313 ULTRA-HIGHSPEED MARKET DATA TICKER SYSTEM

FINAL REPORT Spring, 2008

Team Members

Oluwaseun Shonubi, *Team leader* Khanh Duong Jose Acuna-Rohter Tarun Anupoju Lance Cooper Martin Kolodziej Konstantin Roytman Jing Kai Tan Jong Su Yoon

Faculty Advisors Wai Gen Yee Ben Van Vliet

1



1. Introduction

IPRO 313: Ultra-Highspeed Market Data Ticker System, Spring, 2008, builds on work done in the previous semester. Last Fall, IPRO 313 conducted background research on the market data systems and outlined a basic design and performance metrics for our system. This semester, we will resume system development, and work on performance improvements, enhanced functionality, and more detailed benchmarks.

2. Background

With the dawn of every New YearDaily, the speed of business is ever increasing. What used to be performed in months, days or minutes is now being done in milliseconds. To stay competitive within their industries, businesses today – particularly in finance – need real time access to large volumes of data. This requires vast improvement in information technology infrastructures.

Introduction of Electronic Trading

The 1975 Security Acts Amendment, enacted by Congress in 1975 gave the SEC the authority to create an efficient national market infrastructure. The reason for this was to eliminate as much of the redundant paper work that slowed the trading process down as possible. As a result, people had access to substantially larger volumes of data with significantly less latency.

The new trading infrastructure changed trading both quantitatively and qualitatively. Quantitatively, much more market data are available with lower latencies. A March, 2007 report by the Tabb group indicated that the number of messages transferred per day in global options and equities markets has been growing by 139% annually from 2005 and is projected to continue until at least 2010. Furthermore, volatility in the market has decreased as any sudden "spikes" in the market can be handled very quickly with the new market technologies.

According to TowerGroup, a research firm, \$480m is likely to be spent in <u>America-the United</u> <u>States</u> this year on developing technology for algorithmic trading. In 2005, providing <u>fastest</u> data to customers <u>guickly</u> had such an impact that the financial industry increased spending on computers and software to \$26.4 billion and in the past years, the compound annual growth rate for algorithm use from 2004 through 2007 was projected at 34%. Such is the focus on speed that even location counts. Servers positioned nearest to a trading venue can shave milliseconds of the timing of a trade and get a better price. <u>Moreover, Chicago, being near the geographic center of the United States, enjoys the advantage of being near both markets as well as both coasts.</u>

There exist several vendors of the technologies necessary to participate effectively in this industry, including Townsend Analytics, Exegy, Reuters, Bloomberg, RMDI and Wombat. These companies provide various services, such as real-time price data as well as market analyses. Their current generations of systems, however, are incapable of handling the impending mountains of data.

3. Purpose & Objectives

The objective of IPRO 313 is to create a high performance data ticker system. Ideally, the data ticker plant has to have a sustained optimal throughput of three million price quotes per second – the current industry state of the art - with minimal latency. Toward this end, the team will create a proof-of-concept data ticker plant that processes real data, <u>demonstrating performance</u> <u>bottlenecks and revealing areas should be the foci of optimization attempts</u>. The initial system will be used as a baseline on which optimizations to specific components will be made.

Based on the result from the past IPRO team, the team has set forth the following objectives:

- · Explore competitors' solutions and available technology
 - Know what is currently on the market
 - Know what new technology is available for implementation
 - Understand what works and what does not
- Develop a functioning ticker plant system
 - Analyze ticker protocols used in previous semester
 - Redesign and refine the ticker plant architecture
 - Code a working system
- · Improve system performance
 - Identify performance bottlenecks in the system
 - Design and implement performance optimizations
 - Record performance improvements
- Determine hardware requirements
 - Test off-the-shelf hardware for system
 - Design custom hardware configurations
 - Compare each solution
- Update the technical user manual
- Create a website explaining project initiatives and successes with an online version of the end-to-end system.

4. Methodology

The research results from the previous semester were usable; however, the system prototype was not functioningneither stable nor fully functional. The team agreed that a stable base system is was needed before optimization work could begin, thus necessitating the formation of a software development team was needed. In order for allow the development team to concentrate on developmenting the base system tasks, a research team was also formed to find investigate potential performance-enhancing solutions to be implemented in parallel. Finally, one of the problems identified from the previous semester was the lack of coordination between sub-teams; therefore a third team was created to oversee the whole project.

