Harvard School of Engineering and Applied Sciences

Technical Report TR01-10

Intention-Disguised Algorithmic Trading

William Yuen[1], Paul Syverson[2], Zhenming Liu[1], Christopher Thorpe[1]

[1]Harvard University, [2]Naval Research Laboratory

January 20, 2010

Abstract

We propose a general model underlying the problem of designing trading strategies that leak no information to frontrunners and other exploiters. We study major scenarios in the market and design a family of algorithms that can be proven to leak no information in important scenarios. These algorithms can serve as building blocks for more challenging real-world scenarios beyond our current scope. In contrast to previous work, the strategies we propose protect trader using the existing trading infrastructure.

1 Introduction

Large market participants (LMPs) must often execute trades while keeping their intentions secret. Sometimes secrecy is required before trades are completed to prevent other traders from anticipating (and exploiting) the price impact of their trades. This is known as "front-running". In other cases, LMPs with proprietary trading strategies wish to keep their positions secret even after trading because their strategies and positions contain valuable information. LMPs include hedge funds, mutual funds, and other specialized market players.

In some cases, large order information is leaked when indiscreet brokers share it, or when rogue traders collude to exploit insider information. In recent years, alternative trading systems and dark pools, such as BIDS, BATS, Liquidnet, Pipeline and POSIT have been developed to reduce this inefficiency and information leak in block trading. Even so, new strategies to "probe" such systems attempt to divine pockets of liquidity and exploit that information. Some firms go further and attempt to reproduce successful funds' strategies by monitoring their disclosures and trying to reconstruct their order flow. Even if information is not leaked to the market, order information can be exposed internally within an ethically questionable broker, especially when the broker also conducts proprietary trading.

However order information is leaked, or why it is sought, traders who exploit others' order information extract value from markets at the expense of the LMP. Thus, hedge funds and other firms take great pains to hide their intentions, even generating "noise" trades to hide their intended positions from other traders [2]. We present trading schemes that disguise an LMP's intentions and positions from *any* other entity, including the brokers that the LMP interacts with.

Various studies [13, 12] have shown abnormal price behavior and significant negative price impact from information leakage prior to a block trade execution. Thorpe and Parkes [16, 17] discuss cryptographic and security research on exchanges and how information can be exploited in financial markets. But, existing research generally proposes new infrastructures or protocols, for which adoption is notoriously difficult. We take a simpler approach. Our specific contributions are: (1) to propose a general model underlying the design of trading strategies that leak no information, (2) to study major scenarios in the market and design associated algorithms *that require no changes to the existing trading infrastructure*, and (3) to prove those algorithms leak no information in those scenarios.

These algorithms can serve as building blocks for more challenging real-world scenarios beyond our present scope. Though our approach is algorithmic, we are not concerned with volume-weighted algorithmic trading. See [4], [5] and [11] for a review of the literature and for insights into the study of automated trading.

In Section 2, we discuss existing trading infrastructure, define three types of adversaries, and present ways they can extract information from orders placed by the LMP. In Section 3, we describe the model for information leakage and address the needed properties for an efficient trading strategy. Section 4 introduces different trading strategies that disguise the intention and holdings of the LMP from exploiters. We also evaluate their defensive performance against each of the three types of "exploiters". Section 5 explores the cost estimates of our strategies.

2 Preliminaries

Existing trading infrastructure and exploiters

We use the term *brokers* to include brokers, dealers, and broker-dealers. *Shares* are units of any security, including equities, bonds, currencies, or derivatives. One can long or short any of the shortable securities represented through brokers. Each transaction is for a nonzero integer number of shares; although LMPs typically trade in increments of at least 100 shares. When a trade is executed, the symbol and quantity is publicly reported by the exchange. Typically, only the broker involved in the trade knows the identity of the LMP and whether the LMP was the buyer or seller. The LMP pays commissions and fees for executed trades, usually proportional to the trade volume. *Exploiters* exploit any information they obtain and can be categorized into three classes, from weakest to strongest:

1. *Curious Observers* are able to see trades printed as they are executed and/or the prices and sizes of orders (requested trades) as they are quoted. These observers can, for example, exploit an LMP who splits a large order into sequential smaller blocks of orders. With sufficient intelligence and experience, curious observers may be able to guess the identity and intention of the LMP.
2. *Individual Curious Brokers* are able to see trade orders by the LMP before they are executed. A corrupt or careless broker can leak the LMP's intentions for exploitation by broker insiders or external agents. Using multiple brokers, the LMP can limit the information a single curious broker can extract.
3. *Colluding Curious Brokers* are able to see trade orders by the LMP and can share their information with each other. If all brokers used by the LMP are curious and collude, any benefits resulting from splitting trades across different brokers would be lost. However, all brokers used by the LMP must collude in order to yield complete knowledge.

