

physicsworld

METROLOGY FEATURE

Time traders

10 Jul 2018

Taken from the July 2018 issue of *Physics World*

In today's markets, every microsecond counts. Jon Cartwright discovers how the UK's National Physical Laboratory is keeping regulators up to speed

Just after 2.30 p.m. local time on 6 May 2010, Wall Street experienced one of the biggest, and briefest, crashes in its history. Within minutes, the Dow, one of the three most-followed US market indices, plunged 9%, while prices of individual shares became intensely volatile, in some cases fluctuating between tens of dollars and cents in the same second. More than \$850bn was wiped off stock values - although by the end of the trading day they had mostly recovered.

What caused the Flash Crash, as it came to be known? Early theories blamed either an error in trading software, or a human at a computer inadvertently selling a large number of shares - the so-called fat-finger hypothesis. Some analysts even claimed the Flash Crash was merely part of the more exaggerated ups and downs we should expect as financial trading becomes more decentralized and complex. But many suspected foul play.

In April 2015, at the request of US prosecutors, Navinder Singh Sarao was arrested at his parents' semi-detached home in Hounslow, west London, UK. A lone trader, Sarao, then 36, was accused of crafting "spoofing" algorithms that could order thousands of future contracts, only to cancel them at the last minute before the actual purchases went through. By exploiting the resultant dips in markets, he allegedly earned some \$40m (£27m) over five years.

High speed; high stakes

Sarao was found guilty of spoofing and wire fraud by a US court earlier this year, but it is still not known whether his actions actually caused the Flash Crash. In fact, based on current financial infrastructure, performing an accurate postmortem on an extreme market event can be nigh-on impossible, because it involves knowing when - precisely when - all the trades took place.

Back in the days when traders called out buys and sells on the bustling floors of stock exchanges, keeping official time records of transactions was hardly a problem. But we now live in an era of automated, high-frequency trading, in which orders are executed in microseconds, if not faster (see "The

art of the algorithm", below). Observers have struggled to keep up. In response to a report by US regulators five months after the Flash Crash, [David Leinweber](#), the director of the Center for Innovative Financial Technology at Lawrence Berkeley National Laboratory in the US, wrote that those same regulators were "running an IT museum", with totally inadequate resources to forensically analyse modern transactions.

One of the main problems is that of synchronization, and the proof of it. In a financial organization, a trade may be timestamped as having occurred in a certain microsecond; but how does that organization, or indeed a regulator, know that that microsecond is the same as everyone else's? With the increasing prevalence of atomic clocks, which can keep monthly time to within a fraction of a nanosecond, you might assume microsecond accuracy is child's play. But even the most precise clock needs to be initially informed of the correct time, and it then needs to transmit that time to anyone who relies on it. Such signalling itself takes time, of course - the question is how much.

[Leon Lobo](#) is well equipped to answer this. Since 2011 he has been working at the UK's [National Physical Laboratory](#) (NPL), which in 1955 developed the caesium atomic clock, the first clock proven to be more reliable for timekeeping than the duration of the Earth's motion around the Sun. The "ticks" of an atomic clock are the oscillations between two specific energy states in an atom; a feedback loop locks the frequency of a light source to that of these electronic oscillations, thereby creating a stable frequency standard.

The art of the algorithm

These days, even the most insightful human trader comes with a big flaw: sluggishness. In the amount of time it takes a human to observe a change in the market, decide on the best response and execute an order, a computer can have processed millions of financial transactions. It is no wonder that modern finance has replaced many human operators with algorithmic, high-frequency traders.

For obvious reasons, precisely how various proprietary trading algorithms work is kept secret, but they share similar goals. One is to exploit slightly different prices of commodities in different markets, buying in the cheaper market and then immediately selling in the more expensive market. Such arbitrage is as old as market trading itself, except in this case the price differences are small, while the volumes of transactions are huge. A difference of \$0.0001 might not seem like much, but if it can be repeated a million times a second, that equates to \$6000 a minute.

Algorithms get predictive, too - they can be designed to expect a commodity to ultimately revert to a mean price, or to a long-term up/down trend, or to some more complex pattern of activity based on an analysis of historical data. Algorithms can even delve deep into company

performance figures and assets in an attempt to determine a company's real value, and therefore whether the market valuation is over- or under-priced.

In general, algorithms are not designed to make rash decisions - moderate gains made often is the name of the game. But they have come in for a lot of criticism. One is that the necessary computing infrastructure can be very expensive, meaning that the rewards go to the big investment companies that can afford it, rather than to smaller companies and lone traders - even if the latter are shrewder. But perhaps more worrying is that when transactions are processed so quickly, it is impossible for humans to oversee they are all being made prudently. When failures occur, they can escalate at lightning speed.