Thus, to accomplish the objectives set forth by the whole team, the IPRO 313 Spring 2008 team was split into three different sub-teams:

Development (programming) Team

This team is responsible for developing the system. Their responsibilities included both designing and <u>implementing building</u> the <u>core</u> system, <u>which implements basic</u> <u>functionality</u>. The development team <u>will incorporate used</u> research results from the Research/Optimization team <u>to makein making</u> design decisions, such as which algorithm, hashing function, and so on, should be used.

A. Tasks

For tThe development team's, the task this semester was goal is to improve on last semester's system prototype and to build an end-to-end market data ticker system that uses real market data. Another goal is The intent was to also provide detailed performance metrics of the system. This enables future optimizations and algorithmic enhancements to the system in order to increase system throughput while minimizing latency and variance.

In order tTo complete the task, the software development team was broken up into separate sub-teams to handle development of each module comprising the system.

B. Sub-Teamsteams

Konstantin Roytman and Jose Acuña-Rohter handled the development, testing, and metric collection of the data feed and head end components. Lance Cooper and Jong Su-Yoon worked together to develop the LVC and Client modules. The two teams collaborated to handle the integration of all the modules. (Note that these architectural components will be explained in Section 7.)

Research/Optimization team

This <u>sub-</u>team's responsibilities <u>included_include_looking_atanalyzing</u> the base system developed and researching solutions to improve and optimize the system for maximum reliability and performance. The research results generated <u>were_used_by_the</u> <u>development team's their</u>-implementation of the system.

Research on hash functions <u>was-is</u> conducted to find the fastest hashing algorithm. The purpose of finding the fastest hash algorithm is increasing the data access speed of the Last Value Cache, which is believed to be the bottleneck of the system. In addition, the hash function has to be stable and generate few collisions of hash values.

Formatted: Indent: First line: 0"

Analysis of real opra-OPRA data (OPRA, the Options Price Reporting Authority, is an industry-standard clearing house for real-time options prices) was is conducted which allowed us to see the loads and how they vary every second. Which showed This analysis reveals patterns in data arrival rates, which may be used to guide the design of the system for optimal performance.how much data may arrive to the given hash function and how the hash function should handle and perform with this data.

IPRO/ Web design team

This <u>sub</u>-team <u>was is</u> responsible for managing the project and handling the creation of IPRO deliverables and the IPRO 313 Website. The <u>whole team <u>must adhere to</u> followed the timeline set forth by the IPRO/Web design team and reports back results.</u>

The relationship between among the three sub-teams are is summarized in the diagram below:



Figure 1. Sub-teams relationship

Each team is responsible for researching and developing solutions within its specific area, but also informs and collaborates with the other teams of their findings. Though ultimately all progress <u>will_beis</u> reported to the IPRO/Web design team, each team <u>will_must_</u>draft its own reports, schedule<u>s</u> and presentations. Additionally, each team <u>will_must_</u>report to the group weekly₇ so that the information is presented in a timely and consistent manner.

Spring 2008

5. Assignments

Team Leader:

Oluwaseun Shonubi

Minute Taker:

Jose Acuna-Rohter

In charge of recording decisions made during meetings including task assignments or changes under consideration.

Agenda Maker:

Tarun Anupoju

Responsible for creating an agenda for each team meeting. This provides structure to the meetings and offers a productive environment.

Time Keeper:

Tarun Anupoju

Responsible for making sure meetings go according to agendas.

Master Schedule Maker:

Khanh Duong

Responsible for collecting schedules from all the team members and developing a master schedule, which tells the team when the members are available and how to contact them.

Team advisors

Dr. Wai Gen Yee

Dr. Yee is currently an assistant professor at the Illinois Institute of Technology's Computer Science Department. His research interests are in distributed databases, information retrieval, data mining, security, and peer to peer and ad hoc networking. He is a member of the Department and University Undergraduate Curriculum committee, which is responsible for ensuring the quality of education for undergraduate computer science majors.

