HSSEDI
Homeland Security Systems Engineering & Development Institute

Prepared for:
Department of Homeland Security

Dynamic Data Map Technical Report

May 8, 2018

Authors:

Jason Veneman
Brian Tivnan

The Homeland Security Systems Engineering and Development Institute (HSSEDI)™
Operated by The MITRE Corporation

Approved for Public Release; Distribution Unlimited.
Case Number 18-1675 / DHS reference number 16-J-00184-08

HSSEDI
Homeland Security Systems Engineering & Development Institute

Homeland Security Systems Engineering & Development Institute

The Homeland Security Act of 2002 (Section 305 of PL 107-296, as codified in 6 U.S.C. 185), herein referred to as the “Act,” authorizes the Secretary of the Department of Homeland Security (DHS), acting through the Under Secretary for Science and Technology, to establish one or more federally funded research and development centers (FFRDCs) to provide independent analysis of homeland security issues. MITRE Corp. operates the Homeland Security Systems Engineering and Development Institute (HSSEDI) as an FFRDC for DHS under contract HSHQDC-14-D-00006.

The HSSEDI FFRDC provides the government with the necessary systems engineering and development expertise to conduct complex acquisition planning and development; concept exploration, experimentation and evaluation; information technology, communications and cyber security processes, standards, methodologies and protocols; systems architecture and integration; quality and performance review, best practices and performance measures and metrics; and, independent test and evaluation activities. The HSSEDI FFRDC also works with and supports other federal, state, local, tribal, public and private sector organizations that make up the homeland security enterprise. The HSSEDI FFRDC’s research is undertaken by mutual consent with DHS and is organized as a set of discrete tasks. This report presents the results of research and analysis conducted under:

HSHQDC-16-J-00184
Next Generation Cyber Infrastructure (NGCI) Apex Cyber Risk Metrics and Threat Model Assessment

This HSSEDI task order is to enable DHS Science and Technology Directorate (S&T) to facilitate improvement of cybersecurity within the Financial Services Sector (FSS). To support NGCI Apex use cases and provide a common frame of reference for community interaction to supplement institution-specific threat models, HSSEDI developed an integrated suite of threat models identifying attacker methods from the level of a single FSS institution up to FSS systems-of-systems, and a corresponding cyber wargaming framework linking technical and business views. HSSEDI assessed risk metrics and risk assessment frameworks, provided recommendations toward development of scalable cybersecurity risk metrics to meet the needs of the NGCI Apex program, and developed representations depicting the interdependencies and data flows within the FSS.

The results presented in this report do not necessarily reflect official DHS opinion or policy.

For more information about this publication contact:

Homeland Security Systems Engineering & Development Institute

The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102

Email: HSSEDI_info@mitre.org

http://www.mitre.org/HSSEDI

Abstract

The Homeland Security Systems Engineering and Development Institute (HSSEDI) assists the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) in the execution of the Next Generation Cyber Infrastructure (NGCI) Apex program. HSSEDI developed a comprehensive data map of an essential subsector of the Financial Services Sector (FSS), namely the capital markets. This data map provides a foundational component for an extensive NGCI testing program.

HSSEDI concludes this report with a set of three recommendations for the NGCI Apex program to enhance its representational testing environment:

- HSSEDI recommends that the NGCI Apex program expand this dynamic data map into an exhaustive depiction of workloads and time criticality for a small set of known market events when the market infrastructure experienced particularly heavy workloads and delays.
- HSSEDI recommends that the NGCI Apex program use these known market events and HSSEDI's Threat Model to inform detailed test scenarios for use in the representational testing environment.
- HSSEDI recommends that the NGCI Apex program integrate this dynamic data map with HSSEDI's previous technical reports on Cybersecurity Risk Metrics Survey and the Financial Systems Mapping to provide a comprehensive treatment of the systemic risk facing the FSS.

Key Words

1. Next Generation Cyber Infrastructure (NGCI) Apex program
2. Critical Infrastructures
3. Financial Services Sector (FSS)
4. Capital Markets
5. National Market System (NMS)

This page intentionally left blank

Homeland Security Systems Engineering & Development Institute

Table of Contents

List of Figures

List of Tables

1 Project Overview

The Next Generation Cyber Infrastructure (NGCI) Apex Program seeks to accelerate the adoption of cyber technologies proven to be effective for mitigating information technology (IT) security risk. Initially, the focal, critical infrastructure for the NGCI Program is the Financial Services Sector (FSS). The FSS is one of the most interdependent of the critical infrastructures, comprised of intensely competing organizations which collectively hold the nation's economic security in their decision-making related to technology implementation. The goals of the NGCI program are to 1) increase financial sector-wide situational understanding of evolving IT security risk and the technology associated with mitigating that risk; 2) improve the ability to understand and link compromises in the underlying cyber infrastructure to sub-sector operations; 3) enable greater information flows between sub-sectors as well as across the entire sector; and 4) enable FSS institutions to detect and neutralize adversaries more quickly and effectively than is currently possible. To achieve these goals, the NGCI program requires an extensive testing program beyond testing at the level of individual institutions.

Therefore, the NGCI Apex Program Management Office tasked the Homeland Security Systems Engineering and Development Institute (HSSEDI) to perform "workload modeling which describes the data dynamics within and between systems to provide a basis for workloads in the representational testing environment." As such, HSSEDI developed a comprehensive data map of the capital markets subsector of the FSS, thereby providing a foundational component for an extensive testing program in support of the NGCI Apex program. Here, the term *map* conveys a guide to the mechanisms of the generation and flow of market activity data from one financial institution to another across an entire subsector of the FSS. The specific objective of this report is to identify and depict the scale and time criticality of essential business functions in a central subsector of the FSS. To achieve this objective, HSSEDI performed extensive analyses on a dataset which is both authoritative and exhaustive. Because the capital markets subsector is often identified as one of the most technologically advanced within the FSS, this report identifies workloads that could appropriately serve as surrogates or upper bounds in the representational testing environment.

In this way, this report complements the objectives of previous HSSEDI products for the NGCI program, namely the *Cyber Risk Metrics Survey and Assessment, and Implementation Plan*[1], the *Enhanced Cyber Threat Model for Financial Services Sector (FSS) Institutions*[2], and the *Financial System Mapping*[3]. The objective of the cyber risk metrics survey and assessment task is to identify risk metrics and assessment frameworks that could be candidates to measure the systemic impact of the NGCI Apex program on the FSS. The objective of the threat models survey and assessment task is to identify threat models and frameworks that could be candidates to inform systemic testing in the NGCI Apex program. Finally, the objective of the financial system mapping task is to identify and depict the intrinsically interdependent nature of the subsectors which comprise the Financial Services Sector.

[1] HSSEDI, "Cyber Risk Metrics Survey and Assessment," The MITRE Corporation, McLean, VA, October 2017.

[2] HSSEDI, " Enhanced Cyber Threat Model for Financial Services Sector (FSS) Institutions," The MITRE Corporation, McLean, VA, March 2018.

[3] HSSEDI, "Financial System Mapping (Final)," The MITRE Corporation, McLean, VA, March 2018.

1.1 Task Overview for the Dynamic Data Map

The purpose of the task is to develop a comprehensive data map of a subsector of the FSS, thereby providing a foundational component for an extensive testing program in support of the NGCI Apex program. This technical report describes and analyzes the data dynamics within and between financial systems to provide a basis for workloads in the representational testing environment.

To accomplish this task and advance the objectives of the NGCI Apex program, HSSEDI:

- Performed workload modeling to describe the data dynamics within and between systems and institutions in order to provide a basis for workloads in the representational testing environment. Examples of this are: distributions (e.g., certain business processes, such as front-end commercial banking, experience diurnal and weekly variance in traffic) and what-when-where (e.g., data in other processes have regulatory constraints; for instance, exchanges must immediately route marketable orders to other exchanges to ensure best prices).
- Developed a scalable approach to depict workloads in the representational testing environment from institution-specific systems and data flows (i.e., micro-prudential) to sub-sector and sector (i.e., macro-prudential) to inter-sector (e.g., energy and telecommunications).
- Avoided information which might be institution-specific and therefore sensitive by using authoritative and comprehensive data available to the NGCI Apex program.
- Developed and delivered this technical report which describes the development and implementation of a Dynamic Data Map, to include quantitative metrics, to inform subsequent testing.

In the following sections, HSSEDI provides a general overview of the capital markets subsector and extensive analyses of this subsector. We conclude this report with a discussion of recommendations for next steps for the NGCI Apex program.

2 Overview of the Capital Markets and Related Infrastructure

In this report, HSSEDI describes and analyzes the data dynamics within and between financial systems, thereby providing a basis for workloads in the representational testing environment for the NGCI Apex program. Initially, the focal subsector of the FSS is the capital markets, and infrastructure of the capital markets which carries stock trading messages. This subsector of the FSS was chosen because of its high bandwidth, time-sensitive applications, and because it is empirically analyzable due to the existence of rich data. Unlike other subsectors, an empirical analysis of the capital markets does not require the disclosure of supervisory data by regulatory bodies, nor the disclosure of sensitive, proprietary data from a financial institution. Instead, authoritative sources make exhaustive datasets commercially available. Comprising all transaction activity in the capital markets, the regulatory bodies use these same datasets. In addition, this subsector attracts wide interest from the public. The high throughput necessary to handle stock trading ensures that systems designed to test this workload will be adaptable to other sectors of the financial system.