Keepers of time



Time sync: Leon Lobo wants to provide financial institutions with time-synchronization hardware accurate to the microsecond. (Courtesy: NPL Management Ltd)

Based in the Teddington suburb of London, NPL is responsible for disseminating time to the rest of the UK via fibre-optic, Internet, radio and satellite links. Now, with Lobo as the strategic business development manager for time, it is applying the fine craft of synchronization to financial trading. "You can have your own atomic clock, but a clock in itself is just a stable oscillator - a regular tick," he says. "It doesn't necessarily give you the correct time. That's where our expertise comes in."

Synchronization involves offsetting for time delays, but it is not so simple as dividing the length of the communication route by the speed of the communication. For starters, the communication route may be unreliable. Many of the big financial institutions rely on satellite navigation systems, such as the US's GPS or Russia's GLONASS, to synchronize their internal systems with Coordinated Universal Time (UTC), the global time standard. But these are weak signals that are vulnerable to failure, either by solar activity or deliberate jamming. In January 2016 the GPS signals themselves strayed by 13 microseconds for 12 hours, according to the UK time-distribution

company [Chronos](#), apparently because they were erroneously fed the wrong time from the ground. The failure triggered system errors in companies and organizations the world over.

The Internet is another option for timing information, but here the length of the communication route is not always the same. For example, a signal could travel more or less directly from the source to the receiver one day; another day, it could be redirected via servers on the other side of the world. Even once it reaches a financial organization, the signal has to find its way to individual computers, which are often spread all over a building. Remember, this is a world in which - rumour has it - traders compete to be closer to their building's mainframe, to be sure that their decision-making is at no infinitesimal disadvantage.

What happens inside computers is no more reliable. Every time a signal is processed, it is subject to some delay, which may or may not be consistent. Most computers keep time according to their clock rate - a two gigahertz processor, for example, usually executes two billion instructions per second. But modern software has vagaries. Programs written in the Java programming language have "garbage collection" routines, which periodically reclaim memory but can interrupt timestamping in the process. Meanwhile, according to Lobo, anyone whose computer is running older versions of Microsoft Windows could be whole seconds fast or slow relative to UTC.

„ At the microsecond level, I would say no-one in the City of London has the same time

"You occasionally have incidents called negative deltas, where if I timestamp some data as it leaves and you timestamp it as it arrives, owing to poor synchronization, the arrival time can actually appear to be before the leaving time," he says. "At the microsecond level, I would say no-one in the City of London has the same time."

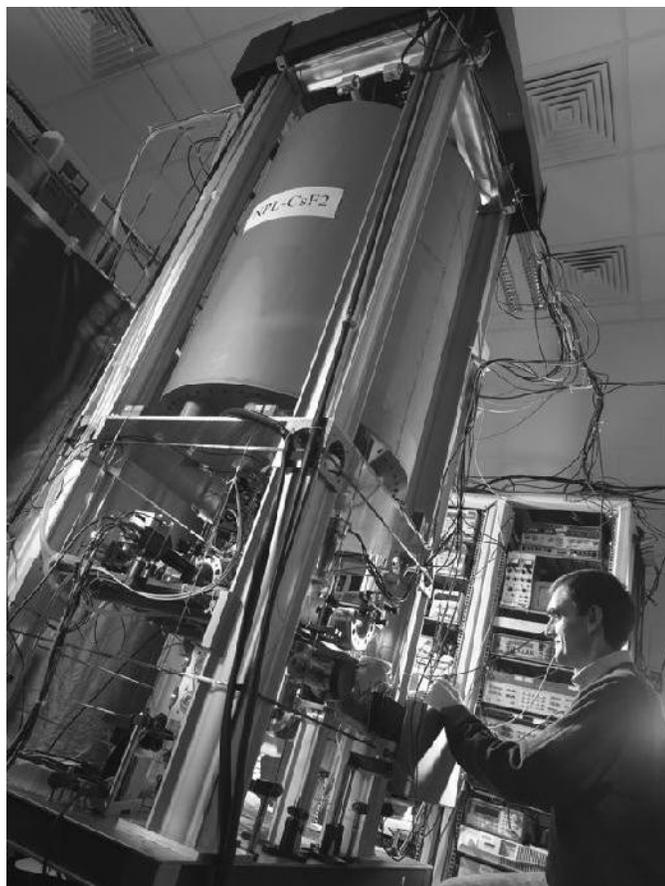
All of which makes the forensics of past market shocks rather tricky. [Neil Horlock](#), a technical architect at the multinational investment bank Credit Suisse, believes ambiguity is the issue. He gives the example of a big investor instructing a broker to offload shares in oil, for no other reason than a new-found concern for the environment. Momentarily before the broker executes this order, another broker also sells his shares in the same oil company. Both sales drive down the market value of the oil company, but the second broker, being ahead of the curve, subsequently re-buys his shares and makes a profit. His timing may have been lucky coincidence. On the other hand, he may have had insider knowledge of the big investor's forthcoming sale - an illegal practice known as "frontrunning".

