sciendo
Financial Internet Quarterly „e-Finanse" 2018, vol. 14/ no. 2, s. 34-46
DOI: 10.2478/fiqf-2018-0010

eFinanse
financial internet quarterly

REGULATIONS AND TECHNOLOGY BEHIND HFT LATENCY, BATCH AUCTIONS AND PAYMENTS FOR ORDER FLOW IN THE US AND EU

Carlos Jorge Lenczewski Martins[1]

Abstract

Since the appearance of high-frequency trading in the 1990s, speed has become one of the key issues in trading and with it, the controversy around High-Frequency Trading. In recent years, there have been many discussions and analyses of how high-frequency trading may affect the financial market – but still without any clear conclusions. Leaving these opinions behind, many adjustments have already been made in the US and Europe - both to regulations and market rules, impacting not only High-Frequency Trading but general electronic trading as well. These rules and regulations are the result of technological developments in electronic trading and more specifically, High-Frequency Trading and the practice of Payments for Order Flow. The question remains as to how deep regulations should go, especially in the case of HFT which can be severely affected by harsh regulatory requirements or procedures. Because two of the most important issues in HFT are time and information, some of the rules and regulations affect aspects such as not only what type of information and how it should be gathered, but also clock synchronisation and time-stamp granularity. Another issue that may be considered controversial in the field of HFT (although it is not a practice limited to HFT) is Payment For Order Flows. Under this mechanism, wholesale market makers pay brokers for their client's order flow – a practice that performed in great amounts and at high speeds may give a considerable level of "inside" information. Regulations, especially from ESMA (MiFID II). try in great part to thus mitigate the practice of Payments For Order Flows.

The aim of this paper is to present technological advancements in the field of trading communications used, not only by HFT firms, but also by exchanges. Additionally, the objective is to underline some challenges regarding regulatory changes that try to adapt to the current level of technology – for example, those related to clock synchronisation and data processing. One last issue brought forward is the possibility of converting markets from continuous-time auctions to discrete-time auctions - a concept that is aimed at liquidating the speed advantage and competition only to price level and hence, eliminating HFT advantages.

JEL classification: D47, D53, G18
Keywords: HFT; MiFID II; batch auctions; granularity; clock synchronisation

Received: 18.04.2017 Accepted: 21.03.2018

1 Institute of Banking and Finance, Finance and Management Collegium, Warsaw School of Economics, Warsaw, Poland, e-mail: clencz@sgh.waw.pl. ORCID https://orcid.org/0000-0003-4595-7976.

Introduction

There has been much controversy around High-Frequency Trading (HFT), ever since its rise in popularity due to technological breakthroughs in the 2000s. This type of trader relies on information and speed to build and implement strategies.[1] Due to the speed involved, there is some confusion on how these traders are classified, sometimes as "*cheetah traders*" - as for example in (Lenzner, 2011), i.e. differentiating them only by speed. The fact is that the *choice* of strategies, and *how* they are implemented, is a much different process than that used by traditional (human) traders. Together with high volumes, they may have different impacts on prices. HFT will not necessarily rely entirely on macroeconomic factors as human or medium/long-term traders do but will rather heavily rely on microstructure elements. Because these microstructure factors are extremely short-termed, any decisions made by HFT traders may generate high volatility in prices – something that *can* but, *does not* necessarily happen. Also, because of the very short half-life of trades under HFT and the potential short-term volatility that is sometimes seen, this type of trader may often be seen as having a negative impact. An example could be the 6 October 2016 sterling flash event that initially was *thought* to be HFT initiated (Ismail & Mnyanda, 2016), (Davies, 2016) but, as the BIS report from January 2017 stated, it was rather that HFT traders were trading to close-out their positions *after* this flash event occurred (BIS, 2017). But the speed of trading shouldn't be of any surprise and also shouldn't be considered harmful. Ever since the NASDAQ[2] and NYSE[3] allowed electronic quotations and electronic order submission, trade has evolved substantially[4]. The use of HFT should be seen as the natural evolution of trading and, one should rather expect the tendency of trading automatization to continue. Controversial practices such as predatory trading or even sporadic illegal practices (sometimes extremely difficult to detect due to trading speeds in HFT) which are excessively highlighted, should not be treated as a general practice of HFT and thus having, as a general rule, negative market impact. Additionally, there is also market manipulation under traditional trading, and that doesn't mean that overall, human trading is "bad" or negative for the markets. HFT should and must be considered a different type of trading – not a typical trader simply using typical strategies implemented faster.

Although poorly defined, HFT can be considered a specific type of algorithmic trading with no human intervention, relying on quick order submission, in time frames of microseconds, or even nanoseconds. In order to achieve these speeds, firms using HFT will resort to very complex and sophisticated computer algorithms that analyse many different types of data (not only macroeconomic data) giving them profit opportunities. Although orders may be submitted in a matter of microseconds, they don't necessarily depend on all data being obtained and analysed at that same time simply because there are many different strategies, some involving the most recent information (e.g. order-book imbalances) while others may rely on "older" information (e.g. statistical arbitrage). Nevertheless, there is always an element of speed involved around the information obtained and the need for fast order submission, something requiring state of the art technology. But another element of trading that is important, apart from powerful computers and programs, is the speed of connection between venues. An advantage, for instance, of 3 milliseconds, for humans may be seem difficult to understand or insignificant while for HFT it is the matter of many trades being performed. This also brings out another feature of HFT: marginal profits. First, due to the high order submission speed involved it is rather difficult for nominal prices to change as much as in normal circumstances, requiring therefore a large amount of orders being submitted. Also, not every order will turn into a transaction, which is directly related to the strategies used by HFT traders. Since the half-life of opened positions is minimal, the marginal profit is also very small. In order to achieve significant total profits, HFT traders need to turn over a considerable amount of capital. In fact, a High-Frequency market maker may turn over its capital even more than five times a day (Easley et al., 2011) generating significant volumes in one or more assets. For HFT traders, speed matters not only in order to quickly submit orders and make trades, but also

1 Although, not a rule, some of these strategies (sniping for example) are considered predatory because they either try to figure out the intentions of other traders or they lead other traders to make certain decisions they wouldn't otherwise make.

