

Do You Trust Your Cryptocurrency Exchange?

Market Abuse as a Security Problem

Slides for talk given at SummerC0n 2019:

<https://www.summercon.org/presentations.html#do-you-trust-your-cryptocurency-exchange>

Disclaimer: Views expressed herein do not constitute legal or investment advice and do not necessarily reflect those of speakers' respective employers

Who are we?

Hayden Melton, PhD

Computer Scientist turned
Electronic Trading Expert

*Quantitative Trading Proposition
Manager at Refinitiv*



<https://sites.google.com/site/haydenmelton/>

Vasilios Mavroudis

Computer Security PhD

*Researcher at
University College London
and
TradeScope*



Supported by Binance Labs



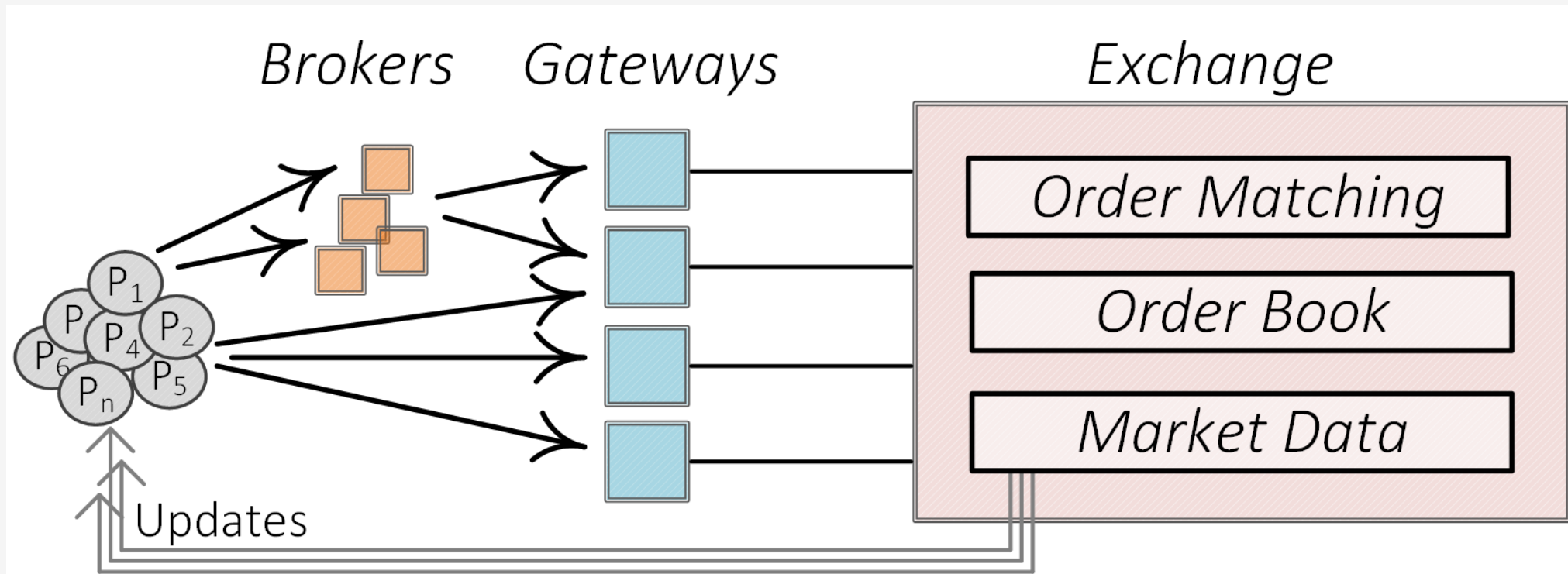
<https://mavroudis/>

<https://tradescope.pe/>

Overview

1. What is an electronic exchange?
 - How does order-matching work?
 - Isn't trading boring?
2. Can markets be hacked?
 - Is there a **threat model**?
 - Are there **exploits**? What about **0-days**?
3. Patching the market

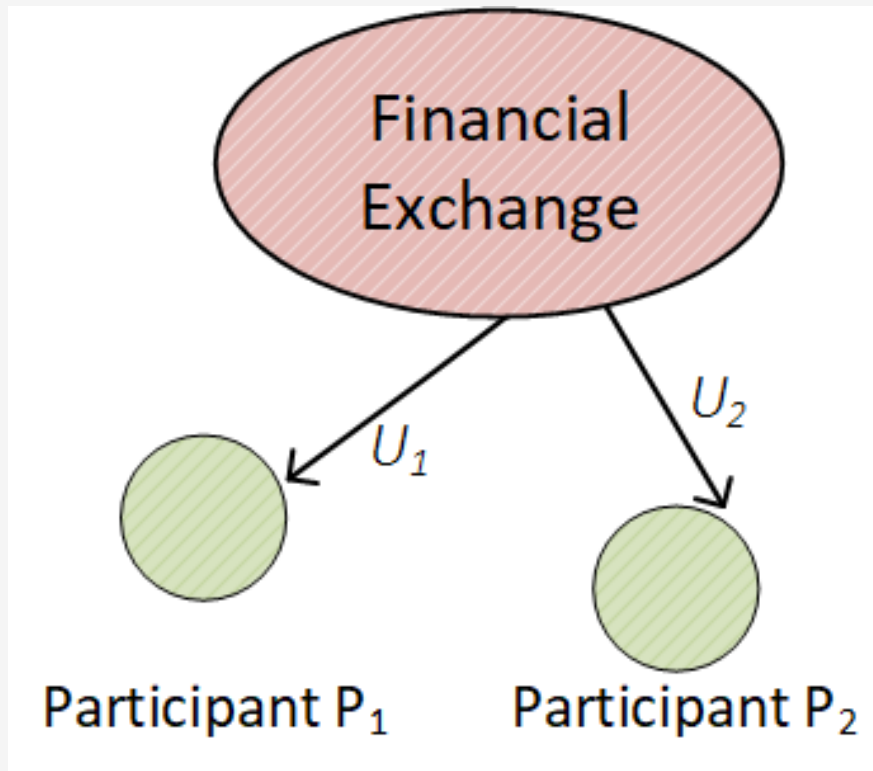
Financial Exchanges: *In a Gold Rush, Sell Shovels*



Central Limit Order Book Interface (Coinbase Pro)



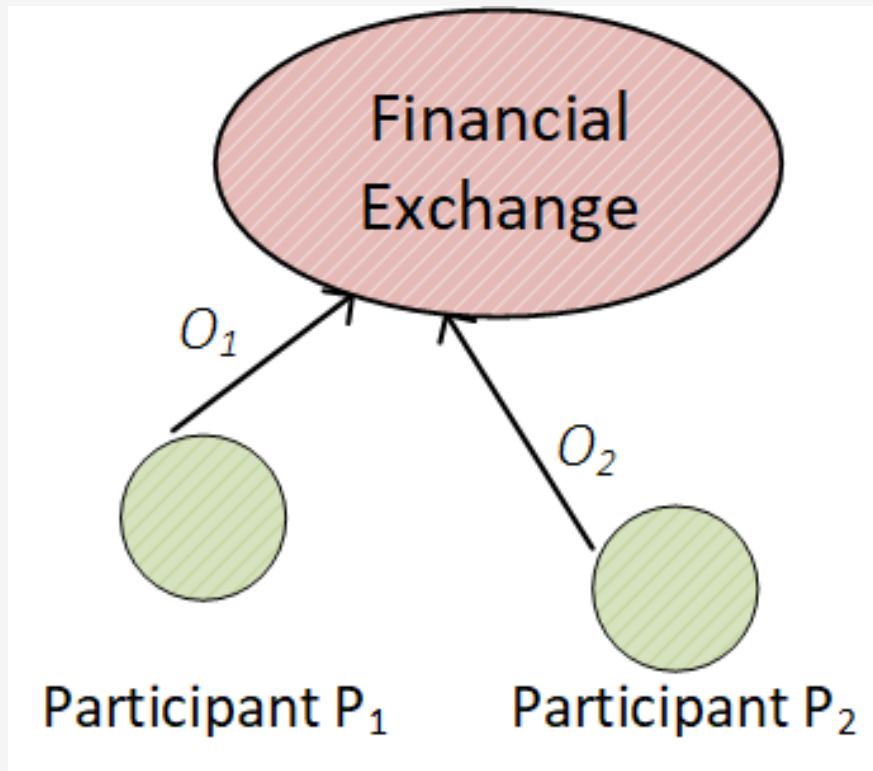
Electronic Trading Feedback Loop



How do participants “see” the prevailing supply and demand in the book?

- Are sent *market data update* messages **U** each time the book changes

Electronic Trading Feedback Loop



How do participants “see” the prevailing supply and demand in the book?

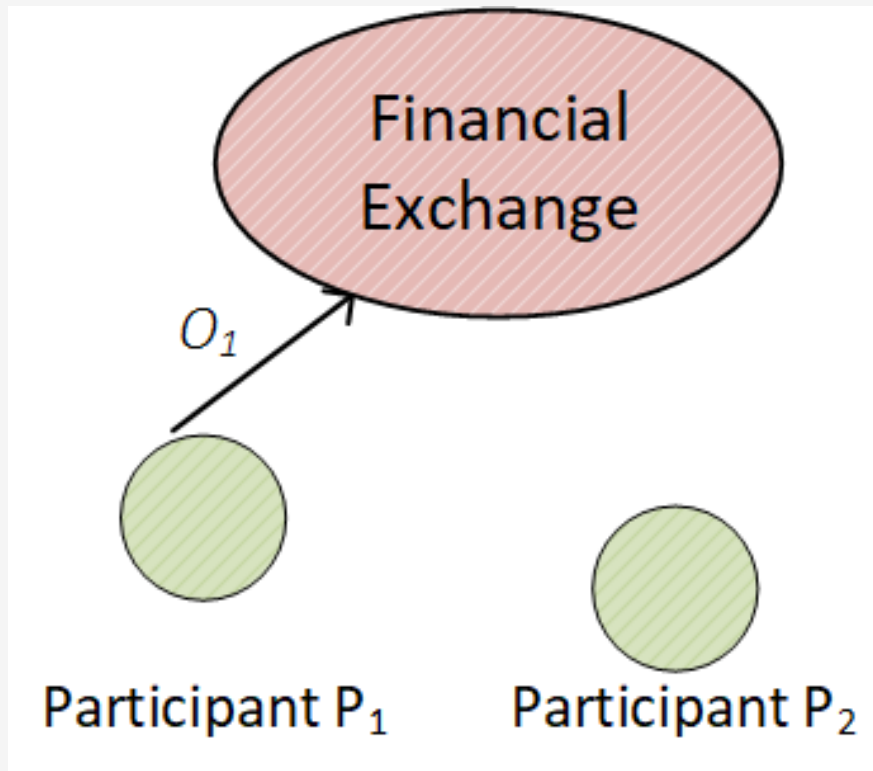
- Are sent *market data update* messages **U** each time the book changes

How do participants alter supply and demand in the limit order book?

- They send order messages **O** to buy or sell an instrument at specified price and qty.