2.1 Overview of the National Market System (NMS)

Regulations governing the capital markets define the National Market System (NMS), colloquially known as the "stock market," as all market centers where investors can buy and sell shares of publicly traded companies. To facilitate the efficient exchange of capital and shares in the NMS, each market center is required to publish both the highest bid (i.e., the price at which an investor is willing to pay for a single share of a given stock) as well as how many shares the investor is willing to purchase at that price. In addition, each market center is also required to publish the lowest offer (i.e., the price at which an investor is willing to sell a single share of a given stock) as well as how many shares the investor is willing to sell at that price.

Across the entirety of the NMS, the highest bid and the lowest offer comprise what is known as the National Best Bid and Offer (NBBO) and the difference between the highest bid and lowest offer is known as the spread. The NBBO reflects a distillation of the order flow across all the stock exchanges comprising the NMS. Figure 1 (taken from a video previously developed by HSSEDI[4]) provides a geographical depiction of the three major datacenters of the NMS, all of which happen to be located in northern New Jersey. The three major datacenters are: (1) Mahwah, which contains the three exchanges comprising the New York Stock Exchange (NYSE) family of exchanges - NYSE Arca, NYSE American (formerly NYSE MKT) and the NYSE itself; (2) Secaucus, which contains both the Chicago Stock Exchange (CHX) as well as the Better Automated Trading System (BATS) family of exchanges consisting of Direct Edge A (EDGA), Direct Edge X (EDGX), BATS Z (BZX) and BATS Y (BYX); and (3) Carteret, which contains the three exchanges comprising the National Association of Securities Dealers Automated Quotations (NASDAQ) family of exchanges – Philadelphia (NQ-Phil), Boston (NQ-Bost) and NASDAQ itself. The newest public exchange, Investors Exchange (IEX), is located in Weehawken New Jersey (near Secaucus) and is shown in Figure 2[5]. As depicted in Figure 2, the communications infrastructure connects the datacenters, both by dedicated, high-speed networks

[4] https://www.youtube.com/watch?v=1ltjnbBaFok

[5] B. F. Tivnan, D. R. Dewhurst, C. Van Oort, J. H. Ring, T. J. Gray, B. F. Tivnan, P. S. Dodds, M. T. K. Koehler, M. McMahon, D. Slater, J. Veneman, and C. M. Danforth. (2018). "Inefficiencies in the U.S. National Market System: Evidence from the Dow 30." In Preparation.

known as “Direct Feeds” depicted in red and by the Security Information Processor (SIP) in blue which consolidates market data to determine and disseminate the NBBO.

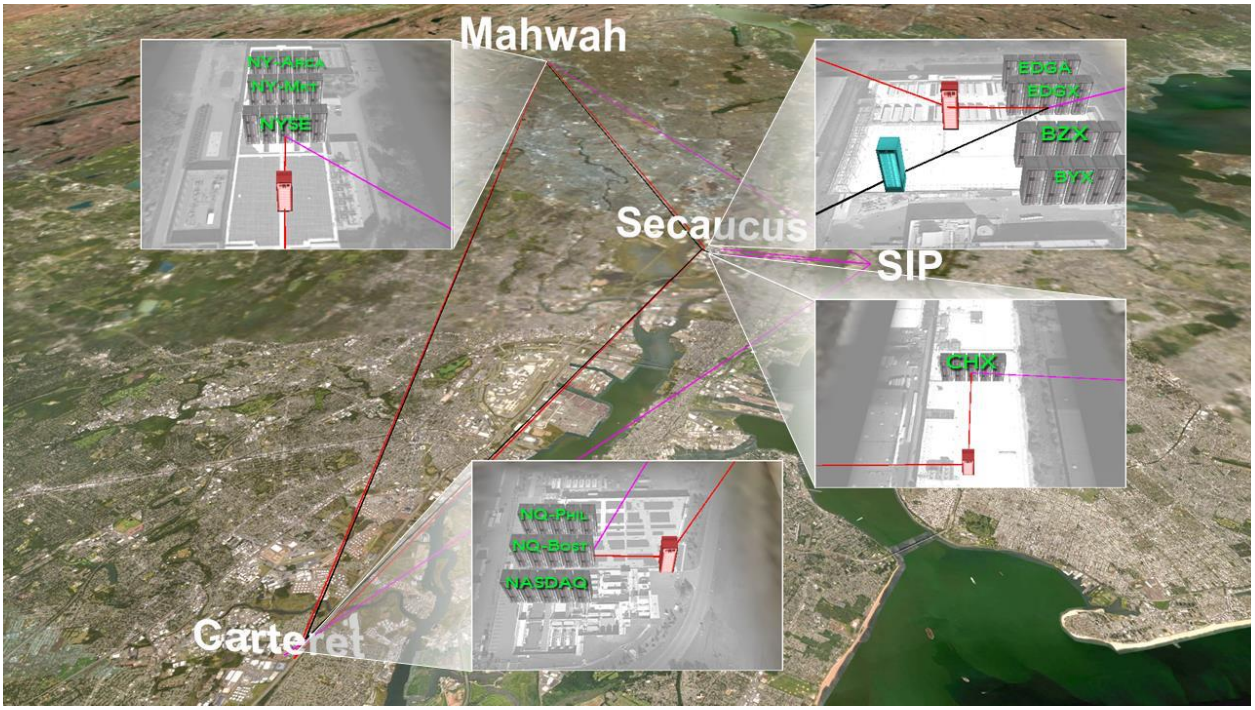

Figure 1. Geographic location of the primary exchange data centers in New Jersey

Financial records flow both between the exchanges and to external parties such as traders and regulators. These external parties are represented by the Observer and Dark Pools in Figure 2. Observers may be physically located in the same data centers as the exchanges, known as co-location, or elsewhere. Dark pools, also known as alternative trading systems (ATSs), are closed exchanges where private parties, such as institutional investors, trade securities. Although the internal information flow within an ATS is unknown, hence the name “dark pool,” they are required to report trades of publicly listed securities. Like observers, an ATS may be co-located with existing exchanges or reside elsewhere. The number[6] of registered ATSs is less than 100 in comparison to the 11 listed exchanges. The number of entities acting in these exchanges is approximately 4,000 institutional investors and 95 million retail investors[7] in the United States. The notional relationships between investors and exchanges is depicted in Figure 3.

