Final Report for Period: 08/2007 - 07/2008
Submitted on: 07/10/2008
Principal Investigator: Liu, Ling .
Award ID: 0306488
Organization: GA Tech Res Corp - GIT
Submitted By:
Liu, Ling - Principal Investigator
Title:
A Peer to Peer Approach to Large Scale Information Monitoring

Project Participants

Senior Personnel

Name: Liu, Ling
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Murugappan, Anand
Worked for more than 160 Hours: No
Contribution to Project:

Name: Bamba, Bhuvan
Worked for more than 160 Hours: No
Contribution to Project:

Name: Zhang, Gong
Worked for more than 160 Hours: Yes
Contribution to Project:
Gong Zhang is developing GeoGrid Peer to Peer information delivery system in the last year of the project. He is partially supported under this grant.

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Activities and Findings

Research and Education Activities: (See PDF version submitted by PI at the end of the report)

Organization of the Project

The PI is leading the PeerCQ system design and development, including the engineering effort throughout the entire project cycle of PeerCQ. The coordination of the efforts include the supervision and the management of the research activities conducted by the following PhD students who are directly or indirectly contributing to this project.

Bugra Gedik - has graduated and joined IBM TJ Watson as a research staff on Aug 1, 2006.

Anand Murugappan - 2nd year PhD student: funded by the project, focusing on dynamic content caching issues in distributed location-based information monitoring system.

Bhuvan Bamba - 3rd year phD student, joined the project in Spring 2007, and working on spatial alarms as applications of PeerCQ system.

Gong Zhang - 2nd year PhD student, joined the project in fall 2006. Currently working on efficient algorithms for large scale location based information monitoring applications built on top of PeerCQ environments. Gong was funded partially by the project.

Research Activities:

Peer-to-peer (P2P) systems are characterized as massively distributed computing systems in which all peer nodes have identical capabilities and responsibilities and all communication is symmetric. The PeerCQ research aims at investigating research issues and evaluation methods involved in building a decentralized Peer-to-Peer system architecture and system-level mechanisms for supporting distributed Internet-scale information monitoring applications.

In the course of the project, we have conducted research in three consecutive phases: Theoretical study of the issues and solution space, engineering issues and techniques, and extending the PeerCQ development to mobile location based peer to peer information monitoring applications. This extension is built on top of PeerCQ and we call it GeoGrid.

Phase I: We have performed a systematic study of research issues in design and evaluation of a scalable and yet secure peer-to-peer protocol and its supporting system architecture.

Phase II: We have investigated the important engineering issues in constructing a PeerCQ prototype system for Internet-scale peer-to-peer information monitoring, emphasizing on load balance among heterogeneous peers and efficient and scalable processing of large numbers of information monitoring requests. An important requirement of PeerCQ is to incorporate peer heterogeneity and information monitoring characteristics of the users into the peer-to-peer service-partitioning and service lookup scheme, while maintaining decentralization and self-configurability.

Phase III: We have designed and developed GeoGrid, a scalable geographical service for location dependent information montiroing utilizing our experience with PeerCQ. We have developed an initial prototype called GeoGrid, a 2-dimensional DHT based P2P network, focuing on load balancing in the presence of hot spot location queries.

Concretely, throughout the course of the project, we have continued to investigate and address the following questions and seeking for methodical approaches and solutions to address them.

(1) How should peers distribute information monitoring tasks (service partitioning) to ensure that the loads on peers are balanced and the overall system utilization is optimized?

(2) What complications might be caused due to the presence of peer heterogeneity?

(3) How can we partition information monitoring tasks among peers while enabling optimizations for efficient processing of triggers with similar structure?

(4) Can we ensure the reliability of peer-to-peer information monitoring systems in the presence of peer joins, departures, or peer failures? or put differently, which mechanisms can we explore to minimize the lost rate of information monitoring requests due to the dynamics of the peer-to-peer network?

(5) Can we defend PeerCQ against distributed Denial of Service attacks?

(6) What security mechanisms can we devise for peer to peer publish-subscribe systems like PeerCQ?

In summary, we are happy to declare the success of PeerCQ. We are proud of PeerCQ in terms of both its fundational research results on decentralized Internet Scale informaton monitoring and the engineering issues involved, we are also happy to see a location-based extension through GeoGrid development on top of the PeerCQ developmemnt, which enables us to focus more on the technical challenges of location-dependent information monitoring and dissemination, and developing optimization techniques for improving system scalabiity and reliability.

Findings:

The major findings of this project include

(1) Location scale information monitoring in an open environment demans for a systematic and serverless approach and a highly scalable and self-configurable architecture.

(2) Developing a decentralized information monitoring middleware system requires both the architecture and the techniques for scalable and reliable processing of large number of CQs over a network of loosely coupled, heterogeneous, and possibly unreliable nodes (peers). A key challenge is how to maintain a good balance between system utilization and load balance in the presence of peer joins, departures and failures.

