

Zheguang Zhao

Brown University
Box 1910, Department of Computer Science
115 Waterman St
Providence, RI 02912
United States of America

Phone: +1-608-630-1677
Email: zheguang.zhao@gmail.com
Homepage: zheguang.github.io
LinkedIn: www.linkedin.com/in/samuelzhao
Google Scholar: goo.gl/DR8pSa

Education

Ph.D. Candidate in Computer Science, Brown University, Rhode Island, expected 2020.
Advisor: Prof. Seny Kamara and Stan Zdonik.
Thesis: Relational database encryption

M.S. in Computer Science, Brown University, Rhode Island, 2016.
Advisor: Prof. Stan Zdonik

B.S. in Computer Science, University of Wisconsin at Madison, Wisconsin, 2012.
Advisor: Prof. Jignesh Patel

Experiences

Los Alamos National Laboratory, NM, Research Intern, 2019 – present. Mentor: Dr. Andrey Lokhov.

Model reconstruction of cyber-physical systems via machine learning, with application to IoT, network infrastructure verification and security.

Aroki Systems, RI, Database Scientist, 2018 – present.

Provably-secure end-to-end database encryption for PostgreSQL and Apache Spark.

Microsoft AI & Research, WA, Research Intern, 2017. Mentor: Kris Ganjam.

Constraint learning for automatic puzzle solving AI.

Intel Labs, CA, Research Intern, 2015. Mentor: Dr. Dulloor Rao.

Efficiency of machine learning algorithms in Apache Spark.

In-memory transactional database VoltDB using non-volatile memory.

Hadapt (Acquired by Teradata), MA, Software Engineer, 2013 – 2014.

Develop the enterprise SQL-on-Hadoop system including query execution, storage engine, high availability and analytics toolkit. Use Agile methodology and continuous integration.

Kosmix (Acquired by @WalmartLabs), CA, Software Engineer Intern, 2012.

Develop an in-memory distributed queue system for the in-house distributed stream processing system in support for data analytics and machine learning.

Great Lakes Bioenergy Research Center, WI, Software Engineer Intern, 2010 – 2012.

Develop biological data management system using .NET and Oracle database.

Research

My main research areas are in cryptography and machine learning. I am interested in both the theories and the applications to real-world problems.

Relational database encryption.

We study how to use Structure Encryption to encrypt SQL databases in the cloud. The challenge is to refrain from any modification of the database, but to design an *emulation* scheme that composes encrypted SQL operators that execute directly on the encrypted data. We leverage the relational database theory to design new cryptographic building blocks. We study query optimization on encrypted data. We develop our scheme using SparkSQL and interface with different database backends. Joint work with S. Kamara, T. Moataz and S. Zdonik at Brown University.

Learning of cyber-physical systems.

Cyber-physical systems are systems that have both continuous and discrete dynamics, where the continuous processes are modeled as differential equations and the discrete processes as finite automata. It is important in many areas from infrastructure networks and IoT to autonomous vehicles and smart homes. We study data-driven approach to reconstruct the hybrid automata of cyber-physical systems, and use the result to further study safety verification and anomaly detection in security settings. Joint work with A. Lokhov, D. Deka and N. Lemon at Los Alamos National Laboratory.

Articles

Behavior of Large Random Graph.

Z. Zhao, supervised by Prof. Paul Dupius,

Randomized Algorithms for Counting, Integration and Optimization, Brown University, April 2017.

Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis.

E. Zgraggen, Z. Zhao, R. Zeleznik, and T. Kraska,

CHI, April 2018.

Signal Search.

J. Engelman, S. Kamara, T. Moataz and S. Zhao,

Software release: <http://github.com/encryptedsystems/Searchable-Signal-Android>.

Press release: <http://esl.cs.brown.edu/blog/signal>, April 2017.

Controlling False Discoveries During Interactive Data Exploration.

Z. Zhao, L. De Stefani, E. Zgraggen, C. Binnig, E. Upfal and T. Kraska,

SIGMOD, May 2017.

