Requirements of Trading Platforms: Fairness and Low latency

Balaji Prabhatkar
Lecture 2, CS 349F
Autumn 2020
Stanford
Exchanges and Trading Platforms

Fundamental Operation Performed by Exchange: Price Discovery!

Stock/Currency/Futures Exchange

Runs on top of Trading Platform

Fundamental Requirement of Exchange Infra: Provide “Fair Access”
Exchanges and Trading Platforms

An Exchange is a meeting place of buyers and sellers along with a set of rules to govern trading; as defined by the SEC, Exchanges shall provide

- **Market Access and Fairness**
  - The Exchange is open to all comers, even small investors; the Exchange will ensure that no participant is unfairly advantaged over another (Securities Exchange Act, 1934)

- **Market Transparency and Coordination**
  - Since traders trade and hedge across multiple markets, it creates strong interdependencies between the markets; to get the best efficiency and pricing, markets need to be transparent and coordinated
  - Gave rise to the National Best Bid or Offer (NBBO) rule (Reg NMS, 1972 and 2005)

- **Market Surveillance**
  - Prevent market manipulation, fraud, etc, by implementing audit trails, surveillance and disciplinary programs. → Big topic for the possible use of AI

- **Market Stability and Systemic Risks**
  - Exchanges shall maintain sufficient capacity to deal with large trading volumes and spikes

Source: https://www.sec.gov/rules/concept/3438672.txt
Requirements: Exchanges and Trading Platforms

As seen in the previous slide, most of the requirements on the Exchange translates to technological requirements when trading is electronic.

These requirements combine with another one imposed by the matching algorithm used by exchanges: Continuous Price-Time Matching

- According to the CPTM, matches are found between buyers and sellers based on *price first*, and then by the *time of arrival* of an order when price is equal.

→ As a consequence, traders are incented to respond quickly so that their order is queued towards the front at their quoted price.

We will take these requirements and convert them into **engineering requirements** for the **exchanges** and for the **market participants** (traders).
Let’s Take a Closer Look at the Trading Platform

Components of the Trading Exchange
1. Market Participants
2. Gateways (where the MPs connect to the Exchange)
3. The Matching Engine (where orders are executed)
The Trading Loop has Two Parts

Components of the Trading Loop

1. **The Exchange Loop:** Gateway → Matching Engine → Gateway
   Orders In, Market Data Out

2. **The MP Loop:** Gateway → MP Servers → Gateway
   Market Data Out, Orders In
The Exchange Loop Must Be **Fair**

**Definition of Fairness**

**Inbound Fairness:** *Orders must be processed in a globally FIFO manner*, regardless of GW through which they enter the exchange.

**Outbound Fairness:** The *Market Data shall reach* all MPs (GWs) at *exactly the same time*
The Exchange Loop Must Be **Fair**

Currently, in on-prem exchange networks

**Inbound Fairness:** Is guaranteed by ensuring the $GW \rightarrow ME$ transit times are *exactly equal* using wires of equal length and identical switches and components on each $GW \rightarrow ME$ path

**Outbound Fairness:** Is guaranteed by *multicasting* the Market Data to the MPs

**Challenge:** This highly bespoke design is not possible to achieve when the exchange moves to the Cloud
**Providing Fairness using Accurate Clocks**

**Inbound Fairness**
- **Synchronize** clocks at all gateways to the ME clock (to nanosecond-level accuracy)
- **Timestamp** orders when they arrive at the gateways to establish order of arrival
- **Resequence** orders according to their gateway timestamps before executing them at the ME

**Outbound Fairness**
- **Timestamp** market data at the ME
- **Hold-and-release** market data at the gateways using the H/R buffers to ensure simultaneous delivery
Whittling a New York Minute
To 100 Billionths of a Second

By JOHN MARKOFF

SAN FRANCISCO — Computer scientists at Stanford University and Google have created technology that can track time down to 100 billionths of a second. It could be just what Wall Street is looking for.

System engineers at Nasdaq, the New York-based stock exchange, recently began testing an algorithm and software that they hope can synchronize a giant network of computers with that nanosecond precision. They say they have built a prototype, and are in the process of deploying a bigger version.

For an exchange like Nasdaq, such refinement is essential to accurately order the millions of stock trades that are placed on their computer systems every second.

Ultimately, this is about money. With stock trading now dominated by computers that make buying and selling decisions and execute them with blazing speed, keeping that order also means protecting profits. So-called high-frequency trading firms place trades in a fraction of a second, sometimes in a bet that they can move faster than bigger competitors.

