



Financial Information Forum Latency Survey Report

February 24, 2011

FIF Latency Survey Overview

Survey Objective: To better understand which latency measurements and statistics will be of value to FIF members.

Survey Questions:

- Which latency would you like to measure (see diagram):
 - Order to Exchange Trading System (See A on Figure 1)
 - Exchange Trading System to Prop Feed Latency (See B on Figure 1)
 - Exchange Trading System to SIP Feed (See C on Figure 1)
 - SIP Quote to SIP message (See D on Figure 1)
- When would you like latency measured?
- Which latency metrics are of interest
 - Max, Min, Mean, Standard deviation etc.
 - Percentiles (90%, 95%, 99.5%, 99.9%)
- Time stamp granularity (Millisecond/Microsecond)
- Do you measure latency on different order types?
- Do you measure latency on different message types (e.g. cancel/replace vs. new order)?
- Would you be willing to share latency statistics on exchange response times (anonymously) with the FIF Market Data Capacity WG?

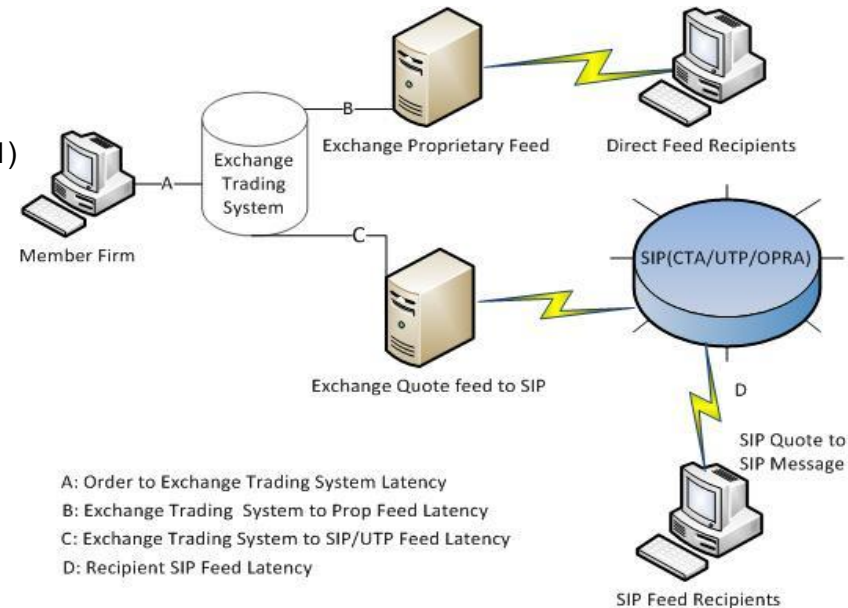
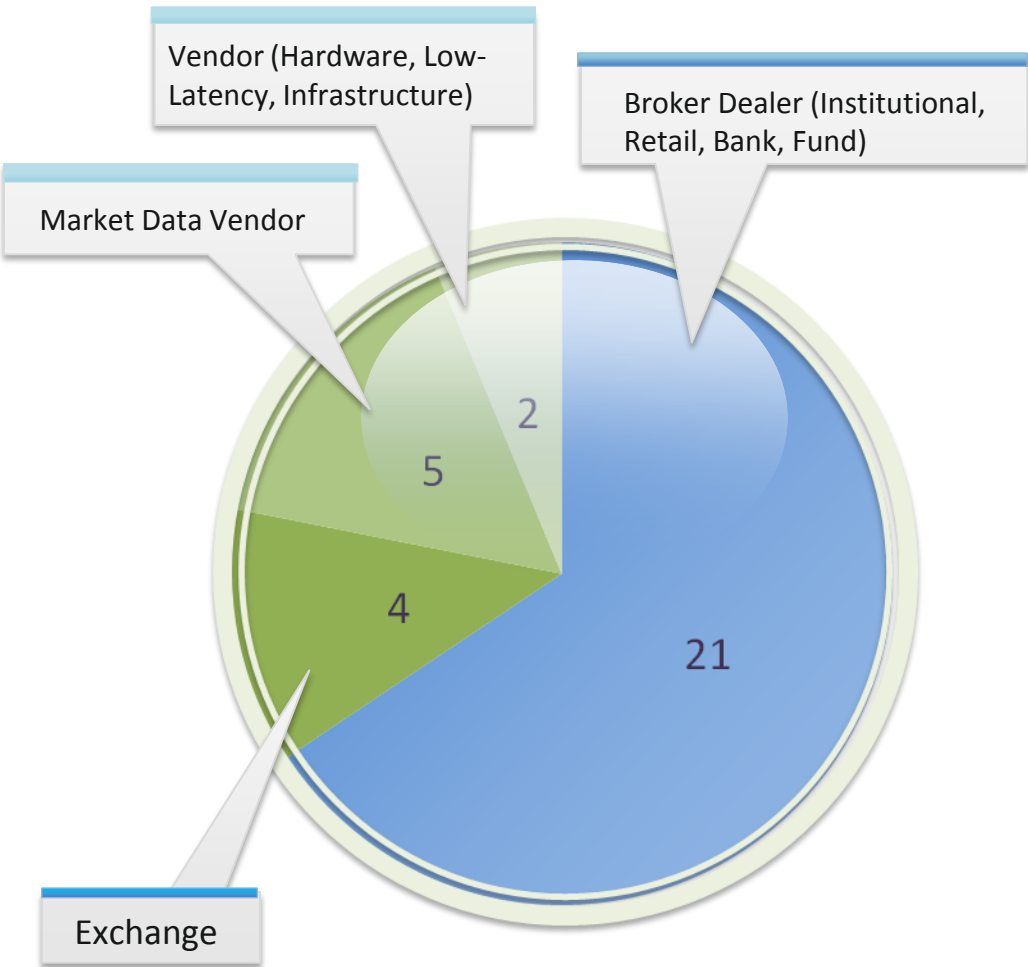