(3) We have reported our results on PeerCQ's service partitioning scheme. Especially we have identified two key issues that are critical for making PeerCQ work in mission-critical applications. The first key issue is to provide a decentralized and yet reliable service on top of a network of loosely coupled, weakly connected and possibly unreliable peers. We addressed this issue by developing a dynamic passive replication scheme to provide reliable service in PeerCQ. The second key issue is the security of the PeerCQ system. We investigated the security issues in PeerCQ through addressing the following attacks and vulnerabilities: (1) Denial of Service attacks, making the service unavailable by infrastructure level attack or host compromise attack, (2) Denial of information attacks, making it difficult to find useful information through spamming, spoofing, and man-in-the-middle attacks.

(4) We have further extended the PeerCQ development to location based P2P system called GeoGrid. Our design focuses on efficient routing and topology constrction and maintenance and the optimization techniques for load balancing and throughput improvement.

Educational Activities and Impacts:

In the course of this project we have engaged in a number of educational activities in the context of PeerCQ.

(1) We used the released the beta version of the PeerCQ P2P protocol software as an internal open source educational device in one of the graduate courses I taught on Advanced Internet Systems. Students in the course has produced a number of interesting projects, includig p2p crawler, p2p map guide, and so forth.

(2) A number of location-based extensions on PeerCQ have resulted in GeoGrid, primarily developed jointly with one PhD student and a couple of MS and undergrad students.

In summary, the PeerCQ project funded by NSF distributed systems program has been serving as an excellent training ground for students at all levels, including Ph.D. research, M.S. special topic or term projects, and undergraduate term projects in existing and new courses on distributed computing systems, distributed information management, and advanced Internet systems and application development. Through the course of this project, I have graduated four PhD students: Lakshmish Ramaswamy now is an assistant prof in University of Georgia. Bugra Gedik is now with distributed stream systems division as a research staff in IBM TJ Watson. Li Xiong is now an assistant professor in Emory University, and Mudhakar Srivatsa is now with IBM network and systems group as a research staff in IBM TJ Watson.

In the final year of the project our research has been conducted along three dimensions.

(1) We develop techniques for enhancing system performance and scalability for decentralized information monitoring applications, addressing questions such as how do information monitors adapt to wide system parameter variations in runtime environments, how does peer-to-peer approach to large scale information monitoring differ from the conventional client-server architecture, in terms of scalability, survivability,

failure resilience, and administrative cost.

(2) We continue our effort on security and trust issues of decentralized networking systems, such as PeerCQ system, and develop defense techniques for the proposed approaches, focusing on addressing the questions such as can we defend PeerCQ against distributed Denial of Service attacks and what security mechanisms can we devise for peer to peer publish-subscribe systems like PeerCQ. We plan to develop the security guards to be incorporated into the final release of PeerCQ as an open source P2P software.

(3) We produced an open source system, GeoGrid, as a successful extension of PeerCQ for mobile location-based event monitoring and dissemination applications and develop specific research challenges utilizing temporal and spatial locality aware optimization techniques.

Training and Development:

In the course of this project, I have a total of 4 PhD students partially funded to work on this project and a total of 4 PhD students working on some issues of this project without being paid as GRA for this project. Among the four students partially paid bu this project, Bugra Gedik is the first of my PhD students working on PeerCQ. Bugra and I have published a good number of papers pioneering the area of decentralized peer to peer information monitoring. Our first paper on PeerCQ service partitioning scheme was appeared in ICDCS 2004 and was awarded as the best paper of the year by ICDCS 2004.

Mudhakar joined the project in Spring 2006 and is responsible for the improvement and extention of the PeerCQ P2P protocol. Mudhakar Srivstsa is not funded by this project due to the shortage of fund, but he continues to work on secure p2P information dissemination networks, studying the vulnerabilities and defense mechanisms for structured P2P protocols and in particular P2P publish and subscribe systems like PeerCQ.

Anand and Bhuvan are PhD students working on developing applications for PeerCQ through Spatial Alarms system.

Gong Zhang is working on GeoGrid, an extension of PeerCQ to mobile location-based applications.

Outreach Activities:

The PI has given tutorials, panels, keynotes at several natioal or international organizations, universities, including panel, tutorial, and research talks at ICWS 2006, WAIM 2006, SST 2007, ICDE 2006, ICDE 2007, and a number of seminars in distinguished lecture series of different US universities and universities in Japan.