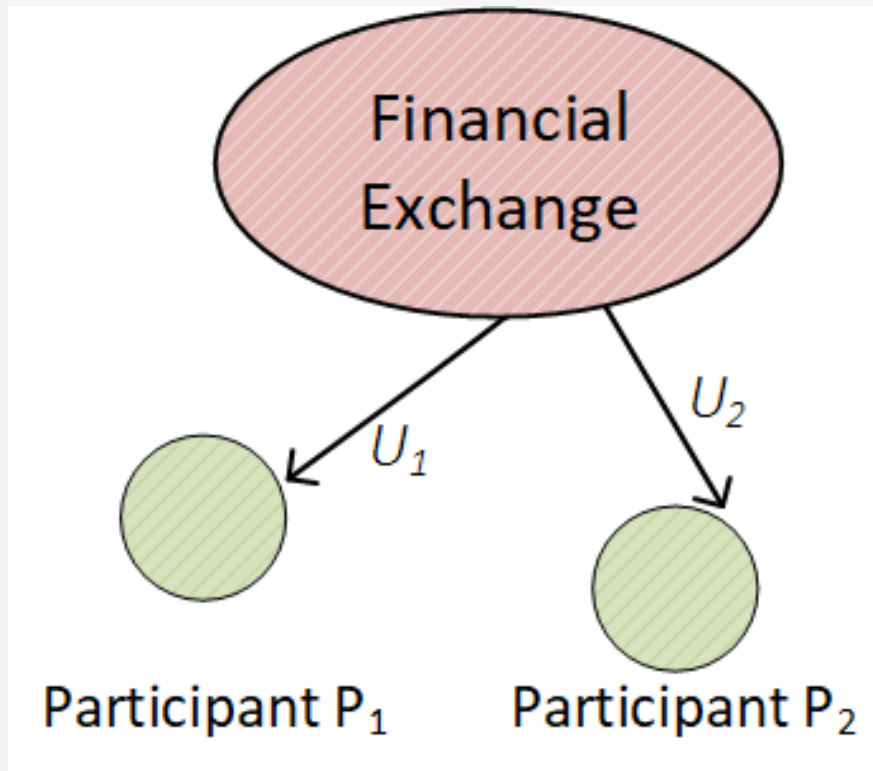
Electronic Trading Feedback Loop

1. A Mkt Participant sends order message to exchange
e.g., *buy 100 BTC @10,000*

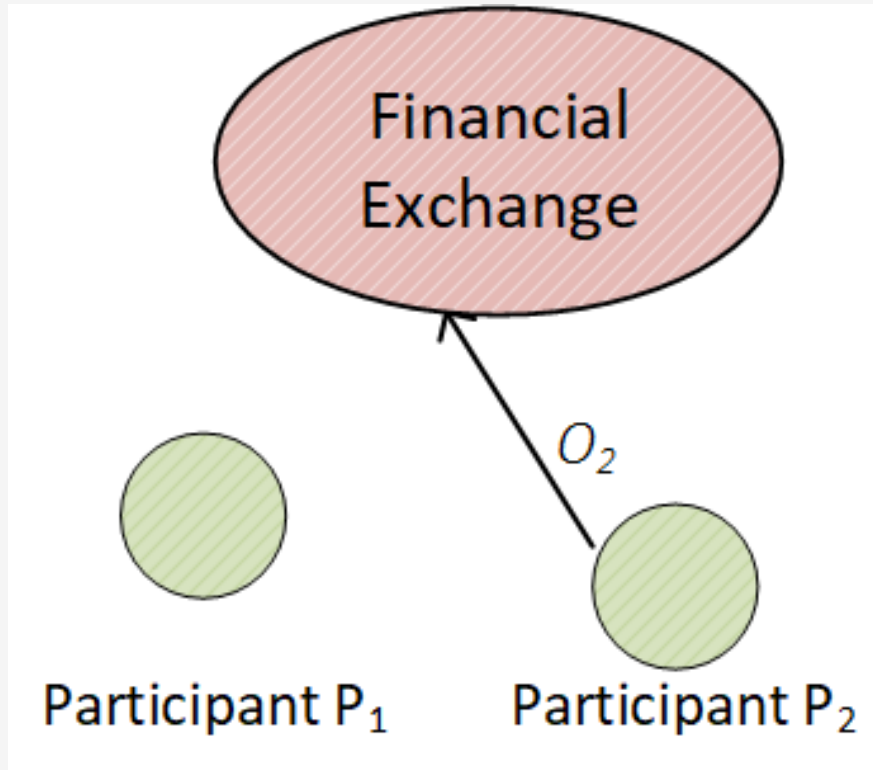


Electronic Trading Feedback Loop

1. A Mkt Participant sends order message to exchange
e.g., *buy 100 BTC @10,000*
2. Order causes change in state of CLOB on exchange; market data update message sent out

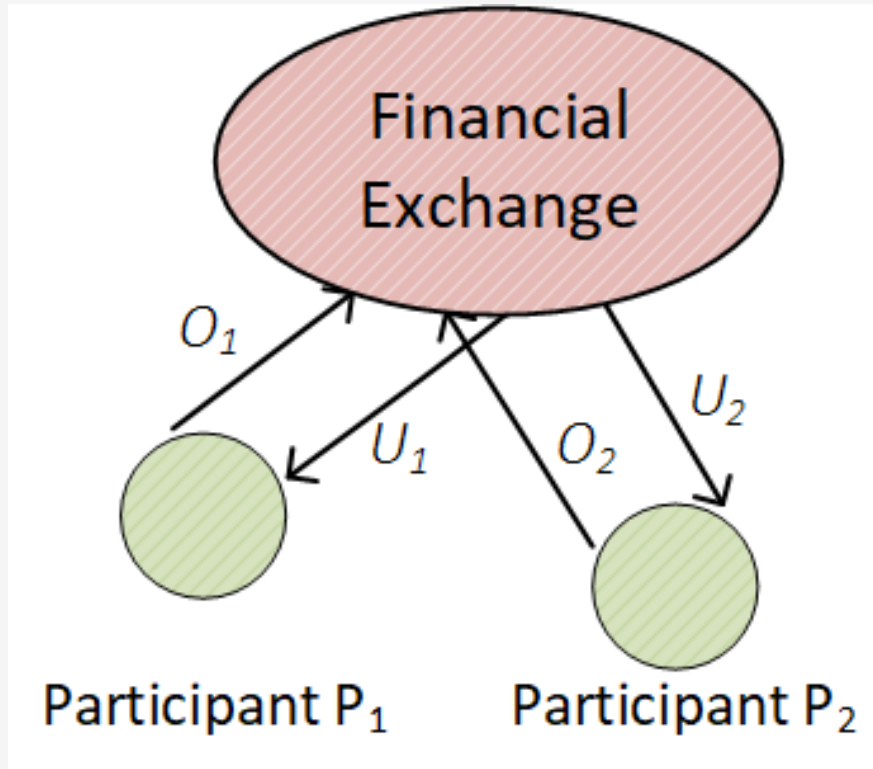


Electronic Trading Feedback Loop



1. A Mkt Participant sends order message to exchange
e.g., *buy 100 BTC @10,000*
2. Order causes change in state of CLOB on exchange; market data update message sent out
3. Another Mkt Participant reacts to new supply and demand schedule in update; sends new order message

Electronic Trading Feedback Loop



1. A Mkt Participant sends order message to exchange
e.g., *buy 100 BTC @10,000*
2. Order causes change in state of CLOB on exchange; market data update message sent out
3. Another Mkt Participant reacts to new supply and demand schedule in update; sends new order message
4. Goto step (2) until market close

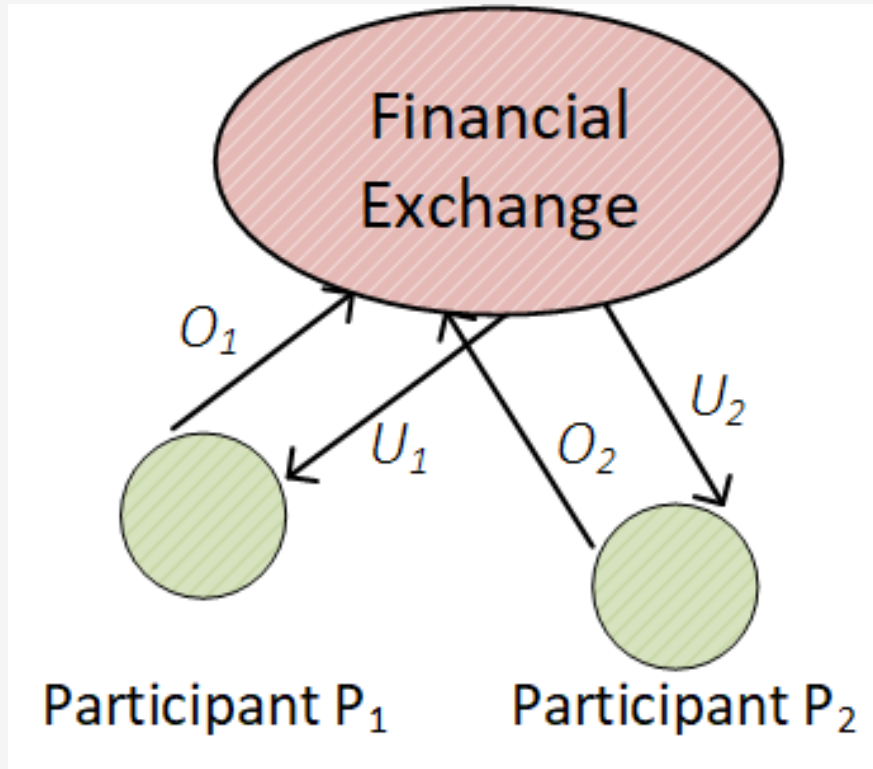
Electronic Trading Feedback Loop

1. A Mkt Participant sends order message to exchange
e.g., *buy 100 BTC @10,000*

2. Order causes change in state of CLOB on exchange; market data update message sent out

3. Another Mkt Participant reacts to new supply and demand schedule in update; sends new order message

4. Goto step (2) until market close



Exchanges observe loads of millions of messages per second

The Continuous Limit Order Book (CLOB)

Total shares offered	Price	Total shares bid for
2200	\$178.30	
300	\$178.29	
8000	\$178.28	
	\$178.27	
	\$178.26	
	\$178.25	5000
	\$178.24	2400
	\$178.23	6000

Above: Instant-in-time snapshot of limit order book for AAPL (Apple stock)

The Continuous Limit Order Book (CLOB)

Total shares offered	Price	Total shares bid for
2200	\$178.30	
300	\$178.29	
8000	\$178.28	
	\$178.27	
	\$178.26	
	\$178.25	5000
	\$178.24	2400
	\$178.23	6000

Above: Instant-in-time snapshot of limit order book for AAPL (Apple stock)

What's the best bid price for AAPL? Best offer price?

If I want to buy 10,000 AAPL shares *right now*, what will it cost me?

The Continuous Limit Order Book (CLOB)

Individual offer orders <i>← decreasing priority</i>	Total shares offered	Price	Total shares bid for	Individual bid orders <i>decreasing priority→</i>
100+100+100+1800+100=	2200	\$178.30		
100+100+100=	300	\$178.29		
7200+800=	8000	\$178.28		
		\$178.27		
		\$178.26		
		\$178.25	5000=1000+200+500+3300	
		\$178.24	2400=1200+100+1100	
		\$178.23	6000=1000+2500+1300+200+1000	

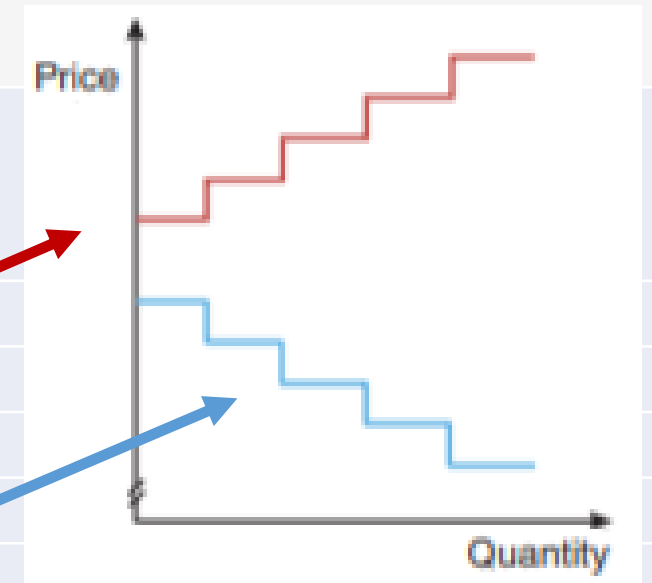
The Continuous Limit Order Book (CLOB)

