A Question of Fairness

How to level the playing field in the financial markets.

By Jerome Simonoff

Today's inequitable computerized trading practices favor a few powerful investors over the majority. But there is an efficient, economical technological solution.

That's the topic of this Fair Trading whitepaper.

Executive Summary

The financial markets and the trading exchanges that support them depend on a public perception of fairness. This perception is essential to maintain the confidence of investors.

At the present time, however, there are a number of technological vulnerabilities in the market that give a small number of ultra-sophisticated traders a decided advantage over others. High-Frequency Trading, Algorithmic Trading, Ultra-Low-Latency Trading and Collocation are the most obvious examples of this destabilizing trend.

In addition, given the current state of the network architecture that supports most of the major exchanges, it is now possible for unethical traders to use technology to delay, disrupt and gain unauthorized access to the orders of other buyers and sellers. Given the windfall profits that can be made, it is reasonable to assume that if the possibility for gainful wrongdoing exists, unscrupulous traders will eventually turn it into an actuality.

Regardless of the scope and nature of the problems, it is the responsibility of self-regulating exchanges and regulatory bodies to correct any trading inequities as soon as they become apparent. Unfortunately, effective oversight on these very important issues is almost non-existent for two basic reasons. The public is not yet fully aware of the transformative dangers posed to the market by inequitable practices. In addition, the current set of regulatory tools does not include a practical, economical solution.

There is, however, an efficient and cost-effective way for exchanges and regulators to address the problems noted above, provide a comprehensive audit trail for both regulatory bodies and the public at large, and level the playing field for all traders and investors.

This solution—an Exchange Gateway Device based on proven network concepts and architecture—can ensure fairness and transparency, moderate volatility, and support the original free market goals of equitable trading and efficient capital formation.

Fairness in the markets depends on equal access to information.

If you visit the website of the U.S. Securities and Exchange Commission (SEC), you will find that the organization's very existence depends on its ability to maintain equal access to information for all investors. An online article entitled "The Investor's Advocate" (http://www.sec.gov/about/whatwedo.shtml) states the following:

"The laws and rules that govern the securities industry in the United States derive from a simple and straightforward concept: all investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it, and so long as they hold it. To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a particular security. Only through the steady flow of timely, comprehensive, and accurate information can people make sound investment decisions."

It's impossible to overemphasize the importance of that last point: The markets need to ensure the transparency of price discovery, a responsibility that was re-affirmed by James Brigagliano, an Acting Director of the Division of Trading and Markets, in testimony to the Senate Banking Subcommittee on October 28, 2009 (http://www.sec.gov/news/testimony/2009/ts102809jab.htm):

"As markets evolve," he said, "the Commission must continually seek to preserve the essential role of the public markets in promoting efficient price discovery, fair competition, and investor protection and confidence."

Mr. Brigagliano also addressed another issue that is having a transformative impact on the markets: technological change.

"The U.S. equity markets have undergone a transformation in recent years due in large part to technological innovations that have changed the way that markets operate," he said. And later in his testimony, he added the following:

"An exchange brings together the orders of multiple buyers and sellers and is required to provide the best bid and offer prices for each stock that it trades, as well as last-sale information for each trade that takes place on that exchange. This information is collected and made public through consolidated systems that are approved and overseen by the SEC. *Any investor in the United States can see the best quotation, and the last-sale price of any listed stock, in real time. This transparency is a key element of the national market system mandated by Congress."* (Note: emphasis added.)

In theory, Mr. Brigagliano's testimony is reassuring. Unfortunately, the critical trading information he describes is not available in real-time. It is not available to all investors at the same time. As a result, there is a disparity in information access that undermines the fairness of the market and provides an opportunity for market manipulation and exploitation.

Recognizing the real-time myth.

In most of the major exchanges today, trading is handled by computers that match buy and sell orders. And the speed of execution is almost beyond comprehension. Trades are now routinely executed in fewer than 200 microseconds. (A microsecond is a millionth of a second.) For comparison, the blink of an eye takes about 350,000 microseconds. That gives you some idea of the turbocharged speed of trading today.

In addition to executing trades, high-speed computerized systems generate and distribute reports of trading activity that heavily influence buying and selling decisions. But it is important to remember that in each case the communication is not instantaneous. It is not real-time.

As the information travels from sender to receiver, an infinitesimal, but measurable amount of time elapses. And geographical distance is a key factor in the timeframe involved. That's why it takes more time to send trading information from New York to London than it does from the Bank of New York Mellon at One Wall Street to the NYSE Euronext exchange down the block. And that communication takes more time than it does to send an order to the exchange from a broker located in the same building.