Journal Publications

Li Xiong and Ling Liu., "PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities", IEEE Transactions on Knowledge and Dat a Engineering, p. 843, vol. 16, (2004). Published,

Ling Liu, Wei Tang, David Buttler, and Calton Pu, "Information Monitoring on the Web: A Scalable Solution", World Wide Web Journal (Kluwer Academic Publishers), p. 263, vol. 5, (2003). Published,

Bugra Gedik and Ling Liu, "A Scalable Peer-to-Peer Architecture for Distributed Information Monitoring Applications", to appear in IEEE Transactions on Computers., p. , vol. , ( ). Accepted,

Lakshmish Ramaswamy, Bugra Gedik, Ling Liu. "A", IEEE), Vol. 16, No. 9, Sept. 2005. pp1-16., "A Distributed Approach to Node Clustering in Decentralized Peer-to-Peer Networks", Transactions on Parallel and Distributed Systems (TPDS), p. , vol. 16, (2005). Accepted,

Lakshmish Ramaswamy, Arun Iyengar, Ling Liu, Fred Douglis, "Automatic Fragment Detection in Dynamic Web Pages and its Impact on Caching", IEEE Transactions on Knowledge and Data Engineering, p. , vol. 17, (2005). Accepted,

James Caverlee and Ling Liu, "QA-Pagelet: Data Preparation Techniques for Large Scale Data Analysis of the Deep Web", IEEE Transactions on Knowledge and Data Engineering - Special Issue on Intelligent Data Preparation., p. 1247, vol. 17, (2005). Published,

Mudhakar Srivatsa, Bugra Gedik, and Ling Liu, "Scaling Unstructured Peer-to-Peer Networks with Heterogeneity-Aware Topology And Routing", IEEE Transaction on Parallel and Distributed Systems (TPDS), p. , vol. , ( ). Accepted,

Bugra Gedik, Kun-Lung Wu, Philip Yu, and Ling Liu., "Processing Moving Queries over Moving Objects Using Motion Adaptive Indexes", IEEE Transactions on Knowledge and Data Engineering (TKDE)., p. , vol. , ( ). Accepted,

Aameek Singh, Bugra Gedik and Ling Liu., "Agyaat: Mutual Anonymity over Structured P2P Networks", Emerald Internet Research Journal (Special Issue on Privacy and Anonymity in the Digital Era), p. , vol. , ( ). Accepted,

Mudhakar Srivatsa and Ling Liu, "DHTGuard: Security Threats in Structured Overlay Networks", the Springer International Journal of Information Security, p. , vol. , ( ). Accepted,

Mudhakar Srivatsa and Ling Liu, "Securing Decentralized Reputation Management using TrustGuard", the Journal of Parallel and Distributed Computing (JPDC) Special Issue on Security in Grid and Distributed Systems., p. , vol. , ( ). Accepted,

James Caverlee, Ling Liu, and Daniel Rocco, "Discovering Interesting Relationships among Deep Web Databases: A Source-Biased Approach", World Wide Web Journal: Internet and Web Information Systems, Springer (WWWJ)., p. , vol. , ( ). Accepted,

Bugra Gedik, Kun-Lung Wu, Philip S. Yu, and Ling Liu, "CPU Load Shedding for Binary Stream Joins", Springer Knowledge and Information Systems. DOI 10.1007/s10115-006-0044-4, September 21 2006, p. , vol. , (2006). Accepted,

Lakshmish Ramaswamy, Ling Liu, Arun Iyengar., "Cooperative Edge Cache Grid: Scalable Delivery of Dynamic Content", IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 19, No. 5, 2007., p. , vol. 19, (2007). Accepted,

Jianjun Zhang, Ling Liu, Lakshmish Ramaswamy, "PeerCast: Churn-Resilient End System Multicast on Heterogeneous Overlay Networks appear.", Journal of Network and Computer Applications (JNCA), Elsevier, 2007., p. , vol. , ( ). Accepted,

Daniel Rocco, James Caverlee, Ling Liu and Terence Critchlow, "Service Class Driven Dynamic Source Discovery with DynaBot", To appear in", International Journal of Web Services Research (JWSR), p. , vol. , ( ). Accepted,

L. Liu, J. Zhang, W. Han, C. Pu, J. Caverlee, S. Park, T. Critchlow, M. Coleman, and D. Buttler, "Providing Multi-Page Data Extraction Services with XWRAPComposer", Advanced Web Services Research and Practices, p. , vol. , ( ). Accepted,

Lakshmish Ramaswamy, Ling Liu, Arun Iyengar, "Cooperative Edge Cache Grid: Scalable Delivery of Dynamic Content", IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 19, No. 5, 2007., p. , vol. 19, (2007). Published,

Bugra Gedik, Ling Liu, and Philip Yu, "ASAP: An Adaptive Sampling Approach to Data Collection in Sensor Networks", IEEE Transactions on Parallel and Distributed Computing (TPDS), p. , vol. , (2007). Accepted,

Bugra Gedik, Kun-Lung Wu, Philip S. Yu, and Ling Liu, "GRUBJOIN: An Adaptive, Multi-Way, Windowed Stream Join with Time Correlation-Aware CPU Load Shedding", IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No.10, 2007., p. pp1361-13, vol. 19, (2007). Published,

Mudhakar Srivatsa and Ling Liu, "Mitigating Denial of Service Attacks on an Overlay Network: A Location Hiding Approach", IEEE Transactions on Parallel and Distributed Systems (TPDS)., p. , vol. , (2008). Accepted,

Aameek Singh, Ling Liu, Mustaque Ahamad., "Privacy Analysis and Enhancements for Data Sharing in *nix Systems", International Journal of Information and Computer Security (IJICS), 2008, p. , vol. , (2008). Accepted,