Ben Van Vliet

Ben Van Vliet is a Lecturer at the <u>Illinois Institute of Technology's Stuart School of</u> <u>Business (IIT)</u><u>Illinois Institute of Technology's Stuart School of Business (IIT)</u>, where he also serves as the Associate Director of the M.S. Financial Markets program. At IIT he teaches courses in quantitative finance, C++ and .NET programming, and automated trading system design and development. Beginning in 2008 he will also teach recurring seminars for The Technical Analyst group in London. He is vice chairman of the <u>Institute for Market Technology</u><u>Institute for Market Technology</u>, where he chairs the advisory board for the <u>Certified Trading System Developer (CTSD)</u><u>Certified Trading</u> <u>System Developer (CTSD)</u> program. He serves also as series editor of the Financial Markets Technology series for Elsevier/Academic Press. Mr. Van Vliet consults extensively in the financial markets industry, primarily on topics related to the mathematics, technology and management of trading systems.

Formatted: Default Paragraph Font Formatted: Default Paragraph Font

Formatted: Default Paragraph Font	
Formatted: Default Paragraph Font	
Formatted: Default Paragraph Font	
Formatted: Default Paragraph Font	

Page 6 of 21

Mr. Van Vliet is the author of three books on trading/investment: "Modeling Financial Markets" with Robert Hendry (2003, McGraw Hill), "Building Automated Trading Systems" (2007, Academic Press). Additionally, he has published several articles in the areas of finance and technology, and presented at several academic and professional conferences, including most recently the INFORMS 2007 Annual Meeting in Seattle, the American Statistical Association 2007 Annual Meeting in Salt Lake City, Microsoft's Advancements in Quantitative Finance conference in New York, and the Automated Trading 2007 conference in London.

Name	Role	Major	Other
Khanh Duong	Master schedule maker	CPE	Web design experience
Jose Acuna-Rohter	Minute taker	CS	IPRO experience
Tarun Anupoju	Agenda maker, Time keeper	CPE	
Lance Cooper		CS	C++ Web design experience
Martin Kolodziej		EE	Web design experience
Konstantin Roytman		CS	
Oluwaseun Shonubi	Team leader	EE	
Jing Kai Tan		EE	
Jong Su Yoon		CS	Project experience
Wai Gen Yee	ProfessorFaculty advisor	Prof., CS	
Ben Van Vliet	AdvisorFaculty advisor	Advisor <u>Pr</u> of, Bus.	

Development (programming) Team

Developing the software and hardware for the system. Responsible for design and implementing the system.

Name	Role	Responsibilities
Konstantin Roytman	Sub-tTeam leader	Overall, data feed
Jong Su Yoon		Client application
Lance Cooper		Data distributor
Jose Acuna-Rohter		Head end

Research/Optimization team

Look at the base system developed, research for solutions to improve and optimize the system for maximum reliability and performance.

Name	Role	Responsibilities
Tarun Anupoju	Sub-tTeam leader	Overall
Jing Kai Tan		Assist and research

IPRO/ Web design team

Responsible for handling the creation of IPRO deliverables and the IPRO 313 Website

Name	Role	Responsibilities
Khanh Duong	Sub-tTeam leader	Overall

Spring 2008

Oluwaseun Shonubi	Deliverable research
Martin Kolodziej	Web development

Tasks and Goals	Team
Explore competitors' solutions & available technology	Research/Optimization
Develop a functioning ticker plant system	Development
Improve system performance	Development & Research/Optimization
Determine hardware requirements	Development & Research/Optimization
Update the technical user manual	Development
IPRO Website	IPRO/Web Design
IPRO Deliverables	IPRO/Web Design
Compile and Practice Presentation	All

Task assignments Development team

	Task	Start Date	Target end date	Lead
	1. Data Generator	2/06/08	4/12/08	Kenny
1.1	Obtain OPRA data.	2/06/08	2/08/08	
1.2	Implement code to stream OPRA data into RAM	2/06/08	4/15/05	
1.3	Implement code to take OPRA data from RAM and send to network sockets	2/06/08		
	2. Head End	2/06/08		Jose
2.1	Create a Decoder for OPRA FAST data	2/06/08	3/20/08	
2.2	Create a template for our data to send out	2/11/08	2/20/08	
2.3	Create UDP socket connection to receive data from CD			
2.4	Get timing metrics on UDP receiving speed and CODEC speed			
	3. Last Value Cache	2/06/08		Jong Su
3.1	Provide ability to use different hash algorithms	3/03/08	3/10/08	
	4. Data Distributor	2/06/08		Lance
4.1	Separate subscription from LVC	2/12/08	3/12/08	
4.2	Interface subscription with Head End	2/18/08	2/25/08	
	5. Client Application	2/06/08		Jong Su
5.1	Improve JAVA client and monitor application	2/17/08	2/24/08	
5.2	implement hashing functions in c++ (SuperFast, CRC32, Adler, tiger)	2/25/08	3/02/08	