Safe Visual Data Exploration.

Z. Zhao, E. Zgraggen, L. De Stefani, C. Binnig, E. Upfal and T. Kraska,

SIGMOD Demo, May 2017.

Bridging the Gap between HPC and Big Data frameworks.

M. Anderson, S. Smith, N. Sundaram, M. Capota, Z. Zhao, S. Dulloor, N. Satish and T. Willke,
VLDB, 2017.

Towards Sustainable Insights.

C. Binnig, L. De Stefani, T. Kraska, E. Upfal, E. Zgraggen and Z. Zhao,

CIDR, January 2017.

Towards a Benchmark for Interactive Data Exploration.

P. Eichmann, E. Zraggen, Z. Zhao, C. Binnig, T. Kraska.

IEEE Data Engineering Bulletin, 2016.

Larger-than-memory Data Management on Modern Storage Hardware for In-memory OLTP Database Systems.

L. Ma, J. Arulraj, S. Zhao, A. Pavlo, S. Dullloor, M. Giardino, J. Parkhurst, J. Gardner, K. Doshi and S. Zdonik,

SIGMOD DaMoN, June 2016.

VisTrees: Fast Indexes for Interactive Data Exploration.

M. El-Hindi, Z. Zhao, C. Binnig and T. Kraska,

SIGMOD HILDA, June 2016.

Data Tiering in Heterogeneous Memory Systems.

S. Dullloor, A. Roy, Z. Zhao, N. Sundaram, N. Satish, R. Sankaran, J. Jackson and K. Schwan,

EuroSys, April 2016.

Open-source Projects

Searchable encryption for mobile messaging in Signal

<https://github.com/encryptedsystems/Searchable-Signal-Android>

Macau: statistical hypothesis testing based on resampling

<https://github.com/zheguang/macau>

Machine learning algorithms in Spark

<https://github.com/zheguang/spark-study/tree/master/study/src/main/scala/edu/brown/cs/sparkstudy>

Consistency control for machine learning algorithms

<https://github.com/zheguang/babel>

R-tree in Rust

<https://github.com/zheguang/rtree>

Spark performance analysis tool

<https://github.com/zheguang/spark-perftool>

VoltDB on non-volatile memory

<https://github.com/zheguang/voltdb>

Certification

Deep Learning Specialization, Coursera / [deeplearning.ai](https://www.coursera.org/deeplearning)

Neural Networks and Deep Learning

Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization

Structuring Machine Learning Projects

Convolutional Neural Networks

Sequence Models

Honors

Eta Kappa Nu

Upsilon Pi Epsilon

Golden Key International Honour Society

Selected Coursework

Abstract Algebra, Prof. Rich Schwartz

Calculus, Prof. Donald Passman, Gheorghe Craciun

Randomized Algorithms for Counting, Integration and Optimization, Prof. Paul G. Dupuis

Cryptography, Prof. Seny Kamara, Joseph Silverman

Probability, Prof. Erik Sudderth, Samuel S. Watson

Computational Linguistics, Prof. Eugene Charniak

Computer Architecture, Prof. Sherief Reda, Mark D. Hill

Distributed Computing through Combinatorial Topology, Prof. Maurice Herlihy

Database Management, Prof. Stan Zdonik, Jignesh Patel, Chris Ré

Microprocessor Synchronization, Prof. Maurice Herlihy

Algorithms and Data Structures, Prof. Eric Vigoda, Ben Liblit

Operating Systems, Prof. Michael Swift

Computer Networks, Prof. Aditya Akella

Physics, Prof. Peter Timbie, Daniel Chung, Ellen Zweibel

Reference

Prof. Stanley Zdonik, Professor at Brown University

Prof. Seny Kamara, Professor at Brown University

Dr. Emanuel Zgraggen, Postdoctoral associate at MIT

Dr. Andrey Lokhov, Los Alamos National Laboratory