Books or Other One-time Publications

Li Xiong and Ling Liu, "Reputation and Trust in Mobile Commerce", (2004). Book Chapter, Accepted
Collection: Advances in Security and Payment Methods for Mobile Commerce
Bibliography: Book Chapter

David Buttler, Dan Rocco, Ling Liu., "Efficient Web Change Monitoring with Page Digest", (2004). , Published
Collection: The 10th International Conference on World Wide Web, May 17-22, 2004. (Manhattan, N.Y.)
Bibliography: Proceedings of the 10th International Conference on World Wide Web

Jianjun Zhang, Ling Liu, Calton Pu, and Mustafa Ammar., "Reliable Peer-to-peer End System Multicasting through Replication", (2004). , Accepted
Collection: IEEE International Conference on Peer to Peer Computing (P2P 2004)
Bibliography: Proceedings of IEEE International Conference on Peer to Peer Computing (P2P 2004), Zurich, Switzerland, Aug. 25-27, 2004.

Mudhakar Srivatsa, Bugra Gedik, Ling Liu., "Scaling Unstructured Peer-to-Peer Networks with Multi-Tier Capacity-Aware Overlay Topologies", (2004). , Accepted
Collection: The Tenth International Conference on Parallel and Distributed Systems
Bibliography: Proceedings of the Tenth International Conference on Parallel and Distributed Systems (ICPADS 20 04), July 7-9 in Newport Beach, California.

Lakshmish Ramaswamy, Bugra Gedik, Ling Liu, "Connectivity Based Node Clustering in Decentralized Peer to Peer Networks", (2003). , Published
Collection: The third IEEE Conference on Peer-to-Peer Computing
Bibliography: Proceedings of the third IEEE Conference on Peer-to-Peer Computing, September 1-, Linkoping, Sweden.

Bugra Gedik and Ling Liu, "Building Reliable Peer-to-Peer Information Monitoring Service Through Replication", (2003). , Published
Collection: The 22nd Symposium on Reliable Distributed Computing (SRDS2003)
Bibliography: Proceedings of the 22nd Symposium on Reliable Distributed Computing (SRDS2003), 6th-8th/October, 2003 - Florence, Italy

Bugra Gedik and Ling Liu, "PeerCQ: A Decentralized and Self-Configuring Peer-to-Peer Information Monitoring System", (2003). , Published
Collection: The 23rd International Conference on Distributed Computing Systems (ICDCS2003)
Bibliography: Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS2003), May 19-22, 2003, Providence, Rhode Island USA. (Best Paper Award)

Mudhakar Srivatsa and Ling Liu, "Countering Targeted File Attacks using LocationGuard", (2005). Conference, Accepted
Collection: Proceedings of 14th USENIX Security Symposium (USENIX Security)
Bibliography: August 1 - 5, 2005, Baltimore, MD.

Jianjun Zhang, Ling Liu, Calton Pu, "Constructing a Proximity-Aware Power Law Overlay Network", (2005). Conference, Accepted
Collection: IEEE Global Telecommunications Conference (GLOBECOM 2005)
Bibliography: Nov. 28 - Dec. 02, 2005, St. Louis, Missouri, USA

Lakshmish Ramaswamy, Ling Liu and Arun Iyenagar, "Cache Clouds: Cooperative Caching of Dynamic Documents on Edge Networks", (2005). Conference, Accepted
Collection: Proceedings of the 25th International Conference on Distributed Computing Systems (ICDCS-2005)
Bibliography: June 6-10, 2005

Mudhakar Srivatsa, Li Xiong and Ling Liu, "TrustGuard: Countering Vulnerabilities in Reputation Management For Decentralized Overlay Networks", (2005). Conference, Published
Collection: Proceed ings of 14th World Wide Web Conference (WWW 2005)
Bibliography: May 10-14, 2005, in Chiba, Japan.

Mudhakar Srivatsa, Li Xiong, Ling Liu, "ExchangeGuard: A Distributed Protocol for Electronic Fair-Exchange", (2005). Conference, Published
Collection: Proceedings of the 19th International Parallel and Distributed Processing Symposium (IPDPS 2005)

Bibliography: April 4-8, 2005 in Denver, Colorado.

Sungkeun Park, Ling Liu, Calton Pu, Mudhakar Srivatsa, Jianjun Zhang, "Resilient Trust Management for Web Service Integration", (2005). Conference, Accepted

Collection: Proceedings of the 3rd IEEE International Conference on Web Services (ICWS 2005)
Bibliography: July 6-10, 2005

Daniel Daniel Rocco, James Caverlee, Ling Liu, and Terence Critchlow, "Domain-specific Web Service Discovery with Service Class Descriptions", (2005). Conference, Accepted
Collection: Proceedings of the 3rd IEEE International Conference on Web Services (ICWS 2005)
Bibliography: July 13-15, 2005

Aameek Singh, Kaladhar Voruganti, Sandeep Gopisetty. David Pease, Linda Duyanovich and Ling Liu, "Security vs Performance: Tradeoffs using a Trust Framework", (2005). Conference, Published
Collection: Proceedings of the 22nd IEEE-NASA Conference on Mass Storage Systems and Technologies (MSST)
Bibliography: April 11-14, 2005, Monterey, California, USA.