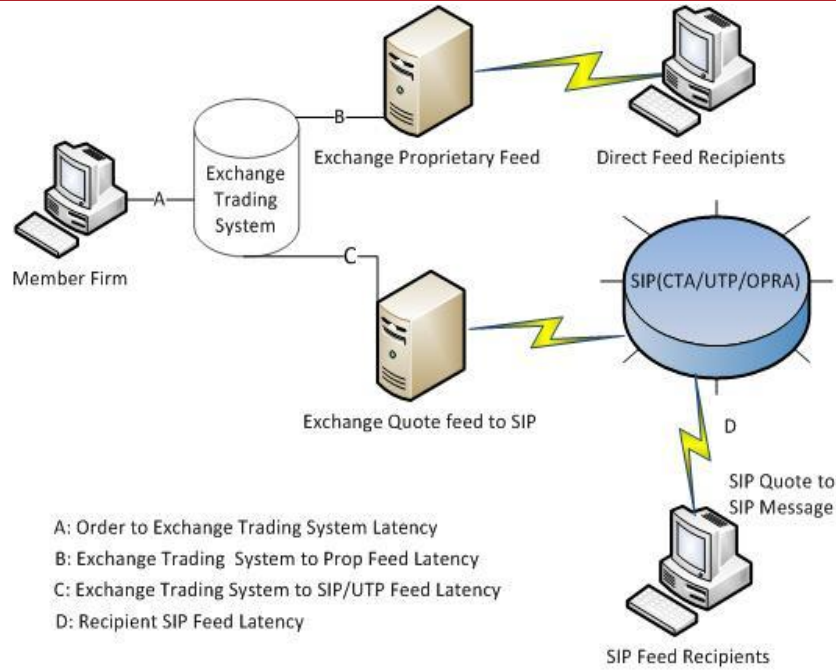
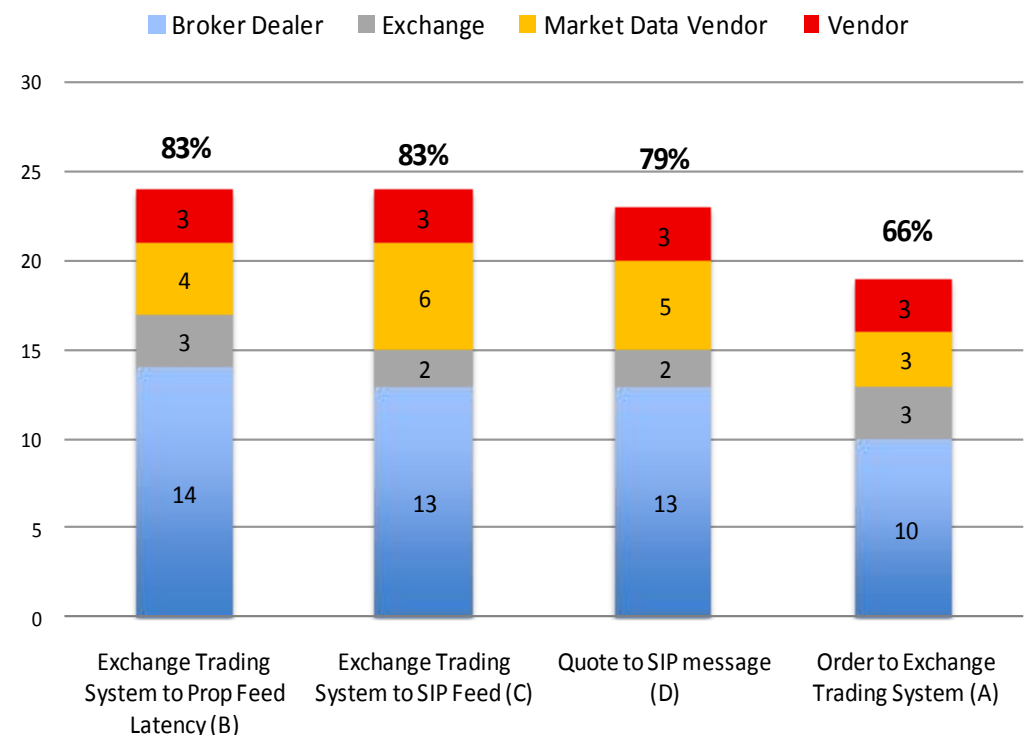
Figure 1 – Latency Network Diagram

Survey Participation – 32 Firms

Firm Type	Firm Name
Broker Dealer	BNP Paribas, Barclays Capital, ETRADE Financial, Fidelity Investments, Knight, Morgan Stanley, Scottrade, UBS and 13 other Broker Dealers
Exchange	BATS Exchange, NYSE Euronext and 2 additional exchanges
Market Data Vendor	Bloomberg, Interactive Data, RealTick, Spryware, Thomson Reuters
Vendor	Exegy, Mantara



Which latency are firms interested in measuring?