5.3	plug hashing functions into LVC module		l
	Baseline End Date		4/25/08

Spring 2008

Research/Optimization team

	Week	
1	2/17/08 - 2/24/08	- In depth view of ADLER32 and CRC32
2	2/24/08 - 3/02/08	- Get code to decode OPRA data
		- Test hash functions using the metrics chosen(time &reliability)
		- Select a hash function
3	3/02008 - 3/09/08	- Implementation of the chose hash function in to the LVC
		- Possibly, presentation on partitioning data sets
4	3/09/08 - 3/16/08	 presentation Presentation on Distributed and parallel processing
		- Hardware support
5	3/16/08 - 3/23/08	- Presentation on hardware support – necessary hardware structure?
		- Analysis of real workloads
6	3/23/08 - 3/30/08	- Presentation on analysis of workloads
		- Handling system failures
7	3/30/08 - 4/6/08	- Presentation on handling system failures
8	4/7/08 - 4/30/08	- Benchmarking hash functions
		- Gathering feed data statistic

IPRO/Web Design team

Week	Task	
02/17/08	Finalize project plan and Code of Ethics deliverables	- Starting web development
02/24/08	Midterm presentation creation	- Establish basic web space
03/02/08	Collect materials for Midterm Report deliverable	- Research options and design
03/09/08	Finalize and revise midterm report	- Layout
03/16/08		- Basic text content
03/23/08	Web	
03/30/08	Development	- Graphics
04/06/08		
04/13/08	Collect materials for IPRO Day and Final report	- Final adjustments and updates
04/20/08	IPRO day materials ready - Poster - Abstract	
04/27/08	- Final Report - Final Presentation	-Power point, poster, information booklet, CD

Page 10 of 21

Spring 2008

6. Obstacles

The IPRO 313 team had obstacles and barriers similar to those all groups experience during the forming and storming phases of team building. Due to conflicting schedules, group meetings outside of class had been difficult to arrange. To solve this problem, one of the two class period of each week was designated to members meeting and working together. Presentations of work and progress have been grouped together on the other class period.

Soft skill challenges that members encountered are:

- Dealing with anxiety when presenting in front of people
- Managing personal schedules with work schedules
- Working with different personalities

Also, our IPRO is a continuation of last semester's project therefore there are problems of understanding what had been done and what needed to be done. The lack of documentation caused delays in the team's progress due to the extra time spent understanding past materials.

Development Team

Our biggest obstacle was the state of the code developed by the previous team being nonfunctional. The majority of the code had to be revised or completely rewritten from scratch. Another technical difficulty encountered during development process was issue of having insufficient knowledge on the raw, sample <u>real world</u> data. The issue was resolved after a team problem solving session. <u>The lesson learned was that after spending a fair amount of time</u> <u>analyzing a problem without success, one must solicit outside perspectives that may help solve</u> <u>the problem. Also, the "bursty" nature of the data proved to be a barrier in collecting consistent</u> matrices.

Another problem encountered was the burstiness of the data. At times, the rate of data couldbe in the tens to hundreds of thousands of messages per second, which overloaded our network capacity. We implemented some load-shedding techniques, but this area of research, known as *conflation*, is still open.

One problem existed through-out the semester was the issue of the functionality of different components inhibiting development. Since components depend on each other, when one does not work properly, others will also not function properly. The team tried to overcaome these this by developing a protocol by which each module should interact with each other as well as protocols for checking out, modifying, and checking in code.obstacles by scheduling the delivery of different components in a manner that would fit the development process. Although we did not develop a rigorous unit testing regime, components were well tested. The team collaborated well for the most part.

Research/Optimization Team

With only The two members of the Research Team majored in Computer Engineering and Electrical Engineering, the team had with little to no background on in some Computer Science specific concepts, issue such as hash functions, which are fundamental to this project. Furthermore, Both before the IPRO, neither team members felt that theirhad the programming skill was not up to par with what the necessary for the project required. To overcome this obstacle, each member tried to put forth extra time researching and learning the necessary skills and knowledge. By the end of the semester team was able to grasp the concepts and delivered significant research results. Formatted: Justified

Formatted: Justified

Formatted: Font: Italic

Page 11 of 21

Spring 2008

IPRO/Web Design Team

Our major obstacle to overcome is coordinating different sub-teams to meet the deadlines forboth IPRO deliverables and system components, which deadlines the team set forth for itself. The major challenge was in scoping the project, breaking up tasks, assigning tasks, and enforcing deadlines among team members. Eventually, the IPRO team learned that by clearly stating motivations, tasks and deadlines, other team members were responsive.

