☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

Whitepaper: Clock Synchronization – Challenge that needs a New Solution

*By **Kelvin To**, Founder and President of Data Boiler Technologies*

March 15, 2017 is the next milestone target date for the U.S. Securities and Exchange Commission (SEC)'s Consolidated Audit Trail (CAT) project – the world's largest data repository. By then, Self-Regulatory Organizations (SROs) and Broker-dealers (BDs) shall comply with 11.6820 (the Rule) to synchronize their clocks in milliseconds with the atomic clock of National Institute of Standards and Technology (NIST). This is huge because the usefulness of CAT data depends on the ability to sequence trade activities on a play-by-play basis to unveil intelligence to address the flash crash. In fact, this seems impossible to ask all SROs and BDs to pay for hefty upgrades to achieve milliseconds (ms) or faster performance. How to handle this clock synchronization (synch) challenge economically and non-invasively, while ensuring the most precise time-stamped data will be collected for CAT? Read on and see if there may be a new solution.

(1) Rationales to push for the extreme clock synch precision

Clock synch refers to technique(s) used to coordinate a common time notion among processes. In general, the Rule calls for a 50 milliseconds maximum divergence from NIST and 1 millisecond time stamp granularity requirements for electronic trade messages. One may say thousandth of a second (millisecond) for scale of detail present in data (granularity) is insane, because very few stopwatches can measure time faster than hundredth of a second.

Reality is – High Frequency Trading (HFT) firms operate in sub-microsecond speed (faster than millionth of a second). Per an article on Markets Media, entitled When 100 Microseconds is Too Long, it uses 3 HFT orders entered at slightly different times with the same time stamp to illustrate the impossibility to decipher what exactly have happened with these trades. Per another article on Tabb Forum, entitled Across Time and Space, it nicely elaborated the growing concerns by regulators about accuracy and synch of the markets' clocks. Both liquidity makers and takers have increasingly relied on technology for best executions. If one player is 50 milliseconds faster, while another player is 50 milliseconds slower than NIST and/or exchange's server, **a lot can happen (including unfair advantages) in between** the 100 milliseconds total time difference. So people in risk analytics are advocating for the extreme precision in clock synch wherever possible.

(2) Are the CAT clock synch requirements being reasonable?

Play-by-play actions will get blur if one isn't pushing clock synch to the extreme for whatever the fastest speed of any HFTs. However, there are huge costs associated with having market surveillance system running at exactly the same ultra-fast speed of HFT. CAT in itself is already a multi-billion dollars project to build and do the yearly maintenance. It is still a myth on **how to fund the outrages CAT ticket price without overly burdening the industry** (see later section for why I think "crisis prevention" won't be a good justification of CAT costs). The ultimate CAT costs may end-up transferred to and bore by end-investors.

Sadly, the CAT billion dollars' budget does not include firms wanting to have NTP (Network Time Protocol) servers upgrade to GPS (Global Positioning System) time source for better clock accuracy. So, for non-HFT market participants

whom currently do not have GPS time capabilities, it is understandable that they prefer to stick with their basic NTP so long as it can meet regulatory requirements. Can NTP satisfy the 50 milliseconds maximum divergence from NIST? Allow me to draw your attention to this report – International Comparison of NTP Servers. This report examines the time transmitted by NTP servers located in North, Central, and South America. Per figure 3 in the report, servers located at Bolivia had "common-view" time differences over 50 milliseconds (in red). In particular, IBMET and INM-1 servers perform the worst (blue columns) when doing a direct comparison with NIST using the "average" method per figure 4 (-34.2ms and +37.2ms respectively). Other than those, **most NTP servers' clocks seem fine synching with NIST,** unless one really gets technical about the occasionally out-of-synch causes (to be discussed in later section).