A: Order to Exchange Trading System Latency
 B: Exchange Trading System to Prop Feed Latency
 C: Exchange Trading System to SIP/UTP Feed Latency
 D: Recipient SIP Feed Latency

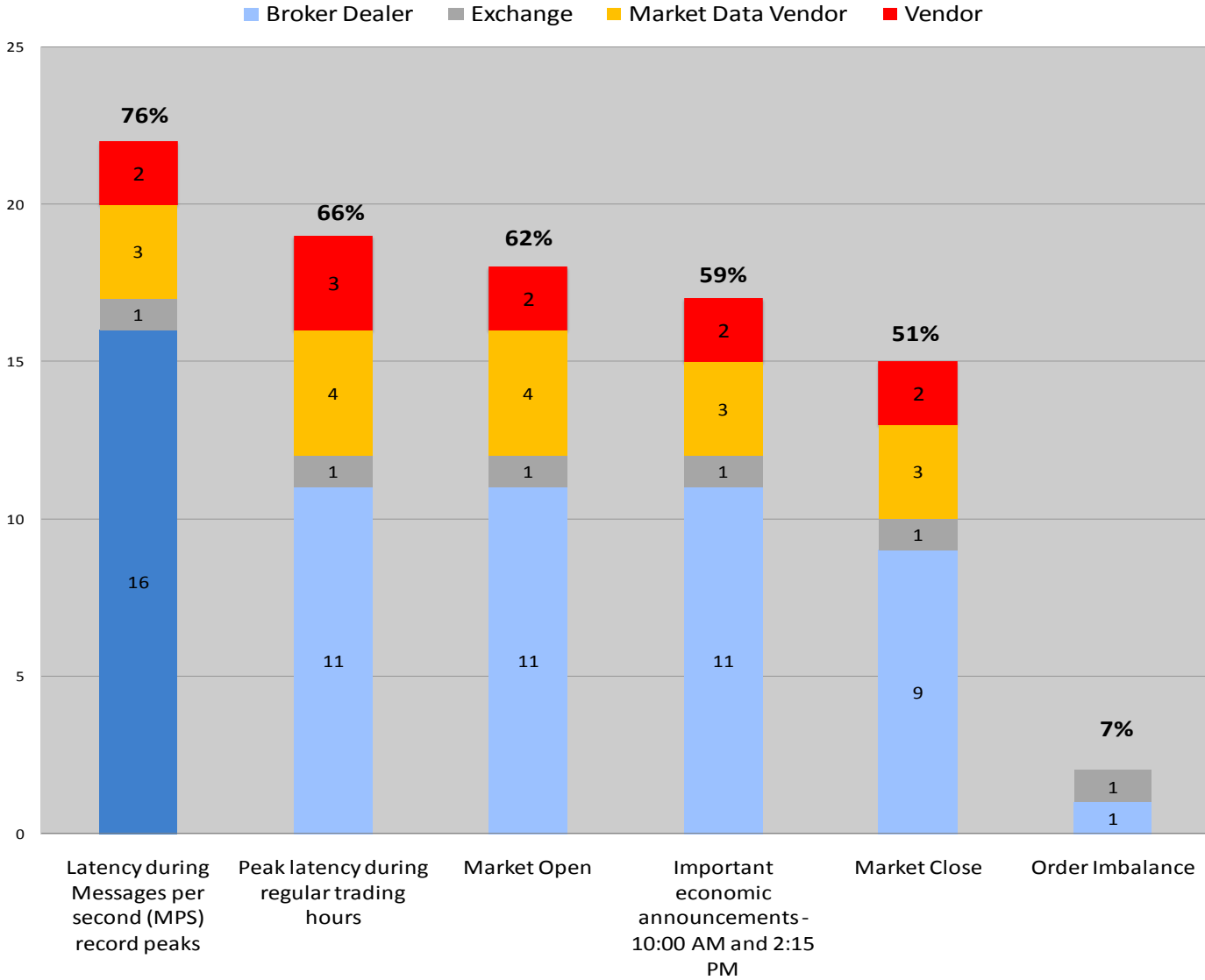
Figure 1 – Latency Network Diagram

Latency Measurement Comments

All of these are very interesting metrics to have. Although the last hop is the most useful, the others are helpful as well. Order to Exchange (A) needs to take into account latency between Member Firm outside network and Exchange Trading System outside network as one stat. Latency within a Member Firm's network should be considered a separate component measurable by the Member Firm. For latency sensitive firms, it would seem Order to Quote latency of proprietary feeds would be of most value as those would be the data feeds they would most rely on given SIP feeds will always require an extra hop through the SIPs and would always incur more overall latency.