Individual offer orders <i>← decreasing priority</i>	Total shares offered	Price	Total shares bid for	Individual bid orders <i>decreasing priority→</i>
100+100+100+1800+100=	2200	\$178.30		
100+100+100=	300	\$178.29		
7200+800=	8000	\$178.28		
		\$178.27		
		\$178.26		
		\$178.25	5000	=1000+200+500+3300
		\$178.24	2400	=1200+100+1100
		\$178.23	6000	=1000+2500+1300+200+1000

Computer science view: `TreeMap<Price, <LinkedList<Order>>> clob;`

The Continuous Limit Order Book (CLOB)

Individual offer orders ← decreasing priority	Total shares offered	Price	Total shares bid for
100+100+100+1800+100=	2200	\$178.30	
100+100+100=	300	\$178.29	
7200+800=	8000	\$178.28	
		\$178.27	
		\$178.26	
		\$178.25	5000 = 1000+200+500+3300
		\$178.24	2400 = 1200+100+1100
		\$178.23	6000 = 1000+2500+1300+200+1000



Computer science view: `TreeMap<Price, <LinkedList<Order>>> limitOrderBook;`

Economics view: aggregate supply and demand chart

Defining Market Abuse

- *Multi-jurisdictional*: what constitutes market abuse in one country may not in another.
- Roughly, Financial Conduct Authority (FCA) in the UK defines it as:
 - **Insider dealing**
 - **Unlawful disclosure**
 - **Market manipulation** (both actual and attempted)
- Can be civil or criminal in nature; both market participants and market operators can be offenders
- Market manipulation: *artificially affecting the supply and demand for a financial instrument*

Defining Market Abuse

- *Multi-jurisdictional*: what constitutes market abuse in one country may not in another.
- Roughly, Financial Conduct Authority (FCA) in the UK defines it as:
 - **Insider dealing**
 - **Unlawful disclosure**
 - **Market manipulation** (both actual and attempted)
- Can be civil or criminal in nature; both market participants and market operators can be offenders
- Market manipulation: artificially affecting the supply and demand for a financial instrument



INTENT!

Is Lying Market Manipulation?

- “Generally it is allowed, encouraged even, for a big market participant to hide its intentions. It is manipulation for a market participant to affirmatively mislead people about its intentions. The space between those two things is very narrow indeed.” –Matt Levine
- **Iceberg** order type
 - My *intent* is to buy 100 units of the instrument
 - Exchange will accept my whole order but I can specify “show only 1 unit”; 99 are hidden

Is Lying Market Manipulation?

- “Generally it is allowed, encouraged even, for a big market participant to hide its intentions. It is manipulation for a market participant to affirmatively mislead people about its intentions. The space between those two things is very narrow indeed.” –Matt Levine
- **Iceberg** order type
 - My *intent* is to buy 100 units of the instrument
 - Exchange will accept my whole order but I can specify “show only 1 unit”; 99 are hidden

Price	Total units bid for	Individual bid orders <i>decreasing priority</i> →
\$99.30	100	[100 hide 0]

VS.

Price	Total units bid for	Individual bid orders <i>decreasing priority</i> →
\$99.30	1	[100 hide 99]

Is Lying Market Manipulation (cont'd)?

- U.S. v. Litvak, U.S. District Court, District of Connecticut, No. 13-cr-00019
 - Convicted by lower court, overturned by appeals court
- Roughly:
 - Trader told institutional customer's "we bought this bond for 90c" —in reality bought it for 85c—to justify selling it to that customer at say 92c.
 - Lied about purchase price to secure better resale price
 - Overturned because customers were sophisticated and should have known better (also on some technicalities)

Front-running

- U.S. v. Johnson et al, U.S. District Court, Eastern District of New York, No. 16-cr-00457.
 - Convicted by lower court, sentenced; currently on appeal
- “Front running” case. Roughly:
 - Agreed to buy 3.5B of British pounds at the 4pm London fixing price
 - Bought some ahead of the fixing window (dealer “buys low”), driving up the fixing price (dealer “sells high”).
 - Arguments on appeal that dealer roughly that the two parties were adversaries, i.e., dealer not acting as agent for customer.
- Jurisdictional issues:
 - Alleged co-conspirator fought and won against extradition from UK to US

Conspiring to Manipulate

- Hayes LIBOR Case (UK)
 - Convicted (2015), permission to appeal so far denied (2019)
- LIBOR is the London Interbank Offered Rate, calculated from *estimates* submitted by leading banks in London.
 - Affects prices of trillions of dollars worth of derivatives.
- Roughly:
 - Asked brokers in interbank market to encourage others to submit rates “high” or “low” with promises of curries, wine etc to benefit his trading book
 - Interestingly, his brokers then charged but acquitted because despite agreeing with him they apparently took no actual action

Spoofing

- U.S. v. Sarao, U.S. District Court, Northern District of Illinois, No. 15-cr-00075.
- Spoofing is causing a false impression of supply and demand by submitting a large *bogus* order and small *bone-fide* contra order
 - Relies on the *herd mentality* of other participants to succeed
- UK-based, traded on CME, convicted and extradited, traded out of parent's basement near Heathrow airport.
- Led to U.S. v. Thakkar, 18-cr-36 (N.D. Ill.), roughly:
 - Software developer of spoofing system specified by Sarao with “back of book” custom functionality acquitted

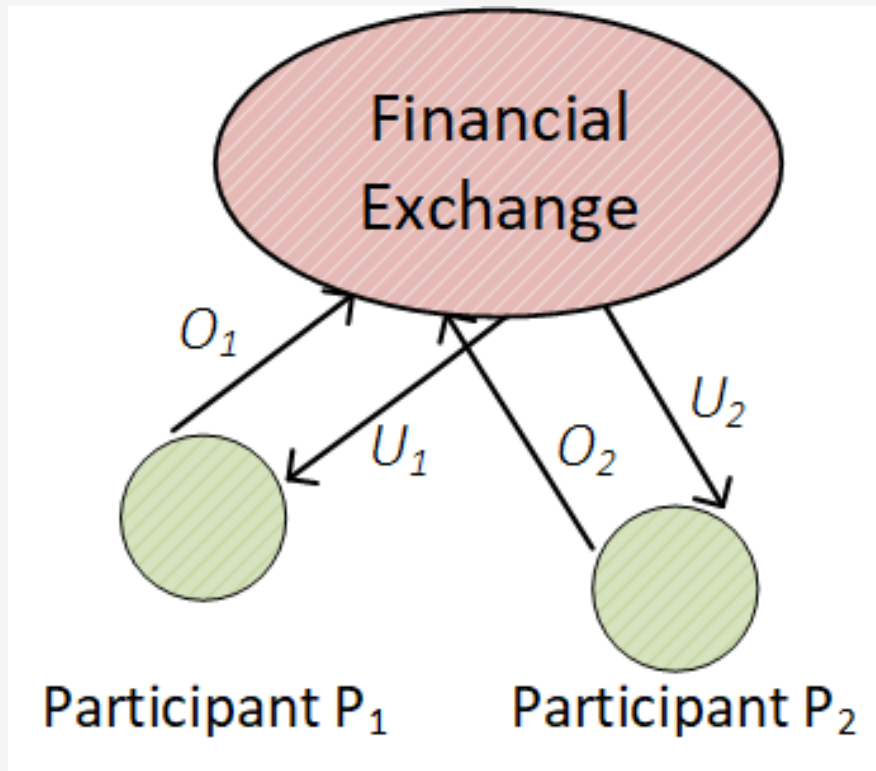
Plenty more examples too...

- Musk at Tesla
 - Different treatment of short-sellers posting negative things vs. him tweeting positive things
- False disclosures (deliberate, accidental, or gradual)
 - Dark pools in equities: no HFT's allowed on the venue; trade against customer orders in the dark pool
 - Last-look in FX: use it for stale pricing, not as a free option to accept only profitable trades
 - Order types not meeting their specification: peg to price levels where no displayed liquidity exists, violating specification of hidden liquidity orders
- “Preferential access” colocation scandal at NSE in India
 - Running example because it's *technical market manipulation*

Technical Market Manipulation

- “A class of market manipulation techniques that **exploit technical details** and glitches in the operation of the exchanges”
- Forms of market manipulation that **would not otherwise be possible without computers**
- Today our focus is on forms of it that pertain to **speed**
 - CLOB: continuous → **the early bird gets the worm!**
 - Many techniques participants can employ to be “faster” than one another
 - Some seem pernicious/manipulative, others seem ok or are just “**socially wasteful**”!
- Relationship between speed, manipulation and fairness
- All through the lens of computer security

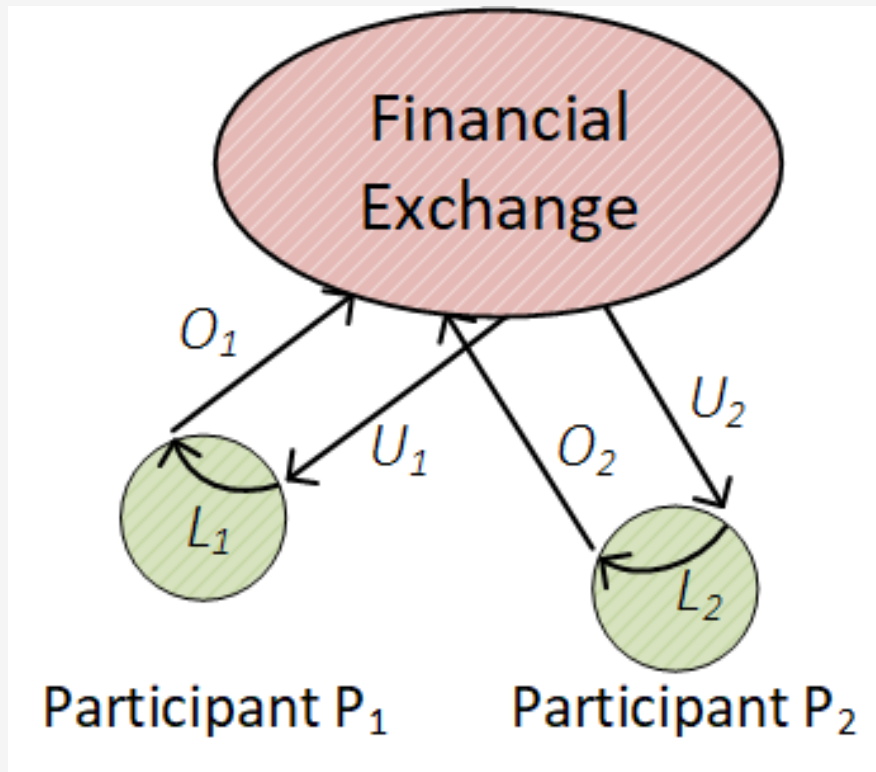
CLOBs Incentivize Competition on Speed



What can market participants do to be faster?

- Minimize *transmit times* of “U” and “O”?
 - Line of sight or microwave comms, co-location, renting office space nearby, increase bandwidth
 - Exploit properties of exchange’s implementation e.g., horizontal scaling

CLOBs Incentivize Competition on Speed



What can market participants do to be faster?