Daniel Rocco, James Caverlee, and Ling Liu, "XPACK: A High-performance Web Document Encoding", ( ). Conference, Accepted
Collection: Proceedings of the International Conference on Web Information Systems and Technologies (WEBIST2005)
Bibliography: 26-28 May 2005 Miami Beach, Florida, USA.

Bugra Gedik, Ling Liu, "A Customizable k-Anonymity Model for Protecting Location Privacy
", (2005). Conference, Accepted
Collection: Proceedings of the 25th International Conference on Distributed Computing Systems (IEEE ICDCS 2005)
Bibliography: June 6-10, 2005

Dan Rocco, "Discovering and Tracking Interesting Web Services", (2004). Thesis, Published
Collection: PhD Dissertation
Bibliography: Oct. 2004

Jianjun Zhang, Ling Liu, Calton Pu, "Constructing a Proximity-Aware Power Law Overlay Network", (2005). Book, Published
Collection: IEEE Global Telecommunications Conference (GLOBECOM 2005)
Bibliography: Nov. 28 - Dec. 02, 2005, St. Louis, Missouri, USA.

Mudhakar Srivatsa and Ling Liu, "Countering Targeted File Attacks using LocationGuard", (2005). Book, Published
Collection: Proceedings of the 14th USENIX Security Symposium (USENIX Security)
Bibliography: August 1 - 5, 2005, Baltimore, MD. pp81-96. AR: 14% (22 out of 148).

Mudhakar Srivatsa and Ling Liu, "Securing Publish-Subscribe Overlay Services with EventGuard", (2005). Book, Published
Collection: Proceedings of ACM Computer and Communication Security (CCS 2005)
Bibliography: November 7-11, 2005, Hilton Alexandria Mark Center, Alexandria, VA, USA. AP: 15% (38 out of 249).

Lakshmish Ramaswamy, Ling Liu, and Jianjun Zhang, "Efficient Formation of Edge Cache Groups for Dynamic Content Delivery", (2006). Book, Published
Collection: Proceedings of the 26th International Conference on Distributed Computing Systems (ICDCS-2006)
Bibliography: Lisbon, Portugal, July 2006. AR: 13.8% (75 out of 540).

Bugra Gedik and Ling Liu, "Quality-Aware Distributed Data Delivery for Continuous Query Services", (2006). Book, Published
Collection: ACM 2006 International Conference on Management of Data (ACM SIGMOD)
Bibliography: Chicago, June 26-29, 2006. AR: 13% (58 out of 446)

Mudhakar Srivatsa, Arun Iyengar, Jian Yin and Ling Liu, "A Client-Transparent Approach to Defend Against Denial of Service Attacks", (2006). Book, Published
Collection: Proceedings of 25th IEEE Symposium on Reliable Distributed Systems (SRDS 2006)
Bibliography: SRDS, UK

Mudhakar Srivatsa and Ling Liu, "Key Derivation Algorithms for Monotone Access Structures in Cryptographic File Systems", (2006). Book, Published

Collection: Proceeding of 11th European Symposium on Research in Computer Security (ESORICS 2006)
Bibliography: ESORICS

J. Caverlee, L. Liu, and J. Bae, "Distributed Query Sampling: A Quality-Conscious Approach", (2006). Book, Published
Collection: Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR '06)
Bibliography: August 6-11, 2006 in Seattle. Acceptance rate: 18.5% (74 out of 399

Mudhakar Srivatsa, Arun Iyengar, Jian Yin and Ling Liu, "A Middleware System for Protecting Against Application Level Denial of Service Attacks", (2006). Proceedings, Published
Collection: Proceedings of 7th ACM/IFIP/USENIX International Middleware Conference (Middleware 2006)
Bibliography: Acceptance Rate: 17% (21 out of 122)

Bhuvan Bamba, Ling Liu, James Caverlee, Vaibhav Padliya, Mudhakar Srivatsa, Mahesh Palekar, Joseph Patrao, Tushar Virender Bansal, Suiyang Li, Aameek Singh, "DSphere: A Source-Centric Approach to Crawling, Indexing and Searching the World Wide Web", (2007). Proceedings, Published
Collection: IEEE 23rd International Conference on Data Engineering (2007)
Bibliography: Istanbul, Turkey; April 17-20, 2007.

Bugra Gedik, Kun-Lung Wu, Philip S. Yu, and Ling Liu, "GrubJoin: An Adaptive Multi-Way Windowed Stream Join with Time Correlation-Aware CPU Load Shedding", (2007). Proceedings, Accepted
Collection: IEEE 23rd International Conference on Data Engineering
Bibliography: Istanbul, Turkey; April 17-20, 2007.