The following example demonstrates what each type of exploiter can see:

Example 1. Assume a trader buys 100k shares of ATK from broker b_1 and sells 200k shares of ATK to broker b_2. Then

- A curious observer sees an unsigned trade volume (ATK: $100k$) from b_1 and (ATK: $200k$) from b_2.
- Broker b_1 sees a signed trade by the LMP of (ATK: $+100k$) from his own transaction and an unsigned volume (ATK: $100k$) from b_2.
- Broker b_2 sees a signed trade by the LMP of (ATK: $-200k$) from his own transaction and an unsigned volume (ATK: $100k$) from b_1.
- If brokers b_1 and b_2 collude, they can infer the net directed volume (ATK: $-100k$) by the LMP.

In all our strategies the LMP places a set of orders for an asset d at one or more brokers in order to yield a net purchase or sale of d. Using minimal resources, we want to prevent reasonably capable exploiters from guessing the net order. Strategies must also stay completely effective even when exploiters are aware that the LMP is using them. We focus primarily on the following scenarios:

- *Multiple brokers with one trader* (nB1T): There is only one trader represented in the market, the LMP. This trader can interact with many brokers.
- *Single broker with multiple traders* (1BmT): Only one broker handles trades, for the LMP and possibly for other market participants.
- *Multiple brokers with multiple traders* (nBmT): Multiple traders and multiple brokers can trade simultaneously, the most general scenario.

Two motivating trading strategies and why they leak information

To hide the net order we first consider a simple approach where the LMP uses two brokers and places an order with each one so that neither broker individually learns about the net order. This simple strategy still allows brokers to

Stock	Shares through Broker b_1	Shares through Broker b_2	Net LMP volume traded
ATK	+500k	−500k	0
SXL	−300k	+400k	+100k
BRCD	−3000k	+2000k	−1000k

Table 1. An example of disguising true trading intentions using two brokers.

extract some knowledge. For example in Table 1, Broker b_1 observes that a large ATK trade has gone through Broker b_2 when Broker b_2 prints the block trade after execution. Broker b_1 is not sure whether the LMP is involved in the trade with Broker b_2, or the direction of the LMP's trade through Broker b_2. But, A knows that his large client has traded one of three net positions: 500k, 500k + 500k = 1M, or 500k - 500k = 0. Similarly, Broker b_1 knows his client's net trade for SXL is 100k, -700k, or -300k shares. Because the number of possible cases is low, Broker b_1 can analyze each scenario and deduce the best exploitation

strategy. For example, a block trade price that is closer to the bid than to the offer is more likely to be seller-initiated [10].

Another simple strategy is to use a single broker but multiple registered traders. An LMP might create several registered trading agents that are not known to be associated with the LMP but trade on its behalf. Thus we could produce the exact same structure as Table 1 except that now instead of a single trader using Broker b_1 and Broker b_2, we have Trader d_1 and Trader d_2, both responsible for the same book of the LMP, trading through a single broker. The broker cannot tell from what he sees if he is dealing with one LMP shopping two blocks or two LMPs. This defends against collusion, unlike the two-broker system. But, if the broker links the two pseudonymous traders together, then he will know everything about their intentions going forward. We can combine the two solutions so that each pseudonymous trader is splitting orders across multiple brokers. This gains both the advantages and the overhead of both approaches.

3 Defining "information leak"

A rigorous definition of "information leak" is needed to understand both the potential threats from exploitation for the LMP and the desired properties of the trading strategies we are seeking. We propose in this section three types of information leak (or rather its absence) so as to formalize the notion: *zero information leak*, *ϵ-information leak*, and *full space strategy*.

Our inspiration is Goldwasser et al.'s [9] notion *zero knowledge*. Roughly, transmitting a piece of information is zero-knowledge if the universe of computations the recipient can perform does not change after receiving the information.

Definition 1. (Zero information leak) *Let $(\Omega, \mathcal{F}, \Pr)$ be a (discrete) probability space that represents all possible positions an LMP might intend to trade and the corresponding a priori distribution over these positions. Let M be the message (which can be a random variable over a probability space other than Ω) that represents the trading records of the LMP observed by exploiters. We say that the LMP's trading strategy leaks zero information with respect to the exploiters if, for any random variable X defined over Ω, we have that the distributions of X and $X \mid M$ are the same.*

Remark 1. When we are only interested in one specific symbol, we shall view Ω be all possible shares an LMP can long or short. It is thus sufficient to assume Ω is a discrete space. The trading strategy we are looking for could be a randomized one, which suggests M should be a random variable. Additionally, the coin tosses of the randomized strategy are necessarily private. The random variable M might be different from Ω. The function X should be viewed as the information related to the LMP's net position that a exploiter wants to extract. For example, X could be a threshold function that indicates whether the shares traded exceed 10,000.

