article are references to PDV.

[2] 1 s = 1000 ms.

[3] In 2011, Emerald Networks announced that they intend to install a new fibre optic cable between London and New York, utilising the shortest connecting route and potentially offering high bandwidth capacity and low-latency connections (advertised at a potential latency of less than 62 ms per "round-trip").