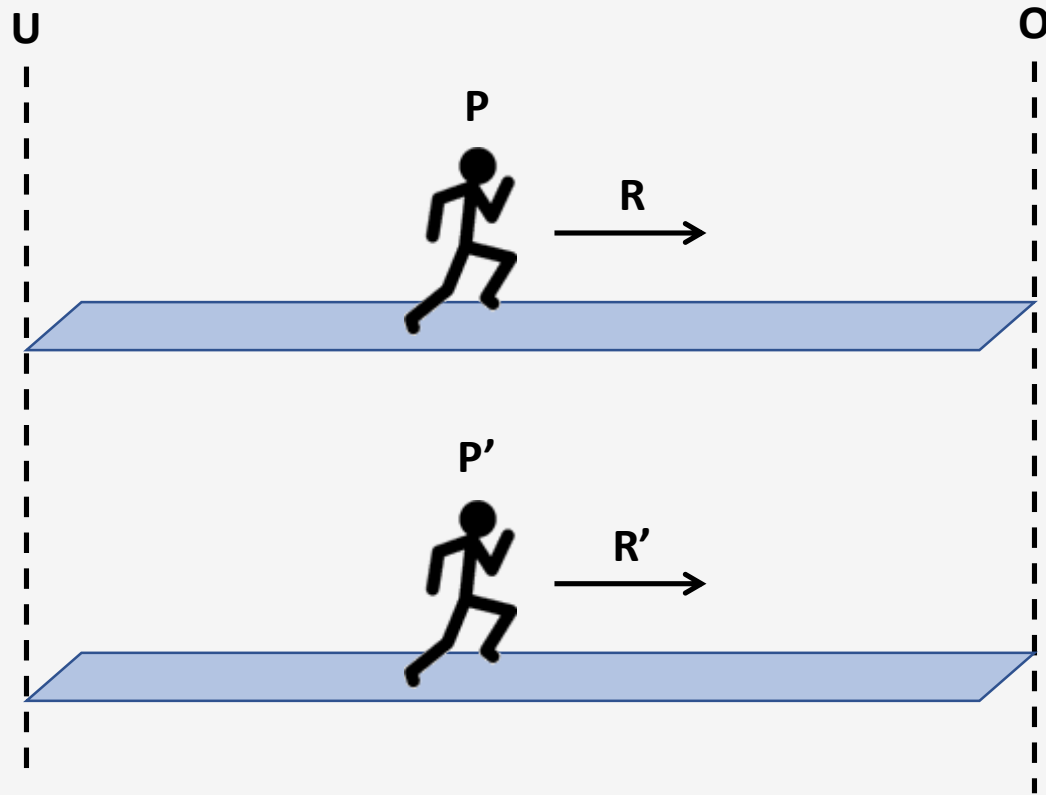
Minimize *transmit times* of “U” and “O”?

- Line of sight or microwave comms, co-location, renting office space nearby, increase bandwidth
- Exploit properties of exchange’s implementation e.g., horizontal scaling

Minimize their *response times*?

- Optimized software, higher performance hardware, software-as-hardware (FPGA)
- Exploit properties of network protocols e.g., TCP/IP

Relationship Between Speed and Fairness



Running race analogy:

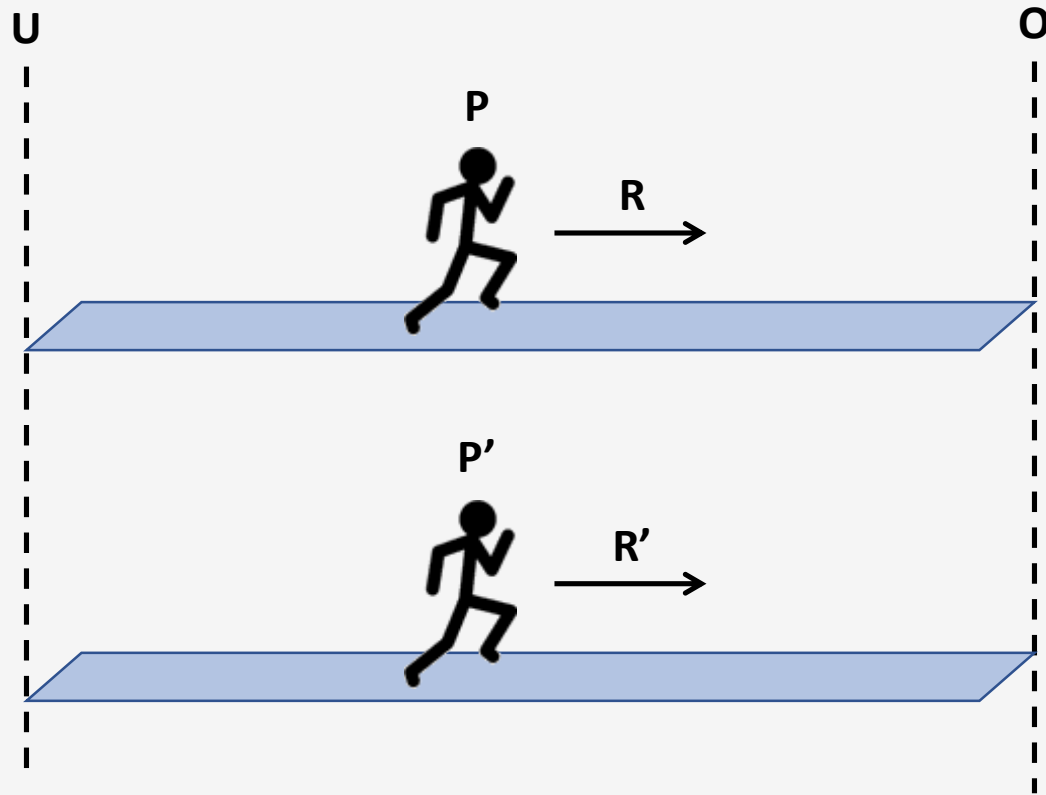
R: response time/runner's speed

U: mkt data transmit time/start line

O: order msg transmit time/finish line

Intuitively: race is fair if no head starts, no shorter tracks → fastest runner wins!

Relationship Between Speed and Fairness



Running race analogy:

R: response time/runner's speed

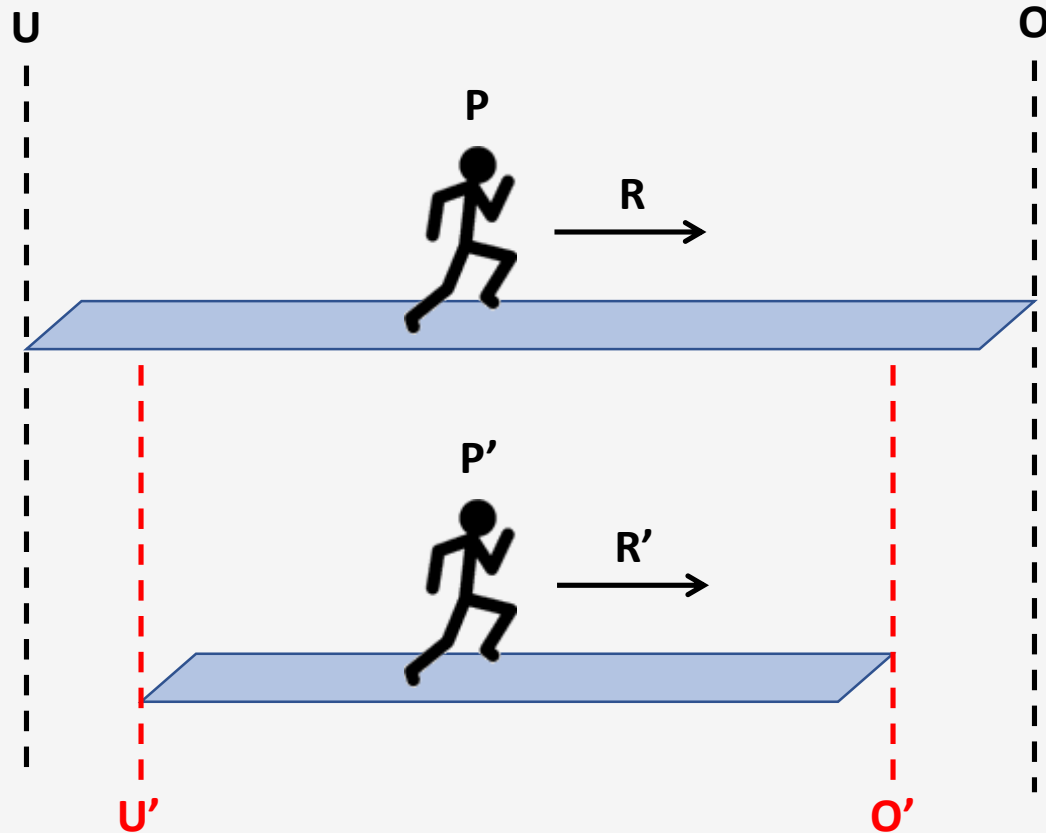
U: mkt data transmit time/start line

O: order msg transmit time/finish line

Intuitively: race is fair if no head starts, no shorter tracks → fastest runner wins!

(Assumes colocation: Market operator is in control of infrastructure relating to U and O by ensuring equal cable lengths, same bandwidth, same media, etc)

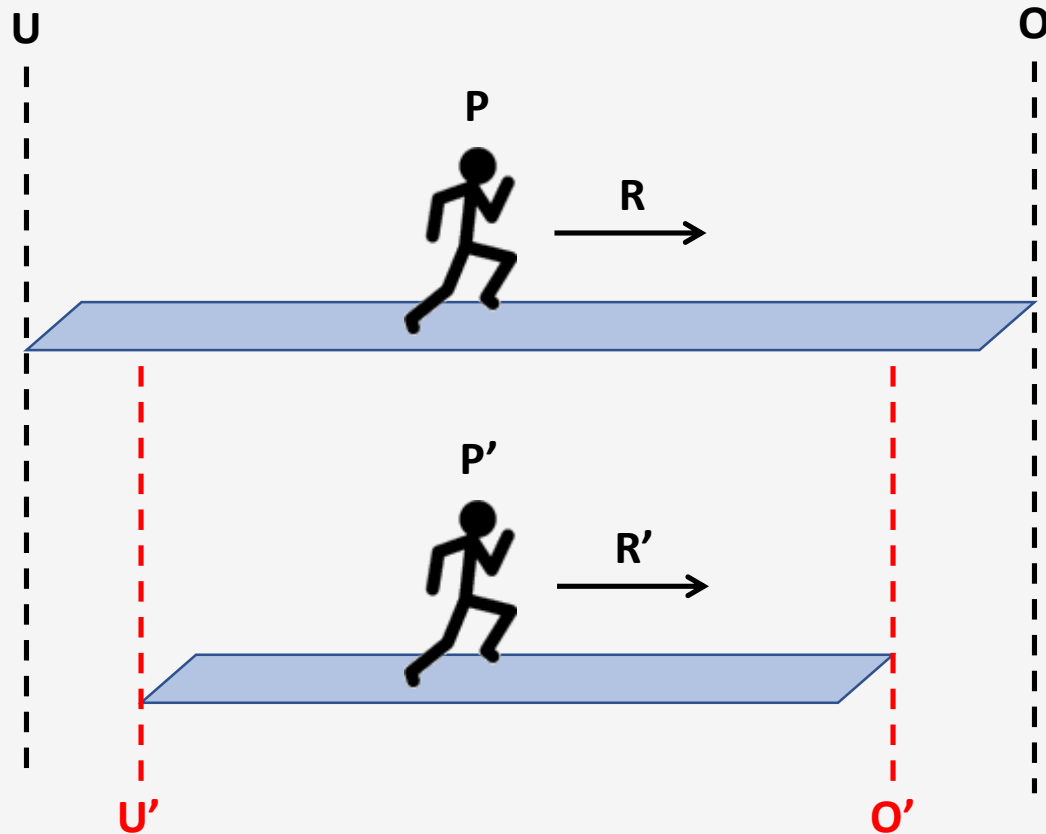
Relationship Between Speed and Fairness



If P and P' are both colocated i.e., both *similarly situated*:

- Not a *fair race*
- Not a *level playing field*
- *Preferential access* given to P' over P

Relationship Between Speed and Fairness



If P and P' are both colocated i.e., both *similarly situated*:

- Not a *fair race*
- Not a *level playing field*
- *Preferential access* given to P' over P