Since Ω is also a subset of the real numbers, the identity function $I(\omega) = \omega$ for all $\omega \in \Omega$ is also a random variable. Furthermore, for any other random

variable f, we have $\sigma(f(\Omega)) \subseteq \sigma(\Omega)$. Therefore it is indeed sufficient for a trading strategy to leak no information that the distributions on Ω and $\Omega \mid M$ are the same. On the other hand, this equality suggests that the random variables M and Ω are independent of each other, which also means that M and $M \mid \Omega$ have the same distribution.

If they are the same, any observer should able to simulate the random variable M on her own without seeing Ω. This may be closer to the traditional definition of zero knowledge: an interactive protocol is said to be zero knowledge if verifiers are able to simulate the transcript on their own [7].

We shall also be able to compute the trading strategy efficiently. The following definition implicitly gives the trading strategy we are seeking.

Definition 2. (Efficient algorithms for zero information leak) *Let* $(\Omega, \mathcal{F}, \Pr)$ *be a probability space that represents all possible intended positions of an LMP and the corresponding a priori distribution over these positions. Let* ω *be a random sample from* Ω*. A trading algorithm* $\mathcal{A}$ *is said to be* perfect-zero-knowledge *with respect to exploiters if the following two conditions hold:*

- $\mathcal{A}$ *can generate an execution plan in polynomial time (wrt a reasonable representation of* Ω*) that ends with the LMP holding exactly* ω *shares.*
- *The exploiters are able to generate the distribution on the random variable* M *on their own without seeing the signal* ω*.*

Remark 2. Regarding the "reasonable representation" of Ω: though it is not specified precisely, in most cases the natural representation for Ω will be well understood. For example, if there is no prior knowledge of Ω, then the size of Ω's representation is $\log_2(|\Omega|)$.

A natural relaxation of zero information leak is to allow ϵ information leak. The definition is essentially the same as this except that exploiters can generate a random variable with a statistical difference[1] from M of at most ϵ. One may think of the difference between perfect zero knowledge and statistical zero knowledge [7] to understand the motivation for this relaxation in security definition.

Definition 3. (Efficient algorithms for ϵ-information leak)*Let* $(\Omega, \mathcal{F}, \Pr)$ *be a probability space that represents all possible positions intended to be traded by an LMP and the corresponding a priori over these positions. Let* ω *be a random sample from* Ω*, a trading algorithm* $\mathcal{A}$ *is said to be an* ϵ-information leak *algorithm with respect to exploiters if the following two conditions hold*

- $\mathcal{A}$ *can generate an execution plan in polynomial time (with respect to a reasonable representation of* Ω*) so that by the end of the execution the LMP holds exactly* ω *shares.*

[1] the statistical difference between two discrete random variables X and Y is defined as $\sum_i |\Pr[X = i] - \Pr[Y = i]|$

– *The exploiters are able to generate a random variable* $\widetilde{M}$ *on their own without seeing the signal* ω*, where statistical difference*[2] *between* $\widetilde{M}$ *and* M *is at most* ϵ*.*

Finally, we propose another way to ensure sufficient noise that an adversary is unable to eliminate any possible values from Ω. Specifically we require that $\Pr[\omega \mid M = p] > 0$ for all q and all ω such that $\Pr[\omega] > 0$.

Definition 4. (Efficient algorithms for full space strategy) *Let* $(\Omega, \mathcal{F}, \Pr)$ *be a probability space that represents all possible intended positions of an LMP and the corresponding prior over these positions. Wolog, assume that* $\Pr[\omega] > 0$ *for any* ω*. Let* ω *be a random sample from* Ω*. A trading algorithm* $\mathcal{A}$ *is said to give a* full space strategy *with respect to exploiters if the following two conditions hold:*

– $\mathcal{A}$ *can generate an execution plan in polynomial time (w.r.t. a reasonable representation of* Ω*) that ends with the LMP holding exactly* ω *shares.*
– *For any message* M *observed by the exploiters,* $\Pr[\omega \mid M] > 0$ *for any* $\omega \in \Omega$*.*

Example 2. (A negative example of full space strategy) Let $\Omega = [-500k, 500k]$. Suppose the trading target is to buy $200k$ shares of ATK and the trading strategy $\mathcal{T}$ is to buy $250k$ from broker b_1 and sell $50k$ to broker b_2. The trading strategy $\mathcal{T}$ does not have *full solution space* because an external observer can immediately eliminate the possibilities from the ranges $[-500k, -300k]$ and $[300k, 500k]$.

Practicality typically limits Ω. For example, when a security has low liquidity, an external observer can simply sum up all the shares that are traded within a specific time window to derive the upper and lower bounds for Ω. Therefore, we view $|\Omega|$ as a security parameter for an LMP. We can increase its value to get a safer system, at the cost of paying higher transaction costs. The security parameter needs not be secret.