To a layman, the incredibly short delay between sending and receiving—a factor network computing experts call "latency"—may seem irrelevant. But in a world where an increasing number of trades are managed by computers using sophisticated algorithms to make unthinkably fast decisions, a few microseconds difference in price discovery or order execution could give one trader a significant and highly profitable advantage over another.

In a March 2, 2012 blog on the *Huff Post Business* section entitled "Business at the Speed of Light: What Is a Millisecond Worth?" Tony Greenberg, the CEO of RampRate, a consultant for IT and cloud computing sourcing decisions, captured his Chief Technology Officer's view of the ultra-low-latency arms race:

"'It can be difficult to imagine how milliseconds or nanoseconds of latency make a significant difference,' says Internet technologist and RampRate CTO Steve Hotz, 'but from the viewpoint of a data transaction making the trip hundreds or thousands of times, that incremental advantage can add up."

Understanding the problems caused by collocation.

This critical split-second advantage is one of the reasons why High Frequency Traders—which account for only about 2% of today's 20,000 trading firms, according to a survey by the AITE Group—are so interested in placing their trading technology as close as possible to the supercomputing systems operated by the major exchanges.

This practice—called "collocation—could involve a building close to the exchange or even space in same building. In fact, some exchanges offer collocation services directly to their customers, a questionable activity that has raised concerns at regulatory agencies like the SEC.

According to a March 1, 2012 article on the website *Wall Street & Technology*, Nasdaq OMX established the lowest latency route from the New York metro region to Brazil's leading exchange in December 2011.

The new connection provided Nasdaq customers with a 2-millisecond round-trip data transmission advantage over the route operated by NYSE Euronext. The difference is less time than it takes for a housefly to flap a wing. But clearly it was meaningful to Nasdaq customers who are willing to pay for the benefits of collocation.

"When trading advantages are measured in mere thousandths or millionths of a second, co-location could be the difference between success and failure," explained David S. Hilzenrath in a February 22, 2012 article on High-Frequency Trading published in *The Washington Post*.

From a competitive standpoint, it makes perfect sense for the exchanges—many of which now operate as for-profit corporations--to offer the best possible services to their customers and to seek meaningful differentiation.

Nevertheless, the trend toward collocation and the dramatic increase in High-Frequency Trading pose significant threats to the fundamental principle of fairness that's a professed goal of regulators and is essential to the long-term health of the financial markets.

In fact, in a 2011 report to Congress recommending major changes in the organization of the SEC, the Boston Consulting Group pointed out that today's computerized, high-speed trading opens the door to market manipulation and potentially creates "an uneven playing field." (The report was cited in *The Washington Post* article noted above.)

As you can see, this concern about inequitable trading is based on commonly recognized market practices. But there are other issues that threaten to undermine the operations of the major international exchanges.
Here are some prominent examples.

The new profit potential of classic arbitrage strategies.

If you can move faster than the competition with Ultra-Low-Latency and High Frequency trading techniques, you can reap enormous profits on virtually every liquid security asset class by exploiting temporary deviations in price on investments offered in multiple markets.

For example, a classic arbitrage strategy like covered interest rate parity in the foreign exchange market profits from the relationship between the price of a domestic bond, the price of a bond denominated in a foreign currency, the spot price of the currency, and the price of a forward contract on the currency. But that's just one application of this approach.

With help from sophisticated algorithms and supercomputers, arbitrageurs can take advantage of increasingly complex models that utilize considerably more than four securities. And they can literally take billions of dollars out of the market that might have benefitted other investors, including institutional investors, endowments and pension funds.

The disruptive force of flash trading.

Hedge funds and other high-powered traders use this controversial computerized practice to issue orders and then cancel them--within the allowable timeframe--before they are filled.

As a result, they are able to see orders from other market participants before the information is available to everyone. This unfair advantage gives them insights into how others are trading and helps them gauge supply and demand. The SEC proposed banning the practice in 2009.

"Concerns have been raised that the use of flash orders by exchanges and other markets may detract from the fairness and efficiency of the national market system," Mary Shapiro, Chairwoman of the SEC, said at the time. "Specifically, flash orders have the potential to discourage the public display of trading interest and harm quote competition among markets.

"In addition," she continued, "flash orders may create a two-tiered market by allowing only selected participants to access information about the best available price for listed securities. Investors that have access only to information displayed as public quotes may be harmed if market participants are able to flash orders and avoid the need to make the order publicly available." (Source: Shapiro Speech Before the SEC Open Meeting, September 17, 2009)

Although some advocates believe that flash trading enhances the liquidity of markets, critics are concerned that it increases market volatility and can cause disruptive "flash crashes" like the one that occurred on May 6, 2010 when the Dow Jones Industrial Average lost almost 1,000 points in a single day.

