May 11, 2010

Exchanges and Data Feeds: Data Theft on Wall Street

A Themis Trading LLC White Paper

By Sal Arnuk and Joseph Saluzzi

Introduction

How many average Americans have been victims of identity theft? A Sun Microsystems survey a few years back actually placed that number at 33%. In that same survey a majority of those victims said that "they are likely to stop shopping and banking with institutions that put their personal data at risk."

We can tell you flat out that if an institution played loosey-goosey with our personal data, they would not be earning our business and we would be lobbying hard to make our outrage known with government regulators and lawmakers, so that they could put an end to the malfeasance.

Now, to take this further, what if you found out that an institution, which should be protecting your information and identity, was providing information regarding your transactions in data reports to their other customers as part of their every day business strategy? Not exactly identity theft, perhaps, but clearly theft of highly proprietary information.

In our previous white paper ("Latency Arbitrage: The Real Power Behind Predatory High Frequency Trading"), we illustrated how the exchanges provide raw data feeds that help high frequency traders (HFTs) figure out market directions. Now, we have discovered that at least two of the exchanges, in addition to providing that information, also provide data that enable HFTs to track specific trade orders, putting institutional trading strategies at further risk.

Put simply, every day, certain market centers are marketing and providing data feeds where they are revealing more information than just the original order, depth of book and trade executions. This is being done legally as a by-product of a market structure that has gone horribly wrong. However, unlike personal identity theft, where nearly all Americans and stakeholders are aware of the practice, the vast majority of institutional and retail traders are not aware that this is being done legally as they enter their order flow.

Thus, instead of traditionally collecting listing and transaction (commission) fees, some market centers are generating revenue by selling co-location, and providing low-latency enhanced market data feeds which contain sensitive trader information. Information in these feeds allows high frequency trading firms to track when an investor changes price on his order, how much stock the investor is buying or selling in accumulation, as well as the ascertaining of hidden order flow. This information assists HFTs in predicting short-term price movements with near certainty.

Exchanges will argue that this information is public and available to all investors. Technically, this may be true, however, realistically, not many retail or institutional investors have the capital to invest in the type of computer systems needed to access this information and most are not even aware that it exists at all. They also are not aware that the information being provided to HFTs is revealing critical information about trading intentions. In this report we have identified two situations where this is occurring: the BATS Exchange direct feed known as BATS PITCH and the NASDAQ direct feed known as TotalView-ITCH.

BATS PITCH Feed
BATS supplies an order ID number that is attached to each order submitted through PITCH. The order ID number is then sent out to subscribers of their PITCH feed. From the BATS PITCH specification:

Themis Trading LLC, 10 Town Square, Suite 100, Chatham, NJ 07928
Phone: (973) 665-9600 / Web: www.themistrading.com

*"If an orders Price or Display values change within the BATS matching engine, a Cancel Order Message will be immediately followed by a new Add Order Message with the **same Order ID** as the original order. An order that changes its Display value from "N" to "Y" will not lose its priority."*[1]

Once a reserve book order (an order that displays only a piece of the entire order) is placed in BATS, the Order ID number tracks cumulative trades over the life of the order. If the quantity or price is changed, the order ID number will show that to all PITCH subscribers. Also, any trade execution related to that order has the same order ID number as the original order. HFTs who subscribe to the BATS PITCH feed could determine how much an order has traded and if it is changing its price. They are able to trace the life of an order and decipher valuable information about that order which helps determine the future price of a stock.

NASDAQ TotalView-ITCH Feed

NASDAQ is supplying information on its TotalView-ITCH product which reveals information about hidden order flow on its exchange. Every time a non-displayed (or hidden) order is executed, NASDAQ sends a message that not only identifies that a trade has occurred, but also identifies if the hidden order was a "buy" or "sell." In addition, the trade order ID associated with that trade is "cumulative." This means that every time a trade executes when part of a hidden order, the same ID number is attached to that trade as the original trade, thereby enabling ITCH subscribers to determine how much of the stock in question the hidden buyer or seller has accumulated.