Although there are more refined notions of knowledge, e.g., that quantify the exact number of bits leaked by a system [8], it is unclear how the amount of leaked information relates to the financial cost of the information. A single leaked bit information can have great value (the sign of an order issued by an insider), but other times even a large information leak may be harmless.

4 Trading strategies

In this section, we design and analyze trading strategies to counter various adversaries in various markets, and in progressively more challenging scenarios.

Multiple brokers with one trader (nB1T)

In order to defend against the three types of exploiters mentioned, we first

[2] the statistical difference between two discrete random variables X and Y is defined as $\sum_i |\Pr[X = i] - \Pr[Y = i]|$

build our strategies using a single trader and n orders placed with n different brokers. We call this *the nB1T platform*. We start with nB1T strategies for the LMP against *curious observers* (the weakest). The following sign flipping game is closely related to a trading strategy that leaks no information:

Definition 5. *(*Sign Flipping Game*) Given an interval $[-q, q]$, find a set of numbers $T = \{t_1, t_2, ..., t_n\}$ such that $\sum_i t_i = q$ and*

1. *For any integer $x \in [-q, q] \cap \mathbb{Z}$ there exists a set of numbers $a_1, a_2, ..., a_n \in \{-1, 1\}$, $t_i \in \mathbb{Z}$ such that*

$$x = a_1 \cdot t_1 + a_2 \cdot t_2 + ... + a_n \cdot t_n. \tag{1}$$

2. *The number n is a function of q. The value of n should be as small as possible.*

Intuitively, for our strategy utilizing Multiple Brokers with One Trader (nB1T), n in the sign flipping game is the number of brokers the LMP interacts with, and $\Omega = [-q, q]$ is the range of net position the LMP wants to hold. By buying or selling volume t_i with broker i, he can construct every possible desired net trading volume, x, bounded between $-q$ and q. Unsigned traded volumes $T_L = \{|a_i t_i|\}$ are printed among other traded volumes W_0 that do not involve the LMP. Observer identification of T_L from $T_L \cup W_0$ depends on market liquidity and other factors. An LMP is always able to set a larger q at the cost of higher transaction costs. When the security parameter q is fixed, a natural goal is to minimize the number of brokers used.

Now, suppose the LMP wishes to buy $x \in [-q, q]$ shares (negative x notated as selling) of a product. She would then be able to execute a sequence of orders $t_1, t_2, ..., t_n$ to each of the brokers such that

$$x = t_1 + t_2 + ... + t_n.$$

From an observer's point of view, he only sees the sequence $|t_1|, |t_2|, ..., |t_n|$. If he does not have information of the LMP's intention a priori, the observer can only attempt to extract knowledge by going through all combinations of the signs for all t_i. Therefore, the LMP's strategy should make the following set as large as possible:

$$S = \{a_1|t_1| + a_2|t_2| + ... + a_n|t_n| : a_1, ..., a_n \in \{-1, 1\}\}.$$

A necessary requirement for a trading strategy being a zero information leak one is $[-q, q] \subseteq S$. Our first goal is to construct $T = \{t_1, t_2, ..., t_n\}$ with minimum possible n such that S fully covers $[-q, q]$. We can find the set T from the following Lemma.

Lemma 1. *There exists a set T with $|T| = \lceil \log_2 q \rceil + 2$ that satisfies the first requirement of the Sign Flipping Game.*

Proof. Let $n = \lceil \log_2 q \rceil + 2$, the following set satisfies the desired property: $T = \{1/2, 1/2, 1, 2, 4, ..., 2^{n-3}\}$.

Now we can prove the statement by induction on n. When $q = 1$ and $n = 2$, we have $-1 = -1/2 - 1/2$, $0 = -1/2 + 1/2$, and $1 = 1/2 + 1/2$.

By hypothesis, if $n \leq N$, and $|x| \leq q$, there exist $a_1, a_2, ..., a_N$ such that $\sum_{1 \leq i \leq N} a_i t_i = x$. And, by definition of n, $2^{n-2} = \lfloor q \rfloor$, so $2^{N-2} = \lfloor q_N \rfloor$

- For any values $x \in [0, 2^{N-2} - 1]$, we thus have that $x - 2^{N-3} \leq q_N$. Thus we can find $a_1, ..., a_N$ such that $\sum_{1 \leq i \leq N} a_i t_i = x - 2^{N-3}$; setting $a_{N+1} = 1$ will make $x = \sum_{i \leq N+1} a_i t_i$.
- If $x = 2^{N-2}$, then $x - 2^{N-3} = 2^{N-3} \leq q_N$. Thus again setting $a_{N+1} = 1$ will make $x = \sum_{i \leq N+1} a_i t_i$.
- For negative values of $x \geq -2^{N-2}$, one of the above two arguments applies, except for some changing of signs and assigning $a_{N+1} = -1$.

