Applying Spam-Control Techniques To Negate High Frequency Trading Advantages

Chris Rose, Capella University, USA

ABSTRACT

High Frequency Traders undoubtedly have an advantage over the average trader or trading desk because they incorporate nefarious devices into their trading schemes such as Flash Trading, Dark Pools and Quote Stuffing. In addition, they usually locate their servers as close to the exchange servers as possible so their data travels the shortest possible distance. However, adding spam-control techniques to control all trades would negate these advantages and return trading to once again being an equitable, free and open market-based system.

Keywords: High frequency trading; spam control; dark pools; quote stuffing

INTRODUCTION

High-Frequency Trading (HFT) accounts for the majority of trading in stocks, commodities and futures in the United States today. For example, HFT firms, only account for approximately 2% of the 20,000 or so trading firms that operate in U.S. markets today, but this 2% account for 73% of all U.S. equity trading volume with aggregate annual profits of $21 billion (Iati, 2009). These sophisticated trading platforms use a combination of proprietary, secret algorithms and the fastest computer hardware available to absorb and analyze massive amounts of market data and news and generate millions of trades timed to the millisecond.

There are basically four distinct advantages that HFTs have over the ordinary trader:

1. Distance - Throughput and latency are important in HFT since they try to place their servers as close to the data source as possible, some servers even being located in the same building as the exchange servers. This is important because since data is roughly travelling at the speed of light, each 186 miles from the exchange servers means an additional millisecond for the data to reach you and in high speed trading, every millisecond counts. Their trade will therefore arrive at the exchange servers before yours every time.
2. Quote Stuffing - HFTs fool other investors into believing that there is interest in a stock when in reality, there isn't. These specialized computer systems allow HFTs to post buy and sell prices they have no intention of actually following through on. For instance, a firm might post a bid for a stock showing they want to buy at a certain price. But by the time investors interested in selling at that price get their order to the market, the false buyer yanks the electronic bid in milliseconds, making it difficult to detect (Kranz, 2010).
3. Flash orders - orders are shown to members of an exchange for a split second before being passed on to the wider market give HFT an advantage, and have become an integral part of HFT. In fact flash orders are sent out as execute or cancel on receipt which means a substantial amount of them self-destruct.
4. Dark Pools - Bigger investors are also moving to dark pools, where orders are anonymously matched so that traders do not alert the wider market to their intentions. Obviously, this means that stock pricing is not transparent and yet dark pools accounted for 9% of the US market in 2008 (Grant, 2009).

SPAM

To be considered spam in the true sense of the word, the e-mail has to be both bulk and unsolicited. Unsolicited means that the recipient has not granted verifiable permission for the message to be sent but it can be normal e-mail e.g. job or sales enquiries. Bulk e-mail means that the message is sent as part of a larger group of messages, all having basically the same content but it can also be normal e-mail such as newsletters, discussion lists etc. but it is only when the e-mail is both unsolicited and bulk that it can be considered spam. Therefore e-mail is spam if:

1. the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients and
2. the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent.

The transmission and reception of the message should also appear to the recipient to give a disproportionate benefit to the sender (The Spamhaus Project, 2010).

SPAM AND HFT

The amount of spam sent daily is enormous, a 2008 estimate was that it makes up 80 to 90% of all emails sent for a total of around 120 billion messages per day (Kleiner, 2008). Recently, a 23-year-old Russian, Oleg Nikolaenko, has been identified by the FBI as the man behind a malicious network of about 509,000 infected computers. This one person was responsible for almost 10 billion spam emails a day (Mail Online, 2010). The frequency of orders in high-frequency trading is also enormous and the volume and method closely resembles spam. Nanex estimates that the average quote volume on the NYSE is 10,000 per second or about 250,000,000 per day. It is this frequency of orders, real and fake that makes HFT and spam very similar.

The difference in volume of spam at more than 120 billion per day and trading at 250 million per day might seem massive but it has to be remembered that spam is directed at millions of computers per day but trades are only directed at a few specific servers. In addition, many trades could, like spam, be considered unsolicited, in that, according to Nanex, many of them result in no trades and they are simply a technique used by HFTs to overwhelm trading servers by sending high volumes of data. At one point during the May 6 flash crash, somebody launched 5,000 quotes at the NYSE for the ticker just one company, Public Storage, all inside of one second. None of those quotes led to a trade. So it is clear that one trader or perhaps more discovered that by blasting the NYSE, they could introduce added latency to the system (Nanex, 2010a).

Hammering is the term used when someone repeatedly tries to connect to an unavailable server with little or no time between connection attempts. It is similar to someone who keeps pushing the redial button on a telephone when a line is busy. However, servers cannot process an unlimited number of requests, so if a server is busy it will deny requests until they have the capacity to fulfill the request. But if a server is already at capacity it also has to send a response to the requesting server that it is busy so that each additional attempt to connect slows down the system even further (Webopedia, 2010).