2 NASDAQ stands for National Association of Securities Dealers Automatic Quotation System. NASDAQ is the first electronic stock market listing (exchange) and, is divided into two markets – one with large company securities (Nasdaq National Market) and the other with emerging companies (Nasdaq Smallcap Market). The NASDAQ was created in 1971 in order to allow for computerized trading and, therefore faster trading

3 NYSE stands for New York Stock Exchange, which dates back to the 18th century and is the largest exchange in the world as to market capitalization with a value of close to 20 trillion USD.

4 The main difference between both exchanges is that NASDAQ is a dealer exchange meaning that trading between investors is made through dealers and not directly between them. The NYSE is an auction market meaning that participants make offers to buy and sell directly between themselves.

for obtaining information even faster than other traders, independently of their HFT status. Sometimes obtaining the information faster may probably even be one of the most important elements in HFT simply because, while HFT traders don't necessarily execute trades, they may repeatedly modify or cancel a significant amount of orders in very short time frames.

The time frames that are currently used in the HFT environment are in the range of microseconds or even nanoseconds. Because this time frame has gradually become smaller and smaller, the increased use of an HFT business model and its technology may eventually lead to a decrease in the time-factor advantage, but as (Easley et al., 2012) mentions: "even if the speed advantage disappears, HFT will evolve to continue to exploit Low Frequency Trading's (LFT) structural weaknesses". Every technological breakthrough gives HFT a chance to become faster and make transactions faster than other traders. This brings many challenges not only to trading firms (and their business models), but also to regulatory issues - some of which presented here.

The need for speed - Technological evolution of HFT trading

High-Frequency Trading is focused on using sophisticated computers and programs that provide the construction of complex algorithms, which in turn submit orders in a matter of microseconds - independently if any trades are actually made or not. Time frames have not always been so small as they currently are, and probably in a few years from now the current time frames will already be too large. It is only a question of hardware and software technology and of course, communications infrastructure. The types of technology available to financial communications are very different, varying in reliability, ease of use, and most importantly, cost and speed of transmission.

Probably the two most widely used technologies for data transmission are microwaves and fibre optic cables. Microwave technology, although being the most popular of HFT connections, is not recent – it has been around for about 70 years. Already in 1949 microwave technology was used to connect two cities in the US - New York and Chicago (Anthony, 2016). Since then, many cities in the US and Europe have been connected through microwave networks. Another option arising in the 20th century was the use of fibre optic cables. At the time, what seemed to limit the future of microwave transmission was a speed that was unreliable, and which could be dropped down for many reasons - the most obvious one being atmospheric conditions, something that would require the increase in the power of the emitters. Fibre optic cables seemed more reliable, and the speed would always be the same. But while the world of data transmission is yet a bit far from reaching the speed of light, microwave networks are still a very reasonable solution for at least three reasons. In the first place, information travels more slowly through a cable than in air even though more information can be carried through fibre optic cables. Second, fibre optic cables are very expensive to implement if the appropriate infrastructure isn't already in place; and in the third place, from a logistical point of view, microwaves seem to be a more effective solution than having cables running everywhere (although many firms already do use them for connections in backup systems). Another important issue is that while two (or more) companies/venues may be connected to each other, when using fibre optics it is rare that these firms will be connected directly - most often there will be certain routers in between them - something that definitely slows the transmission. As estimates presented by (Anthony, 2016) show, the speed, or rather latency, between London and Frankfurt by using fibre optic cables and through many different routers, might be close to 17 milliseconds. If, on the other hand, fibre optic cables were replaced by microwaves directly between both these venues, latency is reduced to around 4.2 milliseconds. This is a significant reduction in time as far as HFT is concerned. Cost-wise, microwave technology transmission requires masts every certain number of kilometres or miles - which depends on the power of the transceiver. Each mast is estimated to cost between 120,000 USD and 250,000 USD while each transceiver between 12,000 USD and 25,000 USD. If in a distance of 400 miles about 20 masts are required, the costs (of only microwave network materials) would be between 3 mln USD and 6.2 mln USD. In the end, total costs may be estimated to be between 10 and 20 mln Euros (Anthony, 2016). Fibre optic networks are therefore cheaper if the infrastructure through which the cables will eventually run already exists. The space for a fibre optical cable may be leased for 1.3 USD per metre (per year) plus the price between 0.5 USD and 1.7 USD for the actual cable (Anthony, 2016). A summary of such a comparison in shown in Table 1.

Table 1: Comparison of fibre optic and microwave technology in the London-Frankfurt distance example

	Fibre optic technology	Microwave Technology
Latency (the less the better)	17 ms	4.2 ms
Mast cost	-	120-250k USD
Transceiver costs	-	12-25k USD
Cable lease per metre/year	1.3 USD	-
Cable price per metre	0.5-1.7 USD	-
General infrastructure costs	1.1-1.8mln USD	3-6.2mln USD
Estimated total costs	2-5mln USD*	10-20 mln USD

*Source: Based on data from (Anthony, 2016); * own estimation*

As mentioned, there may be a place for a significant price (cost) improvement - but only when the infrastructure exists. Sometimes, the actual benefit (in terms of speed) may be significant, compensating the very large costs. An example is the fibre optic cable network which was planned to connect London and Tokyo even though costing an estimated 1.5 bln USD, it would bring speed up (or latency down) by 60 ms (Anthony, 2012).