Interestingly, the number n in the above Lemma is nearly optimal:

Lemma 2. *For any set T that satisfies the first requirement of the Sign Flipping Game, we have $|T| \geq \lceil \log_2 q \rceil + 1$.*

Proof. The total number of possible assignments to $a_1, ..., a_n$ is at most 2^n. Therefore, we need $2^n \geq 2q + 1 \Rightarrow 2^{|T|} \geq 2q + 1 \geq 2q \Rightarrow |T| \geq \lceil \log_2 q \rceil + 1$.

Next, notice that Lemma 1 already suggests a binary flavored algorithm that runs in poly $\log(q)$ time.

Lemma 3. (Efficient coefficient computation) *Let q, n, and T be values and set defined in Lemma 1, for any integer $x \in [-q, q]$, there exists an $O(n)$ (i.e., $O(\log_2(q))$ algorithm finding a_is such that $\sum_i a_i t_i = x$.*

Finally, since the T set is fixed, it is not difficult to see that the strategy derived from the sign flipping game is optimal.

Theorem 1. (Zero information leak strategy) *There exists an efficient algorithm generating trading strategies that leak zero information to curious individual brokers on the nB1T platform.*

Proof. Let $\Omega = [-q, q]$. The strategy is defined in the intuitive way. Let x be the actual shares an LMP intends to trade. Let T be the set defined in Lemma 1. Then the LMP shall find a_i for $1 \leq i \leq |T|$ such that $\sum_{i \leq |T|} a_i t_i = x$ and she will trade $a_i t_i$ volume with the ith broker. Observe that the set T is fixed and *any* external observer will consistently observe b_i trading $t_i = \max(1, 2^{n-1})$ volumes. This trading record can clearly be simulated.

Corollary 1. *There exists an efficient algorithm for full space strategy against curious individual broker on the nB1T platform.*

The above strategies no longer work against *curious individual brokers who do not collude*. For example, curious broker b_n, upon seeing the value a_n of the order, would know that the LMP is intending to buy from the range $[-q, 0]$ if $a_n = -1$ or $[1, q]$ if $a_n = 1$.

If instead, we randomly split the designated value x into two parts x_1 and x_2 such that $x = x_1 + x_2$, and use two set of brokers $b_1, b_2, ..., b_n$ and $b'_1, b'_2, ..., b'_n$, where b_i trades for x_1 shares and b'_i trades for x_2, then intuitively information gained by b_n will not suffice to narrow down the search range $[-q, q]$ because x_2 can cancel out the order she sees. The following theorem formalizes this idea.

Theorem 2. (Full space strategy for a "curious broker" market) *There exists an efficient algorithm that generates trading strategies that are full space against curious broker in the multiple-brokers-with-one-trader market.*

Proof. Let $\Omega = [-q, q]$ and x be the net number of shares to be traded. Consider the following strategy, namely $\mathcal{S}_1$: first let us randomly split $x = x_1 + x_2$, where $x_1, x_2 \in [-q, q]$. Next we use brokers $b_1, b_2, ..., b_n$ to execute $\mathcal{S}_1(x_1)$ and brokers $b_{n+1}, b_{n+2}, ..., b_{2n}$ to execute $\mathcal{S}_1(x_2)$. We claim that the strategy $\mathcal{S}_1$ is a full space strategy.

Let b_j be a curious broker who learns a sign $s \in \{-1, 1\}$. Let the integer $x \in [-q, q]$ be the designated net shares we wish to trade. We need to show that we can use the rest of the brokers to construct any x in $[-q, q]$. Without loss of generality, assume $j \leq n$. Let $x = \sum_{1 \leq i \leq n} a_i \cdot t_i$. If $a_j = s$, then let $x_1 = x$ and $x_2 = 0$; otherwise, let $x_1 = \sum_{1 \leq i \neq j \leq n} a_i t_i + s t_j$ and $x_2 = (a_j - s) t_i$. Then we will have $x = x_1 + x_2$.

When using the strategy proposed above, however, we will have some degree of information leakage due to the fact that x_1 and x_2 cannot be chosen randomly from $[-q, q]$, which is probably not desirable. For example, when $x = q$, the value x_1 can only lie between $[0, q]$ otherwise the value x_2 will fall out of $[-q, q]$. Turning it the other way, the probability the broker b_n gets a positive sign clearly depends on the input x. And by using simple Bayesian statistics, a curious broker can infer the value x.