Today in HFT systems microseconds matter. If one HFT can process information faster than a competitor that is the winning edge. "If you could generate a large number of quotes that your competitors have to process, but you can ignore since you generated them, you gain valuable processing time. This is an extremely disturbing development, because as more HFT systems start doing this, it is only a matter of time before quote-stuffing shuts down the entire market from congestion" (Nanex, 2010c). The system has shown big delays more than once since the May 6 flash crash. It seems that whenever the NYSE receives more than 20,000 quotes per second, its Consolidated Quote System (CQS) feed falls behind (Comstock, 2010). Obviously, some HFTs have discovered that hammering a system will sufficiently slow down a response to allow them to do things they couldn't normally do.

SPAM CONTROL

There have been many attempts to introduce a computing cost to sending e-mail and years ago, Microsoft had a project called the Penny Black project which investigated several techniques to reduce spam by making the sender pay. They considered several currencies for payment and CPU cycles, memory cycles, and Turing tests (proof that a human was involved) were the leading candidates. Basically, this system worked on the principle that if you don’t know me and want to send me e-mail then you have to expend some effort to send me that e-mail. What was generally considered the best was a computational cost. If the effort to send the e-mail is measured in CPU cycles, then since there are approximately 80,000 seconds in a day, a computational cost of just ten seconds per message would limit a spamming computer to at most 8,000 messages daily. So spammers would have to invest heavily in hardware in order to send high volumes of spam. To achieve a ten second delay, it would be a simple matter to make a slight modification to the email SMTP protocol to force each computer to compute a particular mathematical algorithm which is known to require a minimum specific amount of time to compute (Rose, 2004).

A spammer admitted that they simply had four computers and two cable modems in his operation and was able to send out 10 million e-mails a day from those computers running 24 hours a day. They had to send out about 500,000 an hour to make any money since it is estimated that the rate of return in spam is less than one-tenth of one percent. A delay system such as this would force spammers to move from sending spam to the more acceptable model of sending targeted e-mail marketing messages, since instead of sending out millions of e-mail messages each day, they would be limited to thousands, unless they invested very heavily in computer equipment and Internet connections. In this particular case, instead of sending out 10 million e-mails a day this spammer would only be able to send out about 32,000 per day, therefore it would be prudent of them to make every e-mail message count (Rose, 2004).The alternative would be to invest in 1,250 computers and the corresponding networking and Internet connections, which would place spam clearly only within the reach of extremely large companies. At that level of investment, the return of one-tenth of one percent would not make economic sense

TIME

Universal Coordinated Time (UTC) is the method by which time is regulated around the world. For computers, the Network Time Protocol (NTP) is defined in Request for Comments (RFC) 1305. "The National Institute of Standards and Technology (NIST) provide authenticated NTP messages using a symmetric-key algorithm that is compatible with the reference implementation of the NTP software. The authentication ensures that the message originated from a NIST time server and was not modified during transit" (NIST, 2010). This time service is provided by independent servers that are synchronized using the same algorithm therefore the accuracy of the time stamps at all servers should be the same.

Although on the surface it appears simple, the system is complex since there are tiers of NTP servers. Tier one NTP servers connected to atomic clocks, sometimes by GPS, and tier two and three servers are networked to the tier one servers to spread the load of time requests across the Internet. The client software is extremely complex and has to take into consideration such factors as communication delays, and adjust the time on the server in a way that does not upset all the other running processes on the server (Ubuntu, 2010). There are also pools of servers, for example, at the pool.ntp.org project, which is a big virtual cluster of over 2,000 timeservers providing NTP service. The time information provided by the service is directly traceable to UTC(NIST). One protocol that actually would help in the process of controlling HFTs is that "all users should ensure that their software NEVER queries a server more frequently than once every 4 seconds. Systems that exceed this rate will be refused service. In extreme cases, systems that exceed this limit may be considered as attempting a denial-of-service attack" (NIST, 2020).

APPLYING SPAM TECHNIQUES

During the Flash crash of May 2010, there were significant delays in processing orders. In Nanex's formulation, the lag occurred because the NYSE bids and offers weren't being time stamped when they were made, but after they got in line behind other quotes and were processed by the NYSE servers, and by the time that whole process was finished and the time stamps applied, the actual prices had already moved on (Flood, 2010). But perhaps this problem associated with the worst one-day loss in trading history also provides a hint at a solution.

