OPENFABRICS ALLIANCE

12th ANNUAL WORKSHOP 2016

MULTICAST USE IN THE FINANCIAL INDUSTRY

Christoph Lameter

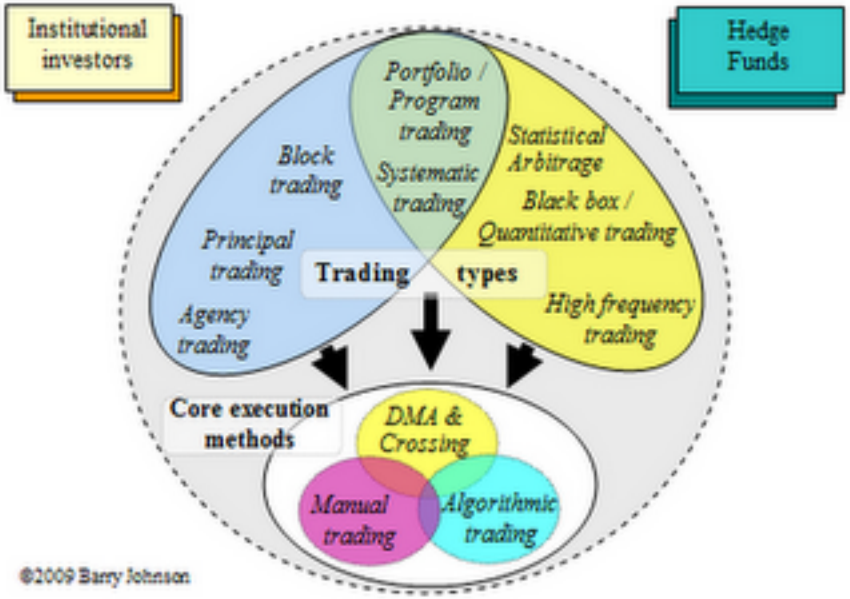
GenTwo

[April, 5th, 2016]

OVERVIEW

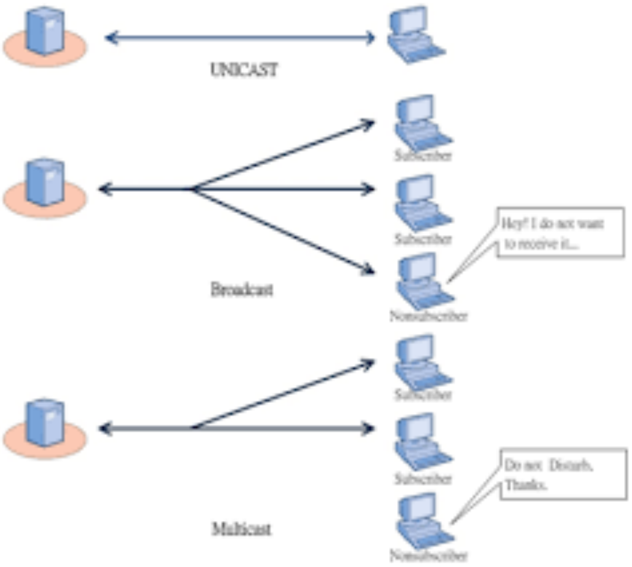
Multicast and the FSI (Financial Services Industry)

- **Short refresher on Multicast**
- **Benefits of Multicast**
- **Rationale for use in Finance**
- **Current Technology deployments**
- **Unusual requirements in Finance**
- **Future Developments**



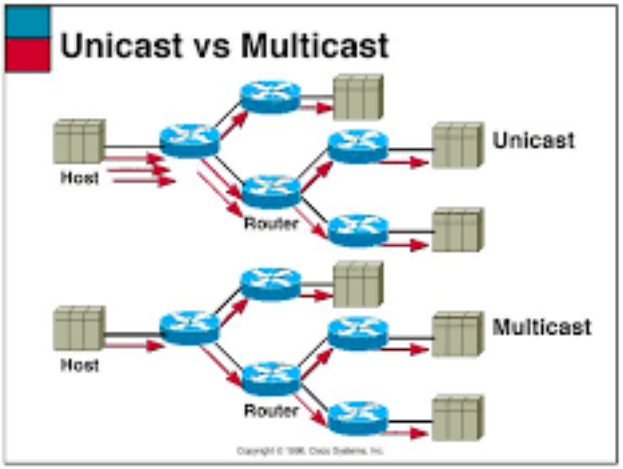
WHAT IS MULTICAST

- Send a single message to multiple receivers
- In “Hardware”. Routers / Switches must support Multicast.
- Publisher / Subscriber model using Channels, Topics, Multicast address to identify data streams.
- Datagrams are network node independent
- Basic Problems
  - Multicast often unknown
  - Off by default in most configurations of routers/switches
  - Complicated configuration
  - Vendors often provide unstable multicast functionality