The technical challenge that the team encountered was the issue of some <u>web-Web</u> browsers[•] being non-standard compliance. The team tried its best to <u>made-make</u> the website capable of displaying correctly on different browser<u>-independent</u>s; however, it is not possible without sacrificing some of the original design elements. With that said, the team was able to make the website display correctly on Mozilla Firefox, Internet Explorer 7 and Safari.

Formatted: Justified

Formatted: Justified

Spring 2008

7. Results

Team Results

System

The current system is a functioning, end-to-end market data ticket system, tested with options data from OPRA. The system is able to operate with real-world market data and is also capable of simulating real-time data processing throughout all system components.



Data Feed

The data feed operates as a UDP server, sending OPRA data to the head-end component. The data feed was outfitted with extra functionality such as constant data throttling, real-time throttling and constant burst sleep control.

Head End

The head-end module operates as a UDP Client receiving data from the Data Feed, an OPRA FAST decoder, and also a UDP server which sends data to the LVC module. The major accomplishment here was getting consistent decoding results with the raw FAST data we were working with.

Formatted: Font: (Default) Arial, 11 pt, (Asian) Japanese (Japan)

Page 13 of 21

I

LastValueCache Implemented and benchmarked multiple hash algorithms in the LVC, system processes real world data, significant net protocol improvements.