Flash trading also has an unfortunate and under-recognized impact on "stop loss orders" which are offered by brokerage houses and online trading venues to unsophisticated, casual investors and the general public.

Here's how it works. Traders use flash trading schemes to discover the price points of stop-loss orders and automatically trigger the sale of the shares involved. Then high-speed computers quickly scoop up the shares at an artificially lowered price, which sabotages the price protection presumably offered by the stop-loss order trading option in the first place.

This questionable practice clearly violates the principles involved in fair trading, especially when you consider that most consumer-oriented computerized trading venues still offer stop loss services.

Emerging opportunities for market manipulation.

As you can see, the activities noted above can have an adverse impact on our financial markets. But there are other threats that may prove to be even more disruptive and abusive.

- *Deliberate interference with data transmission.* The network topology and transmission techniques used by the financial system today are vulnerable to intentional interference. Based on an Ethernet-like model, this communication system uses a Corruption Recognition scheme similar to the Collision Detection protocol that serves as an international standard for network management according to IEEE 802.3 and ISO 8802.3.

  This widely used protocol helps networks ensure data integrity when interference (sometimes called "jitter") causes packets of data to be corrupted. For example, corruption can occur when two devices on a network try, either deliberately or other normal circumstance, to send data over the same channel at the same time. The signals collide, and that collision causes data corruption.

Network protocol will recognize that corruption of a packet has occurred and so the network quickly recognizes the problem. After a specified amount of time, the transmitting device is instructed to re-send the data.

In normal operating environments, most data corruptions are not a cause for concern, since the incremental delay in transmission time has little or no effect on the parties involved. But in a high-speed trading environment, the increased latency of a corrupted communication could have a major impact on the ultimate failure or success of a trade. After all, corrupted market communications that necessitate re-transmission would operate at a decided disadvantage to uncorrupted trades with their lower latency.

Here's something else to consider. Since it is technically possible to deliberately cause interference that will corrupt communications coming from another address on a shared network, it is only reasonable to assume that unethical traders have adopted—or will eventually adopt—this underhanded practice as a way to gain an unfair advantage over competitors using the same wire or optical fiber communication path.

What makes the situation even more unsettling is this: It is very difficult to prove that data corruptions are deliberately caused. It is also nearly impossible to identify the responsible party based on current IT forensic capabilities. As a result, perpetrators will run virtually no risk of detection until a solution is found.

- *The potential for "cracked trades."* Even though most current Internet activity is protected by encryption, the possibility exists that trading communications could be "hacked" and decrypted as well as delayed through intentional data packet corruption.

  Certainly, the infrastructure for cracking trades exists today with the rapid proliferation of extremely fast computers and the growing number of expert hackers with the skills and tools necessary to develop sophisticated decryption algorithms.

  In addition, there are a number of related activities like industrial espionage, cyber terrorism, "hacktivism," and Internet-based crimes like "phishing," malware insertion and "man in the middle" eavesdropping attacks that are steadily advancing the art and science of cybercrime, hacking and decryption.

  In evaluating this increasing threat, it's important to remember two additional points.

Cyber criminals do not have to crack encryption keys in real time to discover "secrets." If a hacker can crack the key at any point during the timeframe that the

key is in use, he or she will have unfettered access to encrypted trade information. As a result, the hacker will have an unfair trading advantage until the "cracked" key is changed.

- Even in cases where encryption keys cannot be cracked, the IP addresses for senders and receivers are not encrypted. This pervasive security loophole exposes some trades to potential exploitation and abuse.

The alarming vulnerability of modern trading systems.

At present, regulators may not have firm evidence of deliberate interference, cracked trades, snooping, and other forms of Internet-based market espionage. But if you examine published data on TCP/IP connectivity and transmission technology, it's clear that the infrastructure used to support trading and other market activities is highly vulnerable to exploitation.

It's also important to remember that leading security vendors like Verisign and Symantec have already been "hacked" by Internet intruders, proving that significant vulnerabilities exist even in best-in-class systems.

In light of all of these issues and threats, the most prudent course for both regulators and self-regulating exchanges is to assume that the possibilities outlined above will eventually turn into probabilities, given the immense rewards that can be reaped by unfair and unethical trading activities.

The increasing challenge for regulators.

In a relatively short span of time, technological developments have revolutionized trading in the world's financial markets. In the beginning of the 21st century, most of the orders in the U.S. were still executed manually and by telephone. Today, only a dozen years later, almost all of the traffic is handled electronically.

