Sarvjeet Singh

2044 Montecito Ave, Apt 8
Mountain View, CA 94043

http://sarvjeet.com
Phone: (765) 637-0537
sarvjeet@gmail.com

Education

Ph.D. in Computer Science **Jan 2009**
Purdue University, West Lafayette, Indiana
Thesis: Database Support for Uncertain Data
Advisor: Prof. Sunil Prabhakar

Applied Management Principles (Mini MBA) **May 2007**
Krannert Graduate School of Management, West Lafayette, Indiana

M.S. in Computer Science **May 2005**
Purdue University, West Lafayette, Indiana

B.Tech. in Computer Science and Engineering **May 2003**
Indian Institute of Technology (IIT) Bombay, Mumbai, India

Experience

Senior Software Engineer **Jan 2009 - present**
Google Inc, Mountain View, California

- Tech Lead of a team of 26 engineers that are working towards maximizing advertiser's ROI through automated real-time bidding in ads.
- Significantly contributed in the design, launch and maintenance of Conversion Optimizer (Target CPA) and Enhanced CPC products. These products manage well over a billion dollars in revenue per year.
- Developed and lead others to develop a number of crucial components for automated bidding: Models, model calibration, back-end infrastructure to process huge amount of data, serving code and front-end.
- Used C++ and Python extensively and expert in a number of technologies such as Bigtable, Map reduce, etc.

Summer Intern **Jun 2008 - Sep 2008**
Google Inc, Mountain View, California

- Proposed and implemented a solution in Java for automatic schema change detection and reconciliation of constantly evolving data sources. The team I worked with later launched Google Public Data product.

Graduate Research Assistant **Jan 2007 - Jun 2008**
Department of Computer Science, Purdue University, West Lafayette, Indiana

- Worked on research issues related to managing uncertain data.
- Developed an open-source project called Orion DBMS which extends PostgreSQL to deal with imprecise or uncertain data.
- Actively involved in developing the theory, core system and indexing structures for Orion.
- Published over 10 research papers at various databases conferences including SIGMOD, VLDB and ICDE.

Graduate Research Professional **Dec 2005 - Dec 2006**
Rosen Center for Advanced Computing, Purdue University, West Lafayette, Indiana

- Developed mobile applications for deployment in intelligent environments using technologies such as J2ME and conducted research on issues related to communication technologies such as RFID and Bluetooth.

Graduate Teaching Assistant **Aug 2005 - Dec 2005**
Department of Computer Science, Purdue University, West Lafayette, Indiana

- Assisted students with their problems and graded lab assignments in the graduate level Operating Systems course.

Summer Intern **May 2005 - Aug 2005**
Amazon.com, Seattle, Washington

- Developed a high performance fault-tolerant business application in C++ utilizing distributed caching and databases.

Graduate Research Assistant **Aug 2003 - May 2005**
Department of Computer Science, Purdue University, West Lafayette, Indiana

- Worked on issues related to Privacy and Security in Databases.

Summer Intern **May 2002 - Jul 2002**
Media Lab Asia (Arm of MIT Media Lab, USA), Mumbai, India

- Developed and implemented a prototype of a Multilingual and Meaning Based Search Engine.

Research Associate **May 2001 - Jul 2001**
Center of Studies in Resources Engineering, IIT Bombay, India

- Developed and implemented algorithms for registration and analysis of satellite images.

Publications

- Y. Qi, R. Jain, Sarvjeet Singh, S. Prabhakar. **Threshold query optimization for uncertain data**. *In proceedings of Special Interest Group on Management of Data (SIGMOD), 2010* **(5 citations)**.
- Sarvjeet Singh, C. Mayfield, S. Mittal, S. Prabhakar, S. Hambrusch and R. Shah. **The Orion Uncertain Data Management System**. *In proceedings of the 14th International Conference on Management of Data (COMAD 2008 Demo), Mumbai, India, Dec 2008* **(5 citations)**.
- S. Prabhakar, R. Shah and Sarvjeet Singh. **Indexing Uncertain Data**. *Chapter 9 of Charu Aggarwal, ed., Managing and Mining Uncertain Data, Springer-Verlag, 2008/9.*
- Y. Qi, Sarvjeet Singh, R. Shah, S. Prabhakar. **Indexing Probabilistic Nearest-Neighbor Threshold Queries**. *In proceedings of Workshop on Management of Uncertain Data (MUD, VLDB), Auckland, New Zealand, Aug 2008* **(9 citations)**.
- Sarvjeet Singh, C. Mayfield, R. Shah, S. Prabhakar and S. Hambrusch. **Query Selectivity Estimation for Uncertain Data**. *In proceedings of 20th International Conference on Scientific and Statistical Database Management (SSDBM 2008), Hong Kong, China, Jul 2008* **(7 citations)**.
- Sarvjeet Singh, C. Mayfield, S. Mittal, S. Prabhakar, S. Hambrusch and R. Shah. **Orion 2.0: Native Support for Uncertain data**. *In proceedings of Special Interest Group on Management of Data (SIGMOD 2008 Demo), Vancouver, Canada, Jun 2008* **(40 citations)**.
- Sarvjeet Singh and Sunil Prabhakar. **Ensuring Correctness over Untrusted Private Databases**. *In proceedings of the 11th International Conference on Extending Database Technology (EDBT 2008), Nantes, France, Mar 2008* **(5 citations)**.
- Sarvjeet Singh, C. Mayfield, R. Shah, S. Prabhakar, S. Hambrusch, J. Neville and R. Cheng. **Database Support for Probabilistic Attributes and Tuples**. *In proceedings of the IEEE International Conference on Data Engineering (ICDE 2008), Cancun, Mexico, Apr 2008* **(48 citations)**.