4.6.1 Trade Message (Non-Cross)[2]

*The Trade Message is designed to provide execution details for normal match events involving **non-displayable** order types…A Trade Message is transmitted each time a **non-displayable** order is executed in whole or in part. It is possible to receive multiple Trade Messages for the same order if that order is executed in several parts. Trade Messages for the same order are **cumulative.***

TRADE MESSAGE (NON-CROSS)

Name	*Offset*	*Length*	*Value*	*Notes*
Message Type	*0*	*1*	*"P"*	*Trade Message*
Timestamp - Nanoseconds	*1*	*4*	*Integer*	*Nanoseconds portion of the timestamp.*
Order Reference Number	*5*	*8*	*Integer*	*The unique reference number assigned to the order on the book being executed.*
Buy/Sell Indicator	*13*	*1*	*Alpha*	*The type of non-display order on the book being matched.* *"B" =buy order* *"S" =sell order*
Shares	*14*	*4*	*Integer*	*The number of shares being matched in this execution.*
Stock	*18*	*8*	*Alpha*	*The security symbol associated with the match execution.*
Price	*26*	*4*	*Integer*	*The match price of the order. Refer to Data Types for field processing notes.*
Match Number	*30*	*8*	*Integer*	*The NASDAQ generated session-unique Match Number for this trade. The Match Number is referenced in the Trade Break Message*

[1] http://batstrading.com/resources/membership/BATS_FAST_PITCH.pdf
[2] http://www.nasdaqtrader.com/content/technicalsupport/specifications/dataproducts/NQTV-ITCH-V4_1.pdf

Themis Trading LLC, 10 Town Square, Suite 100, Chatham, NJ 07928
Phone: (973) 665-9600 / Web: www.themistrading.com

We called NASDAQ and they confirmed that they are providing information about hidden order flow. They also confirmed that a "buy" or "sell" indicator is attached and the trade messages are cumulative.

We are not the only market participants who have noticed that the exchanges are revealing much more information on their data feeds than most investors realize. In a comment letter to the SEC on their concept release dated April 29, 2010, the Securities Industry and Financial Markets Association (SIFMA) stated:

> *"It may, however, be appropriate for the Commission to give greater consideration to the manner in which direct market data feeds may be used by market participants. As noted, direct market data often is faster and more detailed than consolidated data. Also, direct data feed recipients generally are able to more easily trace orders they submit to an exchange or electronic communications network ("ECN") using such feeds – facilitating, for example, their ability to analyze the implications of a particular trading strategy. But some SIFMA members believe that direct market data feeds may be used by third parties to generate more implicit information about the markets. For example,* ***member firms state that direct market transaction information may be linked to particular displayed quotations and, in some instances, direct market data may be used to help discern the presence of reserve orders.*** *As discussed below, SIFMA does not believe that the use of trading strategies used to identify potential liquidity in various markets, whether displayed or undisplayed, necessarily requires a regulatory response. However, it might be beneficial for market participants to have a better understanding of the ways in which their market data, if provided to a trading center publishing direct market data, might be used by other market participants."*[3]

Conclusion

Most institutional and retail investors have no idea that the private trade information they are entrusting to the market centers is being made public by the exchanges. The exchanges are not making this clear to their clients, but instead are actively broadcasting the information to the HFTs in order to court their order flow. The exchanges are likely to counter that when a subscriber signs up to their exchange they then allow the exchange to use this data as they see fit. However, how many investors would have signed that agreement knowing that their hidden orders were being exposed? This practice has been going on for years but not many investors have read the market data specifications. Every day high frequency traders are using the information that some exchanges are supplying to disadvantage unsuspecting investors.

Every time a trader places an order in certain market centers, whether at the market centers directly, or through a third-party DMA, those market centers are collecting data regarding the trader's order flow. They are supplying the information to HFTs that allows them to track when an investor changes price and how much stock has been accumulated. This information is helping HFTs predict short term price movements. Institutional as well as retail footprints are being detected, and "modus operandi" and trading profiles are being created. Traders believe that their trading strategies are protected, when actually their strategies (personal data) -- including variables such as displayed quantity, time stamp, side, revisions, reserve orders, linked executions, order id numbers, accumulations, number of shares -- are being misappropriated by the market centers.

The exchanges believe that they own the data and have the right to distribute it. Why are the exchanges allowed to do this? This is an outrage and demands immediate action by the SEC. How can the public trust the very organizations that are supposed to be protecting them, when these organizations are turning around and providing their personal data. The only difference between personal data theft, and the data-feed issue we highlight in this paper, is the degree of public awareness.

[3] http://sec.gov/comments/s7-02-10/s70210-167.pdf

Themis Trading LLC, 10 Town Square, Suite 100, Chatham, NJ 07928
Phone: (973) 665-9600 / Web: www.themistrading.com