Obviously, the marriage of computing technology and trading has brought many benefits to the markets, including increased speed and efficiency and substantially reduced trading costs. But the use of increasingly sophisticated technology has led to a number of new dangers and risks that regulators and self-regulating organizations have yet to control.

Here's one noteworthy example. Technological advancements have made it possible for new trading venues to emerge outside of the traditional exchange infrastructure. These venues, in turn, have given rise to arcane trading practices like "dark pools" that help market participants disguise large volume trades and maintain more favorable pricing than would exist in a truly transparent market.

In addition, it is important to remember that our complex, interconnected, computer-driven global trading environment makes it very difficult for regulators to perform the forensics necessary to detect market manipulation and identify the perpetrators.

After all, regulators—who have the primary responsibility for ensuring market transparency and fairness—often operate with IT systems that are decidedly inferior to the sophisticated, state-of-the-art platforms used by High Frequency and Algorithmic Traders.

Of course, regulators also have to deal with a crushing volume of traditional crimes like insider trading and fraudulent investment schemes that tax their limited resources. And their efforts to keep pace with all of these market manipulators suffer from the same budgetary constraints that affect all governmental organizations.

Possible conflicts of interest complicate self-regulation.

The exchanges and other self-regulatory organizations also struggle to maintain a public perception of fairness. After all, some large, profitable customers use High-Frequency Trading strategies, collocated technology platforms and Flash Trading techniques to gain a technological advantage over the competition.

In many cases, these customers operate in gray areas of the law where there is no clear path to enforcement. But even in cases where there is a clear violation of the rules or principles of ethical conduct, the exchanges have to think twice about imposing punishment, because they run the risk of losing these highly profitable customers to other trading venues.

When you consider all of these issues and trends, it's easy to understand why it's so difficult for regulators, exchanges and other self-regulating organizations to maintain fairness and transparency in price discovery and trading.

Nevertheless, it is essential to find a practical way to resolve these problems to maintain the credibility of regulatory organizations and the markets themselves. Otherwise, more and more investors will gradually discover that they are operating at a decided disadvantage to a small number of ultra-sophisticated traders. And that will inevitably lead to a heightened demand for new legislation and regulation designed to ensure a level playing field.

In addition, many investors may decide to terminate their relationships with exchanges that do not enforce fair trading policies when offered a choice of exchanges that do enforce these policies, and seek out more equitable venues.

This looming "flight to fairness" could have a wide-ranging impact on markets, exchanges, regulators and the capital formation process.

The search for a practical solution.

Clearly, technology has opened the door to a number of unfair and unethical trading practices that currently outpace the best efforts of regulatory organizations to control them.

The ultimate challenge, however, is not in recognizing the problems. The ultimate challenge is to find a practical and economical solution that addresses these inequitable activities:

- Some participants use their technological superiority to make trades at a faster pace than other investors.

- Some traders use technological trickery to gain unfair advantages in terms of market manipulation and price discovery.

Using technology to address inequitable trading practices.

In theory, the solution is readily apparent. You equalize the latency of trades and ensure that buyers and sellers meet in an open marketplace ruled by a transparent price discovery process. But in practice, those goals seem to lie beyond our reach in today's overtaxed regulatory environment.

What many people fail to realize, however, is this: It is possible to use current technology to address the problems noted above and implement a reliable mechanism for regulatory oversight. These steps, in turn, will help restore the principle of fair trading for all investors regardless of the scale and scope of their financial resources.

Bringing more control to the management of trading traffic.

Exchanges today operate in a digital world that links together two basic communication systems: the Local Area Networks that operate inside the firewalls of a trading organization and the Wide Area Network that lies outside their control. The Internet is, of course, the world's largest Wide Area Network.

In many topological configurations, network engineers use an Internet Gateway Device—which is a hardware and software system—to add functionality to the connection between a Local Area Network and a Wide Area Network.

This technology has great potential in terms of the network architecture of an exchange, because of its ability to provide more control over incoming trading traffic. So for the sake of simplicity, let us call this particular market-focused application an "Exchange Gateway Device" or EGD.
In terms of implementation, the EGD will be integrated into the Local Area Network of an exchange inside the exchange's security firewall. By design, it will be able to use the exchange's encryption keys to decrypt incoming traffic from the Wide Area Network. As a result, it will be able to break down the contents of incoming messages and identify the trading symbol, price, type of trade request, and other pertinent information.

Once this step is completed, the EGD will store each trading message in an electronic message queue. After a defined Queue Interval timeframe, these stored trade requests will be transferred to a Release Batch where individual trades will be sorted in an order determined by the Fairness Algorithm. Trades will then be released to the exchange's trading computers for final fulfillment.