HFTs can exploit not only differences in prices but also differences in response time between different exchange servers. For example, since the NYSE typically runs slower than trading on BATS then time arbitrage is possible. A HFT could flood orders through the NYSE (quote stuffing), forcing a few second slowdown in trading, simultaneously they could establish a trade in BATS which is then sent to the NYSE. Therefore, in reality, since HFTs have the ability to execute numerous trades simultaneously, they can take advantage of time-based bid and offer differentials across exchanges (Malmgren & Stys, 2010).

To bring equity back into trading, it should be required that every trader install a tier-two time server and then each trade would be stamped when it is made with the official NIST stamped time. Suppose that trade is next sent to the exchange servers, but instead of processing the trades, those servers institute an official delay of perhaps five seconds. Now all trades, regardless of where in the world they were made would now have an equal chance at the exchange since there would be no more advantage in locating your servers in the same building as the exchange servers. All data around the world which is already travelling at about 186,000 miles per second should reach the exchange servers, even if they encounter bottlenecks at some router along the way. If attempts are made to tinker with the time servers or by forcing queries more often than every four seconds, the time server algorithms already have a built in protocol to counter that and they would be blacklisted.

Suppose the exchange servers now sort all trades by their NIST time stamps and institute trades based on when they were made, since all traders would now have the exact time no matter where the trade were made. The exchange servers, because of the built-in delay, would also be waiting for trades delayed in transit, and since time can't be successfully controlled by anyone, more fairness would now be in the system and a trader in Kansas or Tokyo would have an equal chance of a successful trade as the HFT who has co-located their server in the same building as the exchange. This should eliminate the distance advantage.

Flash orders are small “immediate or cancel” orders, valid only for microseconds, that carry little risk for HFTs. By probing and trying to find buyer limits to buy or sell, HFTs "have vastly greater knowledge of all aspects of the markets’ depth and breadth than individuals or passive investors like pension plans" (Malmgren & Stys, 2010). Flash orders are obviously very profitable to HFTs because "NYSE Euronext, the only one of the top four U.S. exchanges that doesn’t use flash orders, has seen its portion of the nation’s share trading slip to 30.3 percent in the second quarter from 35.5 percent a year earlier" (Ortega, 2009). On the other hand, Direct Edge, in Jersey City, New Jersey, an early leader in using flash trading techniques and is the fastest-growing equity market in the U.S. "Helped by its three-year-old Enhanced Liquidity Provider program, which handles the most flash trades. Even excluding flash orders, Direct Edge matched 11.2 percent of U.S. stock trades in July, making it the third- largest U.S. equity market by volume"(Ortega, 2009).

In fact since thousands of flash orders are sent out per second and flash orders are sent out as execute immediately or cancel on receipt, this means a substantial amount of them self-destruct, and perhaps only result in a successful trade at the same rate as the one-tenth of one percent success rate in spam. But suppose, using the time-based spam control method, these orders also had to be kept valid for a minimum period of time, perhaps ten seconds, the HFT trick of placing and cancelling orders in microseconds would be negated. This would eliminate the flash order advantage.

In the HFT technique called quote stuffing, the HFTs are in fact spamming the exchange servers by blasting them with up to 5,000 quotes per second for one specific stock without a trade being made. You could also apply a spam control technique used for denial of service attacks to these types of orders, namely, a temporary blacklist. If a computer is found to be hammering the exchange servers, they should be blocked for a minimum of perhaps 30 seconds before being made to once again join the queue. All this would be an attempt to make trades more focused and not just a computerized blast hoping that something will stick. This would eliminate the quote stuffing advantage.

ALL EXCEPT DARK POOLS

There is however, one HFT advantage that cannot be contained without regulation and this is dark pools. A dark pool is a source of liquidity that is not displayed to the general public since it is not publicly quoted. However, regulators are looking closely at dark pools because "recent increases in volume, a lack of transparency for price discovery, and the perception that the retail investor is disadvantaged" (Caplan, 2010). There are numerous questions about "the basic nature of “hidden” liquidity, access, and cost/rebates are being reviewed by the Securities and Exchange Commission and the Financial Services Authority in the United Kingdom—which may chart the direction of these execution venues" Additionally, there are no standardized requirements for reporting the activity in dark pools and this has led to inconsistent and unreliable volume statistics. The SEC has suggested that reporting practices, or lack thereof, present a critical transparency issue that should be given more attention (Caplan, 2010).

CONCLUSION

The proponents of HFTs claim that they in fact are good and provide liquidity to the market but HFT providing liquidity to the market is really an illusion since it is totally dependent on the willingness of HFTs to remain active under all circumstances during all trading hours. HFTs can stop trading at any time during the day when they want and in fact all HFTs close out trades at the end of the day. When several HFTs stopped trading on the afternoon of May 6, the result was a temporary implosion of liquidity. There have been suggestions that some alteration of incentives might be achieved by introduction of a sliding scale of fees or taxes according to volume and speed of trading. However, some believe that HFTs should be required to be available to trade during all trading hours and they should be banned from trading for a period of time if they disengage from trading. That would reduce the risk of a market plunge resulting from HFTs decision to halt trading (Malmgren & Stys, 2010).