In short, **the 50 milliseconds maximum divergence is not unattainable. It is already more lenient than Europe** – MiFID II RTS 25 Article II requires "maximum divergence from UTC to be either 1 millisecond (gateway-to-gateway latency of > 1ms) or 100 microseconds (gateway-to-gateway latency =< 1ms). The granularity of the time stamps can be either 1 millisecond or 1 microsecond, also depending on the gateway-to-gateway latency time".

Note 1: per 11.6860(a)(2), "if the Industry Member's order handling or execution systems **utilize time stamps in increments finer than milliseconds**, then they should record and report data to CAT with time stamps in such finer increment". Take for example, NASDAQ upgrades the Securities Information Processor (SIP) from 225 milliseconds a decade ago to 500 microseconds (µs) today, then soon streaming the speed to 50µs, then their CAT time stamps should improve from 500µs to 50µs accordingly. Time stamp granularity requirement **should go hand-in-hand with how fast a market participant is allowed to conduct their HFT activities**, and how fast exchanges are matching up/ processing trade orders. I applaud the SEC for not letting those sub-millisecond SROs/ BDs off the hook with the generic 50 millisecond standard.

Note 2: per 11.6895(b)(2), it provides a later compliance commencement date **on or before Feb 19, 2018** for industry member whose **business clocks do not capture time in millisecond**". So there's still about a year if any server upgrades/ migration need to be done.

The only trouble I have is regarding **manual orders**. The Rule does separately crafted out those manual order events (i.e. non-electronic communication of order-related information) will have times tamp granularity requirement of 1 second. MiFID II ESMA in Europe also has a 1 second requirement for voice trading systems that do not allow algorithms. **This 1 second may be artificially represented** because it probably takes more than a second to measure and record time manually. Should 1 minute or 1 second be the appropriate time stamp granularity for "manual order"? It depends if it may introduce too much noise and/or overly distorted signals for market surveillance and manipulation detection purposes. Bottom line is: **any time stamp granularity requirements must be reasonably enforceable, or else the standard would be meaningless** (see next section for an elaborated discussion).

(3) Out-of-synch causes and how strict the enforcement should be?

If anyone remembers the Y2K problem during the millennium, then this out-of-synch problem should be treated with equal respect. Given rotation of the earth isn't exactly 24 hours a day (thus scientists introduce a leap second to compensate for the moon's tidal effects and other causes), NIST or UTC need occasional adjustments in order to keep its time of day close to the mean solar time. **Should firms require adjustments to their business clocks for leap**

seconds cumulative differences, I am not aware of any guidance by the SEC. If the SEC said yes, it would mean all SROs and BDs have to upgrade NTP servers to GPS (Global Positioning System) time source, which can be expensive and invasive to the industry.

Asides from the leap second issue, let's look at other out-of-synch causes. A network consists of primary, secondary servers and clients. Then there are NTP message packets traveling in between through multiple communication routing nodes. An inconsistent routing of NTP packets would result in network asymmetry. According to CISCO, "asymmetry routing is not a problem in itself, but will cause problem with Network Address Translation (NAT) and firewalls are used in the routed path". Given the capital markets commonly use sophisticated routing algorithms to control the lag (e.g. IEX 350 microseconds delay / speed bump) of the packet at the router and the next leg of the route, the delay from server-to-client can differ significantly from the delay from client-to-server. Such non-reciprocity of the paths of the NTP packets can cause significant time errors. I would **assume any analysts at the SEC would understand all these routing nuances proliferated by Reg NMS** (good luck with that).