While currently most of the literature acknowledges time frames to be in the range of milliseconds or microseconds, the reality is that HFT traders and also stock exchanges are pushing the limits even further. In 2014, a new connection between NYSE and NASDAQ became available, not relying on fibre optics or microwave technology, but on laser technology. The aim of this network was to reduce latency when compared to a microwave network by 0.18 ms (Durden, 2015). This is for a human trader quite irrelevant but again, crucial for an HFT trader. Another example is from 2015 when the BSE stock exchange was planning to increase the speed of trade in its platform by 20 microseconds from the 200 microseconds at the time (Times, 2015). But this is by now still slower than the current possibilities traders have because technology already allows for a full trade (order sent, processed and information sent back) to be performed within 85 nanoseconds (Sprothen, 2016). When comparing through the years, trading speed has substantially increased (decreased latency). For example, already from 2005 to 2011 the trading speed increased from 97 ms, to 7 ms (Sprothen, 2016), in 2015 it was already less than 1 ms, and in 2016 less than 1 μs as shown in Table 2.

Even if HFT companies don't have the available technology to trade at speeds of 85 nanoseconds, stock exchanges still have a processing time which is higher than the speed at which HFT traders work. Nevertheless, exchanges are always trying to upgrade the technology used - and speed for that matter, in order to adapt to the necessities of market participants - which they wouldn't do if speed was somehow harmful for the markets. But this gap in technology is not something new, and probably will always exist since technology is updated more quickly by commercial trading firms than by stock exchanges. Be that as it may, this gap brings advantages to some participants. A study performed by Nanex LLC on quote delays in the NASDAQ has shown that in the US, the time gap in the quotation mechanism provides HFT traders with a substantial advantage of 500 microseconds over other participants being the result of quotations from different venues being sent to what is called - SIP (McKenna, 2015) described below.

Latency and the Securities Information Processor

United States law prevents each venue from streaming quotes directly, *before* sending them to an SIP or Securities Information Processor. This is a live-streamed central aggregator of every exchange's best bids and ask quotes. Every exchange is required to send all quotations to the SIP, which gathers and processes them into one consolidated data form. This is particularly important

Table 2: Trading speed changes through the years (values in milliseconds)

2005-2010	2011	2015	2016
97 ms	7 ms	< 1ms	< 0.001 ms

Source: Based on data from (Sprothen, 2016) in the United States since it is from this aggregated information that the National Best Bid and Offer (NBBO) is calculated and widely used - not only by traders but also by regulators.

Because every exchange needs to send quotations to the SIP[5] *as fast or faster* than it sends to its subscribers, Nanex LLC has found that due to this process, HFT traders have an advantage of more than 500 μs over companies that do not use direct feeds from the exchanges to quote prices (McKenna, 2015). This is due to the time that is needed for quotes to be processed and then distributed in the SIP. This doesn't happen, on the other hand, if companies get a direct feed from the exchanges. The head of Nanex LLC, Eric Hunsader, states that this delay of 500 μs makes the HFT business model [of latency arbitrage] essentially riskless (McKenna, 2015).

Slowing down the time-clock – the introduction of batch auctions

The time-clock advantages that are seen for HFT traders have raised some questions as to whether these participants should actually benefit from such time advantages as seen with the NASDAQ (or even other venues). Slowing down the trading of HFT firms at a national scale, where HFT alone accounts for more than 50% of the turnover in US equity market (BIS, 2011), may prove itself to be difficult - not to mention the possibility of such measures to even induce some risk into the market because HFT are also very important liquidity providers.

Leaving aside any potential barriers and risks, at least two actions have already been taken to make it more difficult for HFT firms to achieve profits over other traders, just on a time or speed basis. One example of such measures is the newest exchange venue in the US - the Investor's EXchange (IEX). This national securities exchange uses a 350 μs speed "bump" (delay), in order to slow HFT orders, aiming for all trades to arrive, more or less, at the same time (Lam, 2016). This is achieved by using 38 miles of coiled fibre optic cable. Although this is not a significant exchange - with only 1.6% of all trades on the US exchanges, it could be seen as an interesting venue for investors looking for "stable" long-term investments. Another example is the proposal of implementing *batch auctions*, something that has been put forward in the past few years in academic research. Probably one of the most popular papers on batch auctions is the one by Budish et al., where a theoretic batch auction market model is proposed (Budish et al., 2015). What is innovative about this type of trading is that there is no actual slowing down of trading but instead, trading runs under defined time intervals and prices are quoted at the end of the period.

Under typical trading practices, participants make buy and sell offers in a continuous form, acting as an auction. Prices are therefore automatically updated, as soon as the buy and sell offers are matched. With state of the art technology, participants have access to the buy and sell offers more quickly than other participants and therefore, get a picture of the market (of the prices) faster than other traders. The idea behind batch auctions is to match best Bid and Ask orders, just as in a continuous limit book but during predefined periods (intervals) instead of these orders being matched continuously. Because under batch auctions, (order) matching time is being changed from a continuous time into a discrete time, this would mean that prices are calculated at the end of each of these intervals. Traders would submit, modify or withdraw any orders during the matching intervals, but if orders are not filled, they would be carried to the next interval. With the introduction of such intervals, participants are simply having a "fresh start" in the bidding process.

Because of the speed advantages of HFT, which may also lead to some predatory strategies targeted at specific traders, what batch auctions aim to achieve is to eliminate those speed advantages. Another reason for introducing batch auctions is to bring out competition at the price level instead of focusing at the speed level. It is worth mentioning that batch auctions also eliminate the need of complex audit trails since prices are dependent on order matching during the occurring intervals. Assuming batch auctions are feasible, the optimal length of these intervals is what may be of some difficulty to establish. One paper by Fricke, D. and Gerig, A. estimate the optimal time interval for batch auctions (for a typical US stock) to be between 0.2 and 0.9 seconds (Fricke & Gerig, 2016).

Batch auctions are a reality and not just a theoretical idea as in (Budish et al., 2015), from the moment some exchanges introduced batch auctions into the trading calendar. The London Stock Exchange for example, on March 21, 2016 introduced a daily 2-minute batch auction on the FTSE 100, run at 12:00. The Chicago Stock Exchange

5 Since the late 1970s the body that oversees the dissemination of real-time quote information is the CTA or Consolidated Tape Association. As such, the CTA is the entity that oversees the functioning of SIP (NYSE, 2017).