Page 14 of 21

Spring 2008

Client

)		Client						Formatted: Keep with next
Symbol	Bid Price	Bid Quantity	Ask Price	Ask Quantity	Last Bid Tick	Last Ask Tick			<u> </u>
PVVL	8.2	00146	8.35	00060	-	+			
OWAF	2.41	00021	2.52	00032	-	+			
BWSNW	1.1	00015	1.35	00010	-	-			
TQAE	1.35	00297	1.45	00171	=	+			
OPJAG	123.45	00011	124.4	00010	=	-			
BMJK	62.7	00011	63.2	00043	-	+			
HUMEL	14.2	00054	14.7	00021	=	+			
	1.15	00021	1.45	00011		+			
ACEMU	13.8	00021	14.3	00021		+	•		
	2.05	00010	2.3	00210					
	1.55	00877	1.7	00032		+			
Connect	to Server I	P: localhos	it	Connect		STMBOL			
		Fi	oure 11 Scr	en shot of Client	interface		•		Formatted: Font: (Default) Arial, Not Italic
		<u></u>	gure 14. Ocre	sen shot of olient	interface.			\prec	Formatted: Font: Not Italic
									Formatted: Caption, Centered
ca ca	apable of s	subscribinge	e for <u>to </u>data	i from the ticke	r system. A		ont was <u>alco</u>		Tormattear caption, centerea
	ntod oo o	Wab bogo	Jove Applet	erected Sec	Figure 1 Figu	web <u>the</u> cill	crit was also		
<u>mplemei</u> bo cliopt	<u>nted as a</u>	Web page	Java Applet	created. See	Figure 1 Figu	re 1 for a sc	creen shot of		Formatted: Font: (Default) Arial, 11 pt
mplemei he client	<u>nted as a</u> t <u>.</u>	Web page	Java Applet	created. See	Figure 1 Fig	re 1 for a sc	creen shot of		Formatted: Font: (Default) Arial, 11 pt Field Code Changed
mplemei he client Develop	nted as a t. ment Tea	<u>Web page .</u> m	Java Applet	created. See	Figure 1 Figu	are 1 for a so	creen shot of		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt
mplemen he client Develop • D	ment Tea ment Tea eveloped o Sysi o Crea insta o Sysi o Impl o Sigr ch/Optim	Web page . m a base syste tem process ated internal bad of a-mes tem capable emented & ificant net p ization Tea	em es real work message for ssage forma of handling and benchm rotocol impr am	d data ormat: By utilizin ti multiple clients narked multiple ovement over p	Figure 1 Fig	ure for k-mest	sage format		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
mplemen he client Develop • D Researc	ment Tea ment Tea eveloped Sysi Crea insta Sysi Impl Sigr ch/Optim	Web page . m a base syste tem process ated internal ead of a-mes iem capable emented & ificant net p ization Tea ieved novel	em es real worl message forma of handling and benchm rotocol impr am	d data ormat: By utilizin ti multiple clients harked multiple ovement over p	Figure 1 Figu ng the structure hash algorith hast semeste	ure for k-mest mms in the LV g algorithms <u>a</u>	sage format		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
Mplemen he client Develop • D Researc • \ A	ment Tea ment Tea eveloped o Sysi o Crea inste o Sysi o Impl o Sigr ch/Optim	Web page . m a base syste tem process ated internal ead of a-mes iem capable emented & ificant net p ization Tea ieved novel statistically	em es real worl message forma of handling and benchm rotocol impr am work by <u>Be</u> significant p	d data ormat: By utilizin ti multiple clients narked multiple ovement over p <u>inchmarked</u> multiple	Figure 1 Figu ng the structu hash algorith hast semeste ltiple hashing <u>erences usir</u>	ure for k-mest mms in the LV or g algorithms <u>a</u>	sage format C		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
Mplemei he client Develop • D Researc • M <u>d</u>	ment Tea ment Tea eveloped Sysi Crea inste Sysi Impl Sigr ch/Optim Ve had act etermined ignificance	Web page . m a base syste tem process ated internal bad of a-mes tem capable emented & ificant net p ization Tea bieved novel statistically + test	em es real work message for ssage forma of handling and benchm rotocol impr am work by <u>Be</u> significant p	d data ormat: By utilizin ti multiple clients narked multiple ovement over p <u>inchmarked</u> multiple	Figure 1 Figu ng the structu hash algorith past semeste ltiple hashing erences usir	ure for k-mest mms in the LV or g algorithms <u>a</u>	sage format C		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
Mplemei he client)evelop • D (esearc • \ d si • G	ment Tea ment Tea eveloped Sysi Creations Sigr ch/Optim Ve had act etermined ignificance Sathered m Che elap	Web page . m a base syste tem process ated internal ead of a-mes tem capable emented & ificant net p ization Tea hieved novel statistically test hany statistic statistics inv sed CPU tin	em es real work message forma of handling and benchm rotocol impr am work by <u>Be</u> significant p volved inform ne and Perc	d data ormat: By utilizir tt multiple clients narked multiple ovement over p <u>enchmarked</u> mul performance diff RA data mation such as entage Frame I	Figure 1 Fig	ure for k-mest arms in the LV or g algorithms <u>a</u> the statistic eccived, fram	sage format C and bal-		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
Researce • Quelop • D • D • C • Q • Q • Q • C • T	ment Tea ment Tea eveloped Sysi Creationsta Sysi Implies Sigr ch/Optim Ve had act etermined ignificance bathered m Cheen	Web page . m a base syste tem process ated internal ead of a-mes tem capable emented & ; ificant net p ization Tea ieved novel statistically Host hany statistic statistics into rent hash all ted the Alph One at a Tin	em es real work message forma of handling and benchm rotocol impr am work by <u>Be</u> significant p volved inform ne and Perc lgorithms to na Numeric, me and Supp	d data ormat: By utilizin ti multiple clients harked multiple ovement over p <u>inchmarked</u> multiple erformance diff RA data mation such as tentage Frame I better optimize Bob Jenkins, er Fast, and othe	Figure 1 Fig	ure for k-mest urs in the LV or g algorithms <u>a</u> the statistic ecceived, fram C32 Parallel, ad hashing alg	sage format C and cal- nes received, Fowler Noll		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
Researce • Q • G • G • T	ment Tea ment Tea eveloped Sysi Creationsta Sysi Implies Sigr ch/Optim Ve had act etermined ignificance Bathered mr elap iested diffe Testivov,	Web page . m a base syste tem process ated internal ead of a-mes tem capable emented & ificant net p ization Tea ieved novel statistically test nany statistic statistics invi- sed CPU tin rent hash all ted the Alph One at a Tin	am es real worl message forma of handling and benchm rotocol impr am work by Be significant p volved inform ne and Perc Igorithms to na Numeric, ne and Supe	d data ormat: By utilizin ti multiple clients harked multiple ovement over p <u>inchmarked</u> multiple erformance diff RA data mation such as centage Frame I better optimize Bob Jenkins, er Fast, and othe	Figure 1 Fig	ure for k-mest ums in the LV or g algorithms <u>a</u> the statistic ecceived, fram C32 Parallel, ad hashing alg	sage format C and cal- nes received, Fowler Noll gorithms.		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic
mplemei he client Develop • D Researc • 44 e e • G • T	ment Tea ment Tea eveloped Sysi Creationsta Sigr Ch/Optim Vo had act etermined ignificance Bathered m elap ested diffe Test Vo, general	Web page . m a base syste tem process ated internal ead of a-mes tem capable emented & ificant net p ization Tea bieved novel statistically test any statistic statistics in sed CPU tim rent hash all ted the Alph One at a Tim	Antiperiod States and	d data ormat: By utilizin ti multiple clients harked multiple ovement over p <u>inchmarked</u> multiple overnent over p <u>inchmarked</u> multiple performance diff RA data mation such as centage Frame I better optimize Bob Jenkins, er Fast, and other cen hashing al	Figure 1 Fig	ure for k-mest ums in the LV or g algorithms <u>a</u> the statistic ecceived, fram C32 Parallel, ad hashing alg	sage format C and cal- nes received, Fowler Noll gorithms.		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic Formatted: Indent: Left: 0" Formatted: Indent: Left: 0"
Researce • D • D • D • D • D • D • D • D	ment Tea ment Tea eveloped Sysi Creationsta Sysi Implies Sigr ch/Optim Ve had act etermined ignificance Bathered m elap ested diffe Testivo, Vo, general lowever, A	Web page . m a base syste tem process ated internal ead of a-mes tem capable emented & ificant net p ization Tea bieved novel statistically test any statistic statistics invi- sed CPU tin rent hash all ted the Alph One at a Tin the different upha Numer	Antiperiod States State	d data ormat: By utilizin the multiple clients harked multiple lients harked multiple lients ovement over p <u>inchmarked</u> multiple <u>overnent over p</u> <u>inchmarked</u> multiple <u>overnent over p</u> <u>inchmarked</u> multiple <u>inchmarked</u> multiple <u>inchmarke</u>	Figure 1 Fig	ure for k-mest ums in the LV or g algorithms <u>a</u> the statistic ecceived, fram C32 Parallel, ad hashing alg	sage format C and cal- nes received, Fowler Noll gorithms.		Formatted: Font: (Default) Arial, 11 pt Field Code Changed Formatted: Font: (Default) Arial, 11 pt Formatted: Font: (Default) Times New Roma Not Italic Formatted: Indent: Left: 0" Formatted: Indent: Left: 0" Formatted: Font: (Default) Arial, 11 pt Formatted: Indent: Left: 0.5"