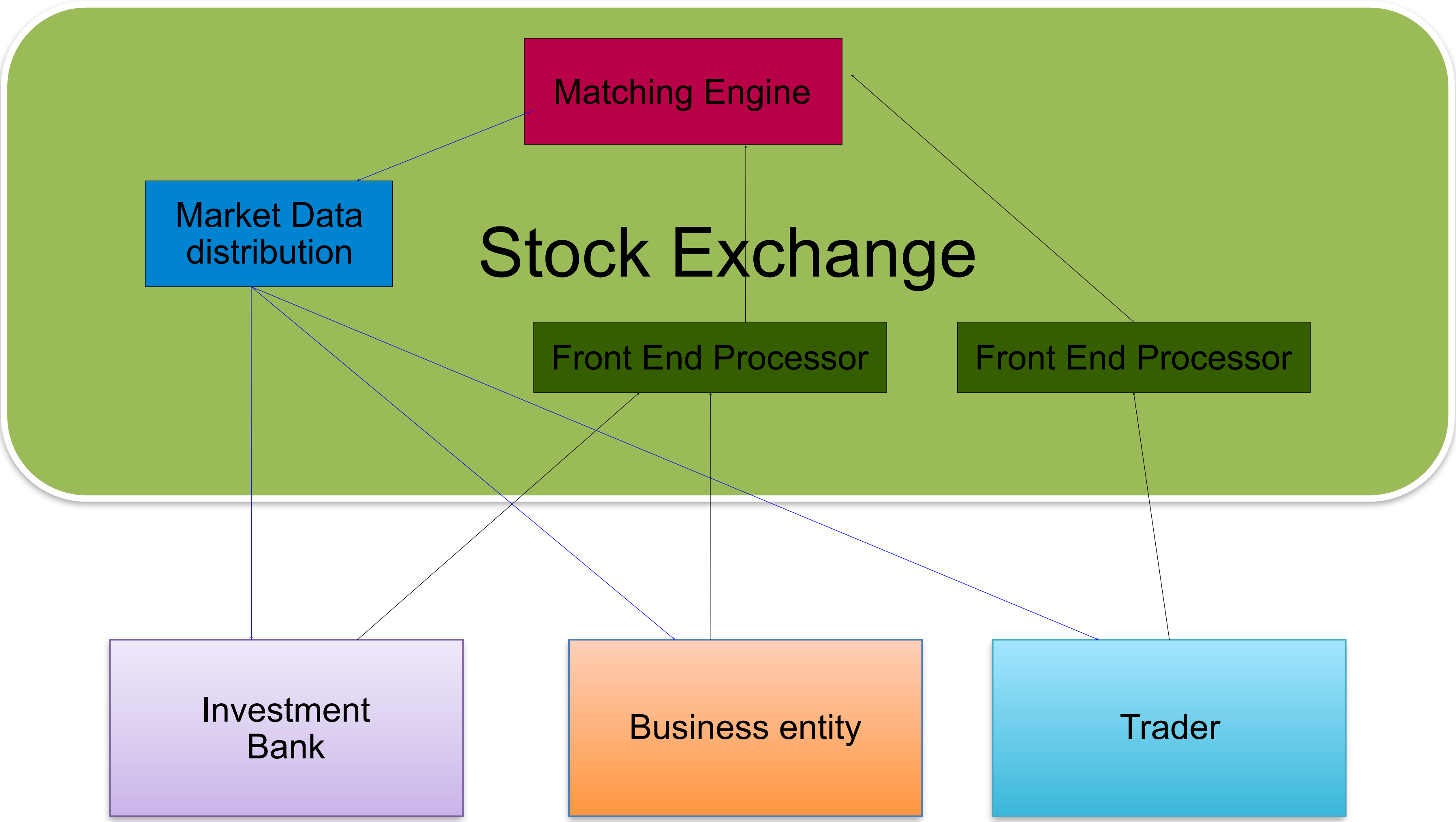


RATIONALE FOR MULTICAST IN FSI

- **Basic: Fast and scalable event notification**
- **100s to 10000s of receivers of multicast traffic**
- **Stock Exchanges distribute Market Data that reflect changes in prices. Systems in FSI need to react to these change fast.**
- **Bandwidth issues with unicast.**
- **Problems with congestion protocols**
- **Tradition: FSI has been using Multicast for decades and its brewed into the Unix Posix standard as well as into the basic design of Ethernet routers and switches.**