(CHX) also introduced a batch auction system June 3, 2016, by the name CHX SNAP (CHX, 2016), and these batch auctions may be initiated either by the exchange or by a participant. Apart from specific requirements regarding size, time or price (details available at (CHX, 2016)) what is important about this batch auction (and others) is that the exchange will not disclose any information but the symbol of the instrument subject of the auction. As such, this mechanism functions as a typical dark pool where participants are not disclosed on price or order size but only the instrument traded. Interestingly enough, for a long time dark pools were criticised by media and regulators on the lack of transparency or the increase in market volatility that they may lead to (Robinson, 2009), (Von Hoffman, 2014), (McCrank, 2014). Now, they are introduced into those same markets in a slightly different form.

Regulatory changes and challenges related to clock-time

The increasing activity of HFT traders has led regulators to introduce mechanisms that better control transactions for algorithmic trading, providing a much higher level of accuracy. Two important market regulators that have made significant changes in trading clock accuracy and granularity is the US (SEC) and the European Union (ESMA). Regulations from both these bodies either have already been introduced or are being introduced and will soon become effective. Even so, many of the regulations are already required by the SEC while, the ESMA regulations (MiFID II) regarding granularity and trading accuracy were effective a bit later, as of January 1st, 2018. Apart from the importance and consequences of specific regulations that try to be up-to-date with the constant changes occurring in the financial markets, the most accurate ones seem to be from ESMA. Probably because ESMA regulations were introduced later than those by the SEC, they have been able to capture the latest practices of the market introducing for example, a time accuracy that goes up to μs while the SEC requirements go up to ms.

Both regulations are different but, the aim is similar. They rely on the introduction of requirements for tracking *orders*, and not necessarily *transactions* as previous regulations did. MiFID II requires, as mentioned previously, time granularity to change from ms to μs, depending on a company's trading activity - if it is considered an HFT trading company, or not. Article 4 of the MiFID II Directive defines a firm engaging in a High-Frequency algorithmic trading, as those firms that (ESMA, Nr 2014/65/EU):

1) use "infrastructure to minimise […] latencies [… such as:] co-location, proximity hosting or high-speed direct electronic access",
2) have no "human intervention [… in either] order initiation, generation, routing or execution […]",
3) have "high message intraday rates which constitute orders, quotes or cancellations".

Also, firms that are classified as having HFT trading activity are required, under MiFID II, apart from the granularity of 1 μs (or less), a maximum divergence from UTC time of 100 μs - where, for firms with human intervention the requirement is 1s.

As for which SEC regulations are concerned, a major rule introduced in 2012[6], is the Rule 613: "Consolidated Audit Trail - National Market System (NMS) plan" (SEC, 2017). This rule introduces comprehensive changes in order to create, implement and maintain a consolidated audit trail. The goal here is for regulators to be able to reconstruct trading activities, and to assure monitoring activities, something that helps detect such manipulation activities as spoofing and layering. Some of the most significant changes this rule bring, may be (Levey, 2013):

1) all types of orders must be submitted (recorded) - this includes not only executed orders but also modifications and cancellations. This enhances a regulator's ability to link every order submitted with every company and across exchanges,
2) requirement for national securities exchanges and financial industry regulatory authorities to submit time-stamps with an accuracy of 1 ms,
3) brokers and dealers must synchronise their clocks to the National Institute of Standards and Technology.

Due to the new reporting requirements, the 613 rule lead to require some 2000 firms and 19 SROs (Self-Regulatory Organizations) to introduce significant changes in operations[7], (catnmsplan, 2014). Such changes affect the front, middle and back office operations, IT

6 Although introduced in 2012, some aspects of this rule were supposed to be gradually introduced - even in 2017 and 2018.

7 SRO - Self-Regulatory Organizations are entities that individually adjust market trading rules (regulations), independently but in line with general national laws and supervision bodies. An example of such an SRO is the NYSE and other exchanges in the US. Although the SEC (Securities and Exchange Commission) is the financial market supervision authority, exchanges have the ability to introduce not only trading rules but also, for example, audit regulations related to trading and its participants.

infrastructure data retention repositories, or may even influence client information.

Apart from data gathering and reporting requirement changes, the SEC also introduced significant adjustments in clock synchronisation. In August 2016, the SEC approved a rule concerning synchronisation (which was an original proposal from FINRA) to reduce, for some firms, clock synchronisation from 1 second to 50 milliseconds of the NIST (National Institute of Standards and Technology) atomic clock (FINRA, 2016). This rule introduces a synchronisation standard of 50 milliseconds for firms that are involved in recording events in NMS (National Market System) and OTC equity securities. Entities required to report in a time span of milliseconds were required to do so by February 20, 2017. As for those entities that didn't fall under the milliseconds reporting requirement, the deadline was February 19, 2018 (Daly, 2016). What this means for those companies that don't fall under the 50 ms "second rule" is that they are allowed a synchronisation margin of 1 second from a source that is linked to the NIST atomic clock.