Known routing nuances are relatively easy to address, but **what about the unknown traffic congestion, Ethernet links availability during severe conditions** (e.g. storm and/or power outage), etc. The American is constantly under cyber-attacks and the impact could significantly affect network traffics. Also, during a business continuity / disaster recovery event (BCP/ DR), don't expect the backup servers will have the same high performance as those Tier 1 servers that firms normally use for order handling or execution. BCP/ DR are only meant to provide limited capacity for a particular duration, but almost never be a 100% replacement of the core for indefinite term (due to **costs vs benefits concern**). Besides, we have seen the exchange abruptly shut down due to glitches without falling back to disaster recovery (see this). Therefore, it'll be hard for the SEC / plan processor to strictly enforce clock synch requirements under such circumstances. Nevertheless, the 1 second tolerance limit for "manual order" (see earlier section) is more a **"spirt of the law" rather than an enforceable standard**. Again, it probably takes more than a second to measure and record time manually, so deal with this 1 second on the best effort basis and I think you'll be fine.

Regarding the SROs, I think the U.S. will or is already adopting the same requirements as the European counterpart – ESMA. To quote from RTS 25 Article 4, ESMA demands "**operators of trading venues shall establish a system of traceability** to [NIST] and define that it is enough to deliver proof by documenting the system design and specs. It is also required to identify the exact point at which a timestamp is applied. Finally, the article specifies that **reviews** of the compliance of the [NIST] traceability system have to be conducted **annually**".

(4) What it's like to achieve sub-microsecond performance?

Millisecond is too slow a speed for HFT. Take Thesys Technologies (the winner of CAT bid and an affiliate of HFT firm Tradeworx) for example. Backed in 2012, Thesys was working with Spectracom to "improve quality of their time synch, as the latency of hardware and trading infrastructure has improved from milliseconds to microseconds." As a result, Thesys servers that initially used NTP have upgraded to GPS time source. The quality of their lock to GPS is consistently within 100 nanoseconds (i.e. 100 billionth of a second), that's the clock precision that HFTs need to go after every tick!

One can watch the documentary film – The Wall Street Code to get some sense about the competitiveness of these HFTs in pursuit of speed. There can be continuous debates over the controversial topics about liquidity impacts of HFT,

but allow me to draw an **analogy from plane design to explain what it's like to be ultra-fast**. Back in the days when plane was first design to fulfill the human dream of flying. Plane designers worked tenaciously to improve the aerodynamics in making lighter and faster planes with honorable intents. However would the designers ever know their well-designed planes would turn into fighter jets during the World War II? **The speed isn't a problem but those abusers with evil thoughts are**. You may ask if there may be unfair advantages in HFT, then I would say: how others would compete with the institutional investors when they are no match in size, and if they can't even use speed?

Given the above analogy, I would argue that the **merits of HFT in a market depend on its "character"**. Per FCA paper #16, it suggests: (1) there is "no evidence in our sample that HFTs can 'see the true market' and trade in front of other participants at a millisecond frequency"; (2) "Do find patterns consistent with HFTs being able to anticipate the order flow over longer time periods (seconds and tens of seconds)". The summary remarks further said "the implications of anticipatory behavior may or may not be detrimental depending on very specific characteristics of the order flow". This is in line with what I said – the **tech in itself is neutral**, it all depends if someone is trying to abuse such power.

In my opinion, Thesys whom built the MIDAS system for SEC is a creditable act in use of technology. They appear to have the best sense about market structure and flash crashes. I think they are better than the 2 other final bidders in terms of embedding an analytical framework in the design of CAT, so I congrats them for being selected as the plan processor to build and operate the CAT. Whereas for other HFT firms, I have lots of puzzling questions: **Is there really no good or bad guys**, but only traders jockeying themselves around the market trying to make money? How would a passive market maker that don't do statistical arbitrage and don't carry inventory, be able to **use speed and market structure advantage to gain edges** in almost every asset class (see this)? Can someone use HFT to do manipulative "just-in-time" process (i.e. sell first and then buy back within lightning speed)? Would the research on quantum computing by Goldman Sachs, RBS, CME, and Guggenheim be an effective solution to cope with flash crashes and related consequences? **No-one knows at this point if any well intent tech development may inadvertently become threats** to financial stability.