If these rules are combined with a delay to negate the location advantage and a minimum time to display a trade and a blacklist for any trader that is found hammering the system, it is entirely possible that trading can once again return to being an equitable, free and open market-based system. Unfortunately, only regulation would be able to prevent the hidden trades and lack of transparency in dark pools.

REFERENCES

1. Caplan, K., Cohen, R., Lenz, J and Pullano, C. (2010, Volume 4, Number 2) Dark pools of liquidity. PriceWaterhouseCoopers. Retrieved 12/6/2010 from http://www.pwc.com/us/en/alternative-investment/assets/NY-10-0105-PwC-alt-Caplan.pdf
2. Comstock, C. (August 31,2010) Business Insider. Nanex: There WILL Be Another Flash Crash Because Someone Will Cause It On Purpose. Retrived 9/6/2010 from http://www.businessinsider.com/nanex-there-will-be-another-flash-crash-because-someone-will-cause-it-on-purpose-2010-8
3. Flood, J. (August 10,2010) The Atlantic. NYSE Tech Delays Contributed to the May 6 Flash Crash. Retrieved 9/7/2010 from http://www.theatlantic.com/technology/archive/2010/08/nyse-tech-delays-contributed-to-the-may-6-flash-crash/61987/
4. Grant, J. (July 10, 2009) 'Secretive' firms dominate US share trading. FT.com. Retrieved August 19, 2009 from http://www.ft.com/cms/s/a5f03366-6d69-11de-8b19-00144feabdc0,Authorised=false.html?_i_location=http%3A%2F%2Fwww.ft.com%2Fcms%2Fs%2F0%2Fa5f03366-6d69-11de-8b19-00144feabdc0.html%3Fnclick_check%3D1&_i_referer=&nclick_check=1
5. Iati, R. (2009, July 10) The Real Story of Trading Software Espionage. Advanced Trading. Retrieved August 19, 2009 from http://advancedtrading.com/algorithms/showArticle.jhtml?articleID=218401501#undefined
6. Kleiner, K. (2008, April 25). Happy spamiversary! Spam reaches 30. Retrieved 12/6/2010, from http://www.newscientist.com/article/dn13777-happy-spamiversary-spam-reaches-30.html?full=true
7. Krantz, M. (2010) Computerized stock trading leaves investors vulnerable. Retrieved 8/2/2010 from http://www.usatoday.com/money/markets/2010-07-09-wallstreetmachine08_CV_N.htm
8. Mail Online (2010, December 1) FBI target 23-year-old Russian man behind a third of world's spam emails. Retrieved 12.1.2010 from http://www.dailymail.co.uk/news/article-1334653/FBI-target-23-year-old-Russian-man-worlds-spam-emails.html?ito=feeds-newsxml
9. Malmgren, H and Stys, M. (2010, Summer) The Marginalizing of the Individual Investor. The International Economy. Retrieved 12/4/2010 from http://www.international-economy.com/TIE_Su10_MalmgrenStys.pdf
10. Nanex, Inc. (2010a) Analysis of the "Flash Crash" Retrieved 9/12/2010 from http://www.nanex.net/20100506/FlashCrashAnalysis_CompleteText.html
11. Nanex, Inc.(2010b, November 29) Flash Equity Failures in 2006, 2007, 2008, 2009 and 2010. Retrieved 12/1/2010 from http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysis_Equities.html
12. Nanex ,Inc. (2010C) Analysis of the "Flash Crash" Date of Event: 20100506 Part 4, Quote Stuffing, A Manipulative Device. Retrieved 9/8/2010 from http://www.nanex.net/20100506/FlashCrashAnalysis_Part4-1.html
13. NIST. (2010) National Institute of Standards and Technology. Retrieved 12.3.2010 from http://www.nist.gov/pml/div688/grp40/its.cfm

14. Ortega E, Westbrook J and Martin E. (2009, August 5) Flash Trading Reversal at SEC May Hit Direct Edge. Bloomberg.com. Retrieved on 12/3/2010 from http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aQDDFnAWMld0
15. Rose C. (2004) Finding a Recipe for Spam. Review of Business Information Systems, Vol. 8, No. 2, pages 19-25.
16. The Spamhaus Project (2010) Retrieved 12/5/2010 from http://www.spamhaus.org/index.lasso
17. Ubuntu, (2010). Help documentation. Retrieved on 12/6/2010 from https://help.ubuntu.com/8.04/serverguide/C/NTP.html

NOTES