All the above-mentioned changes are challenging, not only from the reporting requirements point of view, or even data repositories requirements, but also from a time point of view. In general, a private individual using a personal computer probably doesn't even think about the source of the time presented. It is not even really that important for time to present some degree of inaccuracy - simply because people don't work in such fractions of time as milliseconds. But for financial companies, the necessity for time accuracy is completely different - especially with the new requirements. As prosaic as it may seem, time synchronisation with an atomic clock may not be that simple and this is related to technological issues. Entities, either under the SEC or ESMA regulations, are required to maintain a synchronisation margin from the UTC time standard. The problem is basically related with the source by which computers are synchronised. There are many different channels for synchronising time: either from GPS satellites, over the internet or from a government time service such as the NIST (Skoog, 2017) or the National Physical Laboratory (NPL). As to using internet time providers, most often it is used the Network Time Protocol (NTP). This is a protocol on which most often Distributed Denial-Of-Service (DDoS) attacks occur so it may not be the best solution for financial institutions. When using GPS satellites as a time source, companies will usually have a dedicated server which will later "distribute" the precise time over the entire network of servers and computers. This requires adequate software and hardware. GPS solutions may be very accurate but still are vulnerable to signal disruptions (independent of their being malicious or accidental) (Skoog, 2017). The reasons for fluctuations of time or their sources for that matter, under SEC or ESMA regulations don't really matter, because financial institutions are required to monitor and maintain an accurate time, by whatever means it takes. The best solution would therefore rely on having one main provider, and a secondary provider such as the national atomic clock providers. Currently it is easier than before to have the required level of accuracy because the UK's NPL offers atomic clock precision using two methods: by telephone (Telephone Time Service) and the second, by using the Network Time Protocol (NTP) through which firms may be linked. Basically, the client wishing to obtain UTC time runs a program software which sends periodic time requests across the internet to one or more specified NPL servers and then receives time-stamped data allowing for clock adjustments.

Taking into account that the most popular mechanism a financial institution uses to synchronise with atomic clocks is through the internet (ie. NTP), in the end, what is important is to assure some sort of server security. Skoog presents eight security features that modern NTP servers should include (Skoog, 2017):

Although, all these time-related requirements are aimed at having a more transparent market, and increasing the ability to reconstruct market orders for such things as detecting market abuse, there are significant challenges related with them. Many market participants face challenges regarding technological and logistical implementation of the time related regulations (Mcdowell, 2016). One challenge involves the moment (time-stamp) a trade is made. As David Murray of Corvil questions: "*should the time-stamp be put in place when the order is sent or when the application [transaction system] confirms?*". FINRA for example, in its OATS clock synchronisation Q&A, states that what basically is necessary is to synchronise the "*order handling system software*" (FINRA, 2017). One could understand this as the moment an order is recorded (not when it is sent). But one of the main problems regarding clock synchronisation is related with the fact that it is required for the clock to be synchronised to the coordinated universal time (UTC)

Table 3: Network Time Protocol server security features

Feature	Function	Benefits
NTP Reflector with hardware time--stamping	100% hardware-based NTP time--stamping operations	Extremely high-accuracy and high--throughput NTP operations. Ensures DoS attacks won't bring down the NTP server. Detects DoS attacks so system administrator can be notified. Bandwidth limits non-NTP traffic. Processes primarily just NTP packets with no server-management access possible.
CPU-based bandwidth limiting	Allows only a predetermined number and type pf packets to reach the NTP Server CPU	Protects against DoS attacks
HTTPS/SSL	Encrypts management traffic between server and its web interface	Protects against data theft and eavesdropping
Password access	Requires a password to manage the server	Ensures authorised-only administrative access
NTP MD% Authentication	Provides key-based hashed NTP packet exchange between clients and servers	Ensures NTP packets are not spoofed as a way to corrupt the time between time server and time client
Access control lists	Limits clients that can access the management interface and select features to an administrator-specified list	Protects against unauthorized server use or entry as an attack vector of the network
TACACS+, RADIUS, and LDAP authentication	Limits who has management access to the server based on whether they have credentials based on industry-leading access-management systems.	Enables the server to be well-managed as a part of a network-wide access-management system. (Admins don't have to manage credentials locally for each device).

Source: (Skoog, 2017)

at a global level - something that requires a different and more accurate IT approach than it is currently used. Of course, this may not represent a challenge to every financial company, because it will depend on the level of granularity. Nevertheless, for those companies that will be affected by these changes, it may mean modifications in communications, hardware, or even software that demand more efficient and fast coding.

The level of precision and amount of information to be submitted which are required by both regulations - independently of whether the company is an algorithmic participant or human participant - raises another interesting issue occasionally mentioned. All this information that is compiled may be an extremely profitable source of information on trading behaviours and strategies of other participants - especially when payments for order flow are present.

Payments For Order Flow

In competitive markets and, for the most liquid instruments, brokers would aim to choose market makers who offer the best execution conditions for their clients' orders. This could mean best price, best speed, etc. Another practice does not rely on the broker finding the best market maker but, on the latter "buying" the broker's clients' orders and executing them in the market - a practice called "Payment for Order Flow". As the CFA institute describes, Payment for Order Flow is the "*practice of wholesale market makers paying brokers (typically retail brokers) for their client's order flow.*" (CFA Institute, 2016). The orders that clients submit are submitted to their broker, but brokers may sell these orders to other wholesale firms. Two institutions will therefore profit independently of the direction of the order: the wholesaler and the broker. There are different sources of profit depending on the institution. The broker will profit from selling these orders and the wholesale institution will profit from executing orders at better prices than those at which clients originally submitted the orders. Under these agreements, the wholesaler is usually required to meet the "best execution" obligation but, this definition is not akin to finding "the best price". It may also mean the fastest trade or the most likely to be executed (Massa, 2016). As such, wholesalers, may find the fastest order execution but not the cheapest price for that order. The wholesale institution may also profit from submitting orders against the clients - this can be considered a form of front-running. Paradoxically, clients may also eventually profit from reduced costs because brokers may reduce commissions by the payments they receive from wholesale institutions, but this is rarely the case.