Spring 2008

Formatted: Indent: Left: 0"

- IPRO/Web Design Team
 Developed the IPRO website that contain all information about the project
 Each deliverables was delivered by deadline without fail

Page 16 of 21

8. <u>Design</u> Recommendations and Future Work

Based on the results of this semester, we recommend that this project be continued during future semesters. The base system is functioning. The next team could then concentrate on optimizing the system. There are also research results that could be utilized.

Data Feed

The large data volumes being dealt with proved to be one of the largest barriers to improvements due to inadequate testing time and inconsistent testing results. Future work on this component should focus on more advanced handling of the real-time data bursts that occur within the data.

Head End

The "bursty" nature of the data proved to be a barrier in collecting consistent metrics. This problem should be the main focus of future work on this module. Creating a custom buffering system along with traffic control between the data feed and the head end would be a major improvement to the reliability of the overall system. A system for shedding excess buffer load (e.g., tossing out all but the last update to a particular option) must be developed. Cloud computing could be used as a potential solution to handle the large data volumes.

Another suggestion would be allow for multiple incoming data feeds, as well as splitting load between multiple LVCs. <u>This may allow better throughput as multiple computing components</u> can be utilized in parallel.

LastValueCache

There should be a more heavily multithread the last value cache and a separate subscription module so that multiple LVC instances can be run. Subscription module needs a login process to send initial data from LVC to client. <u>The subscription module should also be able to distribute data to multiple clients using IP multicast</u>. This type of data transmission promised to improve price information dissemination scalability.

Client

The basic functions are done. To get cached data, it needs a login process. <u>The client</u> requires the ability to retrieve data directly from the head end rather than going through the LVC, which is a bottleneck.