- Sarvjeet Singh, C. Mayfield, S. Prabhakar, R. Shah and S. Hambrusch. **Indexing Uncertain Categorical Data**. *In IEEE 23rd International Conference on Data Engineering (ICDE 2007), Istanbul, Turkey, Apr 2007* **(80 citations)**.
- R. Cheng, Sarvjeet Singh, S. Prabhakar, R. Shah, J. Vitter and Y. Xia. **Efficient Join Processing over Uncertain Data**. *In proceedings of the ACM 15th Conf. on Information and Knowledge Management (ACM CIKM 2006), Arlington, USA, Nov 2006* **(58 citations)**.
- Reynold Cheng, Sarvjeet Singh and Sunil Prabhakar. **U-DBMS: A Database System for Managing Constantly-Evolving Data**. *In Very Large Databases Conf. (VLDB 2005 Demo), Trondheim, Norway, Aug 2005* **(76 citations)**.
- M. Surve, Sarvjeet Singh and P. Bhattacharyya. **Agro-Explorer: A Meaning based Multilingual Search Engine**. *International Conference on Digital Libraries (ICDL), New Delhi, India, Feb 2004* **(2 citations)**.
- P. Bhattacharyya, Sarvjeet Singh, U. Gundevia, U. Misra and T. Chandra. **A Multi-Lingual, Meaning Based Search Engine**. *International Conference on Universal Knowledge and Language, Goa, India, Nov 2002.*
- S. Mukherji, K.S. Rao and Sarvjeet Singh. **Registering SAR Interferometric images using cross-correlation**. *National symposium on recent advances in Remote Sensing and GIS technologies for Natural Resource Management, Mumbai, India, Dec 2001.*

Selected Projects

Please visit `http://sarvjeet.com/projects.shtml` for more details about these projects.

- **Orion: A Database System for Managing Uncertain Data**
  Orion is a DBMS for handling imprecise or uncertain data. This kind of data arise naturally in many applications such as sensors, data integration, bio-informatics etc. Orion is an extension of PostgreSQL and is written in C and PL/pgSQL. I contributed significantly to the theory, core system and indexing structures for Orion.

- **Zecosystem: An infrastructure for context-aware neural environments**
  The aim of this project was to provide users/students with an intelligent, ubiquitous digital environment for education. As part of this project, I developed Cyber Infrastructure (CI) Portal, which is a client/server application that allows a Mobile phone/PDA to control the resources offered by a CI Gateway. I developed several applications inside CI Portal for demo purposes. Technologies used: J2ME/Java, Bluetooth, Avetana, Java Native Interface, X11 Events.

- **Multilingual and Meaning Based Search Engine**
  I developed a complete multilingual meaning based search engine system with modules like crawler, preprocessor, search, page rank etc. Initially, it supported English, Hindi and Marathi

languages. The search was carried on using the meaning of the keywords entered as opposed to keyword based search engines like Google, giving it a broader applicability and greater accuracy than existing models. The development environment was Linux/Unix and Windows using C++, MYSQL database, mysql++ (library), shell scripting, PHP and HTML.

- **Phoebes: An Artificial Creativity Engine for Generating Music**
  Developed a new model for creativity, based on Minsky's model. Specifically for the domain of music, we developed Phoebes in Java, an engine that can be trained with music samples and then asked to generate music on its own, based on a consensus between several mindless agents, each one representing a particular aspect of music.

- **Distributed Banking System**
  Implemented a multi-threaded distributed banking system in Java/RMI supporting distributed transaction processing, concurrency control, logging and recovery for distributed database. Two phase commit was implemented as the final distributed commitment protocol. The distributed nature of transactions coupled with the possibility of failure of any server at any point of time made it a challenging project.