Although the practice of Payment For Order Flow may seem unethical due to conflicts of interests, there may also be found some advantages. One advantage relies on reducing costs for the broker related to the search of market makers while another advantage is the clear fact that the broker gets a financial "improvement" by selling the client's order flow to the wholesalers. A concern brought in the Guidance on the practice of "Payment for Order Flow" issued by the FSA (currently the Financial Conduct Authority) (FSA, 2012) is the fact that the market maker is receiving order flow that otherwise it wouldn't be receiving if it did not offer a payment for it. This means that in order for the market maker to recover the value of payments made to the broker, it is not possible for the market maker to offer the best price available in the market. As such, the client will practically always receive a price that is worse than the best in market. Another issue is involved with the market maker - broker relationship. Not all market makers want/can fulfil the requirements presented by the broker. In theory, the best price that could be offered to the brokers' clients would be the ones from market makers that do not buy order flow - because they do not incur additional costs. In the end, order flow is being directed to a limited amount of market makers which offer a price especially designed for the brokers' clients. One might think that this is not much different from the typical banking or financial service where a client chooses a service provider that offers the best price (or lowest costs). The problem with Payment For Order Flows is that there isn't a direct relationship between the service provider and the client. The broker is making a decision in the name of the client which, may not be in the best interest of the client - if clients don't receive better prices or lower commissions and if the broker sells these orders to the highest bidder.

When considering High-Frequency Trading businesses, Payment For Order Flows may be especially controversial. This is due to the ability to not only submit extremely fast orders, but also by paying for Order Flow and therefore, coming into possession of extremely valuable "inside information", giving them a more effective ability of trading against the uninformed retail clients (Marchisi, 2016). Based on what the former US Stock Exchange Official John Marchisi states, one may understand that Payment For Order Flow significantly changes the market (Marchisi, 2016). Currently, or in a traditional manner, in order to achieve profits traders are rather interested in two elements of trade: volume and volatility. These elements give the basic potential of generating profits. What the presence of Payment For Order Flow does is to take away the need for volatility. The profit will be generated from volume (order flow) which is available to a limited number of wholesalers. It is clear that this could lead to a structure similar to an oligopoly.

Market makers (HFTs especially) find small trades the most attractive (Levinson, 2016). So, even that up to this moment the general idea would be that *all* clients could be affected, in reality, only the small retail clients are the key interest of HFT wholesalers. This is due to two motives that become quite obvious after specifying them. One, retail traders are not so well informed as larger traders. Although this could be important under *normal* trading conditions, when wholesalers possess clients' trading information up-front then this may become somewhat less relevant. The second reason is that the value of trades is significantly smaller for retail traders than for institutional ones. This means that the price impact will also be substantially smaller, allowing for more reliable profits (Levinson, 2016) - it is definitely better to handle large numbers of small trades, than a smaller portion of them that may affect prices unfavourably.

Although Payment for Order Flows is not entirely prohibited, regulations are becoming stricter when conflict of interests arise - either between financial institutions or between them and financial clients. Already in 2012, the UK's supervision body (at the time FSA) introduced guidelines regarding Payments For Order Flow considering them to generate conflict of interest.

Also, the recent ESMA's MiFID II tightens regulations regarding conflicts of interests thus limiting the practice of Payment of Order Flow.

Payment For Order Flow (POFP) is differently regulated in different markets. In the US, this practice is not yet considered illegal but there are more requirements for even more information disclosure coming from financial institutions to their clients. Under the rule 606, it is required for brokers-dealers to disclose quarterly their handling of selling order flow (CFR, 2018). Currently, it affects retail orders - ie. up to $200,000 for equities and up to $50,000 for options but there are proposals to expand this rule for institutional orders as well (SIDLEY, 2016).

Even as the PFOF is legal in the US, there is the possibility of litigations being brought by clients if there is some sort of wrongdoing by brokers or the HFT wholesalers. But in practice this is not as easy as it seems. One reason, as previously mentioned, is the fact that wholesalers prefer retail orders/clients. This means that orders are obviously also small. This in turn means that individual litigations bear disproportionately high costs. Also, the potential damages a retail client incurs are small when taking into account the entire value of orders sold by the broker to the wholesaler. A potential solution could be class action litigations. The obstacle lies in the fact that most (if not all) clients when opening a new brokerage account are required to accept a class action litigation waiver (2014). There have been situations though where the US Supreme Court has sustained the possibility of maintaining mandatory class action waiver clauses on the base (amongst others) of the Federal Arbitration Act that favours the freedom of contract for arbitration agreements (Peralta, 2016). But the issue is more complex because brokers are members of the Financial Industry Regulation Authority which in turn prohibits the use of mandatory class action waivers in pre-dispute arbitration clauses - rules that must be accepted by the SEC acting under a Congressional mandate based on the Securities Exchange Act of 1934 (Peralta, 2016). This shows that there are contradicting rules and regulations, both in favour and against class action litigation waivers in agreements affecting clients of brokerage firms. In the end, while the PFOF is legal in the US, there are situations where order execution is against the best interests of the client; it may still be difficult for the clients to exercise their rights.

As far as the European continent is concerned, guidelines and regulations from both FCA and ESMA try to restrict the practice of Payments For Order Flow as much as possible since, as a FSA guidance rule on Payment For Order Flow states, this is a harmful process for which its arrangements "*create a clear conflict of interest between the clients of the firm and the firm itself*" (FSA, 2012). MiFID II also prohibits any practices that may create a clear conflict of interest between companies and their customers. The key issue relies on how to prove that this practice, or other similar ones, generate conflicts of interests - even with the introduction of MiFID II and tighter regulations on PFOF practices as of 3rd of January 2018. Under the FSA Guidance Rule from 2012, PFOF is not literally prohibited but, according to the Conduct of Business Sourcebook for companies with activities carried in the United Kingdom, companies must not pay or accept any payments, commissions, or non-monetary benefit for that manner, other than situations for which (ESMA, 2017):

1) "the payment […] does not impair compliance with the firm's duty to act in the best interests of the client, and",

2) details of the inducement are disclosed to the client "[…] in a manner that is comprehensive, accurate and understandable, before the provision of the service", and

3) the inducement is "[…] designed to enhance the quality of the service to the client".