Spring 2008

Spring 2008

9. References

Jullienne Walker, "The Art of Hashing" http://eternallyconfuzzled.com/tuts/algorithms/jsw_tut_hashing.aspx

Paul Hsieh, "Hash functions" 2007 http://www.azillionmonkeys.com/qed/hash.html

Ross N. Williams, "CRC: A paper on CRCs" <u>http://www.ross.net/crc/crcpaper.html</u>

Hash collision http://en.wikipedia.org/wiki/Hash_collision

Bob Jenkins, "A Hash Function for Hash Table Lookup" http://burtleburtle.net/bob/hash/doobs.html

Stefan Waner, Steven R. Costenoble, "Confidence Intervals: Miscellaneous on-line topics for Finite Mathematics 2e" http://people.hofstra.edu/stefan_waner/Realworld/finitetopic1/confint.html

Joy Ying Zhang, "Confidence interval and the Student's t-test" http://projectile.is.cs.cmu.edu/research/public/talks/t-test.htm

t test calculator http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=C

ActiveState Programmer Network http://aspn.activestate.com/ASPN/

Nanex Compression on an OPRA Direct Feed http://www.nanex.net/OPRADirect/OPRADirect.htm

Kernighan, Ritchie. "The C Programming Language", 2nd Edition. Prentice Hall, 1988.

Option Price Reporting Authority <u>http://www.opradata.com/</u>

Fix Protocol http://www.fixprotocol.org/

Spring 2008

10. Acknowledgements

Options Price Reporting Authority (OPRA) OPRA provided the team with real work data to workwith . Enabling us to create a base system capable of processing real world market price.

11. Team Roster

Faculty Advisor Wai Gen Yee Ben Van Vliet

Team Members Oluwaseun Shonubi, Team leader Khanh Duong Jose Acuna-Rohter Tarun Anupoju Lance Cooper Martin Kolodziej Konstantin Roytman Jing Kai Tan Jong Su Yoon

Page 19 of 21

Spring 2008

12. CD-ROM Table of Contents

Directory of D:\

table of content.txt Website.zip

Directory of D:\Class presentation

<u>Code of ethics Write up.doc</u> <u>Code of ethics Exam presentation.ppt</u> <u>Code of ethics write up 2.doc</u> Code of ethics.ppt CRC32 Presentation.ppt CRC32 Write Up.doc Deadlines 01.ppt Deadlines 02.ppt FAST write up.rtf FAST.ppt Hashing Technique Write Up.doc Hashing Technique.ppt IPRO313_Presentation_20070130_v1.0.ppt LVC Overview.pdf options presentation writeup.doc Project Plan - Write up.doc Software team status report 03-24-2008.ppt Statistic Analysis.ppt Stock options presentation.ppt Unit Testing.ppt Website progress.doc What we did at last semester.doc

Directory of D:\Deliverables

ABSTRACT.doc FINAL PRESENTATION.ppt IPRO 313 - Code of ethics.doc IPRO 313 - Final Report.doc IPRO 313 - Midterm Report.zip IPRO 313 - Minutes.zip IPRO 313 - Project Plan.doc IPRO 313 - Website.zip POSTER.ppt

Directory of D:\Softwares

code.zip Hashing.zip ipro313_source_code.tar.gz README.txt

Directory of D:\Softwares\OPRA data analysis code

analysis.py avgprice.py consec.py graphanalysis.py

Page 20 of 21

Spring 2008

keystat.c message.c uthash.h

Directory of D:\Technical documents

ClientDocument.doc DataFeed_HeadEnd_Integration_Metrics 03-24.xls DataFeed_HeadEnd_Integration_Metrics 04-23.xls Fast for OPRA.pdf Graph documentation.pdf Graph summary.pdf Hash Function Benchmart and Collision - Result.xls Hashing Function Final Report.doc Hashing results.pdf High_Level_Architecture_1.pdf IPRO 313 ? Message Structure.pdf IPRO313_TechManual_v005.doc Level 2 Architecture - Decoder.pdf Level 2 Architecture-LVC.pdf Level 2 Architecture - Client.doc Level 2 Architecture - DataGenerator.doc Level 2 Architecture - Decoder.doc <u>opra10.zip</u> opra9.zip optimistic_case_benchmarks.xls pessimistic_case_benchmarks.xls realistic_case_benchmarks.xls Two most frequent symbol APVVL_graph.pdf Under construction

Page 21 of 21