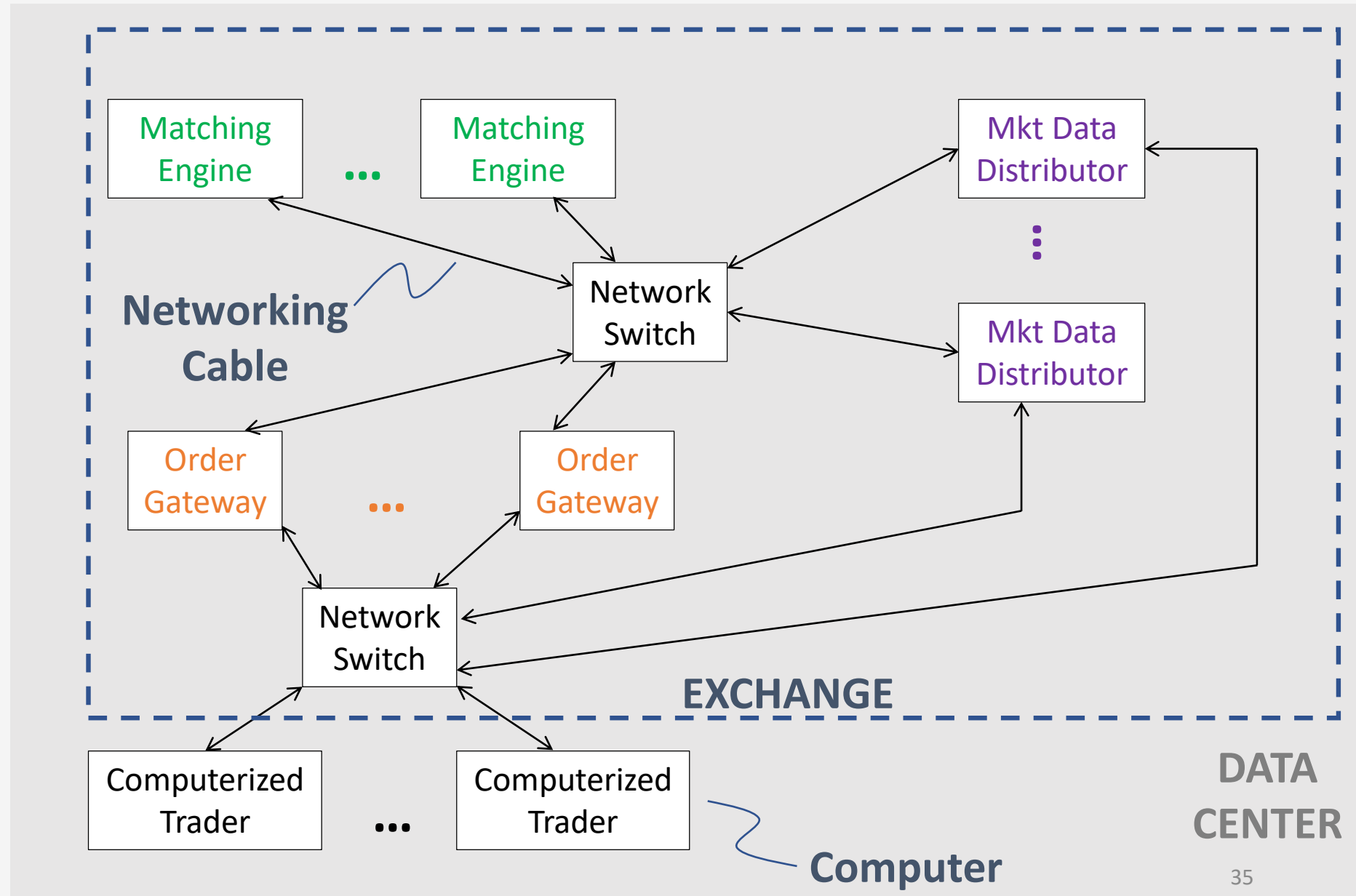
NSE of India case is instructive in understanding how this might happen

https://www.sebi.gov.in/enforcement/orders/apr-2019/order-in-the-matter-of-nse-colocation_42880.html

Threat Model for Unfair Colocation

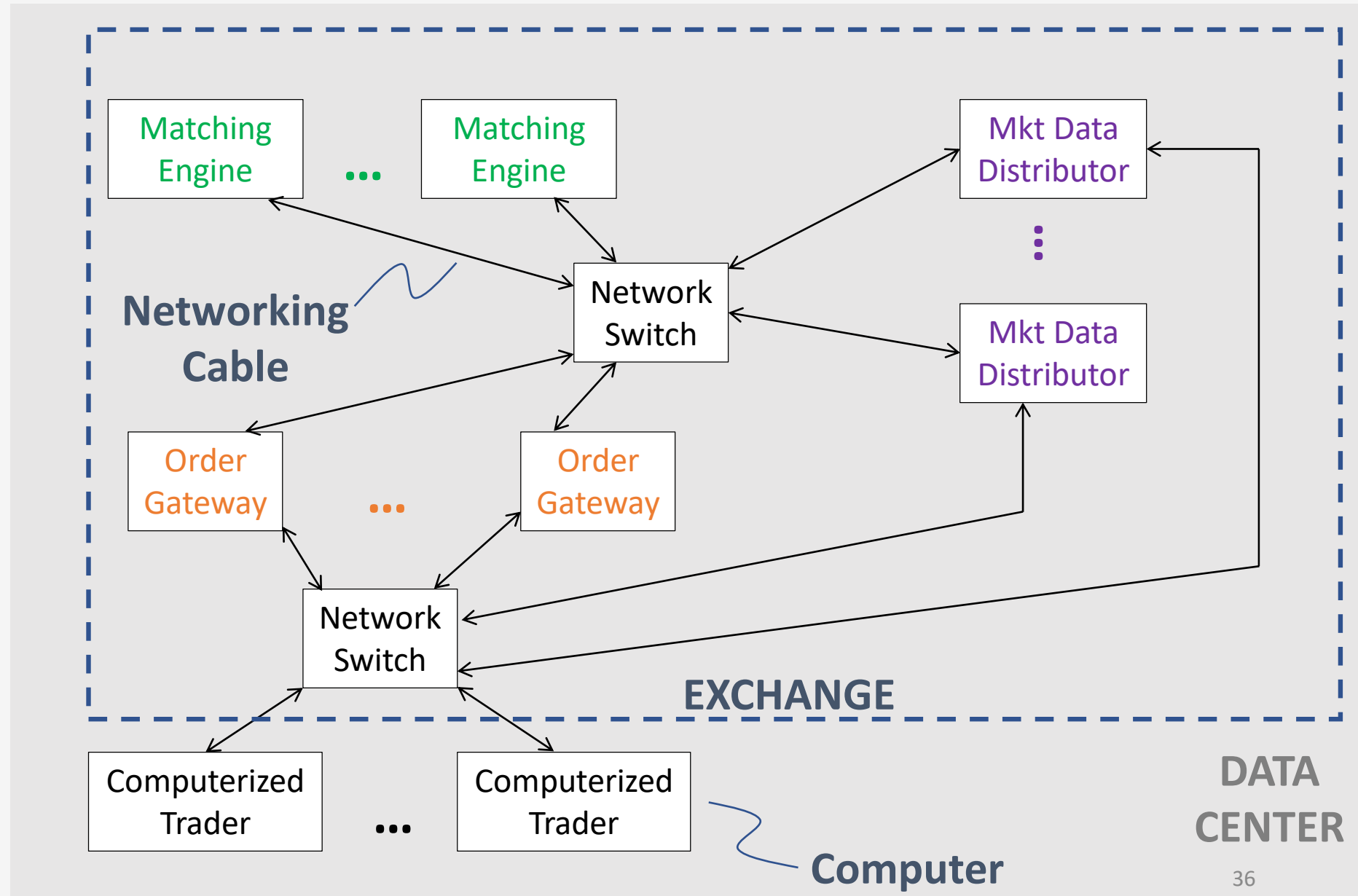
Actor	Threat
Data center operator	<ul style="list-style-type: none">• Margin of error differences in cable lengths (+/- 4m → 20ns)
Exchange insiders	Selectively divulging information about: <ul style="list-style-type: none">• Less loaded/higher performance servers• Ordering in which mkt data updates sent• IP addresses of backup servers
Exchange software designers	Design for: <ul style="list-style-type: none">• High-availability and high-throughput and <i>not</i> fairness (horizontal scaling)• Reliable mkt data delivery over same-time delivery (TCP/IP vs multicast)• Lack of enforcement policy for active-standby servers so both can be used at once
Market Participants	<i>(Subsequent slides...)</i>
Third party software and hardware vendors	<ul style="list-style-type: none">• Changes to arbitration schemes in switches/OS network stack etc not consistent with first-come-first-served style of fairness, but perhaps that result in better throughput fairness• Jitter: time variation for the same operation due to e.g., speculative execution

Horizontal Scaling in Exchange Implementations

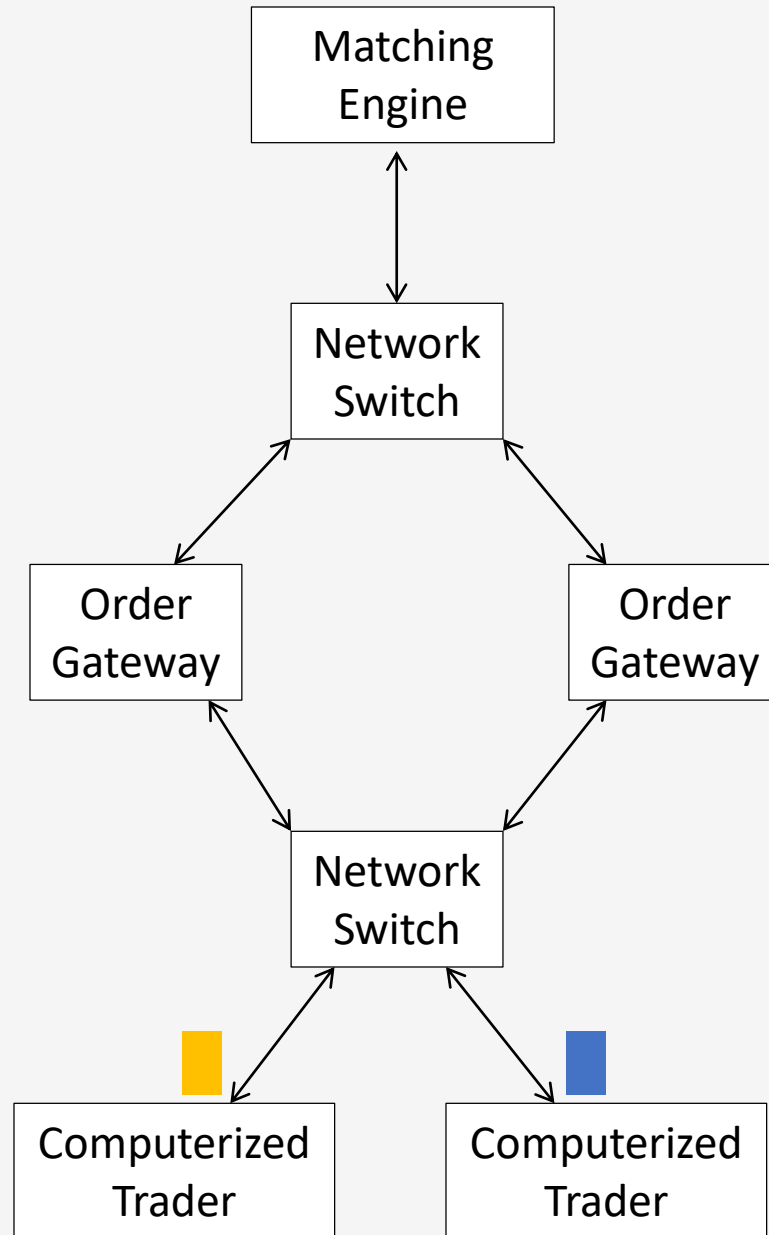


Horizontal Scaling in Exchange Implementations

Same components duplicated over-and-over to improve *performance* and provide *redundancy*