Next, we turn to the algorithm for ϵ-information leak, which could be viewed as a generalization of strategy $\mathcal{S}_1$. The algorithm is described as follows: instead of using only 2 set of brokers, we use $2/\epsilon + 1$ set of brokers. Now let $2/\epsilon$ set of brokers to do random canceling ordering (i.e., the 1st set buys x_1, the 2nd set buys $-x_1$, the 3rd set buys x_2, and so forth, where $x_1, \ldots, x_n$ are fully random from $[-q, q]$). This yields the following ϵ-zero knowledge strategy:

Theorem 3. (ϵ-information leak strategy in a "curious broker" market) *There exists an efficient algorithm that generates trading strategies characterized by an ϵ-information leak against "curious individual" brokers on the nB1T platform.*

Proof. Let $\Omega = [-q, q]$. Let us consider the following strategy, namely $\mathcal{S}_2$: let $b_{i,j}$ be the brokers we need to use, where $1 \leq i \leq 2/\epsilon + 1$ and $1 \leq j \leq \lceil \log_2 q \rceil + 2$. Uniformly generate random numbers $x_1, x_2, ..., x_{1/\epsilon}$ from $[-q, q]$. Let $\{i_1, i_2, ..., i_{2/\epsilon}$

be a random subset of $\{1, ..., 2/\epsilon + 1\}$ (i.e., randomly removing one number out of the latter set). Then for $1 \leq j \leq 1/\epsilon$, we let $b_{i_j,1}, ..., b_{i_j,n}$ execute $\mathcal{S}_1$ to buy x_j shares; and let $b_{i_{j+1/\epsilon},1}, ..., b_{i_{j+1/\epsilon},n}$ executes $\mathcal{S}_1$ to sell x_j shares. For the rest n brokers, they execute $\mathcal{S}_1$ to buy x shares. Anyone can generate the absolute volumes for the broker $b_{i,j}$ because this value is deterministically set to $\max(1, 2^{j-1})$. Therefore, the statistical difference comes from the sign a broker gets assigned. On the other hand, the probability that $b_{i,j}$ falls into the group that buys x shares is ϵ. And if $b_{i,j}$ is not in the group buying x, the probability of his/her sign being positive is exactly $1/2$. Hence, for any broker, the sign $s \in \{-1, 1\}$ he/she gets satisfies $|\Pr[s = 1] - 1/2| \leq \epsilon$. As a result, a broker can simulate the sign by letting $\Pr[s = 1] = 1/2$. For this simulation, the statistical difference is at most ϵ.

Finally, note that when $1/\epsilon$ is a constant or a polynomial in n, the above strategy has a $\Omega(\text{poly}(n))$ expansion the number of brokers.

Countering collusion

The above nB1T strategic platform does not yield strong defense against curious colluding brokers: they can share knowledge with each other, including the identity of the LMP and the sets a_i and t_i. If colluders know the total number of brokers used and can find all of them, the value x can be trivially extracted.

Even if n is not known or not all n brokers collude, certain possible values for x can be eliminated: Suppose, for example, the LMP uses two sets of brokers $R = \{b_1, b_2, ..., b_n\}$ and $R' = \{b'_1, b'_2, ..., b'_n\}$, and that $R \cap R' = \emptyset$. Let $B = R \cup R'$. Suppose brokers $B_c \subset B$ collude and share the information $T_c \subset T$ and $A_c \subset A$ with each other, and let J be the set of indices corresponding to colluding brokers. With enough colluders they can learn significant information. For example, if $\sum_{j \in J} a_j t_j > \sum_{i \notin J} |a_i t_i|$, colluding brokers would know that $1 \leq x \leq q$.

To maximize the collusion resistance for a given n, it is clearly optimal to split q uniformly across all n brokers. In other words, every broker is used to trade q/n shares, either buying or selling. (Let some brokers be allowed to receive no order when n is odd to hide a zero position.) This of course leaks n (easily countered by randomization). Also, note that, even if n is known, the colluding brokers B_c can never learn more than their proportion of the LMP's position.

Single broker with multiple traders (1BmT)

To defend against broker collusion, we now examine utilizing m registered trading agents (hereafter referred to as traders) by the LMP to create the desired net position. The mathematics behind this 1BmT platform is very similar to the nB1T strategy: Simply substitute m traders placing orders at one broker in place of one trader at n brokers (where $m = n$). The same theorems hold for 1BmT as for nB1T. In practice, the additional redundant positions held by the traders add ongoing carrying and transaction costs. Also, changing brokers, especially in a developed market, is generally easier than changing registered traders.

We next consider strategies against the *curious individual broker*, assuming he is unable to identify the traders associated with the LMP. Suppose the LMP places a set of orders $\{a_k t_k\}$ at the broker via m different traders. Let $W_0 = \{w_1, w_2, ..., w_z\}$ be the normal market interest seen by the broker; i.e., the set of

orders the broker receives from clients not affiliated with the LMP. The broker thus sees total market interest $W_t = \{a_k t_k\} \cup W_0$. In a very liquid market, $\exists w_i \in W_0 \ni |w_i| = |a_k t_k|$ for $k = 1, ..., m$. In this case, the broker cannot identify any $a_k t_k$ from W_t, and the 1BmT platform does not leak information to him. This is not so when liquidity is low and the broker knows the LMP is employing 1BmT, however. For example, if there is no corresponding surge in activity in the overall market or at other brokers, he can infer that all market interests may originate from the LMP. Furthermore, if $\nexists w_i \in W_0 \ni |w_i| = |a_k t_k|$ for some k, $a_k t_k$ can be identified as originating from the LMP. Thus, elements in the set S can be eliminated, similar to the nB1T platform under collusion. These potential information leaks on the 1BmT platform in an illiquid market motivate our next strategy platform.