Bugra Gedik, Ling Liu, Kun-Lung Wu, Philip S. Yu, "Lira: Lightweight, Region-aware Load Shedding in Mobile CQ Systems", (2007). Proceedings, Published
Collection: IEEE 23rd International Conference on Data Engineering (Istanbul, Turkey; April 17-20, 2007)
Bibliography: Istanbul, Turkey; April 17-20, 2007.

Sangeetha Seshadri, Vibhore Kumar, Brian F. Cooper and Ling Liu, "Optimizing Multiple Distributed Stream Queries Using Hierarchical Network Partitions", (2007). Proceedings, Published
Collection: Proceedings of the 21st IEEE International Parallel and Distribute d Processing Symposium (IPDPS), Long Beach, 2007.
Bibliography: Long Beach, 2007. Acceptance rate: 26% (109 out of 419).

James Caverlee, Steve Webb, and Ling Liu, "Spam-Resilient Web Rankings via Influence Throttling", (2007). Proceedings, Published
Collection: 21st IEEE International Parallel and Distributed Processing Symposium (Long Beach, 2007)
Bibliography: Long Beach, 2007. Acceptance rate: 26% (109 out of 419).

Sangeetha Seshadri, Lawrence Chiu, Karan Gupta, Paul Muench, Ling Liu and Brian F. Cooper, "A Fault-Tolerant Middleware Architecture for High-Availability Storage Services", (2007). Proceedings, Accepted
Collection: IEEE International Conference on Services Computing (SCC) 2007.
Bibliography: July 9-13, 2007, Salt Lake City, Utah, USA.

Aameek Singh, Mudhakar Srivatsa, Ling Liu, "Efficient and Secure Search of Enterprise File Systems", (2007). Proceedings, Accepted
Collection: Proceedings of IEEE International Conference on Web Services (ICWS 2007)
Bibliography: July 9-13, 2007, Salt Lake City, Utah, USA.

Mudhakar Srivatsa and Ling Liu, "Secure Event Dissemination in Content-Based Publish-Subscribe Networks", (2007). Proceedings, Accepted
Collection: Proceedings of 27th IEEE International Conference on Distributed Computing Systems (ICDCS 2007)
Bibliography: Toronto, Canada, June 24-29, 2007.
Acceptance Rate: 13% (71 out of 528).

Jianjun Zhang, Gong Zhang, Ling Liu, "GeoGrid: A Scalable Location Service Network", (2007). Proceedings, Published
Collection: Proceedings of 27th IEEE International Conference on Distributed Computing Systems (ICDCS 2007)

Bibliography: Toronto, Canada, June 24-29, 2007.
Acceptance Rate: 13% (71 out of 528).

James Caverlee and Ling Liu, "Countering Web Spam with Credibility-Based Link Analysis", (2007). Proceedings, Published
Collection: ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC) 2007
Bibliography: Aug 12-16, Portland, OR.
AR: 15% (32 out of 204)

Padliya, Mudhakar Srivatsa, Mahesh Palekar, Joseph Bhuvan Bamba, Ling Liu, James Caverlee, Vaibhav Patrao, Tushar Virender Bansal, Suiyang Li, Aameek Singh, "DSphere: A Source-Centric Approach to Crawling, Indexing and Searching the World Wide Web", (2007). Proceedings, Published
Collection: 23rd International Conference on Data Engineering.
Bibliography: Istanbul, Turkey; April 17-20, 2007

Kipp Jones and Ling Liu, "Map-matching: Towards Improving Wireless Positioning", (2007). Proceedings, Published
Collection: Proceedings of the 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (Mobiquitous 2007).
Bibliography: August 6-10, 2007, Philadelphia, PA.

Jianjun Zhang, Ling Liu, Gong Zhang, Lakshmish Ramawamy., "A Utility-Aware Middleware Architecture for Decentralized Group Communication Applications", (2007). Proceedings, Published
Editor(s): Proceedings of
Collection: ACM/IFIP/USENIX 8th International Middleware Conference.
Bibliography: Newport Beach, Orange County, California, USA. November 26th - 30th, 2007

Bugra Gedik, Kun-Lung Wu, Philip S. Yu, and Ling Liu, "MobiQual: QoS-aware Load Shedding in Mobile CQ Systems", (2008). Proceedings, Published
Collection: Proceedings of the 2008 IEEE International Conference on Data Engineering (ICDE 2008)
Bibliography: April 7-12, 2008, Cancun, Mexico.

Aameek Singh and Ling Liu., "SHAROES: A Data Sharing Platform for Outsourced Enterprise Storage Environments", (2008). Proceedings, Published
Collection: Proceedings of the 2008 IEEE International Conference on Data Engineering (ICDE 2008)
Bibliography: April 7-12, 2008, Cancun, Mexico.