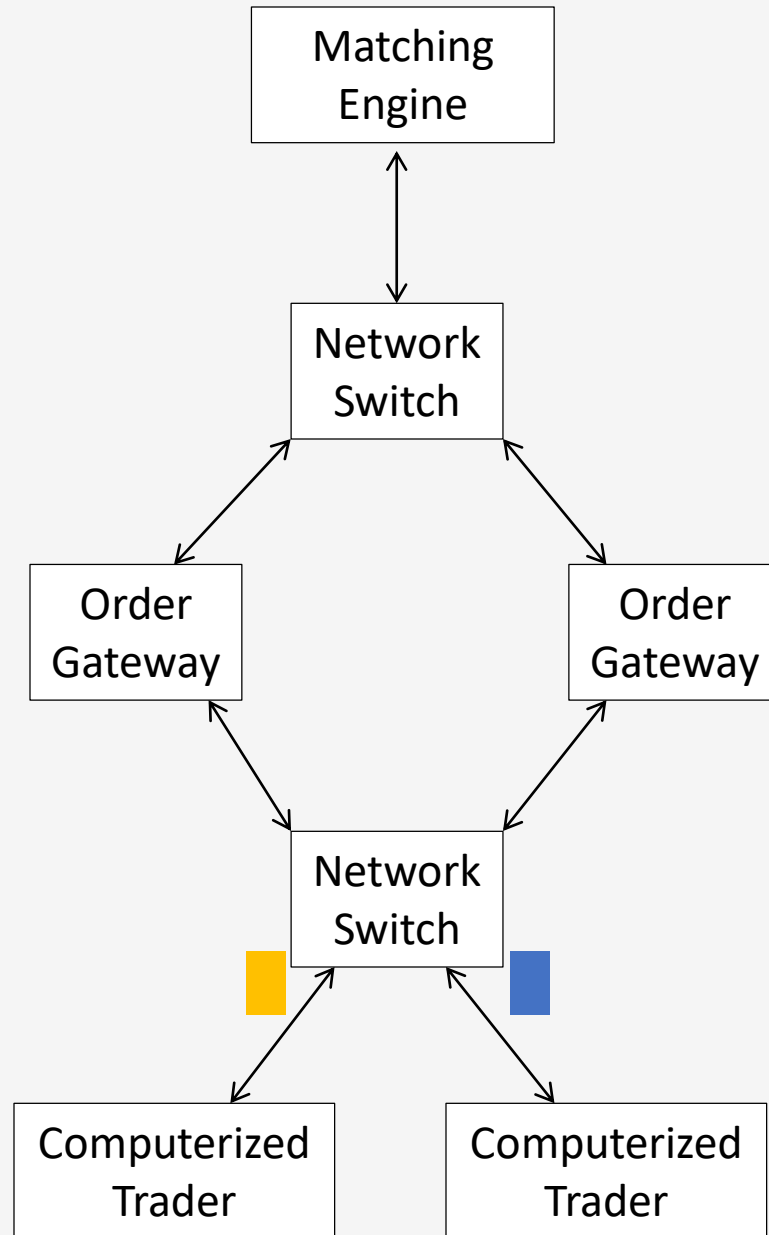


Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

Horizontal Scaling and Computer Network Effects

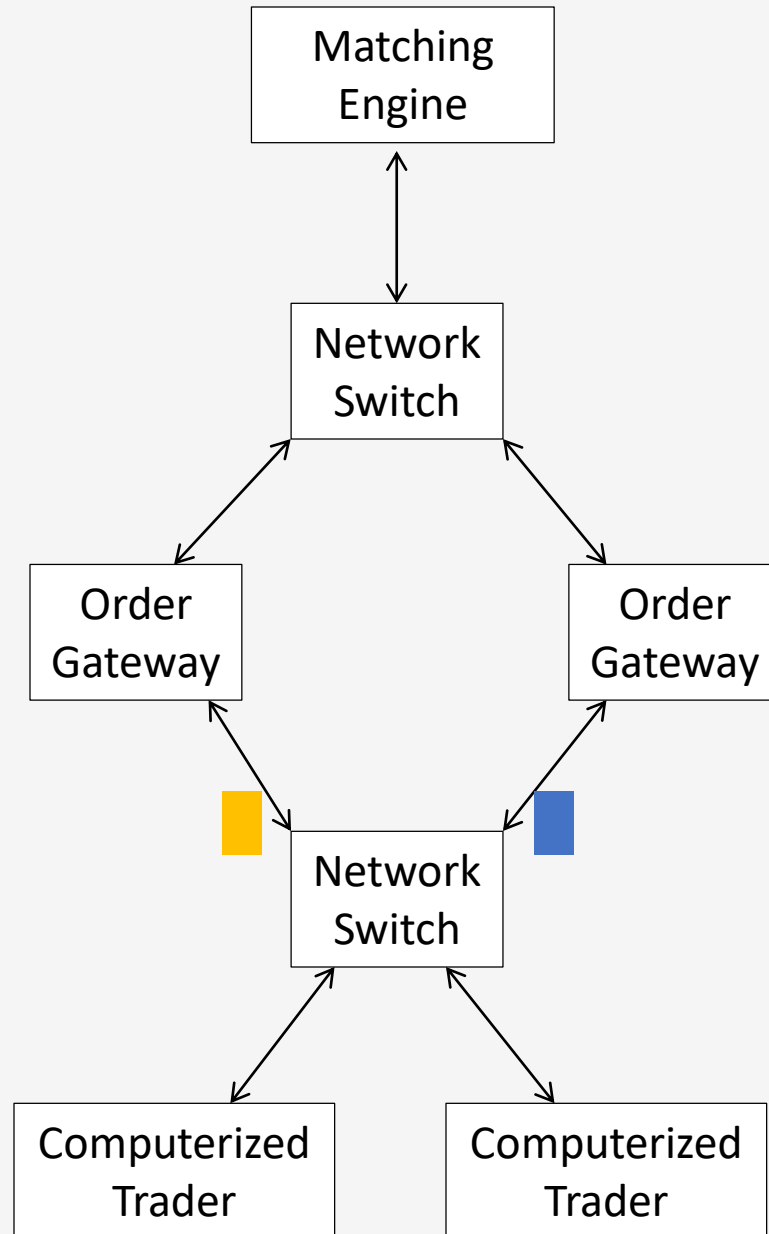


T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

Horizontal Scaling and Computer Network Effects



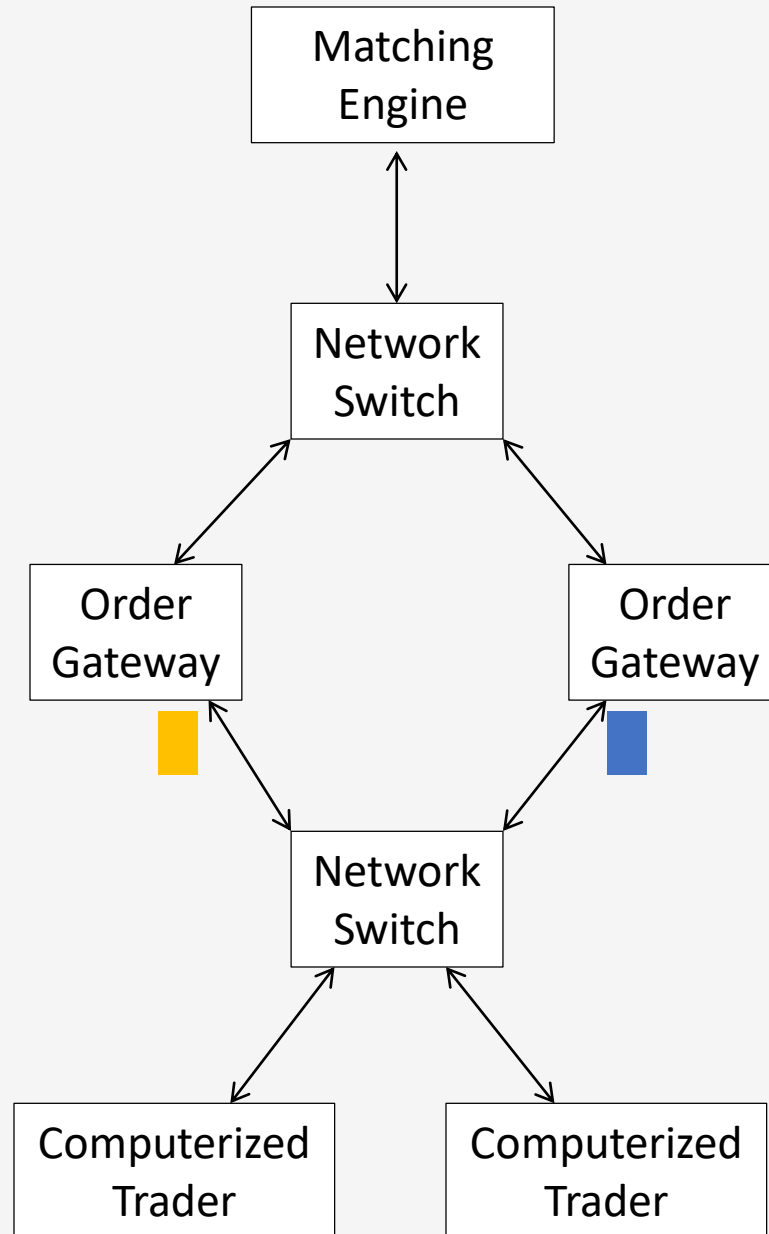
T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

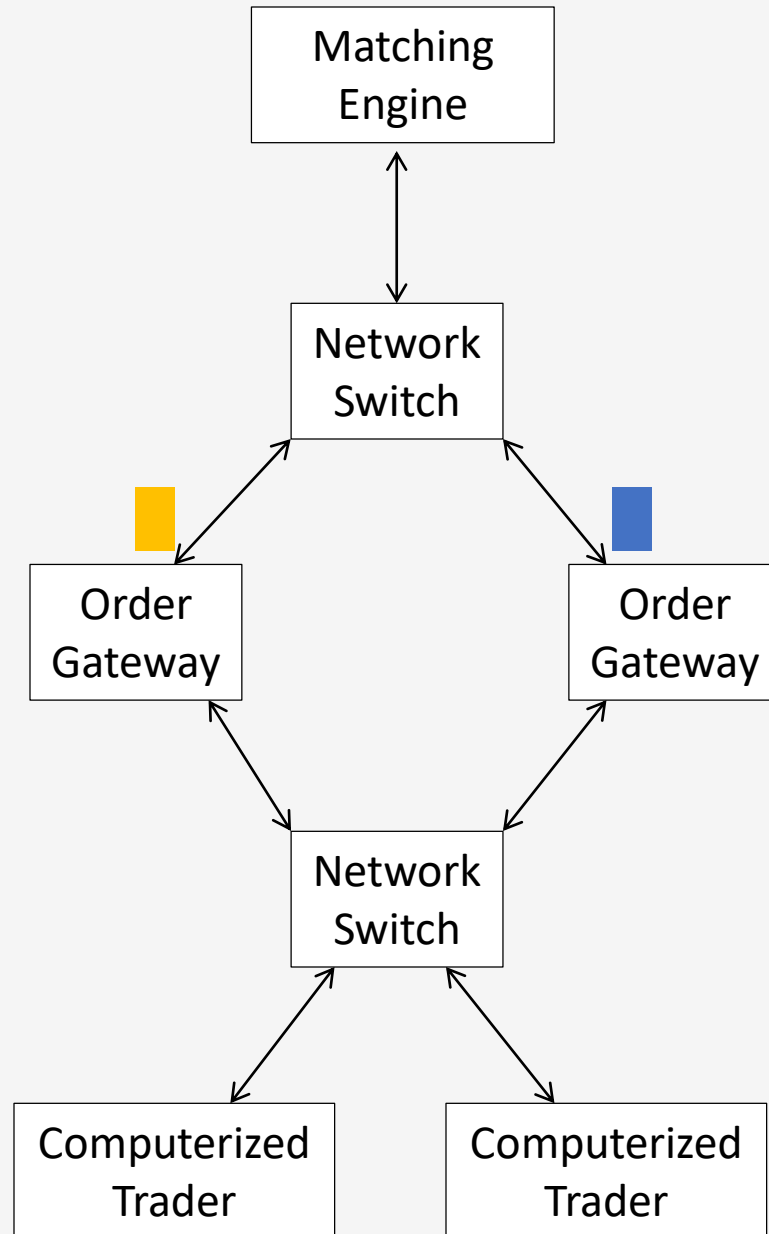
T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