[6] https://www.sec.gov/foia/docs/atslist.htm

[7] https://www.ici.org/pdf/2017_factbook.pdf

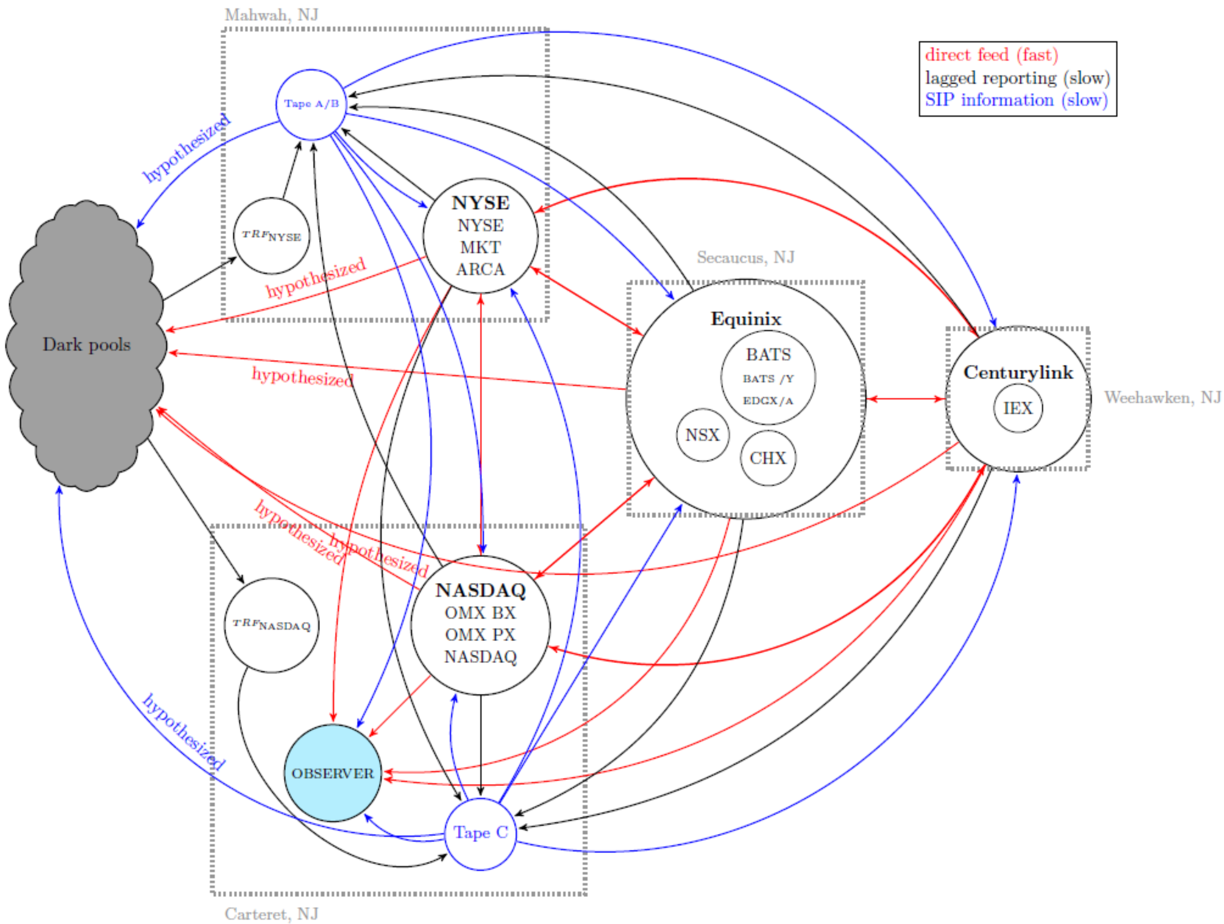

Figure 2. Interconnections between the exchanges

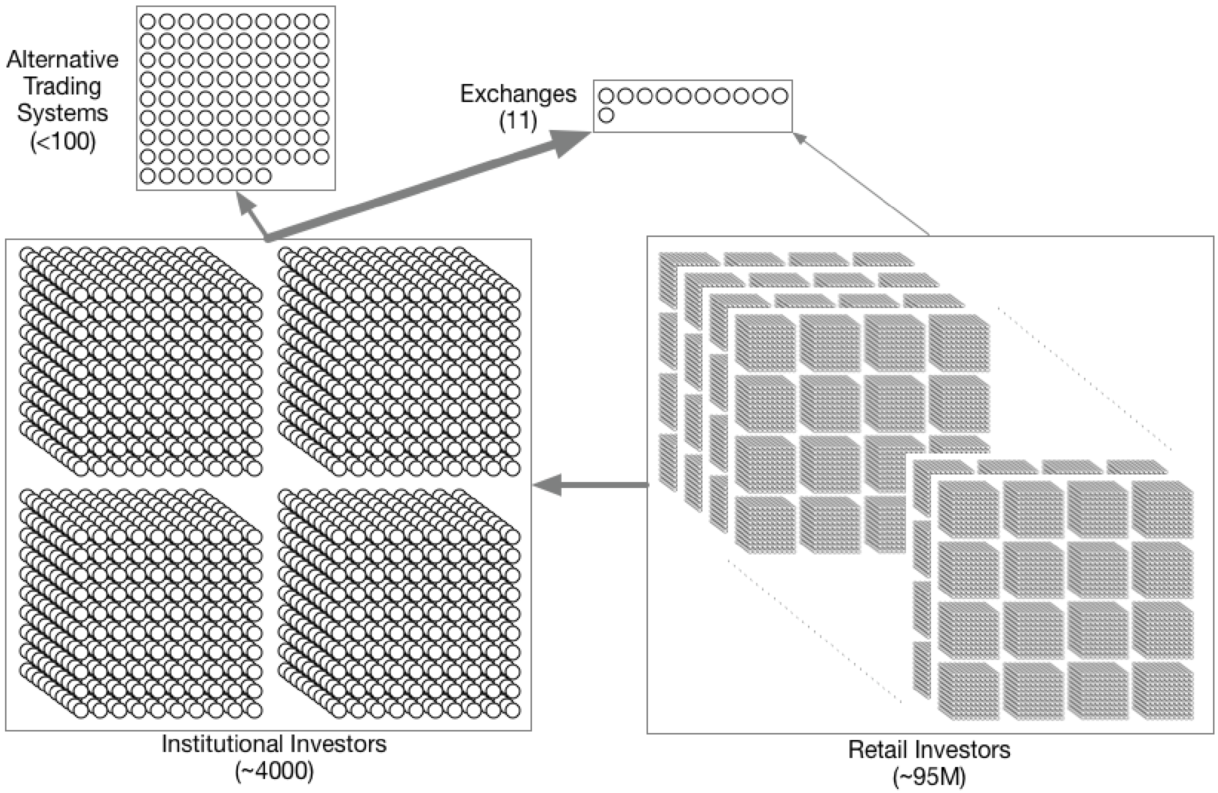

Figure 3. Relationships and numbers of investors and exchanges

2.2 Interdependence Across the Financial Services Sector

In a prior report, HSSEDI described the interdependence intrinsic to the FSS[8]. In Figure 4, HSSEDI provides a graphical depiction of this interdependence, by tying in Figure 1 with the "maps" from the aforementioned HSSEDI report. In Figure 4, Component C is a detailed depiction of a Bank/Dealer[9], a central element of the Financial Services Sector often serving as an intermediary in many financial services. The Trading Desk provides core functions of the Bank/Dealer. Most notably, the Trading Desk executes securities trades in the National Market System, depicted as Component B in Figure 4. Component A of Figure 4 is the multi-layered map of the financial system from the Department of the Treasury Office of Financial Research (OFR)[10] that HSSEDI recommends to use to depict the interdependent nature of the financial system. For simplicity, Component A contains two Bank/Dealers while the National Market System is merely depicted as a single exchange in Component A. In this way, Figure 4 not only depicts connectivity and information flows from individual financial instructions, to subsectors and sectors, but also illustrates the complementarity of this report with other HSSEDI technical

[8] HSSEDI, "Financial System Mapping (Final)," The MITRE Corporation, McLean, VA, 2018.

[9] Bookstaber, R., Paddrik, M., & Tivnan, B. (2017). "An agent-based model for financial vulnerability." Journal of Economic Interaction and Coordination, 1-34.

[10] Bookstaber, R., & Kenett, D. Y. (2016). "Looking deeper, seeing more: a multilayer map of the financial system." OFR Brief, 16(06), page 7.

reports, namely, the *Financial Systems Mapping* and the *Enterprise Threat Model Technical Report.*[11]

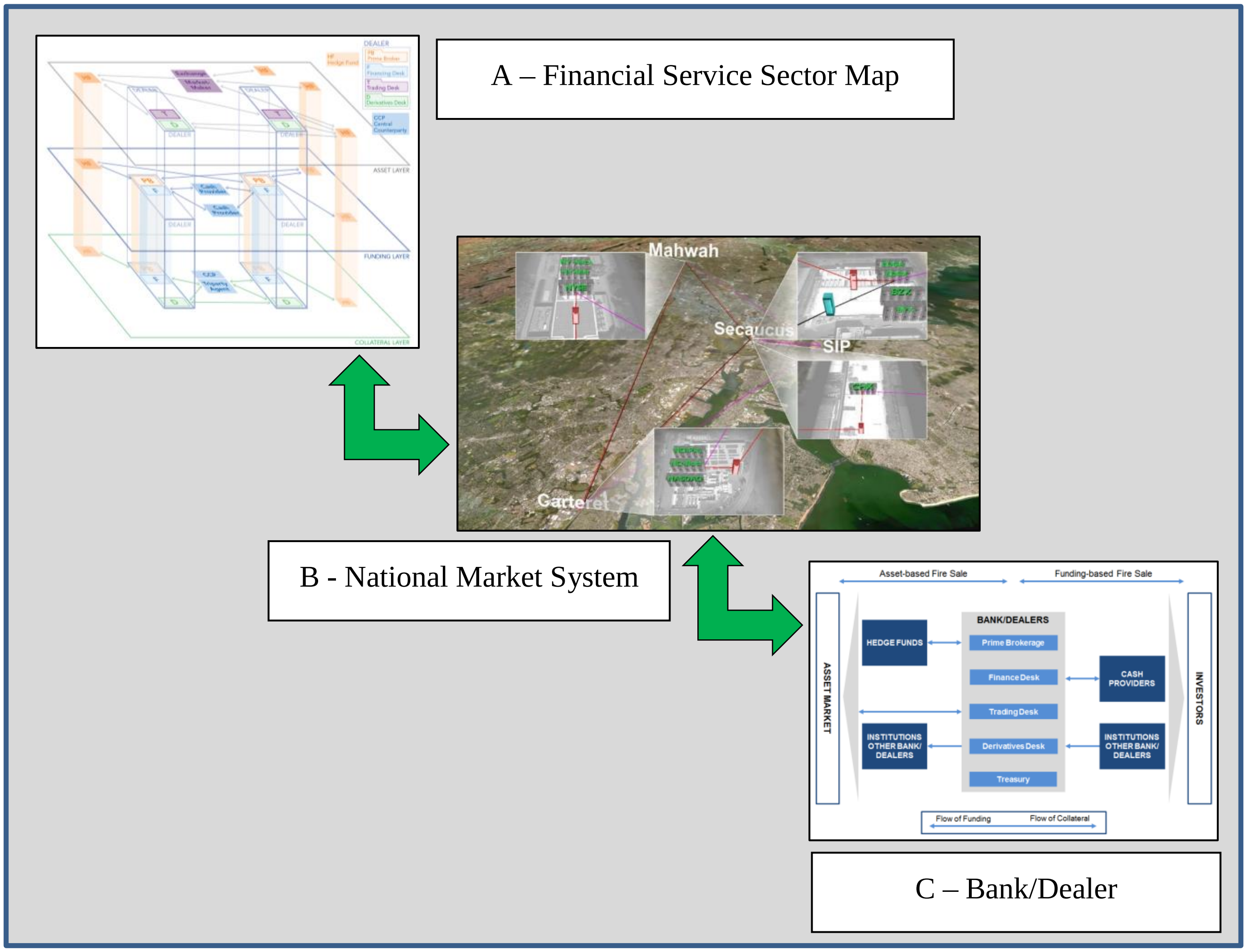

Figure 4. Interdependence across the Financial Service Sector

[11] HSSEDI, "Enterprise Threat Model Technical Report," The MITRE Corporation, McLean, VA, 2018.