Response Rate: 91%

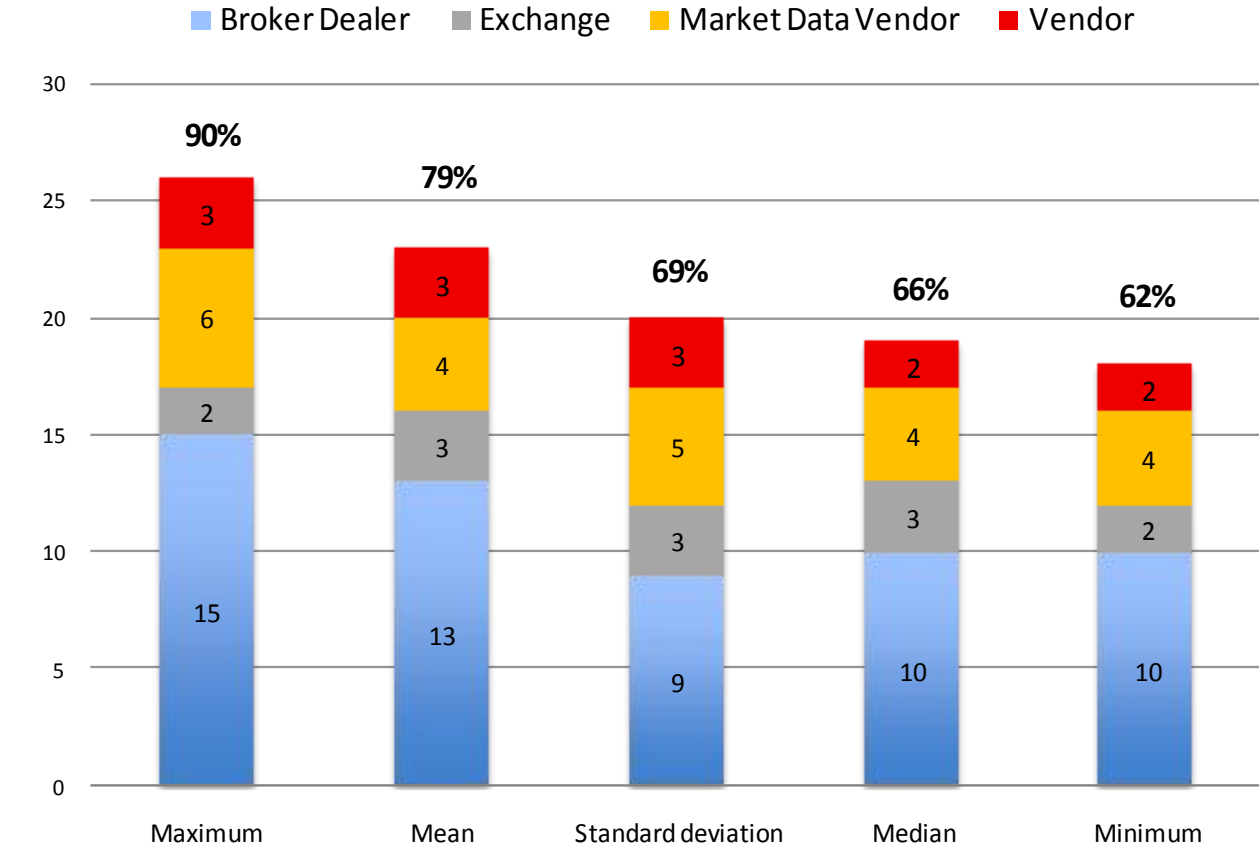
What time of the day is best for latency measurement?



Additional Comments
 An average system latency (normal operating) would also be useful, so as to determine deviations during market open, market close and at the record peaks.
 An overall number for the day.