Multiple brokers with multiple traders (nBmT)

We can extend the above strategies by using n brokers (with index j) and m traders (with index i), with the security parameter q remaining the same. In general form, the LMP uses the set of traders $\{d_1, d_2, ...d_m\}$, each of the trader d_i places orders with a subset of brokers $\{b_{i1}, b_{i2}, ..., b_{in}\}$. In total, a maximum of $n \cdot m$ orders are placed with a maximum of $n \cdot m$ unique brokers. In practice, some of the b_{ij}'s are the same broker. One possibility is to split the net order x that the LMP wishes to place into m different orders, $\{a_1 t_1, a_2 t_2, ...a_m t_m\}$, for m traders as in the sign flipping game in Definition 5. Each trader d_i can then place its individual single order $a_i t_i$, with broker b_{i1}. In this case, the total number of orders placed is $m = \lceil \log_2(q) \rceil$.

Curious observers cannot see the identities of the traders. Thus, nBmT would look the same as nB1T to curious observers. So, as under nB1T, the external observers cannot extract the trade order made by each trader, thus cannot extract any knowledge about the LMP. Another variant (nBmT2) of this strategy is to divide x into m sets of orders $\{a_1 t_1, a_2 t_2, ...a_m t_m\}$ for m traders according to the sign flipping game. Strategy nBmT2 would appear like m LMPs encrypting their trade volumes using hte sign flipping game with security parameter q. Thus, the external observers cannot extract the trade order made by each trader as derived in the base case of nB1T and therefore cannot extract any knowledge about the LMP.

We now study the performance of nBmT against *curious individual brokers*. Each trader d_i places its order $a_i t_i$ at a different broker. Let W_i be the set of orders each broker b_{i1} receives from his clients not affiliated with the LMP, or normal market activity. Broker b_{i1}, sees total interest $W_{ti} = a_i t_i \cup W_i$, and he cannot identify $a_i t_i$ from W_i since he he does not know that d_i is affiliated with the LMP. This case is different from 1BmT because the market activities W_v at other brokers $b_{v1}, v \neq i$, are also increasing due to the activity of the LMP in nBmT. Thus, even in a low liquidity environment, broker b_{i1} cannot determine whether the increase in $|W_{ti}|$ is due to the activity of the LMP (the presence of $a_i t_i$), or due to increased general market volume (an increase in $|W_i|$). Similarly, for nBmT2, each broker b_{ik} sees the total activity $W_{tik} = \{a_{ik} t_{ik}\} \cup W_{ik}$ but

cannot identify any $a_{ik}t_{ik}$ from W_{tik} even in an illiquid market. Thus, the nBmT strategies can guard against curious individual brokers.

Under total collusion, nBmT2 collapses to nBmT. For nBmT, collusion does not benefit brokers if traders $\{d_i\}$ are not revealed to be affiliated with the LMP. Colluding brokers do not know which orders are affiliated with the LMP and therefore would act at worst as a single broker in the 1BmT scenario. Thus, the nBmT strategy can guard against total broker collusion.

5 Cost estimates

We now estimate the costs of our strategy. The costs of our trading strategy depend on the security parameter q, which depends on the maximum net position we wish to take on any given trade in our strategy. We will assume that the maximum number of shares we want to buy or sell is 3 million shares on any trade. Assuming we trade at increments of 100 shares, we would need $n = 15$ brokers and $q = 2^{15} \times 100 = 3,276,800$. Thus, our fixed commission cost is $\$32,768$ assuming $\$0.01$/share of commission.

Estimating bid-ask spread is more difficult, but we believe that there should be minimal impact on bid-ask spread. If the LMP is trading with itself, then it can happily create a competitive bid or offer and trade at it. It doesn't "lose" the spread we trade with itself. For instance, assume the top bid (LMP's) is $\$11.09$ and the bottom offer is $\$11.25$. The LMP can fill its own bid at $\$11.09$: the net to it is no different than making an offer at $\$11.24$ and then buying that. What is more difficult is estimating the impact of these activities on other prices.

The cost of being front-run depends on the efficiency of the market and the relative size of the order. For example, in Kumar's study [13] on data dating 1983 (when the market was arguably less efficient than in 2007), an average $15bps$ market impact was seen 15 minutes prior to a downtick block, but Keim and Madhavan's study found much more. Moreover, there is no way to calculate the cost of an LMP having its positions or strategy stolen in terms of basis points; given that firms create noise trades to avoid this, we can be sure it is of great value.