Mudhakar Srivatsa, Arun Iyengar, and Ling Liu, "Preserving Caller Anonymity in Voice-over-IP Networks", (2008). Proceedings, Published
Collection: Proceedings of the IEEE Symposium on Security and Privacy
Bibliography: May 18-21, 2008 The Claremont Resort, Berkeley/Oakland, California, USA

Yiming Zhang, Ling Liu, Dongsheng Li and Xicheng Lu, "Distributed Line Graphs: A Universal Framework for Building DHTs Based on Arbitrary Constant-Degree Graphs", (2008). Proceedings, Published
Collection: Proceeding of IEEE International Conference on Distributed Computing (ICDCS 2008)
Bibliography: June 14-20, 2008, Beijing.

James Caverlee, Ling Liu, Steve Webb, "SocialTrust: Tamper-Resilient Trust Establishment in Online Communities", (2008). Proceedings, Published
Collection: Proceedings of the International Joint Conference on Digital Libraries (JDL 2008)
Bibliography: June 16-20, 2008 - Pittsburgh, Pennsylvania

Sangeetha Seshadri, Vibhore Kumar, Brian Cooper, Bhuvan Bamba, Gong Zhang, Ling Liu, Karsten Schwan, "Grouping Distributed Stream Query Services by Operator Similarity and Network Locality", (2008). Proceedings, Published
Collection: Proceedings of the 2008 IEEE Int. Conference on Service Computing (SCC)
Bibliography: July 8-11, 2008, Honolulu, Hawaii, USA.

Bhuvan Bamba, Sangeetha Seshadri, Ling Liu, "Scaling Location-based Services with Dynamically Composed Location Index", (2008). Proceedings, Published
Collection: Proceedings of the 2008 IEEE Int. Conference on Service Computing (SCC)
Bibliography: July 8-11, 2008, Honolulu, Hawaii, USA.

Anand Murugappan and Ling Liu., "A Energy Efficient Approach to Processing Spatial Alarms on Mobile Clients", (2008). Proceedings, Published
Collection: Proceedings of the ISCA 17th International Conference on Software Engineering and Data Engineering (SEDE-2008)
Bibliography: June 30 - July 2, 2008, at the Omni Los Angeles Hotel at California Plaza, Los Angeles, California, USA.

Web/Internet Site

URL(s):
http://disl.cc.gatech.edu/GTPeers/
Description:
PeerCQ and a number of research efforts related to PeerCQ are listed at http://www.cc.gatech.edu/projects/disl/GTPeers/.

Other Specific Products

Product Type:
Software (or netware)
Product Description:
The beta release of the PeerCQ P2P protocol source code is now available at
http://wwww.cc.gatech.edu/projects/disl/PeerCQ/downloads/download.html
Sharing Information:
We have made all PeerCQ papers available online. We also have made the recent beta release of the PeerCQ P2P protocol available at PeerCQ website, and GeoGrid is available from GeoGrid website.

Contributions

Contributions within Discipline:
Since the first paper on PeerCQ appeared, we have received many requests asking for (1) the technical report of the paper, (2)questions relatd to the PeerCQ project, and (3) the possibility of download the PeerCQ software. These requests demonstrate that the techniques developed under the PeerCQ project are of general interests to the public and research communities and are benefical to many who work in the same or related discipline.

In the last year of the project, we have developed GeoGrid on top of PeerCQ for location dependent monitoring. We just released the GeoGrid first version as open source. The system incorporates both client map based GUI with the decentralized topology and lookup methods. With GeoGrid, we are experimenting large scale decentralized information monitoring through mobile location based applications and services such as Spatial Alarms.

Contributions to Other Disciplines:

Contributions to Human Resource Development:
Bugra Gedik has graduated in July 2006 and is now a reserch staff at IBM TJ Watson. His involvement in PeerCQ opens a new horizon for him to continue his research in a new direction.

Mudhakar Srivatsa graduated in April 2007 and is now a reserch staff at IBM TJ Watson. His involvement in PeerCQ is on distributed network security, which opens a new horizon for him and graduated with a system security dissertation.

Jianjun Zhang is another PhD student who has been partially supported by this project. Jianjun graduated in July 2006 and is now a McKinsey&Co.

Contributions to Resources for Research and Education:

We have used both PeerCQ prototype, GeoGrid prototype and PeerCrawl to provide the senior undergraduates and graduate students some handon code for a course based project.

Contributions Beyond Science and Engineering:

PeerCQ also creates a community building technique to organize a set of users with certain common interests. We have built an location-based extension, called GeoGrid, to monitoring the location information.

We are currently building a secure overlay service infrastructure as a part of the PeerCQ development and plan to release this software as an open source in early 2009.

Categories for which nothing is reported:

Organizational Partners

Contributions: To Any Other Disciplines

Research Activities:

Peer-to-peer (P2P) systems can be characterized as massively distributed computing systems in which all peer nodes have identical capabilities and responsibilities and all communication is symmetric.