3 Analysis

The structure of the data flows of the financial institutions in the NMS can be a difficult topic to understand even for those immersed in it day-to-day. HSSEDI conducted its empirical analyses of authoritative data from U.S. stock markets, which illustrate the architecture of the system. The analysis begins by breaking down the data associated with trading a single asset then moves to multi-market, multi-asset analyses to get a system-wide view. This approach provides a layered view of activity in this system giving details of important considerations and constraints along the way.

3.1 Single Asset Analysis

These examples will start with Apple stock (AAPL) which was the highest market capitalization asset traded on the day chosen for this analysis, August 11, 2015. This day is representative of a typical trading day with no major news events triggering increased activity in the market. Figure 5 shows the stock price of Apple over the course of a single day – including both pre-market (04:00 to 09:30 hours or 0 to 19,800 seconds in the figures below) and after-hours trading (16:00 to 20:00 hours or 43,200 to 57,600 seconds in the figures below). Figure 5 is typical of the price chart most people see in news reports on the stock market. With data at the microsecond level, even in this typical view one starts to see some anomalous behavior with the spikes at around 44,000 seconds. Those spikes are after-hours trades which, along with pre-market, are not usually shown when one looks at a stock on Yahoo Finance or similar sites. This begins to illustrate the complexity of considering data flows when looking at finance markets.

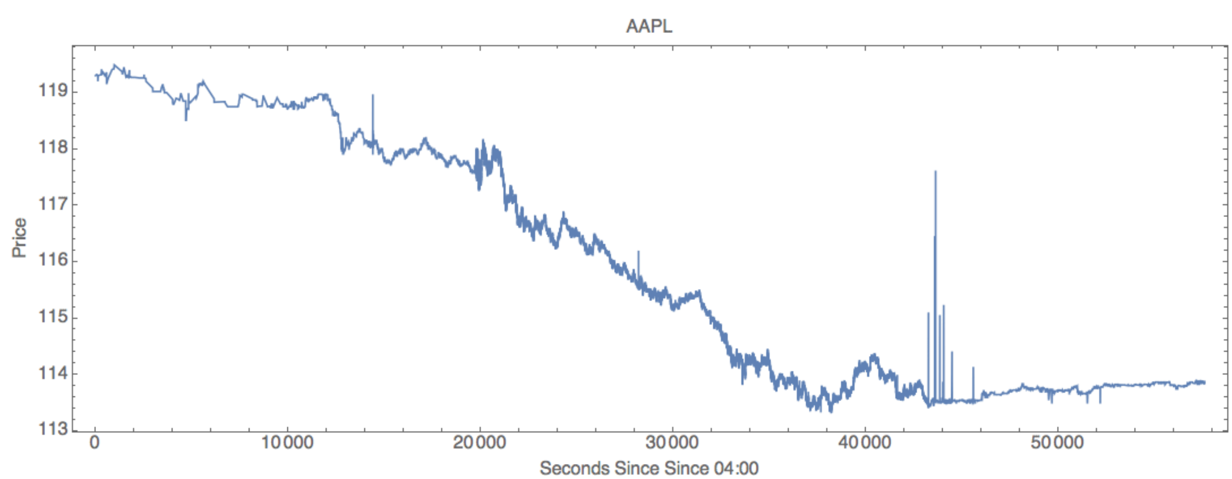

Figure 5. Apple stock price on August 11, 2015

Dynamics become more complex when one considers the number of trades per second for a single stock, as seen in Figure 6. Here, one will notice the clear difference in activity during pre-market, regular market (09:30 – 16:00), and after-hours trading (16:00 – 20:00).

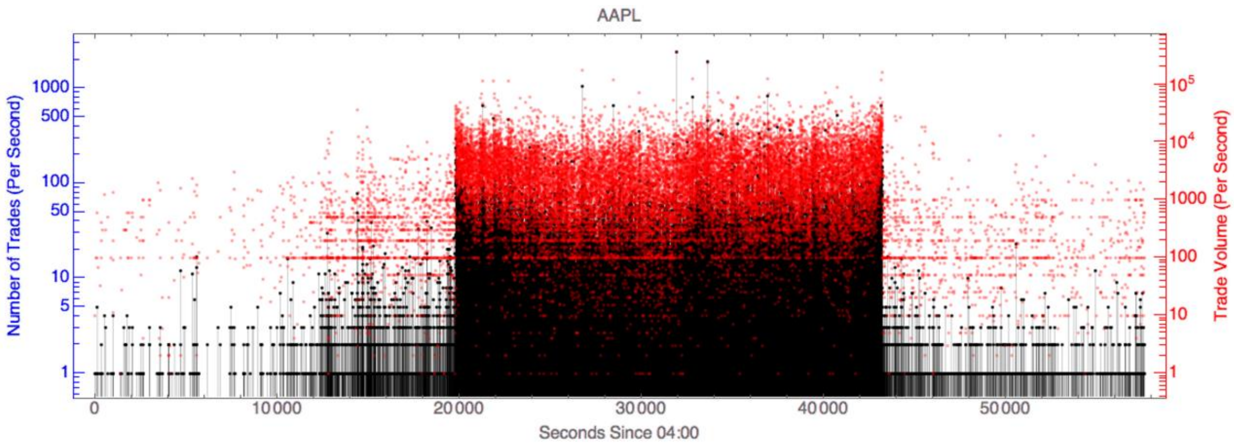

Figure 6. Apple stock trades per second on August 11, 2015

To get a sense of how much capital is flowing across the NMS, a look into the number of dollars traded per second in Figure 7 reveals some astounding numbers. Here one sees that at the highest spike $40 million in Apple stock is traded in one second.

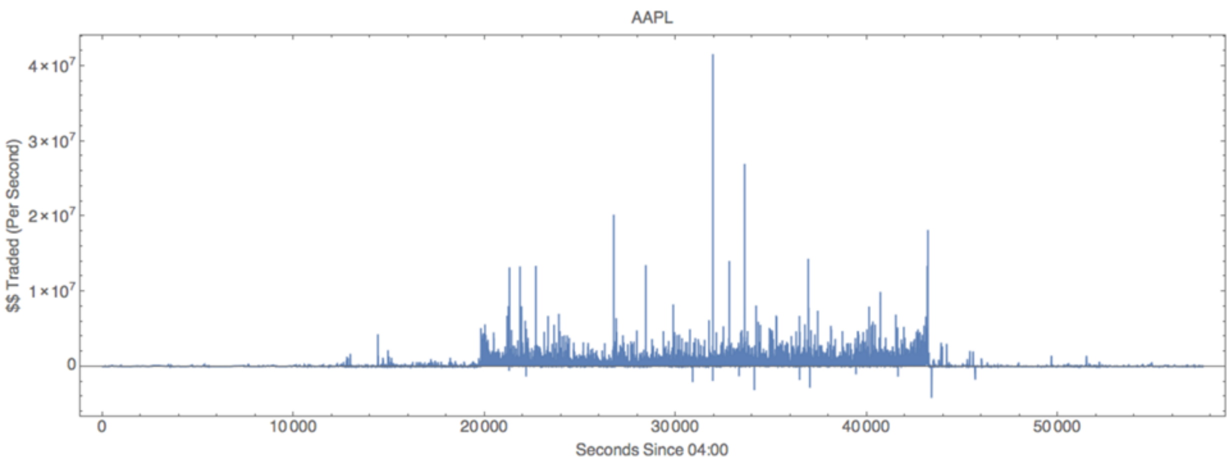

Figure 7. Apple stock dollars traded per second on August 11, 2015

To put things in perspective, the cumulative dollars traded, Figure 8, shows that there's a steady climb to around $10 billion traded in a single day for this asset. At such a high rate, the spikes seen above $40 million traded in one second hardly register.