The pressure to manage these high-speed trades grows when the stock market becomes more volatile, as it has been in recent months, in part to prevent the fastest traders from taking unfair advantage of slower firms. High-frequency traders typically ac-

Continued on Page A17
Market Capitalization by Geography

- 60 major exchanges, total value: $69 trillion
- NYSE: $18.5 trillion market cap, single largest
- 16 exchanges in $1 Trillion club, 87% of total value: NYSE, Nasdaq, LSE, Deutsche Borse, TMX Group (Canada), Japan Exch Group

Over 93% of global stock value is divided between three continents

The 16 exchanges that comprise the "$1 Trillion Club" account for 87% of global market capitalization

The NYSE itself is bigger than the world's 50 smallest major exchanges

16 exchanges, each with a total market capitalization over $1T, can be considered to be in the exclusive "$1 Trillion Club"

Source: Visual Capitalist: Exchanges by Size
Let’s Look at the Market Participant Loop
Traders Always Seek an Advantage (Arbitrage Opportunity)

Traders cut down to size
July 28, 2000: 1:15 p.m. ET

Chicago Mercantile Exchange orders traders to take off platform shoes

NEW YORK (CNNfn) - If David Bowie, members of the rock band Kiss or any other glam-rock icons ever decide to trade futures at the Chicago Mercantile Exchange, they'll have to take off their platform shoes and wear flat shoes like any other peon.

On Thursday the exchange circulated a memo reminding traders of its dress code, which requires they wear safe shoes, and added this specific: shoes with heels higher than two inches are a no-no, starting Monday, July 31.

Male and female traders have taken to wearing platform shoes, in response to fashion trends and a need to tower over their competition on the hectic, crowded trading floor. Though the Chicago Mercantile Exchange (CME) couldn't confirm injuries, Reuters reported Thursday that some traders had suffered twisted ankles and foot injuries.
The Market Participant Loop has to have **Low Latency**

Ensuring Low Latency of the MP Loop

1. **The Market Data** is sent to the MP servers and received from them using **Layer 1 switches**
2. **FPGAs or Smart NICs** are used at the MP servers to respond to a market data item (called a market event) **ASAP**
More Generally, Accurate Timing is Foundational for the Financial Industry

<table>
<thead>
<tr>
<th>Exchanges and Trading Platforms</th>
<th>• Ensure “fairness”: in-order execution of transactions • Enable execution of “time-synced smart contracts” • Make hybrid and cloud-only models possible</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Conventional exchanges: Nasdaq, NYSE, SGX, LSE, TSE, HKSE, Euronext, Chicago ME)</em></td>
<td><em>(Digital currency exchanges: Coinbase, Binance, GDAX)</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investment Banks and Hedge Funds</th>
<th>• Sync with exchanges as reference for market data timestamps • Run realistic simulations for better trading algorithms • Regulatory compliance: MIFiD II in the EU and CAT in the US</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(e.g., BofA, Citi, Goldman Sachs, JPMC, Morgan Stanley, UBS, Wells Fargo)</em></td>
<td></td>
</tr>
</tbody>
</table>

| Payments and Retail Banking | • Accurately timestamp transactions across disparate geographies for transaction tracking/debugging • Recreate network dynamics for 100s of apps |
|-------------------------------|-------------------------------------------------------------------------------------------------
| *(e.g., Visa, Mastercard, Paypal, eBay, Wells Fargo, HSBC, Sumitomo, Barclays)* | |
In Summary

• Carefully-engineered networks are expensive to scale and maintain
  – So a clock-based approach to provide fairness is desirable
  – Clocks solve several problems: (i) inbound fairness can be provided in public clouds, (ii) IP multicasting for outbound market data is hard to provide, but 1-to-many unicasting with hold-and-release buffers provides outbound fairness

• The low latency craved by traders can be provided in the cloud using
  – High-speed I/O (although this will still be at 2—3 microsecond-level)
  – Or, when cloud providers make Smart NICs available to traders (nanoseconds possible)

• But, fundamentally, with the advent of AI/ML techniques
  – Trading will move from a speed game to a combination of speed and accuracy of mining market data game
In Summary

• Deutsche Borse will describe the low latency achievable in today’s exchanges and how timing/clock sync play a fundamental role

• Cisco ExaBlaze will describe how an extremely low latency is achievable in the Market Participant Loop (tick-to-trade latency)

• Nasdaq will describe the motivation for moving to the Cloud

• We will also see how to synchronize clocks (in the Cloud) and how to use it solve a variety of problems

• In the next class, Mendel will describe virtualization and how this enables the Cloud
Demo of Clock Sync and Fairness in the Cloud