Response Rate: 91%

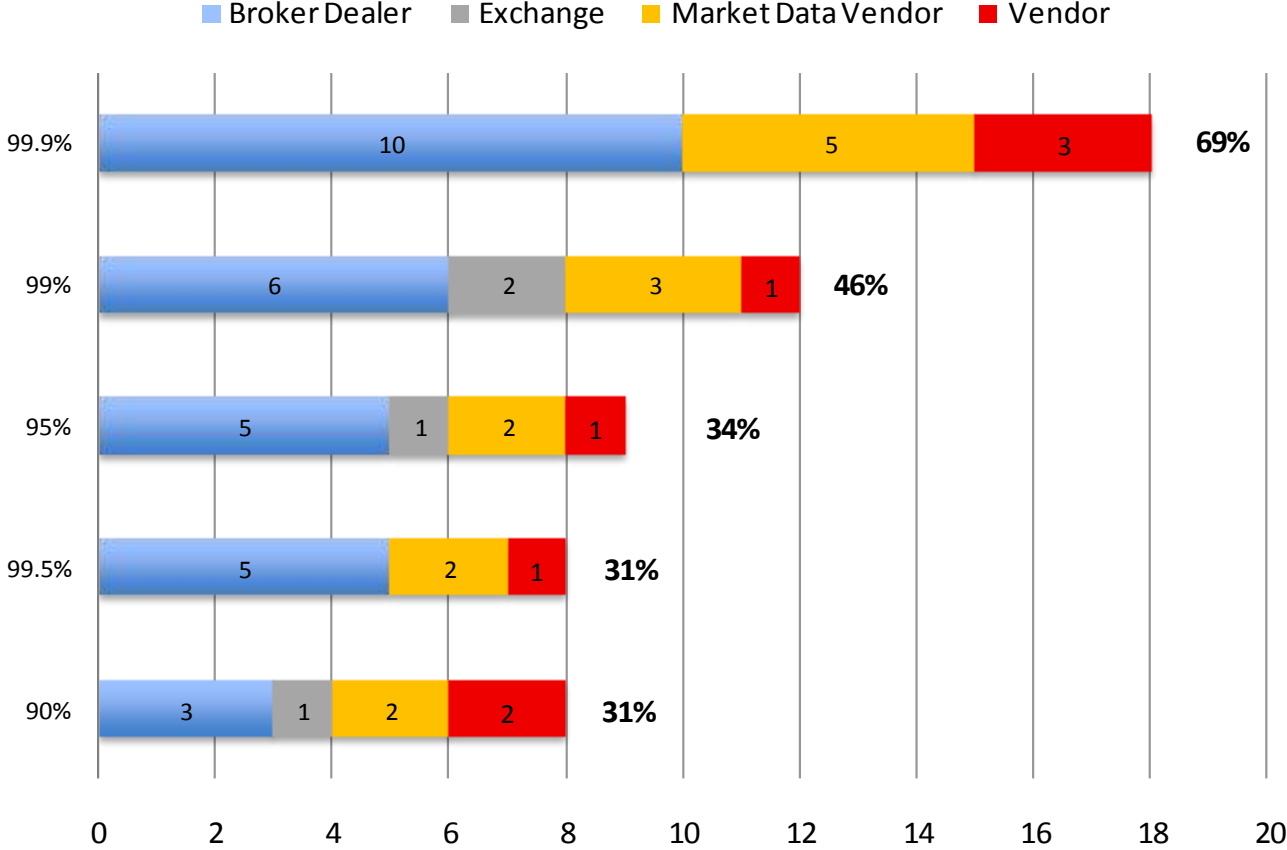
Latency Measurement – Metrics of Interest



Additional Comments
 Mode is also of interest
 Max and Min can be single outliers which are not indicative.
 It would be nice to have access to the raw data.

Response Rate: 91%

Latency Measurement – Percentiles of Interest

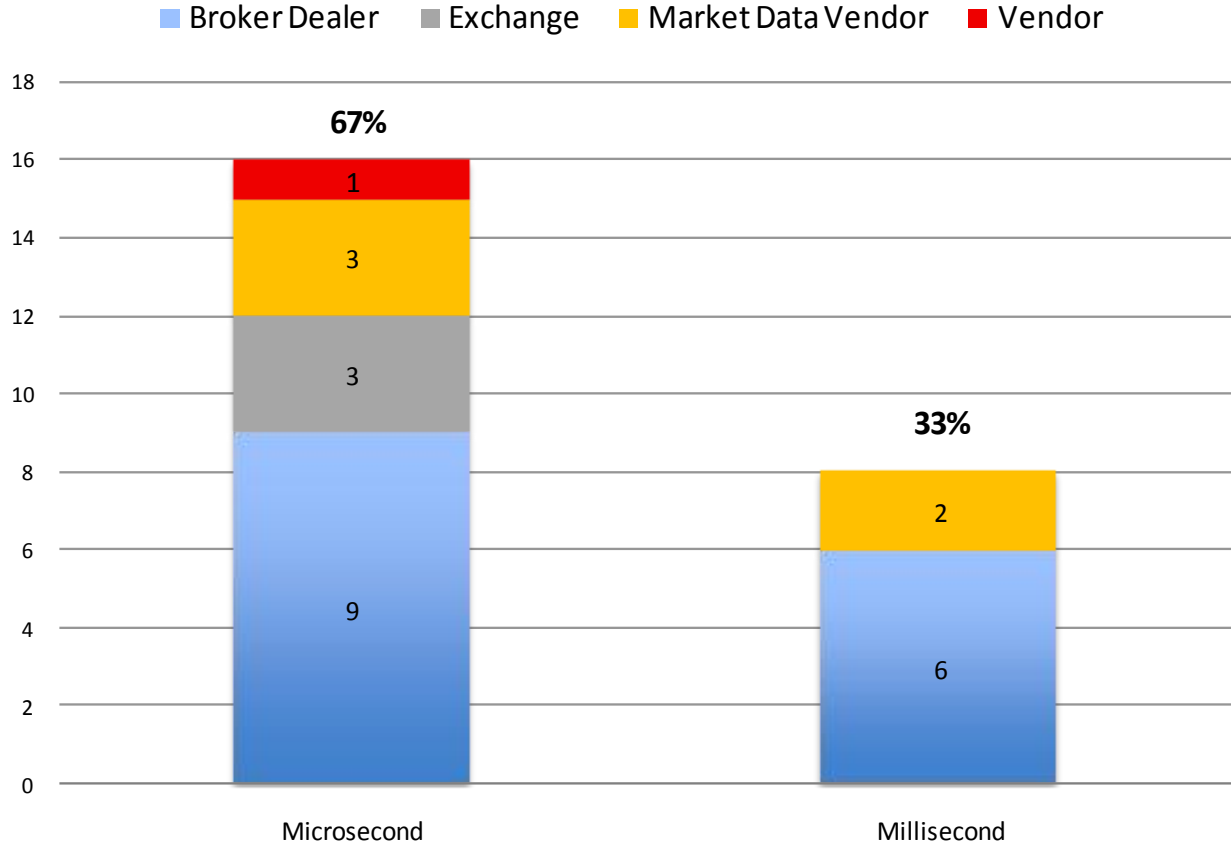


Percentiles of Interest - Comments

By breaking orders up into buckets (e.g. 10 microseconds) and collecting a count of order per bucket, you get a histogram that can easily be used by any firm to evaluate for any percentile that would be of interest to them.