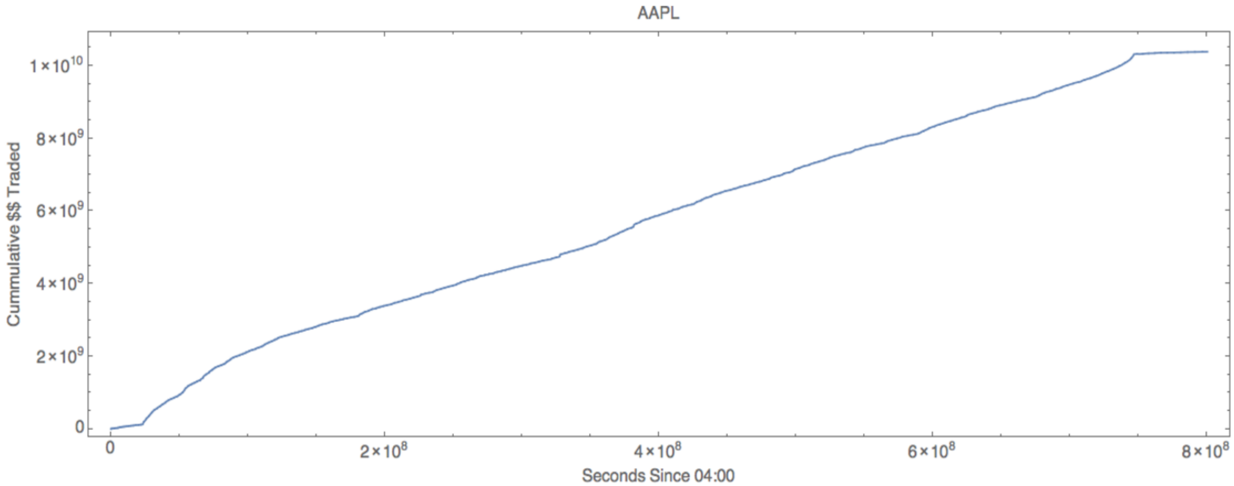

Figure 8. Apple stock cumulative dollars traded on August 11, 2015

Figure 9 shows how assets are traded at a high volume at the multiple exchanges that comprise the NMS. The opening and closing hours differ per exchange and are the reason that some lines start late or stop short. The start of regular trading at 09:30 (19,800 seconds since 04:00) is seen in Figure 9 as the sharp increase in volume at all markets at opening time.

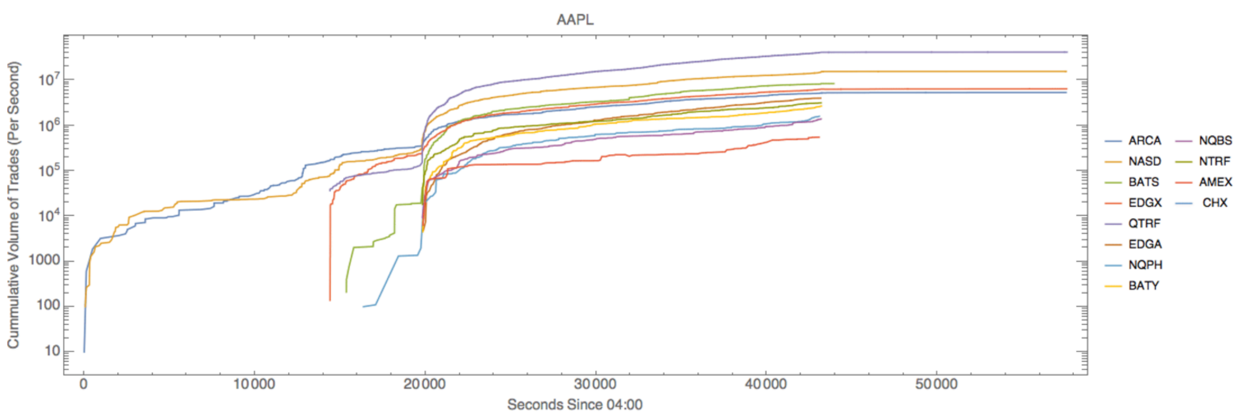

Figure 9. Cumulative volume by exchange for AAPL shares

3.2 Multi-asset Analysis

The analysis will now examine what is experienced in this sector by looking at the combined characteristics of all assets so get a broader picture of activity. Tying the dynamic fluctuations of market activity with communication networks and notions of bandwidth or capacity, it may be more intuitive to think of trade volume and quote lots (a quote lot is typically 100 shares) in terms of messages, where a message represents an atomic unit of communication that describes a number of shares or lots for a particular asset (e.g., 8 shares of Apple). Since there are far more quote messages than there are trade messages (roughly 10 to 1), the following figures of quote messages per day represent the high end of traffic on communication networks. In Figure 10, one can again clearly see the opening and closing of regular trading hours. Figure 10 shows messages that are recorded at each “Tape” or SIP of which there are three. The SIPs link “the U.S. markets by processing and consolidating all protected bid/ask quotes and trades from every trading venue into a single, easily consumed data feed."[12] SIP A and B are operated by the Consolidated Tape Association and contain stocks listed on NYSE on SIP A and stocks listed on NYSE Arca, NYSE MKT, BATS and regional exchanges on SIP B. SIP C is operated by NASDAQ for NASDAQ listed stocks.[13]

During regular trading hours there are very few times that less than 1,000 messages are seen per second while there are times when message traffic exceeds 100,000 messages per second. Similar to Figure 6, Figure 10 shows that the highest traffic occurs right after markets open or before they close.

Figure 10. Quote messages per second in seconds since midnight

Several analyses were performed to determine the data profiles in the NMS. Dates for the analysis were picked based on the number of trades on a date, notional dollar value of trades, and volatility. A date, June 24, 2016, with features on the high-end of these criteria was chosen for additional analysis. This date had the highest dollar value day of 2016 and the sixth highest in the last eight years. The high activity seen on this date was due to a combination of the Russell indices rebalancing[14] (a known periodic occurrence) with the results of Great Britain's vote to leave the EU.

The following tables, based on an analysis of daily summary data[15], provide visibility into the high activity days and exchanges. The top 10 high trade and notional value activity days are shown in Table 1 and Table 2. Table 3 shows the amount of trade activity by exchange since

---

[12] https://www.ctaplan.com/index

[13] http://www.utpplan.com/overview

[14] http://www.ftserussell.com/index-series/index-resources/russell-reconstitution

[15] http://markets.cboe.com/us/equities/market_statistics/historical_market_volume/

2010. Since this tally only includes trades and trades are about 1/20th of all messages then it is estimated that NASDAQ processed approximately 260B messages from 2010 to February of 2018.

The average day sees somewhat less than half the number of trades that are seen on the highest activity days with a mean of 30,403,170 and a standard deviation of 7,765,933 trades. The minimum number of trades seen since 2010 is 7,556,055.

Table 1. Top 10 trading days from January 2010 through February 2018

Date	Number of trades
2011-08-08	**74,671,169**
2015-08-24	**72,140,693**
2011-08-09	**70,402,718**
2011-08-05	**69,257,795**
2011-08-10	**67,424,274**
2018-02-06	**66,967,915**
2010-05-07	**66,174,403**
2010-05-06	**65,791,659**
2018-02-09	**63,584,408**
2016-01-20	**63,193,966**

Table 2. Total notional value of trades January 2010 through February 2018 by date

Date	Notional value of trades ($)
2018-02-06	6.998334e+11
2018-02-05	6.379111e+11
2015-08-24	6.358204e+11
2018-02-09	6.355986e+11
2011-08-08	5.741051e+11
2016-06-24	5.736876e+11
2011-08-09	5.601463e+11
2018-02-08	5.576612e+11
2011-08-05	5.439415e+11
2010-05-06	5.320244e+11

Table 3. Total trades from January 2010 through February 2018 at each exchange

Date	Market Participant
NASDAQ (Q)	13,116,478,027
NASDAQ (DQ)	12,330,177,635
NYSE Arca (P)	8,040,151,615
BATS BZX (Z)	5,972,705,249
NYSE (N)	5,424,407,590
EDGX (K)	3,655,105,299
BATS BYX (Y)	2,237,758,958
EDGA (J)	1,998,183,818
NASDAQ BX (B)	1,958,891,509
NYSE (DN)	1,608,910,746

At up to nearly $700 billion of trades passing between the exchanges on a high activity day we can see the importance of reliable and secure communications connecting them. The notional value of trades is just part of the story though. From an engineering perspective, the rates of messages passing between the exchanges is most important for the design of testing capabilities.

The messages that pass between exchanges are not just trade messages as tallied in Table 1. Far more numerous are the quote messages that detail the price and number of shares a party is willing to buy or sell. On the direct links between exchanges, quotes are referred to as *add* messages. Prices change very often throughout a day and a party who submitted an add message often wants to cancel their quote when the parameters of it no longer suit their means. Canceling quotes is done through *modify* messages. As seen in Figure 11, messages break down to roughly 49% add, 46% modify, and 4 % trade messages. The remainder of message types, at less than 1% of the total, will not be covered here.