T=3 Gateways receive messages

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

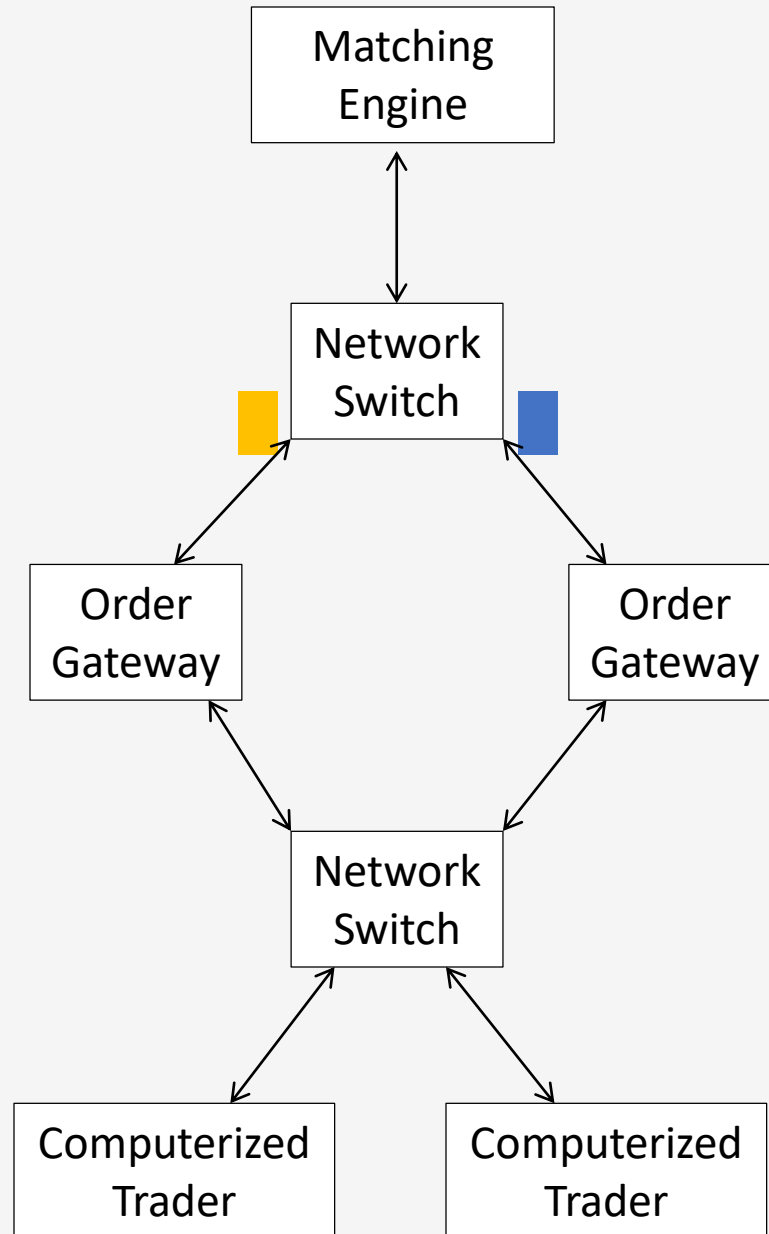
(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

T=3 Gateways receive messages

T=4 Messages leave gateways
(if subject to same delay)

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

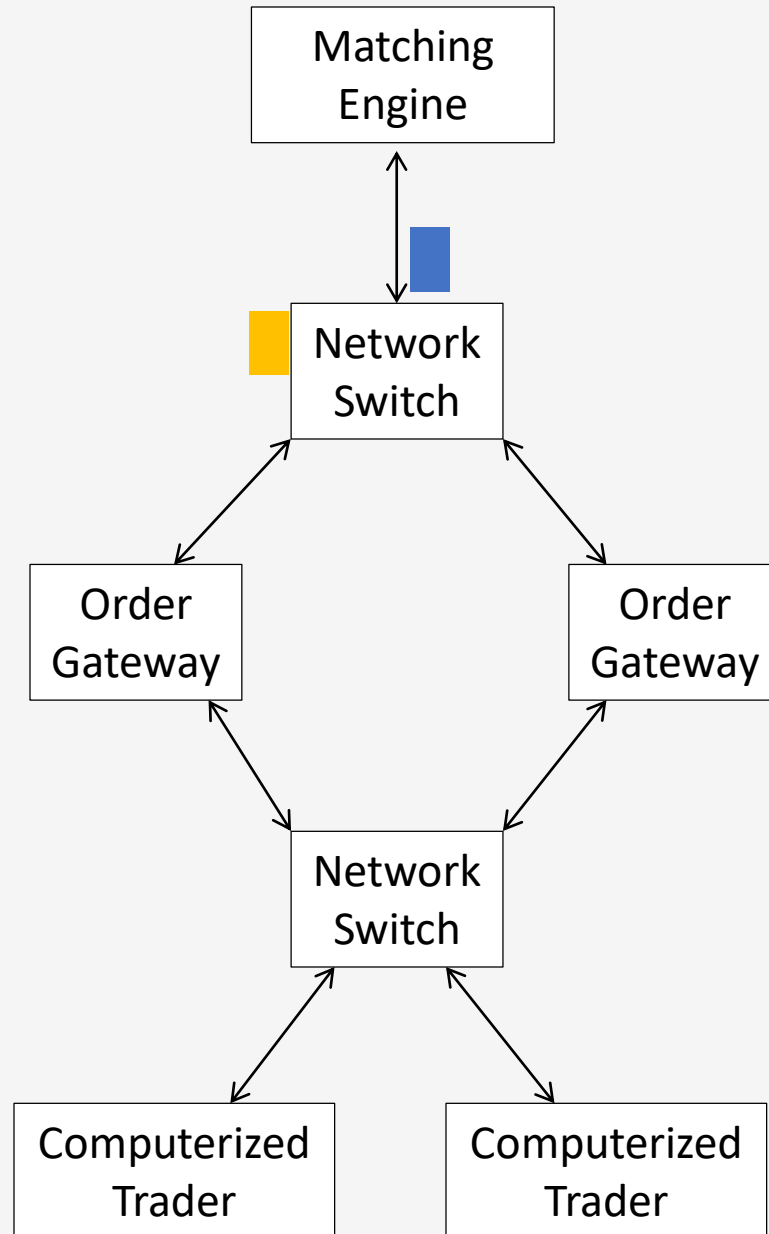
T=2 Network switch forwards to gateways
(if subject to same delay)

T=3 Gateways receive messages

T=4 Messages leave gateways
(if subject to same delay)

T=5 Messages reach second switch

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

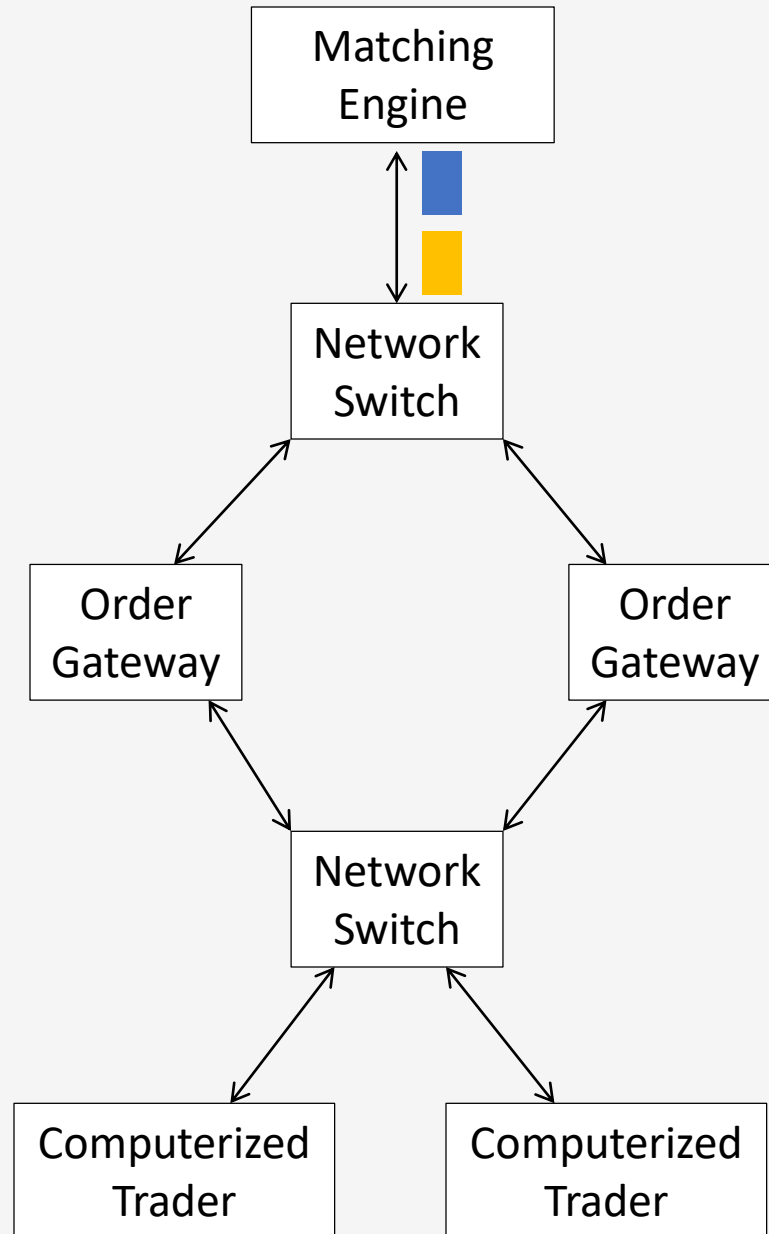
T=3 Gateways receive messages

T=4 Messages leave gateways
(if subject to same delay)

T=5 Messages reach second switch

T=6 First message leaves switch

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

T=3 Gateways receive messages

T=4 Messages leave gateways
(if subject to same delay)