Response Rate: 81%

Latency Measurement – Time Stamp Granularity



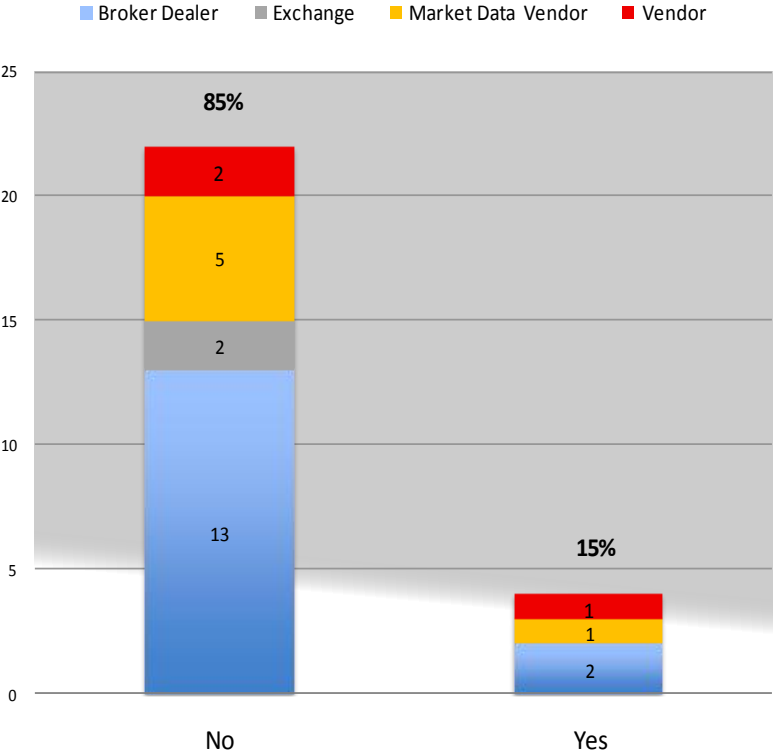
Time Stamp Granularity - Comments

Though microsecond is preferable, whatever granularity is available or appropriate for the system being measured.

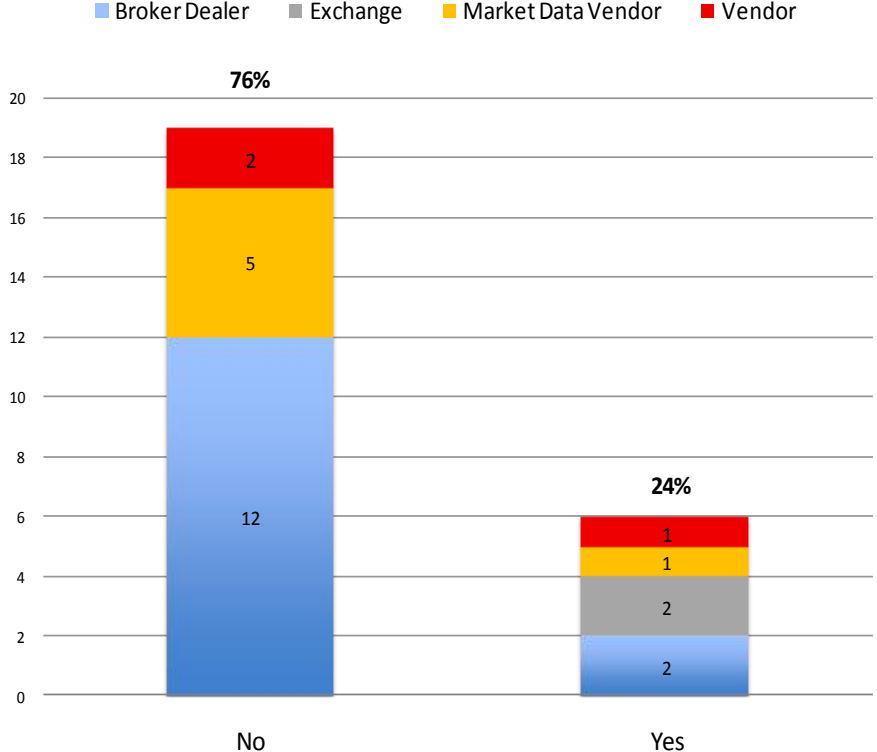
Response Rate: 63%

Do you measure latency on different order and message types?