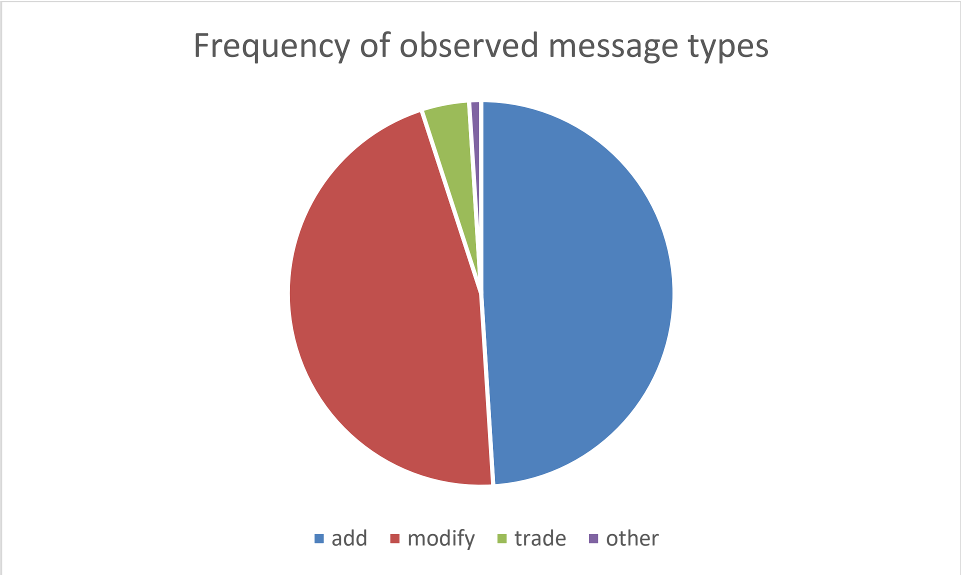

Figure 11. Frequency of observed message types

Rates of messages across all exchanges on a single day are shown in Figure 12. This figure shows the aggregated counts of all message types across all 8,000+ tickers on a single day. The messages are counted in single second intervals from 04:00 in the morning to 20:00 at night. The regular trading day (9:30am - 4:00pm ET) is clearly shown with the abrupt increase and decrease of activity. Message rates peak at 2,154,769/sec and average 117,919/sec during the trading day. In all, this day saw 2,858,836,964 messages on the direct data feeds.

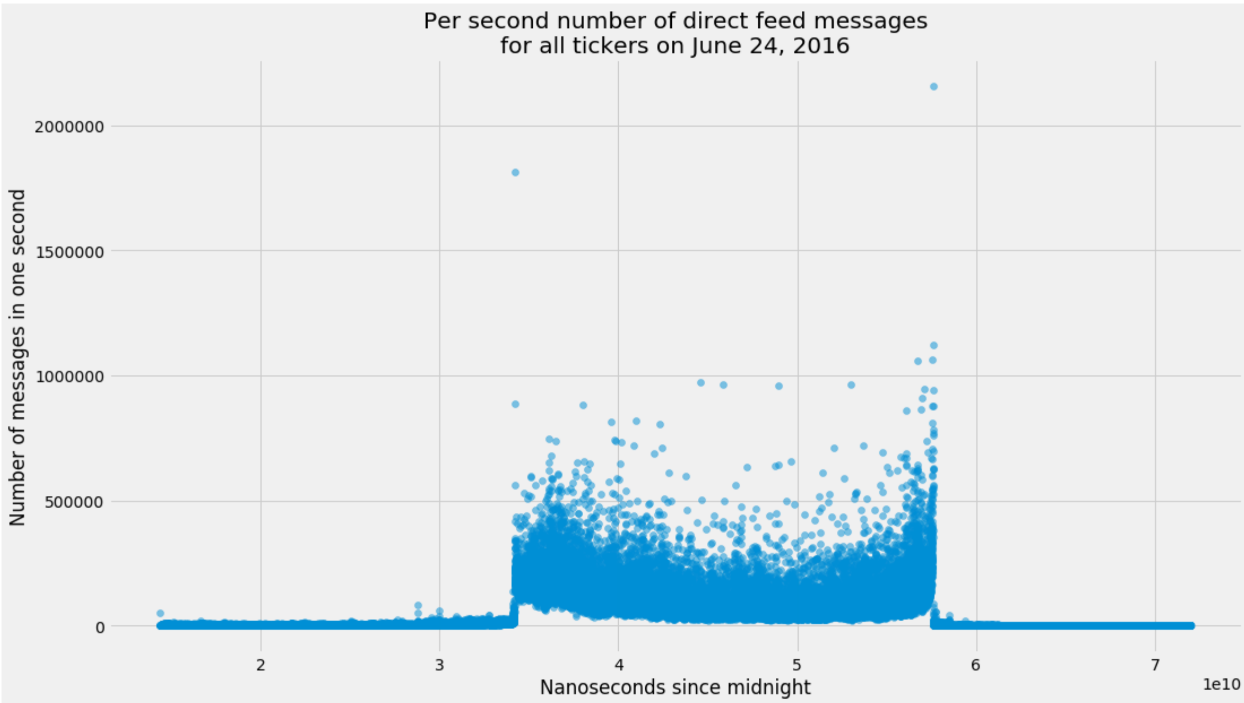

Figure 12. Rates of messages for all tickers in a single day

For comparison, the Visa payment network processes 150 million transactions per day and is architected to handle a maximum of 24,000 transactions per second[16]. The messaging app WhatsApp handles 55 billion messages per day[17]. While WhatsApp handles more messages, they are at the other end of the spectrum from the time critical nature of payment networks and asset trading.

3.3 Bandwidth

To convert message rates to bandwidth needs we can look at how the messages are formatted and transmitted. Exchanges use different market data feed protocols with the primary ones being Financial Information Exchange (FIX), ITCH[18], and PITCH[19]. These protocols share similar

[16] https://usa.visa.com/run-your-business/small-business-tools/retail.html

[17] https://blog.whatsapp.com/10000631/Connecting-One-Billion-Users-Every-Day

[18] Protocol developed by the Island Alternative Trading System

[19] https://www.batstrading.com/resources/membership/BATS_PITCH_Specification.pdf

characteristics in that for each message type they encode similar information. For add messages they encode the message type, unique message identification number, timestamp, order type, number of shares, stock identifier, and price. The importance of speed and quantity of messages pushes these to be small with an ITCH add order comprising 40 bytes, cancel modify orders at 23 bytes, and trades at 44 bytes. Combining the different order types the bandwidth calculation that follows will use 37.5 bytes per message for bandwidth calculations in congruence with the BATS exchange connection examples[20]. On the high activity day shown in Figure 12 an observer would need a connection with a minimum bandwidth of 646 Mbps (million bits per second) to intake the direct feed data at the maximum message rate with no delay or packet loss in transmission. However, the 646 Mbps figure does not account for intra-second bursts which can reach as high as 2,700 Mbps from a single exchange[10]. Due to the high message rates, exchanges recommend market participants connect to their data feeds at rates of 1 Gbps or higher. The average rate observed across all exchanges based on the data in Figure 12 is a factor of about 18 less than the maximum rate. This observed average rate scaling factor is in line with the stated average rate scaling factor by BATS of 19[21].

3.4 Infrastructure

The infrastructure that connects the exchanges is differentiated by bandwidth, latency, and price. Connections between the exchanges are offered in four main ways[10]:

- Virtual Private Network
  - Low bandwidth, high latency, low price
  - This runs through the public, internet backbone and is for applications which are not time-critical.
- Co-location
  - High bandwidth, low latency, high price
  - 1 Gbps and 10 Gbps connections for when the receiver is in the same facility as the exchange.
- Extranet
  - Medium bandwidth, medium latency, medium price
  - Access exchange data through a third-party connection.
- Private line Ethernet
  - Direct point-to-point connection via fiber-optic or microwave/millimeter wave antennas between exchanges or between an exchange and a participant
  - Fiber-optic
    - High bandwidth, low latency, high price

[20] http://cdn.batstrading.com/resources/membership/BATS_Connectivity_Manual.pdf

[21] http://cdn.batstrading.com/resources/features/bats_exchange_Latency.pdf

- Microwave/millimeter wave wireless[22]
  - Medium bandwidth, very low latency, very high price

These connections are not mutually exclusive for exchange data customers. A trading firm may be co-located in an exchange and have with both fiber and wireless connections to other exchanges. While the wireless connections are considered supplemental[23] to fiber-optic they command a premium price due to their speed. The speed difference for the wireless millimeter-wave connections can be significant with a Carteret to Secaucus, NJ latency of 153µs compared to 93µs on a fiber optic line[24] - 40% faster.

3.5 Dynamic Message Traffic

The image in Figure 13 and Figure 14 shows freeze frames taken from a message visualization application[25] developed to explore the dynamics of financial data. They show individual trades passing between the exchanges to an observer at Carteret, NJ, one of the main exchange locations. Each rectangle represents a trade with the width being proportional to the number of shares in that trade.

Figure 13 depicts two stocks, Bank of America (BAC) in blue and AAPL in green. Trades of BAC are recorded by the SIP in Mahwah, the location of NYSE which is the listing exchange for BAC; while AAPL trades are sent to the SIP in Carteret, the location of NASDAQ which is the listing exchange for AAPL. The trades shown in the image are from the first few hundred microseconds after the markets opened on February 5th, 2018, the day with the fourth highest volume in the last eight years. The straight paths represent messages from exchanges to the respective SIPs, while the circular paths represent messages that originated and terminated in the same location. Each trade is represented by a rectangle with its width proportional to the size of the trade (i.e., number of shares). The width of each path depicts the proportion of the number of trades currently passing through that path. Trades begin their journey at the exchange executing the trade and then move to their respective SIP location (i.e., listing exchange). The average time this takes is approximately 500µs. While this is just the activity of two tickers, it illustrates that all of the exchanges are producing trades and each link provides the necessary information for the immediate action of market participants. Any delay due to capacity constraints or service issues will therefore have an immediate impact on all subsequent activity.