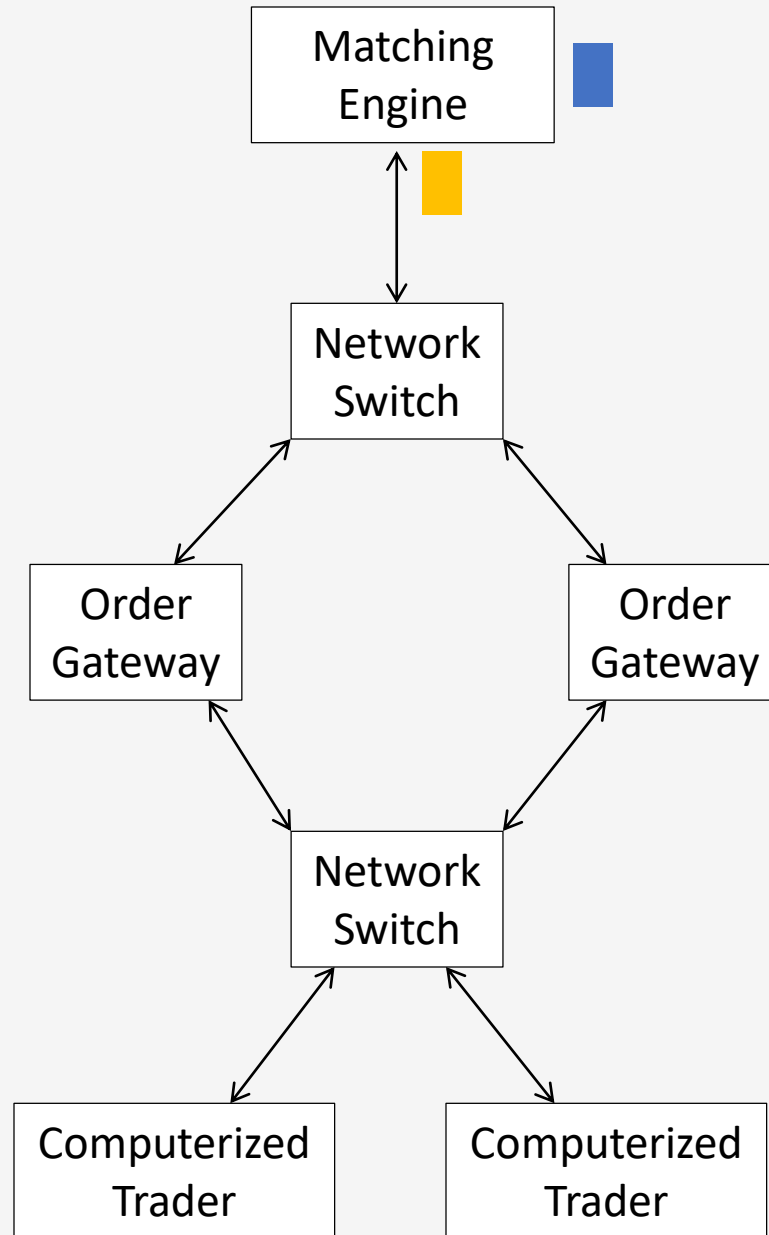
T=5 Messages reach second switch

T=6 First message leaves switch

T=7 Second message leaves switch
(serialization over single link forces total ordering)

First message reaches matching engine

Horizontal Scaling and Computer Network Effects



T=0 Both traders send a message to venue

T=1 Both messages reach first network switch

(if equal cable length & same bandwidth)

T=2 Network switch forwards to gateways
(if subject to same delay)

T=3 Gateways receive messages

T=4 Messages leave gateways
(if subject to same delay)

T=5 Messages reach second switch

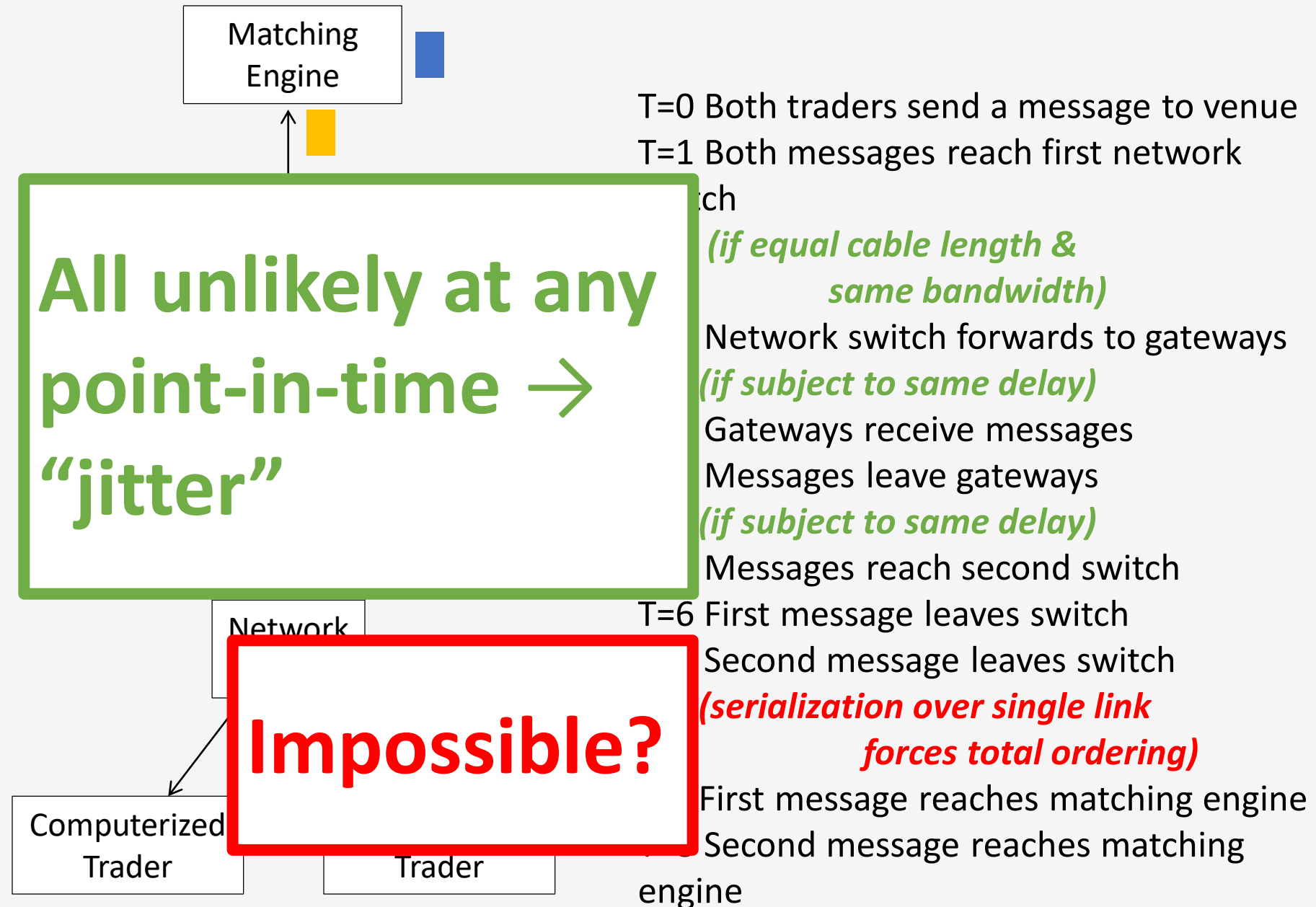
T=6 First message leaves switch

T=7 Second message leaves switch
(serialization over single link forces total ordering)

First message reaches matching engine

T=8 Second message reaches matching engine

Horizontal Scaling and Computer Network Effects



Adversarial Model (1)

- The adversary uses **all the means** at their disposal to gain a competitive **edge**.
- They have **full knowledge** of the exchange's design, infrastructure and processes.
- They are able to exploit advantages that last only a fraction of a **microsecond**.
- They may **collude** with exchange insiders who can provide them with information.
- Use sophisticated **optimization techniques** for software, firmware and hardware.

Adversarial Model (2)

They write their own firmware

“One of the cool things he did was hack the Sangoma cards we were using. Those cards used a fairly small Xilinx FPGA, a Spartan 3, to do most of their work. He hacked a firmware register twiddler into the FPGA and found some registers we could drive.

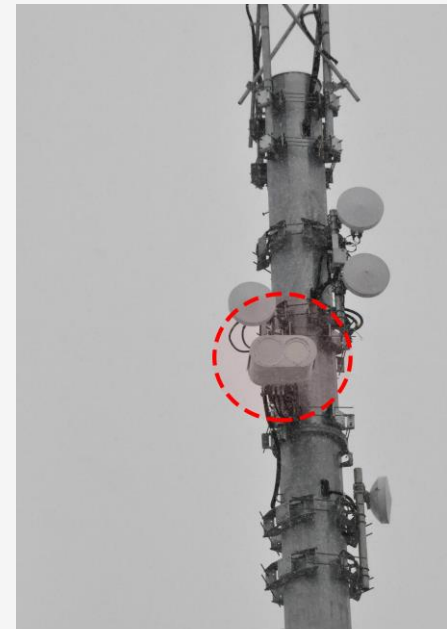
*Before long we had **our own working firmware** and could bypass the Sangoma firmware... ..the Sangoma stack was costing us nearly a whole **millisecond**. The new stack was around ten microseconds.”*

Design their own high performance hardware

“So we did our own schematic and PCB layout to build our own simple PCIe network board. It was direct SFPs to Xilinx FPGA with nothing much else”

<https://meanderful.blogspot.com/2018/01/the-accidental-hft-firm.html>

Deploy their own network links



Known Exploits

Optimistic Messaging

Send all but one of an order's TCP segments before an event. Once the event occurs, depending on its outcome either send or invalidate the last segment. In an 1Gbps feed, for every twenty five bytes you speculate on, you save 200 nanoseconds. (CME)

Denial of Service

Actively congest gateways used by others to slow them down. (NSE of India)

Networking Protocol Violations

Drop order-message characters to reduce latency on a 10Gig/sec network link (CME)

Wire Tapping

Get as close to the data source as possible. Tap into feed wires and bypass the exchange's modems and network devices. Then dump the bits as they come. (KRX)

Bit-by-Bit Reads *(combined with WT)*

Instead of reconstructing whole packets, just read bits. If the first few bits have the info you are interested in then you can submit orders before the whole packet has arrived (KRX)

0-days are sought-after and are exploited until the exchange/others find out!

State of Cryptocurrency Exchanges

- Not much different than “traditional” centralized exchanges.
 - No colocation yet
- Little regulation; many academic studies of price manipulation emerging.
- Low barriers to participant entry; can connect directly to the market
 - E.g., \$25k min balance to day trade equities in US
 - KYC onboarding
- Centralized exchanges do the matching off-the-chain.
- Decentralized exchanges (DEXs) attempt to avoid centralization
- Slow permissionless DEXs suffer from exploitable delays
- Permissioned DEXs are not truly decentralized

Fixing the Markets

Goal:

- Eliminate the effects of varying transmission times among participants that are colocated.
- Put another way, no advantage to technical market manipulation in pursuit of speed beneath some threshold e.g., 1ms

Fixing the Markets

Goal:

- Eliminate the effects of varying transmission times among participants that are colocated.
- Put another way, no advantage to technical market manipulation in pursuit of speed beneath some threshold e.g., 1ms

Solution:

- Accumulate order messages and process them in a (generally) different order to that in which they were received. Obvious!?
- Economists: replace *continuous market* with *batch-style market*

Fixing the Markets

Goal:

- Eliminate the effects of varying transmission times among participants that are colocated.
- Put another way, no advantage to technical market manipulation in pursuit of speed beneath some threshold e.g., 1ms

Solution:









- Accumulate order messages and process them in a (generally) different order to that in which they were received. Obvious!?
- Economists: replace *continuous market* with *batch-style market*

GOTCHA: Mechanism itself needs to be resistant to manipulation!

Design of the *Ideal Latency Floor* Mechanism

Accumulate orders in buffers before they are processed against LOB:

- One buffer for each distinct resource in contention (*resists manipulation*)
- Timer for each buffer; first order starts timer for say 1ms
- Lottery-style release from buffer; one lottery ticket per firm (*also for resistance*)
- Buffers are “sealed”

Limit Order Book for instrument NZD/USD				
Offer orders		Prices	Bid orders	
 [C,2Mio][D,1Mio][B,10Mio][A,1Mio]		0.7343	}	
 [C,2Mio][A,1Mio]		0.7342		
 [C,2Mio][B,5Mio][E,1Mio]		0.7341		
		0.7339	}	[D,1Mio][B,5Mio][A,1Mio][E,3Mio] 
		0.7338		[A,2Mio][C,1Mio][D,5Mio] 
		0.7337		[C,2Mio] [A,5Mio] 

Conclusions

- Market abuse is a complicated topic!
 - Regardless, markets now are probably the most efficient through all history
- Lots of *published* technical market manipulation techniques
 - Disciplinary notices, court cases, insider blogs, academic papers,...
- Looked at ameliorating effects of speed, and techniques in its pursuit
 - Ideal Latency Floor deployed on Refinitiv Spot FX Matching
 - Not all techniques *manipulative*; some just “socially wasteful”
- Speed and traditional security perspectives:
 - Confidentiality, Integrity (parties to trades), Availability (nanosecond DoS)
 - Prevention vs Detection Mechanisms

Can cryptocurrency exchanges can learn from traditional exchanges?

Questions!