Dirty tricks

This is a simple example: dirty tricks in modern finance can be much more sophisticated. In any case, as Horlock explains, distinguishing guilt from innocence involves reconstructing a precise chronology of

events, which can be microseconds apart. "Having better clock synchronization means that, if there is an investigation, [the investigators] can ask for all the logs, to demand proof that that the trading was completely above board," he says. "If the trades are uncorrelated, they can see that quite often by the timestamps."

This year, in order to bring more protection to markets, the European Union introduced the second version of its Markets in Financial Instruments Directive (MiFID II), which imposes strict requirements on the accuracy of clock synchronization, in some cases down to 100 microseconds. At NPL, Lobo has developed a solution to help big financial organizations achieve that level of accuracy - and more.



NPL's Krzysztof Szymaniec (right) works on the caesium-fountain frequency standard, which realizes the SI definition of the second. (Courtesy: NPL Management Ltd)

case of a fibre breakage.

Precision and traceability

There are, in fact, two aspects to synchronization. One involves precision, which means choosing hardware that operates as quickly and deterministically as possible. The other aspect is traceability. In

Rigour is key. Delivered with various industry partners, NPL's solution involves sending timing information through dedicated fibre optics - as well as channels shared with select other service providers - of known length and proven resiliency. Every piece of hardware along the way is chosen for its deterministic behaviour, so that NPL engineers can calculate with great precision how much delay will be incurred. At the client side, timing software is eschewed in favour of hardware-based, NPL-certified time-protocol units. "We manage the time end to end," says Lobo.

Every minute, these timing units are "pinged" by NPL to ensure that their time is in sync with its record of UTC. The timing protocol accounts for the basic travel time of the pings, but to account for any unknown latencies, one of NPL's live caesium atomic clocks is transported to the far end of the fibre, where the actual received signal can be calibrated against UTC. At all times, NPL keeps an atomic clock at a hub away from Teddington, to make sure that the timing of the units is correct in

other words, Lobo and his colleagues can guarantee that they know the path their timing signal took, and its associated accuracy, every step along the way. "Most important is confidence in the infrastructure," says Lobo.

Currently, the NPL solution guarantees synchronization to UTC to an accuracy of one microsecond - two orders of magnitude better than that stipulated by MiFID II. As an extra layer of confidence, NPL itself knows the accuracy to yet another order of magnitude (that is, 100 nanoseconds) - Lobo claims they could make it even more accurate, but believes there is no point if the demand is not there. The system has already been supplied to London data centres including Equinix, TeleHouse and Interxion, he says, and foreign organizations are next on the list.

Will a hi-tech solution such as NPL's become the standard

in finance in years to come? Horlock is unsure. "Here's the thing," he says. "While better clocks make our business less risky and better regulated, margins across the industry are smaller than they have ever been, and costs are a major factor in selecting any solution. So long as 'free' services such as GPS are deemed good enough - something that the European regulators are keen to assure us - then more expensive solutions will have to justify their introduction by demonstrating a return on that investment, such as enabling better analytics."

One thing is for certain, Horlock says - big financial institutions will have to mitigate against the risks of GPS if they are going to fall in line with MiFID II standards. But there are alternatives to GPS besides NPL's. Based on technology from the Second World War, eLORAN is a navigation system similar to GPS, using long-wave radio transmitters on the ground rather than up in space. It does not require a fixed communications channel and is cheaper than NPL timekeeping, although it is still ultimately fed by GPS and is not traceable. "Everyone will end up having an alternative to GPS," Horlock concludes. "Leon's is a very high-quality alternative, but it comes at a price. And it can only really be delivered to the major financial centres, whereas a lot of us are moving out."

Lobo counters that an alternative method of timekeeping is of little benefit unless it is calibrated and continuously monitored, to know that it is accurate, stable, resilient and auditable. "The key benefits our solution offers includes end-to-end traceability, audit capability, resiliency, and effectively a trusted time for [a client's] infrastructure, allowing them to measure internal systems against a stable, accurate reference at the ingress point."

It is not the first time anyone has had to argue for an improved time standard. The large clock on the former corn exchange in Bristol, UK, where *Physics World* is published, has two minute-hands: one for "Bristol time", and one for "London time", which in the early 19th century was a little over 10 minutes ahead. That was before Bristol's reluctant adoption of Greenwich Mean Time (GMT) in 1852, five years after its adoption over the rest of Great Britain as a universal time.

GMT made life a lot easier for rail operators of the day, who had until then struggled to coordinate the arrival and departure times of trains in different cities. Lobo believes the NPL system could be just as useful for modern finance. "It's similar to that unification involving GMT," he says, "but at the microsecond level."

- Enjoy the rest of the June 2018 issue of *Physics World* in our digital magazine or via the *Physics World* app for any iOS or Android smartphone or tablet. Membership of the Institute of Physics required

Jon Cartwright is a freelance journalist based in Bristol, UK, jcartwright.co.uk

Copyright© 2018 by IOP Publishing Ltd and individual contributors