[22] http://www.nasdaqtrader.com/content/Productsservices/trading/CoLo/ExpressConnectFS.pdf

[23] http://qnasdaqomx.com/WirelessConnectivity

[24] http://anova-tech.com/sample-page/map/

[25] In additional to this technical report, HSSEDI has also developed a visualization application to animate these dynamic flows. Contact the NGCI Apex Program Management Office or HSSEDI for access to this visualization application.

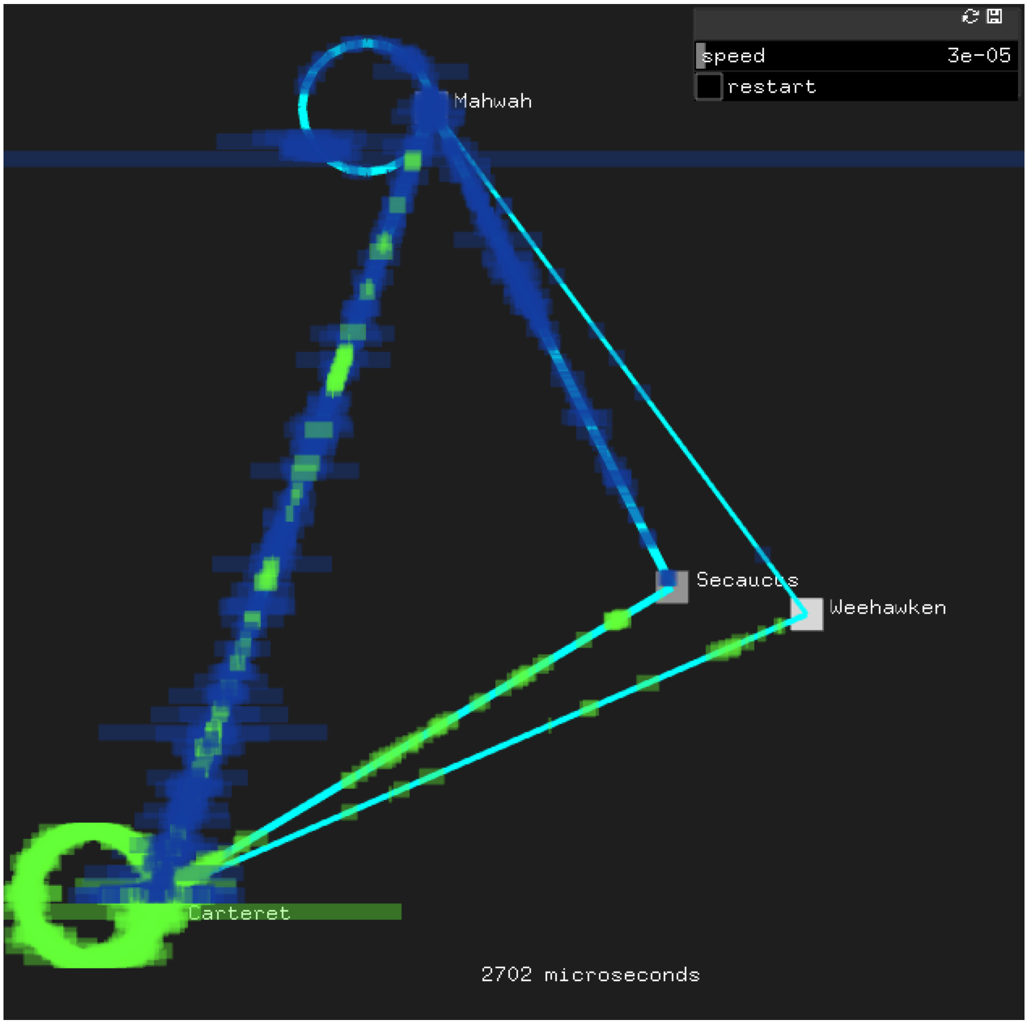

Figure 13. Visualization of messages passing between exchanges and to an observer

Figure 14 depicts some of the dynamics of this system, which has irregular bursts of market activity and significant differences in speed. This figure is composed of a series of sequential frames taken from the message visualization application. The sequence goes from left to right then top to bottom. In the first frame, top left, the market has just opened and trade messages for AAPL from the exchanges start being sent to the SIP in Carteret. As in the previous figure, the width of the rectangles represents the size of the trade. On the bottom of the second frame a large trade appears originating and terminating in Carteret and covering the width of the frame. Most of the initial trades occurring away from Carteret reach Carteret by the 13th frame – approximately 230µs after the market opening. Beyond the short sequence of events captured in Figure 14, a dynamic depiction of the message flows via the visualization application shows bursts of activity and message hopping where one can see some messages pass others at significantly greater speeds on the same path – a consideration for the development of the representational testing environment.

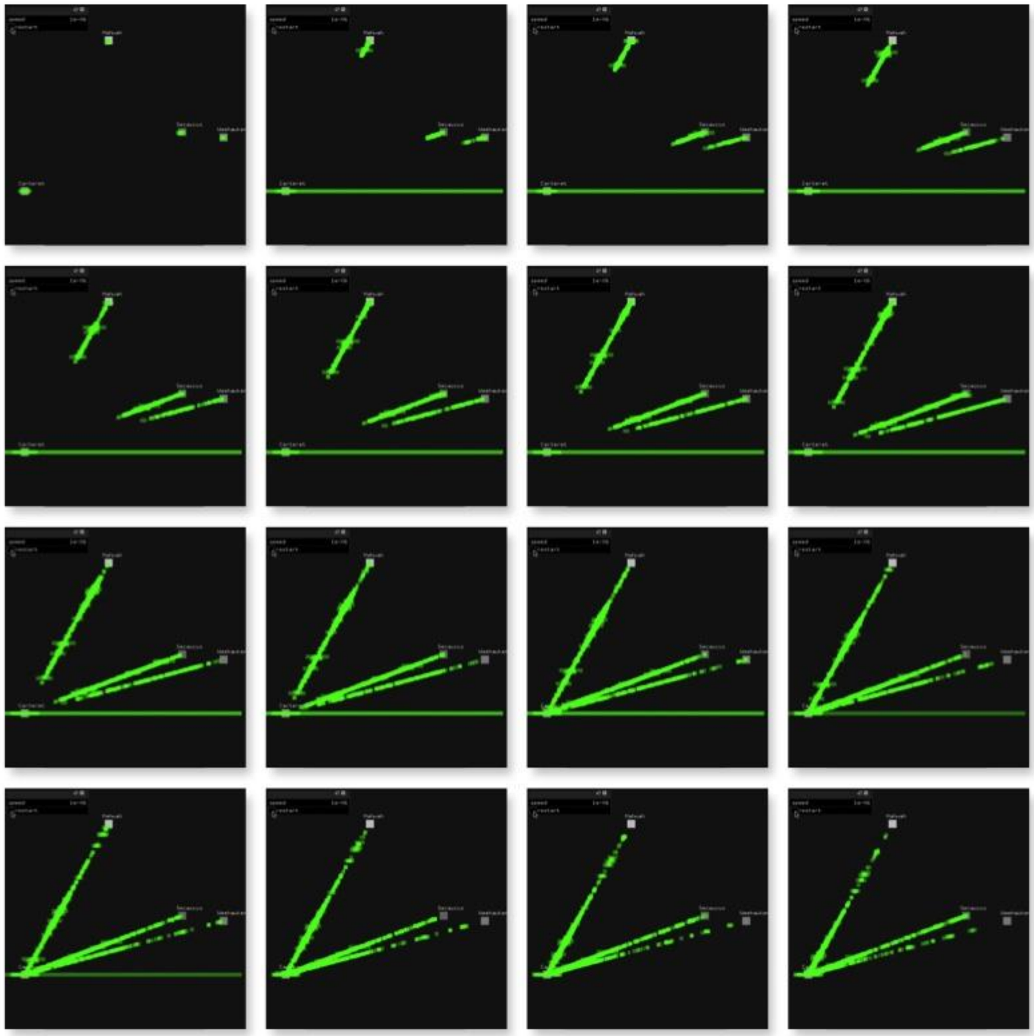

Figure 14. Sequential frames from a video of market dynamics in action

3.6 Data Source

The message level analyses in this report are derived from authoritative data from the same source from which the Securities and Exchange Commission (SEC) gets their Market Information Data Analytics System (MIDAS) data[26]. This data comprises both the direct from all the exchanges as well as SIP data feeds for quote and trade messages.

The intake of the data occurs in NASDAQ's co-location facility – its datacenter in Carteret, New Jersey. This co-location enables the data provider to add their own timestamp of when messages were received when they ingest the market feeds. Timestamps from the exchanges and the data provider are captured at the microsecond level. This additional timestamp provides further insight into market dynamics from the position of a market participant.

[26] https://www.sec.gov/marketstructure/midas.html

3.7 Dynamic Data Behavior - Impacts of High Traffic

Figure 15 shows the number of times that a stock is apparently locked (i.e., a spread of 0) at the SIP. This also gives a preliminary indication that increasing message traffic for a given asset leads to more apparent locks in that asset. This suggests capacity and/or computational constraints on processing incoming messages.