In the traditional approach, trading messages received by an exchange are forwarded to trading computers for fulfillment in the order in which they were received. But this process ensures that traders who successfully exploit latency will gain an unfair advantage.

The EGD, on the other hand, provides the capability to negate the effect of the latency of trades and establish an equitable queue order determined by the controlling Fairness Algorithm—a flexible algorithm that can be configured in a number of ways.

Here's one example of a practical configuration.

To ensure trading fairness and efficiency, an exchange could configure the Fairness Algorithm to set the Queue Interval at approximately one second. The decision could also be made to establish a sort sequence in which "buys" in the Release Batch would be processed first followed by the "sells."

In each buy or sell sub-group, trades would be arranged in the appropriate “limit order” price sequence and processed in descending order according to price.

This approach would ensure that buys with the highest price get matched with the highest priced sells. Market orders would be ranked at the top of each group in the sort sequence.

Practical experience and an analysis of trading activity will provide the insights needed to fine-tune both the Fairness Algorithm and the Queue Interval to achieve the optimal level of performance and fairness for all of the stakeholders involved: exchanges, investors and regulatory organizations.
Naturally, these controlling algorithms can be formulated in a variety of ways. But the goal would be to establish a “Fairness Algorithm” that would take latency out of the factors in trading activity and empower trading from all locations however distant from the exchange while maintaining the optimal level of trading efficiency we would in fact expand the market opportunities in this shrinking worldwide marketplace. In addition, the exact nature of the algorithm and any changes to the algorithm should be made public at all times to ensure transparency.

Because of its ability to negate latency as a significant factor, the EGD’s Fairness Algorithm will also serve as an effective countermeasure against snooping, the intentional corruption of packets, and “front-running” Since market participants wouldn’t be able to take advantage of reduced latency, they will have no incentive to engage in these unethical practices.

Of course, the development of the Fairness Algorithm would initially be a challenging task requiring the active involvement of all market stakeholders. It would also undoubtedly involve some trial and error.

But in a market environment that is spinning out of control, it is important to remember this point: It is technically possible to achieve the goal of global latency equalization with an EGD and an effective Fairness Algorithm.

It's really not that difficult from a technical standpoint.

An effective tool for regulatory oversight.

Another key advantage of the EGD is that it can be designed to automatically compile a comprehensive audit trail of all trading activity. This audit trail can then be analyzed by regulators to ensure equitable trading, maintain transparency, and identify any anomalous activities.

In fact, regulators could develop automated systems to scrutinize trading on an continual basis and make sure that the published Fairness Algorithm is, in fact, the one in use and is achieving its stated goals.

Thanks to these capabilities, the EGD will increase the administrative control of self-regulating organizations, improve the ability of regulators to do their job, and make trading more transparent—all of which will help restore confidence in our financial markets and the exchanges that support them.

A flexible solution for today and tomorrow.

Since it is based on a technological concept that's widely used today, the EGD will easily interface with existing communication and trading platforms without requiring major changes or re-engineering. In fact, the entire implementation would be efficient and economical.

Thanks to its inherent flexibility, the Fairness Algorithm will also be easy to adjust to maintain efficiency during changing market conditions. As a result, the EGD will be able to adapt to peaks and valleys in trading volume, different times of day, and other factors that affect exchange operations.

In addition to meeting the operational requirements of today's volatile markets, the technological flexibility and adaptability of the EGD will ensure its long-term relevance, since trading on the financial markets will continue to be affected by new IT developments and innovations.

Restoring the principle of fair trading to today's volatile financial markets.

The world's leading exchanges and other high-volume trading venues will reap a number of benefits from the implementation of the EGD and its Fairness Algorithm.

They will be able to prove their commitment to equitable trading practices. They will significantly improve the transparency of buying, selling and price discovery. They will also help minimize the disruptive volatility that can generate windfall profits for market manipulators.

Over time, as investors' perception of unfair trading heightens, these "fair trading" exchanges may also be able to gain a powerful competitive advantage by winning the public trust, earning the approval of regulators, and attracting the vast majority of people who want to participate in a free market on a level playing field.

The Exchange Gateway Device is not the complete solution to the problem of abusive trading practices. But it represents a significant step forward in the right direction.

For more information, visit www.fairtrader.us

About The Author

Jerome Simonoff is an investor, a technological innovator, and the President of CHAX, Inc., an independent software development company specializing in funds transfer software.

In addition to inventing solutions for the financial services market, he is a highly regarded IT consultant and entrepreneur. He can be reached by emailing jerry@chax.com