

Lei Zhang

Website: geraldleizhang.com
Tel: +1 (404) 662-8268
Email: geraldleizhang@gmail.com

EDUCATION

Emory University, Atlanta

- Ph.D. student in Computer Science Aug 2018 – May 2021 (Expected)
 - Advisor: Prof. Ymir Vigfusson

Georgia Institute of Technology, Atlanta

- Master of Science (M.S.) in Computer Science Aug 2015 – Aug 2018
 - Once Ph.D. candidate. Advisor: Prof. Karsten Schwan

Tsinghua University, Beijing

- Bachelor of Engineering (B.E.) in Computer Science and Technology Aug 2011 – Jul 2015
 - **Thesis:** Panku: a matrix-centered distributed computation framework

RESEARCH INTERESTS

Distributed In-Memory Caching, Distributed Systems, Storage Systems

RESEARCH EXPERIENCE

Formalized Data Placement Optimization in Multi-cloud Storage

- Triones is an erasure-code based systematic model, focusing on formalized data placement optimization in multi-cloud storage where current methods can only fulfill ad-hoc cloud storage requirements
- We propose a non-linear programming model for data placement configuration with complex requirements
- We design and implement a set of APIs for automatically generating optimized multi-cloud storage configurations

Sparse Matrix Partition in Graph Processing

- Panku is a matrix-centered distributed computation framework, aiming at optimization of distributed storage for real-world large scale computing tasks like graph processing and machine learning, where data are always expressed in sparse matrix format
- I design a sparse matrix partition algorithm which takes both local and global storage optimization into consideration
- I implement the proposed computation framework, which achieves trade-off among computational skew, single node memory consumption, and network traffic in multi-core computing environment, providing up to 2 times improvement on overall computing performance

Developing Message Transport Middleware Layer Transaction in HPC Environment

- EVpath is designed to be an event transport middleware layer, which is maintained by CERCS Gatech
- I implement a publish/subscribe library to expand transaction module
- I experiment on expanding EVpath's scalability over InfiniBand, working on fabric communication frameworks like Libfabric and testing message passing and memory management component on EVpath

Distributed Storage for Sensitive Data

- This work aims at achieving distributed, secret shared, and obfuscated data storage, which can provide trade-off between security of sensitive data and storage resources.
- I design and implement two novel schemes for deception within both XOR secret sharing and polynomial-based threshold secret sharing that jointly encode multiple secrets, together with quantitative security analyses for these schemes
- I build a distributed prototype implementing proposed schemes, achieving notable enhanced security with limited and reasonable infection on system performance including storage space, latency, and throughput

Understanding Causality Between System and Network in Cybersecurity

- We are trying to find causality between system behaviors and network behaviors, for achieving malware attribution and for further purposes like actor, tool, and intent identification
- I parse raw system and network logs for a real malware execution dataset, which includes clearing up useless traces, and identifying unique system patterns for further defining malware's unique features
- I design a new LSM-Tree liked data structure for effectively storing large scale malware execution traces and efficiently clustering malware families

Understanding Distributed Cache Warm Up Process

- We want to investigate how long a cache takes to become useful after being suppressed or lost cache items
- I give a definition of a cache server getting warmed-up by cache hit rate, and with the definition provide a formula to estimate the cache warm-up time.
- I measure the accuracy of provided formula through real world CDN traces from Akamai, non-CDN traces from CloudPhysics, and file system cache traces from MSR. Experimental results show that our provided formula can give an accurate estimation of cache warm-up time for most of the traces.

WORK EXPERIENCE	Research Assistant , Georgia Tech Teaching Assistant , Gatech CS 3210: Design Operating Systems Software Engineer Intern , Facebook Inc. Research Assistant , Emory University	Aug 2015 – Aug 2018 Aug 2016 – Dec 2016 May 2018 – Aug 2018 Aug 2018 – present
PUBLICATIONS	Maomeng Su, Lei Zhang , Yongwei Wu, Kang Chen, Keqin Li, Systematic Data Placement Optimization in Multi-Cloud Storage for Complex Requirements, <i>IEEE Transactions on Computers (TOC)</i> , 2016, Vol. 65, Issue 6, pp.1964-1977 Lei Zhang , Douglas M. Blough, Deceptive Secret Sharing. <i>Dependable Systems and Networks (DSN)</i> , 2018 48th Annual IEEE/IFIP International Conference on. IEEE, 2018 Lei Zhang , Juncheng Yang, Anna Blasiak, Mike McCall, Ymir Vigfusson, How Long Does it Take to Warm up a Distributed Cache, Under preparation Reza Karimi, Lei Zhang , Gregory Chockler, Ymir Vigfusson, FIREHOSE: Highly Scalable Publish-Subscribe, Under preparation	
PROFESSIONAL SERVICE	External reviewer for SRDS 2018, SOCC 2018, EuroSys 2019	
OTHERS	Bronze medal, 24th, 25th China Mathematical Olympiad Experienced in C++, C, C#, Python, GoLang Familiar with Linux Kernel, Parallel Programming, Assembly Programming, LLVM, Apache Kafka, MySQL, MongoDB	