Latency measurement for different order types



Latency measurement for different message types (e.g. cancel/replace vs. new)

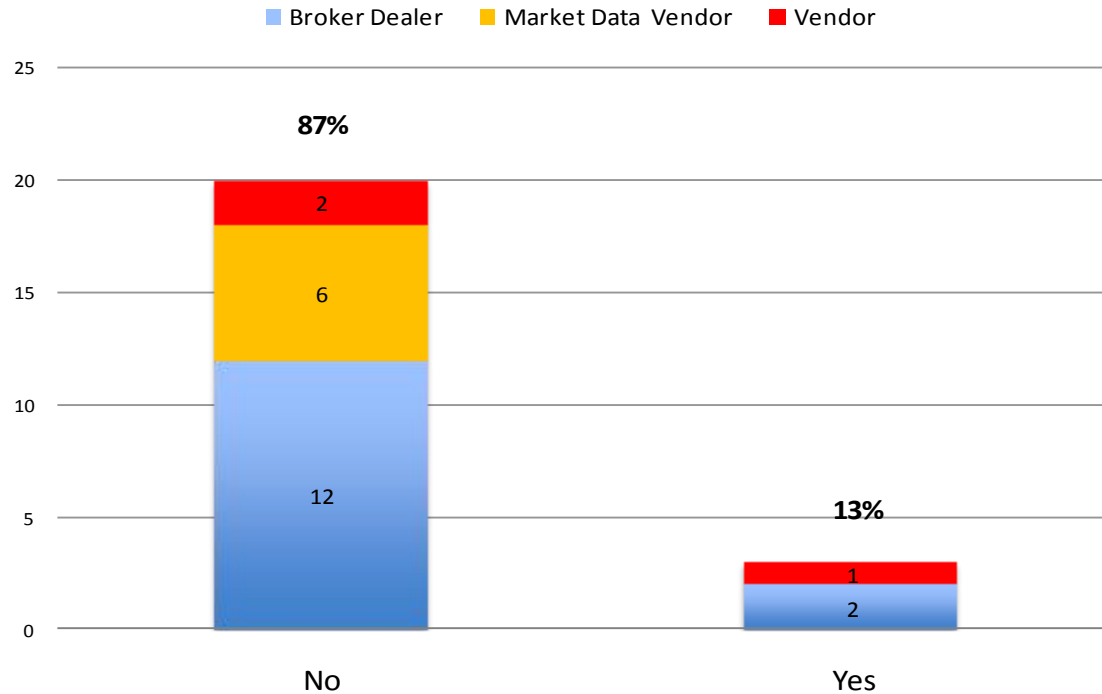


Different Order/Message Types - Comments

We have not specifically concentrated on different order types. Something we need to do.
 Cancels generally have higher latency than order accepts (accepts can be faked). Out on IOC is meaningful as well (accept + implied cancel). Order Ack, Cancel Ack, and IOC out are all useful.
 Since all orders follow the same path, there should be little difference in latency for different order types.
 Generally, orders that matches in our matching system have a higher latency than cancels.

Response Rate:
 Latency Measurement for different order types: 72%
 Latency measurement for different message types 69%

Would you be willing to share exchange latency statistics anonymously with the FIF MD Capacity WG?



Comments - Share Statistics

Only for market data, not order entry.

We have a excellent metric of exchange response times, for 50+ market centers.

We are open to sharing our markets latency stats, but at this time are not planning to redistribute specifics regarding access time to other markets.

Response Rate: 66%

Survey Conclusions and Next Steps

Conclusions

- ❑ Over 80% of respondents are interested in FIF collecting market data latency statistics on both SIP and Proprietary feeds
- ❑ Over 70% of respondents want FIF to collect latency statistics during messages per second record peaks and throughout the trading day
- ❑ 90% of respondents are interested in obtaining Maximum latency statistics, followed by 79% for Mean and over 65% for Standard Deviation, Median and Minimum
- ❑ 69% of respondents would like to see 99.9% percentile statistics and over 35% indicated an interest in 90%, 95%, 99%, 99.5% percentiles
- ❑ Where possible, FIF should capture latency statistics in microseconds
- ❑ Less than a quarter of respondents are considering order or message types in various latency measurements
- ❑ Sharing exchange latency data not favored amongst 87% of broker dealer and vendor respondents

Next Steps

- ❑ Follow up with exchanges and SIPs to determine their ability to provide FIF data
- ❑ FIF to work with exchanges to establish latency statistic collection methodology

Appendix 1: Additional Comments

1. Which latency would you like to measure (see diagram):
 - ❑ Order to Exchange Trading System (See A on Figure 1)
 - ❑ Exchange Trading System to Prop Feed Latency (See B on Figure 1)
 - ❑ Exchange Trading System to SIP Feed (See C on Figure 1)
 - ❑ SIP Quote to SIP message (See D on Figure 1)