- **An Online Stock Exchange**
  As the conventional trading floor system disappears from Stock Exchanges, an online approach is adopted to manage the massive amount of data and flow a stock exchange has to handle. We designed a complete system that distributes services over the internet and allows a huge number of concurrent biddings. The system then deals between the users and updates its massive database of equity ownership. Also implemented are safety features and other real life stock market mechanisms.

- **GUMNAM: A two-way anonymous communication protocol**
  The aim of this project was to design and implement a two way anonymous communication protocol. In this protocol, unlike the other protocols used for anonymity, identity of both the sender and receiver is to be protected. As, a communication cannot be initiated without the knowledge of the second party involved in it, we use a concept of public handle to identify the party that we are trying to reach. We used a semi-trusted third party for this protocol and found a minimal trust model for the communication between entities. This project was implemented in C using Remote Procedure Calls (RPC).

- **Voronoi Diagrams for Moving Object Databases**
  Many algorithms are proposed for constructing and maintaining Voronoi Diagrams in static or small dynamic environments. Moving object databases pose additional challenges such as huge number of objects (which cannot fit in memory) and arbitrary object movements which make the current algorithms unsuitable for these databases. Constructing and maintaining a Voronoi diagram over moving objects enable us to answer many queries such as Nearest-Neighbor (NN), Reverse Nearest-Neighbor (R-NN) etc. very efficiently. As part of this

project, I designed and implemented an efficient algorithm for constructing and maintaining a Voronoi Diagram over moving object databases.

- **Library for threshold signatures**
  We implemented a library for RSA Threshold Signatures based on protocol proposed by Shoup in Practical Threshold Signatures. Threshold signature schemes enable a group of n entities share a private signature key in such a way that, for some parameter k, any subset of k entities can collectively create a valid signature on a message, whereas any collection of k-1 or fewer entities cannot. To the best of our knowledge there is no open source Threshold Signature toolkit available. This toolkit was developed in C using OpenSSL Crypto Library.

- **A Distributed Peer to Peer File Sharing System**
  Developed a Java driven Linux based system for decentralized sharing of files on a Peer to Peer basis. Emulating the paradigm of Network Neighborhood of the Windows platform, the system was extensible enough to identify new hosts wanting to join the community. It also supported an intelligent file search for fishing through files of peers and a smart download management approach to optimize multiple downloads.

Professional Activities

- Program committee member for ICDE 09 demo track.
- External Reviewer for many conferences including SIGMOD 07, VLDB (05, 06, 07), ICDE (05 - 09), DASFAA (05 - 08), ACM SAC (05, 06), EDBT 09, COMAD 05, MobiDE 05, SSDBM 07 and SSTD 05
- Computer Science Graduate Committee representative and Webmaster for Graduate Student Board, Purdue University (2005 - 06)
- Treasurer for Upsilon Pi Epsilon (Computer Science Honors Society) (2005 - 06)

Computer Skills

- **Programming languages**: Proficient in C/C++, Python, Java, Unix shell programming. Working knowledge of Perl, Scheme, LISP, ML, Ruby, Pascal, Fortran, VHDL, 8085/86 Assembly language
- **Operating Systems**: Unix/Linux, Windows, Mac OS X
- **Databases**: SQL, PostgreSQL, PL/pgSQL, Oracle 9i, XPath, XQuery, JDBC
- **Miscellaneous**: Map reduce, J2EE and EJB, PHP, Servlets, OpenGL, HTML, XML/XSL, Java Script, CSS, J2ME, BREW, Bison/Yacc, Flex

Honors and Awards

- Awarded the Bilsland Dissertation Fellowship for academic excellence by the Department of Computer Science, Purdue University, 2008-2009
- Selected by College of Science, Purdue University for Applied Management Principles (AMP) program offered by Krannert Graduate School of Management, 2007
- Recipient of Verizon Foundation Scholarship for academic excellence, 2005-06
- Member of Upsilon Pi Epsilon (Computer Science Honors Society)
- Awarded Misys Foundation Fellowship for academic excellence in IIT, 2000-03
- Received Merit Award for academic excellence from IIT Bombay for the academic year 1999-2000
- Awarded travel grant for attending DMSN 05 Workshop held in Trondheim, Norway
- Ranked among the top 0.1% of about 150,000 students appearing for the IIT JEE-99
- Recipient of National Talent Scholarship, India, 1997
- Scored 97% in National (Indian) Mathematics Olympiad
- Ranked 10/700 in Bitwise 2K++ Algorithm Intensive Programming Contest (IIT Kgp 2001)