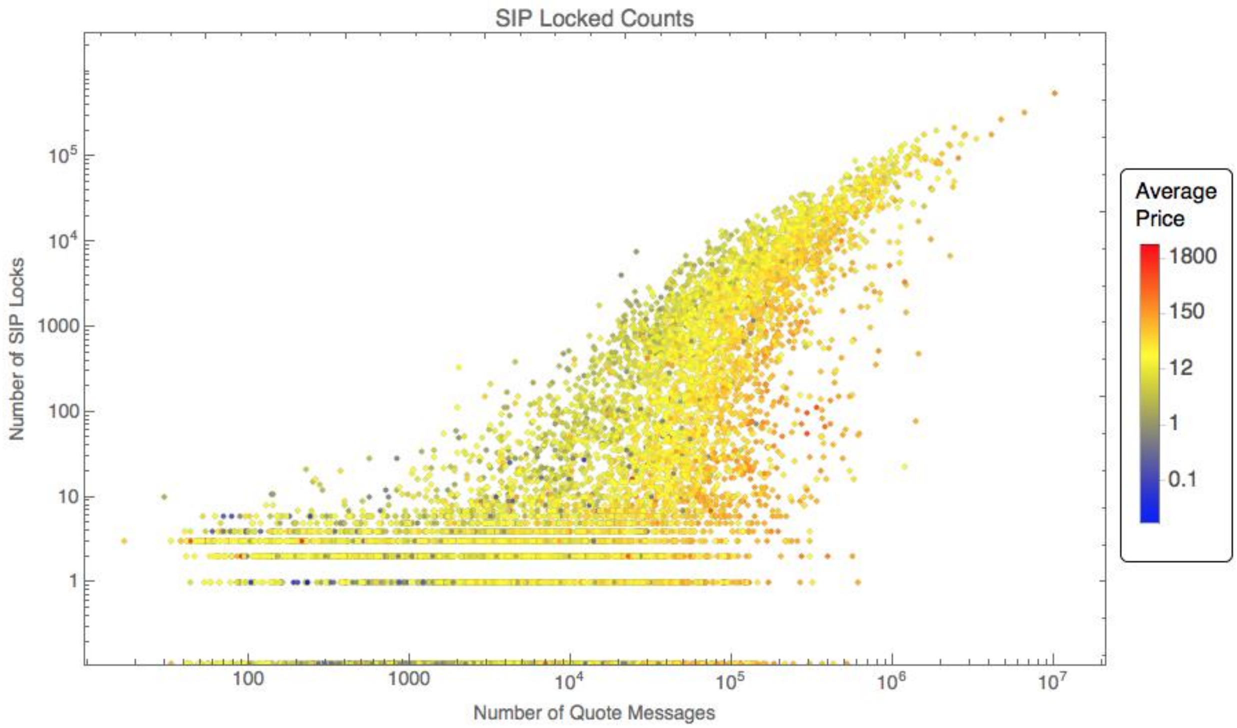

Figure 15. Number of SIP locks as a function of capacity

The SIP crosses in Figure 16 indicates that as the number of quote messages increases there is an increasing number of apparent market crossings as seen at the SIP. In this figure, the price of the stock is given in the color chart to the right. Stocks that trade for less than one dollar (i.e., penny stocks, blue to blue-green in the figure) operate under different market rules and appear to follow a different relationship between number of messages and crosses.

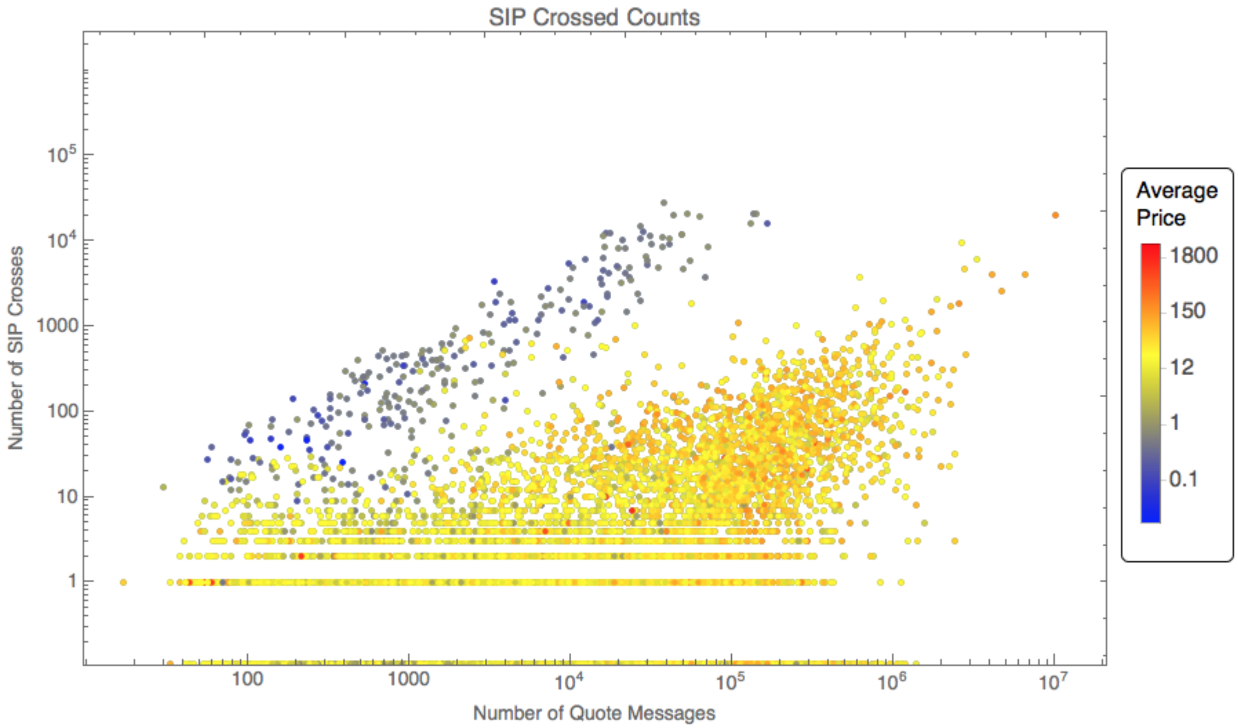

Figure 16. Number of SIP crosses as a function of capacity

4 Conclusions and Recommendations

As described in this technical report, HSSEDI developed a comprehensive, data map (i.e., a guide to the mechanisms of the generation and flow of market activity data from one financial institution to another across an entire subsector) of an essential subsector of the FSS, namely the capital markets. This data map provides a foundational component for an extensive testing program in support of the NGCI Apex program.

This technical report describes several analyses of the data dynamics within and between the financial systems and infrastructure comprising the NMS. The analysis provides empirical support to the nature of background traffic present in a highly active, highly valued, and highly engineered sector of the financial system. Possible uses of this are for testing of red/blue exercises focused on the financial sector. The findings presented here show that data flows range from zero to millions of messages per second. Single messages, expressing transactions of hundreds of thousands of dollars, may only be valid for such short times that capacity and computational resources need to continually operate with high reliability. The analysis shows, and gives the boundaries of, data activity that fluctuates across several orders of magnitude for many aspects of this system including day, time-of-day, asset, location, and message type. The data map, with its boundaries and structure, may inform real-time determinations of normal and abnormal performance as well as ensure future FSS resiliency. Lastly, these analyses and the ensuing data map may provide a basis for workloads in the representational testing environment for the NGCI Apex Program.

Finally, HSSEDI concludes this report with a set of three recommendations for the NGCI Apex program to enhance its representational testing environment.

- HSSEDI recommends that the NGCI Apex program expand this dynamic, data map into an exhaustive depiction of workloads and time criticality for a small set of known market events (e.g., “Flash Crash,” “Manic Monday” and February 5th and 8th, 2018) when the NMS infrastructure experienced particularly heavy workloads and delays.
- HSSEDI recommends that the NGCI Apex program use these market events and HSSEDI’s Threat Model to inform detailed test scenarios for use in the representational testing environment.
- HSSEDI recommends that the NGCI Apex program integrate this dynamic, data map with HSSEDI’s previous technical reports on Cybersecurity Risk Metrics Survey and the Financial Systems Mapping to provide a more comprehensive treatment of the systemic risk facing the FSS.

List of Acronyms

Acronym	Definition
AAPL	Apple Stock
ATS	Alternative Trading System
BAC	Bank of America Stock
BATS	Better Alternative Trading System
BYX	BATS Y Exchange
BZX	BATS Z Exchange
CHX	Chicago Stock Exchange
DHS	Department of Homeland Security
EDGA	Direct Edge A Exchange
EDGX	Direct Edge Exchange
FFRDC	Federally Funded Research and Development Center
FIX	Financial Information Exchange
FSS	Financial Services Sector
HSSEDI	Homeland Security Systems Engineering & Development Institute
IEX	The Investors Exchange
IT	Information Technology
MIDAS	Market Information Data Analytics System
NASDAQ	National Association of Securities Dealers Automated Quotations
NBBO	National Best Bid and Offer
NGCI	Next Generation Cyber Infrastructure
NMS	National Market System
NQ-Bost	NASDAQ Exchange -– Boston
NQ-Phil	NASDAQ Exchange – Philadelphia
NYSE	New York Stock Exchange
S&T	Science and Technology Directorate

Acronym	Definition
SIP	Security Information Processor