TYPICAL FSI DATA FLOW



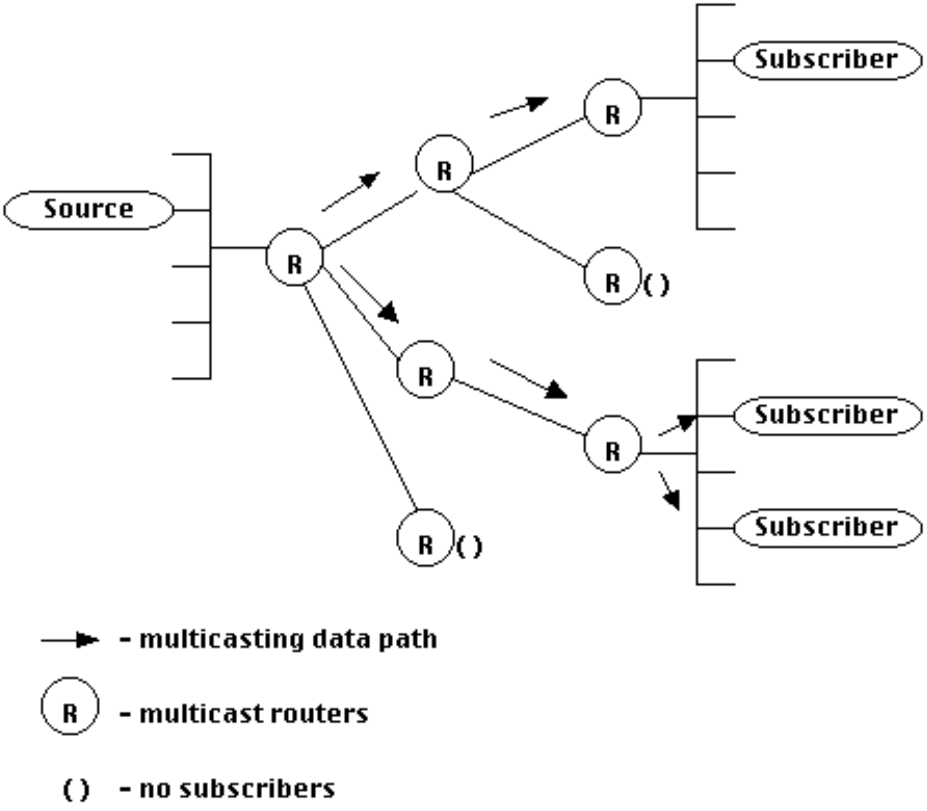
REQUIREMENTS IN DIFFERENT AREAS OF FSI

- Stock Exchange
  - Fairness of access
  - Reliability
  - Capacity / Bandwidth
- Bank
  - Reliability
- Investment Fund
  - Reliability
- Algorithmic Trading Companies
  - Latency
  - Capacity

PARTICULAR USE CASES WHEN INTERACTING WITH A STOCK EXCHANGE

Classic mission critical data flow in FSI

- **Market Data widely distributed from the Stock Exchange**
- **Processed and logic applied which results in sell or buy orders.**
- **Unicast transaction with the Stock Exchange which leads to the creation of market data describing the transaction.**
- **Alternate event sources: News Analysis, Twitter and misc sources.**



Market data is price and trade-related data for a financial instrument reported by a trading venue such as a stock exchange. Market data allows traders and investors to know the latest price and see historical trends for instruments such as equities, fixed-income products, derivatives and currencies.

OPENFABRICS ALLIANCE

MULTICAST IN AN RDMA BASED FABRIC

CHARACTERISTICS OF FSI MULTICAST TRAFFIC

How the Infiniband API is used

RFC 3208

PGM uses the concept of Negative Acknowledgements (NAKs). A NAK is sent unicast back to the host via a defined network-layer hop-by-hop procedure whenever there is a detection of data loss of a specific sequence. As PGM is heavily reliant on NAKs for integrity, when a NAK is sent, a NAK Confirmation (NCF) is sent via multicast for every hop back. Repair Data (RDATA) is then sent back either from the source or from a Designated Local Repairer (DLR) at some point closer to the destination.

- Short messages (within Ethernet Frame size)
- Smaller messages are used for faster communications
- Unconnected mode. UD datagrams.
- Typically confined to the same subnet or partition of the fabric. Often a need to cross to Ethernet.
- No one sided transactions. There is no multicast implementation for RDMA in the sense of remote memory operations.
- Multicast must use messaging through the RDMA APIs.
- FSI requires reliability. The Fabric offers actual reliability through flow control for UD. Its possible to have reliable datagram transmission through hardware facilities.
- This “reliability’ sometimes does not work in corner cases. Thus often traditional Multicast recovery protocols are used.
- Reliability through a counter in each message to be able to check for missing packets and protocols that allow retrieving missed packets. See f.e. the PGM Protocol which is the basis of many of the reliability through software middle ware available. See RFC 3208 for details.

FUTURE / WISHLIST

IGMP operates between the client computer and a local multicast router. Switches featuring IGMP snooping derive useful information by observing these IGMP transactions. Protocol Independent Multicast (PIM) is then used between the local and remote multicast routers, to direct multicast traffic from the multicast server to many multicast clients.

IGMP operates on the network layer, just the same as other network management protocols like ICMP.

The IGMP protocol is implemented on a particular host and within a router. A host requests membership to a group through its local router while a router listens for these requests and periodically sends out subscription queries.

- Time Stamps on receive and send
- Multicast loopback prevention
- Packet Aggregation
- Offloads for Checksums, Segmentation
- Support for variety of multicast protocols (IGMP, PIM sparse, Mrouter, Bridging to other fabrics like Ethernet)
- Enhanced reliability
- Complexity of bridging the difference between multicast protocols on Infiniband and Ethernet
- Multicast Routing and failover when traversing gateways
- Multicast traffic load balancing
- Better solutions for the slow consumer problem

OPENFABRICS ALLIANCE

12th ANNUAL WORKSHOP 2016

THANK YOU

Christoph Lameter

GenTwo