This research aims at investigating research issues and evaluation methods involved in building a decentralized Peer-to-Peer system architecture and system-level mechanisms for supporting distributed Internet-scale information monitoring applications. We carry out the PeerCQ project in two consecutive phases. First, we have performed a systematic study of research issues in design and evaluation of a scalable peer-to-peer protocol and its supporting system architecture. Second, we have investigated the important engineering issues in constructing a PeerCQ prototype system for Internet-scale peer-to-peer information monitoring, emphasizing load balance among heterogeneous peers and efficient and scalable processing of large numbers of information monitoring requests. An important requirement of PeerCQ is to incorporate peer heterogeneity and information monitoring characteristics of the users into the peer-to-peer service-partitioning and service lookup scheme, while maintaining decentralization and self-configurability.

Concretely, we have been working on the following questions in the first year of this funded research:

(1) How should peers distribute information monitoring tasks (service partitioning) to ensure that the loads on peers are balanced and the overall system utilization is optimized? What complications might be caused due to the presence of peer heterogeneity? How can we partition information monitoring tasks among peers while enabling optimizations for efficient processing of triggers with similar structure?

(2) How can we ensure the reliability of peer-to-peer information monitoring systems in the presence of peer joins, departures, or peer failures? Which mechanisms can we explore to minimize the lost rate of information monitoring requests due to the dynamics of the peer-to-peer network?

The first question and the research results of our investigation has been documented in the first PeerCQ paper. A short version of 10 pages has been published in ICDCS 2003 and won the best paper award, out of 405 submitted papers and 70 accepted papers. PeerCQ uses Continual Queries (CQs) as its primitives to express information-monitoring requests. The PeerCQ development has three unique characteristics. First, we develop a systematic and serverless approach to large scale information monitoring, aiming at providing a fully distributed, highly scalable and self-configurable architecture for scalable and reliable processing of large number of CQs over a network of loosely coupled, heterogeneous, and possibly unreliable nodes (peers). Second, we introduce an effective service partitioning scheme at the P2P protocol layer to distributed the processing of CQs over a peer to peer information monitoring overlay network, while maintaining a good balance between system utilization and load balance in the presence

of peer joins, departures and failures. A unique feature of our service partitioning scheme is its ability to incorporate strategies for handling hot spot monitoring requests and peer heterogeneity into the load balancing scheme in PeerCQ. To evaluate the effectiveness of PeerCQ's service partitioning scheme with respect to system utilization and load balancing, we have designed a series of experiments using a simulator. The first set of experiments evaluates the effect of different grouping factor values on the performance of PeerCQ. The second set of experiments intends to measure the effectiveness of the PeerCQ service partitioning scheme in terms of load balancing and system utilization. It demonstrates how well the load is balanced and the system is utilized in PeerCQ. The third set of experiments evaluates the effect of the utilization function used in the relaxed matching on the performance of PeerCQ. For further detail see the ICDCS 2003 paper or our technical report, currently under consideration for a journal publication.

A key challenge in peer-to-peer computing systems is to provide a decentralized and yet reliable service on top of a network of loosely coupled, weakly connected and possibly unreliable peers. The second question and the research results have been documented and published in the 22nd Symposium on Reliable Distributed Computing (SRDS2003), titled ì Building Reliable Peer-to-Peer Information Monitoring Service Through Replicationî . The main result can be summarized as follows. First, we propose an effective dynamic passive replication scheme designed to provide reliable service in PeerCQ, a decentralized and self-configurable peer-to-peer Internet information monitoring system. We first describe the design of a distributed replication scheme, which enables reliable processing of long-running information monitoring requests in an environment of inherently unreliable peers. We have also developed an analytical model to discuss its fault tolerance properties. In addition, we have conducted a set of initial experiments using our simulator, showing the feasibility and the effectiveness of the proposed reliability approach.

Project Educational Activities

There are several educational activities related to the PeerCQ project. First, our first beta version of the PeerCQ P2P protocol software has been made available as an internal open source educational device in the graduate course on Advanced Internet Systems to experiment with peer-to-peer computing and real-life update tracking applications. We also included a 2 hour lecture on PeerCQ in our undergraduate' database system implementation course. It stirred a lot of interests from the students. Furthermore, three PhD students have been engaged in concrete research tasks under the PeerCQ project.

We believe that the PeerCQ project will serve as an excellent training ground for students at all levels, including Ph.D. research, M.S. projects, and undergraduate term projects in existing and new courses in topics such as distributed computing systems, distributed information management on the net, and advanced Internet application development. Furthermore, the PeerCQ technology produced in this funded project will have a significant impact on the development of human resources, by means of direct training of

Ph.D. students, and offering a variety of research seminars and new innovative courses at Georgia Tech.
In the subsequent years, we plan to continue our research on the first two questions and further investigate the following questions.

(3) What mechanisms are best candidates for providing efficient location (lookup) service across a peer-to-peer overlay network? How do we guarantee the quality of the lookup service?

(4) What are the performance metrics for evaluation of the proposed peer-to-peer solutions, how do information monitors adapt to wide system parameter variations in runtime environments? How does peer-to-peer approach to large scale information monitoring differ from the conventional client-server architecture, in terms of scalability, survivability, failure resilience, and administrative cost?

In addition we will continue our effort towards the development of the first PeerCQ prototype, aiming at making the prototype system available as an open source P2P software at end of the second year.