Prior to the approval of CAT, I have said "the CAT is bad" because I despise locking valuable data in a centralized vault and it lacks real-time scrutiny of the markets. Regulatory access of CAT data in **T+5 days simply won't be fast enough to catch rogue traders whom operate in nanoseconds**. Should regulators slow down the industry in pursuit of speed? This is like asking why the U.S. doesn't have freeways with no speed limit (i.e. a moot point). Preventing the next financial crisis isn't an appropriate justification for overly invasive regulatory requirements that put an undue burden on everyone. Over suppressing the activities of HFTs may demolish the openness of our capital market. In order to keep the market vitalize, one should be creative in seeking new solution.

(5) Is it necessary to perfect the exactitude of clock synch?

I have heard complaints about "technology only benefits the rich people" (see this), but there is also a wave of rebellious movements to liberalize the super-computer market at a much cheaper costs (see my earlier thread). The trick is: identifying invalid assumption(s) and experimenting with non-conventional thoughts to discover new solutions.

It is obviously ultra-expensive for HFTs to constantly compete for the fastest speed in the market. Co-location and microwave towers for ultra-low latency transmission are out of reach for most firms other than those elite HFTs. I am

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

not commenting whether everyone should jump into the speed race, but it is not wrong for the SEC and other regulators around the world to encourage SROs and BDs to up their risk management capabilities. **Risk and compliance need to match up with the super-fast trading algorithms** by equipping themselves with much faster surveillance tools. Would the regulator be requiring all SROs and BDs to upgrade their systems to synch with a GPS time source some day? I think that's probable but not a specific requirement in this current moment.

Given we live in a less than ideal world and not all firms have GPS time source, we had to make the best out of this difficult situation. Like BCP/ DR that we had to think about the recovery objectives under limited budget, we need to consider the **analytic and control objectives** for CAT in here. In terms of sequencing CAT reportable events, it's all about "onset detection". In other words, the **likelihood of irregular activities detect within a time interval**. The troubles with onset detection under the prevailing practices are: (1) there bound to be a lot of false positives/ negatives (i.e. errors) by way of determining the shape/ trend of a pattern – the inaccuracy of technical analysis/ time series forecast; (2) perfecting the exactitude of clock synch and measuring vectors graphically is a resources-draining exercise.

No matter how one try to accelerate the GPU and/or boosting the computing power, **the synching and vector measurement processes are highly complex and time consuming**. As a result, risk management practices in the capital markets are mostly done in post-trade instead of pre-trade. Since measuring vectors graphically is such an erroneous process, **can we stop doing it**? If we can **find a much faster way to do onset detection**, would it enable preventive risk controls to be done in real-time rather than after-the-fact investigations? If such new alternate process **wouldn't introduce too much noise and/or overly distorted signal** for market surveillance and manipulation detection purposes, does that mean there is no necessity to perfect the exactitude of clock synch? These are no hypothetical questions, but smart engineers approaching a financial engineering problem with an engineering mind set – i.e. balancing the processing speed with fulfillment of the analytic and control objectives.

In our case, we found a **new solution that inspired by concepts of music plagiarism**. It is part of our patent pending invention to sequence trades in musical formats. Not only does the compressed musical format save CAT storage space, it **eliminates the costs and troubles to perfect the exactitude of clock synch**. Imagine someone's singing is a bit out of tune, out of beat and rhythm. You may still be able to recognize what song he or she may be singing. That's the beauty of audio pattern recognition. It is **much simpler than visual recognition techniques, much faster and effective to pick up signals** and filter out noises. It preserves the big picture of cross asset classes and venues analysis. This is an effective way to **curb the rogues from using synthetic created trades to circumvent the controls**. Sadly CAT doesn't include Futures data (jurisdiction under the CFTC instead of the SEC), or else multiple instruments can be streamed into a symphony orchestra for comprehensive analysis. It is the fastest method for active surveillance in the capital markets. I envisage that it'll **transform the entire capital market risk management practices to be more agile**!

*** END ***