While MiFID II also does not literally prohibit order flow payments it certainly inserts more information and rules related with it. The most significant rule on routing client orders is the Article 27 of the MiFID II Directive (ESMA, 2014). This Article addresses especially, the process of order flow routing, by setting rules for the wholesaler and the broker. Previously in MiFID I, there was not a clear definition of "best execution". Firms should take "*all reasonable steps*" when executing orders on behalf of clients. Under MiFID II, the conditions of an order execution on behalf of clients are more precise, and specified under Article 27 (1) where, "*[…] firms must […] take all the sufficient steps to obtain […] the best possible result for their clients […] taking into account price, costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order*" (ESMA, 2014). It is therefore imperative for firms to take all the mentioned conditions into account before executing third party orders. It will now be more difficult for wholesalers to offer other best conditions apart from price. Aside from the more precise definition of order execution conditions, under Articles 24 and 27 (and others) firms have also new requirements regarding execution policy which should be reviewed on a regular basis, and information policies.

It is worth underlining that Article 27 and other articles linked with this one are nevertheless supported on the conflict of interest foundation. For example, Article 27 (2) states "*An investment firm shall not receive any remuneration, discount or non-monetary benefit for routing client orders to a particular trading venue or execution venue which would infringe the requirements on conflicts of interest or inducements set out in paragraph 1 of this Article and Article 16(3) and Articles 23 and 24.*" (ESMA, 2014). Article 23 addresses conflicts of interests while, Article 24 sets principles in providing information to clients. This means that while Article 27 regulates payments for order flows, it is based on firms not violating conflicts of interest rules. But even when MiFID II limits most of the PFOF practices, firms may start to offer other types of services - as noted in an FCA review, by providing "*an 'arranging service' and not the execution of clients orders*" (FCA, 2014).

The issue of "Payment For Order Flow" seems therefore, far from being over - even with the many efforts from ESMA for the EU or FINRA for the US.

Conclusion

High-Frequency Trading by some might be seen as controversial. Some academic papers refer to HFT as having a negative influence on the financial markets and prices, while in others, as having a positive influence on price volatility and spreads. Ever since the increase in their trading activity, one may see fast technological changes that not only affect the way trading is performed but that have also led to important regulatory changes. During the last few years in both the US and Europe adjustments have been made to regulations, reflecting how the financial market functions and which market practices have been present for many years – such as HFT and Payments for Order Flow. The question remains as to how far regulation is going or should go. Simple changes already made in regulations, such as clock synchronisation, lead to significant modifications in internal procedures, hardware and/or software. These changes not only affect the back office, but also the middle and front office - influencing in the end the way the market functions and the presence of HFT. It is definitely becoming more difficult for HFT traders to achieve profits as a result of many regulations, which gives the impression that this business model may be harmful to the financial market - something that essentially has not been answered. If the reason for having a negative influence is speed, then it seems strange that for years and years now, everyone that was faster than other traders could submit orders without any problems. Yes, speed is a little different than it was 50 years ago but, so is technology. Also, there was not a significant problem when exchanges started to introduce electronic trading at a time when only some traders did take advantage of the speed involved. In the end, most traders did adapt to electronic trading and it was only a question of time. Maybe this should be looked at in the same manner when considering HFT.

The aim of this paper was to first present the technological changes involved with trading especially related with HFT, along with their respective regulatory changes (and challenges). One of the possible changes is the general introduction of batch auctions (already available in such places as in FTSE 100 or CHX SNAP) – a mechanism aimed to knock out HFT advantages by introducing discrete-time auctions instead of continuous-time auctions. Finally, the last objective of this paper was to present the current state of regulations on the controversial practice of Payments For Order Flows – a practice seen as extremely controversial especially if performed by HFT entities. All of these objectives aim to bring attention to regulations that seem to target the HFT business model and, in such a way to eventually eliminate it from the market.

References

Anthony, S. (2012). *$1.5 Billion: The Cost of Cutting London-Tokyo Latency by 60 Ms*, Extremetech. Retrieved from https://www.extremetech.com/extreme/122989-1-5-billion-the-cost-of-cutting-london-toyko-latency-by-60ms (05-03-2017).

Anthony, S. (2016). *The Secret World of Microwave Networks*, Arstechnica,.Retrieved from https://arstechnica.com/information-technology/2016/11/private-microwave-networks-financial-hft/ (05-03-2017).

BIS (2011). *High-frequency Trading in the Foreign Exchange Market*, BIS.

BIS (2017). *The Sterling "Flash Event" of 7 October 2016*, BIS.

Budish, E., Cramton, P., Shim, J. (2015). The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. *The Quarterly Journal of Economics 130* (4), 1547-1621. Doi:10.1093/qje/qjv027.

Catnmsplan (2014, February 27). *SEC Rule 613 - Consolidated Audit Trail (CAT)*, Proceedings from Industry Outreach Event.

Caproasia (2017). *2016 Stock Exchange Market Capitalization.* Retrieved form https://www.caproasia.com/2017/01/27/2016-stock-exchange-market-capitalization/ (12-07-2018).

CFA Institute, (2016). *Payment for Order Flow,* CFA Institute.

CFR (2018). Disclosure of Order Routing Information. *Commodity and Securities Exchanges 242*, 242-606.

CHX, (2016). *CHX SNAP, Chicago Stock Exchange*. Retrieved from http://www.chx.com/snap/ (31-03-2017).

Daly, R., (2016). *SEC Tightens Clock-Sync Mandate,* Marketsmedia. Retrieved from http://marketsmedia.com/sec-tightens-clock-sync-mandate/ (17-03-2017).

Davies, R. (2016, October 7). What Caused the Pound's Flash Crash?, *The Guardian*, (01-04-2017).

Durden, T. (2015). *Meanwhile, Over at the "New York" Stock Exchange: Even More Lasers*, Zerohedge. Retrieved from http://www.zerohedge.com/news/2015-12-27/meanwhile-over-new-york-stock-exchange-lots-lasers (05-03-2017).