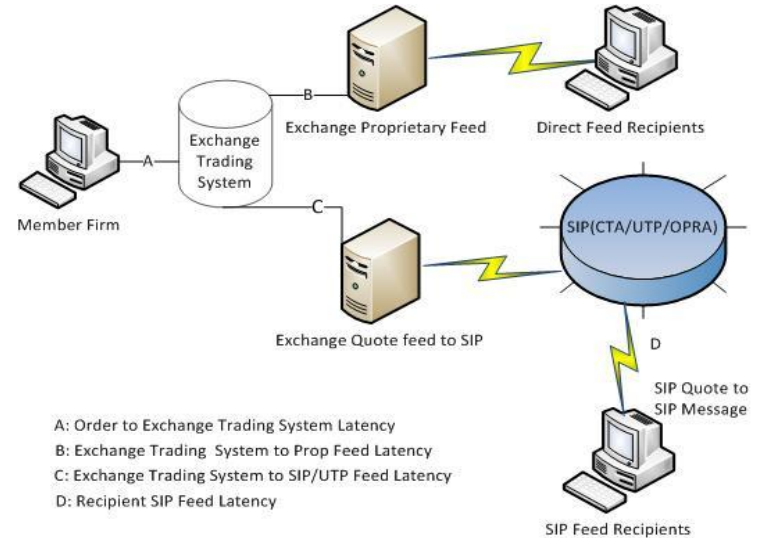


Figure 1 – Latency Network Diagram

Latency Measurement Comments

My priority would be: D, C, A, B

Order to ETS is critical. The current exchange timestamp on NASDAQ and BATS report the acknowledgement time (order is in the book), which might or might not be well after the receive time (Order to ETS). However this measurement can not be detected in software by an exchange application. To be useful the ETS time should not include exchange queuing time.

What I would really like to see, would be a combination of the last 3 - Exchange trading system order receipt, to SIP message departing the SIP data-center (or possibly, even a arrival time at a common reference point, such as Savvis NJ2 Or Equinox NY4, via best available telco (10 gig). All metrics should be documented in reference to how they were collected, and granularity of timestamps, load conditions, time of day, sample size.

All of these are very interesting metrics to have. Although the last hop is the most useful, the others are helpful as well.

Order to Exchange (A) needs to take into account latency between Member Firm outside network and Exchange Trading System outside network as one stat. Latency within a Member Firm's network should be considered a separate component measurable by the Member Firm. For latency sensitive firms, it would seem Order to Quote latency of proprietary feeds would be of most value as those would be the data feeds they would most rely on given SIP feeds will always require an extra hop through the SIPs and would always incur more overall latency.

Appendix 1: Additional Comments

2. What time of the day would be the best for latency measurement?
- Publication of Order Imbalance
 - Market Open
 - Market Close
 - Latency during Messages per second (MPS) record peaks
 - Peak latency during regular trading hours
 - Important economic announcements 10:00 AM and 2:15 PM
 - Specify time window (e.g. five or ten minutes before/after market close)

Additional Comments

Every 10 minutes

An average system latency (normal operating) would also be useful, so as to determine deviations during market open, market close and at the record peaks.

An overall number for the day.

BATS provides real-time latency statistics that allow Members to monitor latency for all of the order entry ports at any point in time throughout the entire trading day.

3. Which latency metrics are of interest (select all that apply)
- Max
 - Min
 - Mean
 - Median
 - Standard deviation
 - Other (Please specify)

Additional Comments

Mode is also of interest

Max and Min can be single outliers which are not indicative.

It would be nice to have access to the raw data.

With respect to real time latency, to generate a meaningful min it is best to pick a window (e.g. 1 minute) and average the mins of all windows. A single min data point for an entire day is less meaningful.

Appendix 1: Additional Comments

4. Percentiles

- 90%
- 95%
- 99%
- 99.5%
- 99.9%
- Other (Please specify)

Additional Comments

By breaking orders up into buckets (e.g. 10 microseconds) and collecting a count of order per bucket, you get a histogram that can easily be used by any firm to evaluate for any percentile that would be of interest to them.

5. Time stamp granularity

- Millisecond
- Microsecond

Additional Comments

Though microsecond is preferable, whatever granularity is available or appropriate for the system being measured.

6. Do you measure latency on different order types? If yes, what have you observed?

Different Order Types

We have not specifically concentrated on different order types. Something we need to do.

Cancels generally have higher latency than order accepts (accepts can be faked). Out on IOC is meaningful as well (accept + implied cancel). Order Ack, Cancel Ack, and IOC out are all useful.

It would be nice to have access to the raw data.

Since all orders follow the same path, there should be little difference in latency for different order types.

Appendix 1: Additional Comments

7. Do you measure latency on different message types (e.g. cancel/replace vs. new order)? If yes, what have you observed?

Additional Comments

All inbound traffic that follows the same path should not show much difference in latency.

Generally, orders that matches in our matchinhg system have a higer latency than cancels.

We measure Order to Ack, Cancel to Out and Order to Quote.

8. Would you be willing to share latency statistics on latency measurements of exchange response times (anonymously) with the FIF Market Data Capacity WG?

Additional Comments

Not at this time. As a private firm, we rarely provide proprietary information, anonymously or not. However, our executive team would consider releasing this information if our competitors do.

Although we collect this information, it is in a rough format and would take some effort to turn into a proper report. So we are not ready to share at this time.

Only for market data, not order entry.

We have a excellent metric of exchange response times, for 50+ market centers.

It's more a case of what format would we provide this in? Is it possible for the raw data that is used by FIF be made available? Perhaps even in real-time :)

We are open to sharing our markets latency stats, but at this time are not planning to redistribute specifics regarding access time to other markets.