The LMP is able to reduce trading costs in our strategies if she is willing to compromise the resolution in x. Suppose in the nB1T case that she can accept $x-r \leq x_c \leq x+r$, where x_c is the coarse net volume she is willing to accept, her trading costs would be decreased by a constant linear with $\lfloor \log_2(2r) \rfloor$. Suppose r is the residual amount that the LMP does not care about, then essentially she can treat her achievable set of orders as the set of non-negative numbers multiplied by $2r = \{0, 2r, 4r, 6r, ... \lceil \frac{q}{2r} \rceil \cdot 2r\}$. For every order x she wishes to place, she can find a x_c which is within $\pm r$ of x. Then, the number of brokers needed is $n_c = \lceil \log_2(\frac{q}{2r}) \rceil$.

6 Conclusions and Future Work

We have examined the problem of placing orders while hiding intention. We presented models of information leakage, and based on these models, we derived three classes of strategies against curious observers, individual curious brokers, and colluding curious brokers.

Though not our current focus, we believe transaction costs of these strategies can sometimes be reasonable, such as when the notional share price is high and/or the bid-offer is tight. We hope this class of intention-disguised algorithmic trading can reduce the profitability of and incentive for exploiting trade information, and alter market behavior as a whole. To that end, understanding these costs, and reducing them, is important.

Open Questions and Future Research

We believe that either finding the lower bound of the brokers that need to be used (in terms of $f(n)$) or finding a better strategy using fewer brokers may be possible. Furthermore, the sign of a trade with any one broker may be inferred by an observer using a trade direction algorithm such as that developed by Ellis, Michaely and O'Hara [6] or Peterson and Sirri [15]. Our strategies are unaffected, assuming that all trades are filled in one round. However, realistically, such trades may take multiple rounds. On the other hand, in practice, there are also often other market participants trading, thus creating cover noise against identifying the trades initiated by the LMP. Even in an extremely illiquid market with no other active trading participants, the orders being worked by brokers are not synchronous in practice. Therefore, even if a broker with malicious intention is able to deduce the signs of other brokers, he cannot front run confidently that he has seen all the relevant trades initiated by the LMP.

Acknowledgement Zhenming Liu is supported in part by NSF CCF-0634923.

References

1. S. Brain, "A front-running smile?" *Traders Magazine*, May 2005, available online at `http://www.wisdom.weizmann.ac.il/~oded/zk-tut02.html`
2. G. Chacko, personal communication, August 2009.
3. G. Di Crescenzo, "Privacy for the stock market" in *Proc. Financial Cryptography and Data Security*, 2002.
4. I. Domowitz, H. Yegerman, "The cost of algorithmic trading: a first look at comparative performance," in *Algorithmic Trading: Precision, Control, Execution*, March 2005.
5. I. Domowitz, H. Yegerman, "Measuring and interpreting the performance of broker algorithms" in *ITG Inc. Research Report*, August 2005.
6. K. Ellis, R. Michaely, M. O'Hara, "The accuracy of trade classification rule: evidence from NASDAQ," *Journal of Financial and Quantitative Analysis*. 2000.
7. O. Goldreich, "Zero-knowledge: a tutorial." accessed through `http://www.wisdom.weizmann.ac.il/~oded/zk-tut02.html`
8. O. Goldreich, E. Petrank, "Quantifying knowledge complexity," in *32nd IEEE Symposium on Foundations of Computer Science*, 1996.

9. S. Goldwasser, S. Micali, C. Rackoff, "The knowledge complexity of interactive proof systems," in *17th Annual ACM Symposium of Theory of Computing*, 1985.
10. L. Harris, "Trading and exchanges: market microstructure for practitioners," *Oxford University Press*, 2003.
11. M. Kearns, Y. Nevmyvaka, A. Papandreou and K. Sycara, "Electronic Trading in Order-Driven Markets: Efficient Execution", *IEEE Conference on Electronic Commerce (CEC)*, 2005.
12. D. B. Keim, A. Madhavan, "The upstairs market for large-block transactions: analysis and measurement of price effects," *The Review of Financial Studies,* 1996.
13. R. Kumar, A. Sarin, K. Shastri, "The behavior of option Price Around Large Block Transactions in the Underlying Security," in *The Journal of Finance*. 1992.
14. A. Madhavan, "VWAP Strategies," *Investment Guides, Transaction Performance.* Spring 2002.
15. M. Peterson, E. Sirri, "Evaluation of biases in execuation cost estimates using trade and quote data," Forthcoming in *Journal of Financial Markets*. 2002.
16. C. Thorpe, D. C. Parkes, "Cryptographic securities exchanges" in *Financial Cryptography and Data Security*, 2007
17. C. Thorpe, D. C. Parkes, "Cryptographic combinatorial securities exchanges" in *Financial Cryptography and Data Security*, 2009