Easley, D., Lopez de Prado, M., O'Hara, M. (2011). The Microstructure of the "Flash Crash": Flow Toxicity, Liquidity Crashes and The Probability Of Informed Trading. *The Journal of Portfolio Management 37*(2), 118-128.

Easley, D., López de Prado, M.M., O'Hara, M. (2012). The Volume Clock: Insights into the High-frequency Paradigm. *The Journal of Portfolio Management 39*(1), 19-29.

Times, E. (2015). *BSE Eyes 10 Times Faster Trading in the Next Few Months.* Economic Times*.* Retrieved from http://economictimes.indiatimes.com/markets/stocks/news/bse-eyes-10-times-faster-trading-speed-in-next-few-months/articleshow/49031646.cms (05-03-2017).

Nr 2014/65/EU, *Markets in Financial Instruments* (MiFID II), 2014, ESMA.

ESMA, (2017). *Conduct of Business Sourcebook*, COBS (Release 14), EMSA.

FCA, (2014). *Best execution and payment for order flow*. Review (TR14/13), Financial Conduct Authority.

FINRA, (2016). *Regulatory notice 16-23*, FINRA. Retrieved from http://www.finra.org/industry/notices/16-23 (17-03-2017).

FINRA, (2017). *OATS Clock Synchronization FAQ*, FINRA. Retrieved from http://www.finra.org/industry/faq-oats-clock-synchronization-faq (17-03-2017).

Fricke, D., Gerig, A. (2016). *Too Fast or Too Slow?* Determining the Optimal Speed of Financial Markets.

FSA, (2012). *Guidance on the Practice of "Payment for Order Flow"*. Finalised Guidance (FG12/13), FSA.

(2014), *High Frequency Trading Litigation - Defenses(1)*, QUAINTON LAW. Retrieved from http://quaintonlaw.com/high-frequency-trading-litigation-defenses-1/ (24-03-2017).

Ismail, N.I., Mnyanda, L. (2016, October 7). *Flash Crash of the Pound Baffles Traders With Algorithms Being Blamed*, Bloomberg. Retrieved from https://www.bloomberg.com/news/articles/2016-10-06/pound-plunges-6-1-percent-in-biggest-drop-since-brexit-result (01-04-2017).

Lam, B. (2016). *A New, Slower Stock Exchange.* The Atlantic. Retrieved from https://www.theatlantic.com/business/archive/2016/06/iex-approved/487898/ (13-03-2017).

Lenzner, R. (2011, October 8). *"Cheetah" Traders Are Dangerously Controlling Our Markets,* Forbes. Retrieved from https://www.forbes.com/sites/robertlenzner/2011/10/08/cheetah-traders-are-dangerously-controlling-our-markets/#544c8f607094 (05-03-2017).

Levey, S. (2013). *SEC Rule 613: Consolidated Audit Trail - National Market System (NMS) plan*, Deloitte.

Levinson, C. (2016, May 10). *Exclusive: U.S. Investigates Market-aking Operations of Citadel, KCG.* Reuters, Retrieved from http://www.reuters.com/article/us-usa-stocks-probe-exclusive-idUSKCN0Y11CJ (24-03-2017).

Marchisi, J.C. (2016, June 13). *Theft of Opportunity - The Impact of Regulation NMS on the Retail Investor*. Retrieved from http://www.nanex.net/aqck2/4713.html (24-03-2017).

Massa, A. (2016, May 20). *Payment for Order Flow - Professionals Pay to Make Trades for Small Investors.* Bloomberg, Retrieved from https://www.bloomberg.com/quicktake/payment-for-order-flow (19-03-2017).

McCrank, J. (2014). *Dark Markets May Be More Harmful Than High-Frequency Trading,* Reuters. Retrieved (31-03-2017).

Mcdowell, H., (2016). *Clock Synchronisation… Time is Ticking.* The trade, Retrieved from http://www.thetradenews.com/Regulation/Clock-synchronisation…time-is-ticking/ (17-03-2017).

McKenna, F. (2015). *Here's the Advantage High-frequency Trading Firms Have Over Everyone Else.* MarketWatch, Retrieved from http://www.marketwatch.com/story/heres-the-advantage-high-frequency-trading-firms-have-over-everyone-else-2015-08-13 (05-03-2017).

NYSE, (2017). *NYSE.* Retrieved from https://www.nyse.com/data/cta (13-03-2017).

Peralta, A.J. (2016, March 4). *The Legal Landscape of Class Arbitration Waivers n Consumer Contracts.* Retrieved from http://apps.americanbar.org/litigation/committees/appellate/articles/winter2016-0316-legal-landscape-class-arbitration-waivers-consumer-contracts.html (25-03-2017).

Robinson, G. (2009, September 22). *Exchanges Warn G20 of Dangers in 'Dark Pools',* FT. Retrieved from https://www.ft.com/content/71069076-a795-11de-b0ee-00144feabdc0 (31-03-2017).

SEC, (2017). *Rule 613 Info,* SEC. Retrieved from https://www.sec.gov/divisions/marketreg/rule613-info.htm (17-03-2017).

SIDLEY, (2016). *SEC Proposes Rules to Enhance Order Handling Information Available to Investors,* SIDLEY. Retrieved from http://www.sidley.com/news/2016-08-08-sder-update (18-03-2018).

Skoog, P. (2017, January 30). *Synchronizing Trading Clocks for Regulatory Compliance - Are Your networks ready?*, Electronic Design. Retrieved from http://electronicdesign.com/digital-ics/synchronizing-trading-clocks-regulatory-compliance-are-your-networks-ready.

Sprothen, V. (2016, August 8). Trading Tech Accelerates Toward Speed of Light. *The Wall Street Journal.* Retrieved from https://www.fnlondon.com/articles/trading-technology-accelerates-toward-speed-of-light-20160808 (05-03-2017).

Von Hoffman, C. (2014). *What are "Dark Pools," and Why are They Dangerous?,* CBS. Retrieved from http://www.cbsnews.com/news/finances-dark-pools-explained/ (August 12).