Zong-Xian Shen

6-Year Experience of Cybersecurity Engineering, Distributed System, and Compiler Optimization
zongxias@andrew.cmu.edu | (650) 770-5999 | https://zsshen.com | https://github.com/ZSShen

EDUCATION

Carnegie Mellon University — **Mountain View, CA**
Master of Science in Software Engineering — Expected Dec. 2020
National Chiao Tung University — **Hsinchu, Taiwan**
Master of Science in Computer Science — Jul. 2011 – Jul. 2013

PROFESSIONAL EXPERIENCE

Appier Inc. — **Taipei, Taiwan**
System Engineer — Nov. 2016 – Jun. 2019

- Developed a new real-time bidding system by implementing high-performance components, such as thread pool, load balancer, and message queue via C++ 14/17, boosted the queries-per-second benchmark by 20 times.
- Developed a unified pipeline using Ansible, PostgreSQL, and Prometheus to synchronize immense volumes of data between Appier's inventory servers, reduced data syncing time by 800 times and achieved zero data loss rate.
- Developed a monitoring and alert system via Kafka and Grafana to monitor the performance and the money flow of inventory servers, successfully reduced the alert response time from a few hours to less than 10 minutes.
- Scaled up the real-time bidding business by integrating 15 new ad publishers in North America and East Asia and supported partners for spec migration, which improved bidding traffic by 12.7% and boosted business revenue.

Trend Micro Inc. — **Taipei, Taiwan**
Software Engineer and Cybersecurity Researcher — Sep. 2013 – Nov. 2016

- Developed a distributed sandbox system using Intel Pin to identify malicious behavior and Cuckoo sandbox for large-scale instrumentation, contributing signatures with 50 times of detection coverage larger than fuzzy hash.
- Refactored the test framework of the virus scan engine by simplifying engine and pattern verification flow and adopted Docker and Kubernetes to balance workload, reducing the test duration from 24 hours to less than an hour.
- Designed a machine learning scheme for .NET malware detection via XGBoost and bytecode n-gram, yielding 98% detection rate for 5 million malware samples and 0.1% false positive rate for 5 million normal files.
- Developed an anti-malware platform by integrating static and dynamic analyzers for discovering new malware attacking trends; helped gain a competitive edge in the market by identifying the future product design trend.

Mozilla — **Taipei, Taiwan**
Compiler Developer and Open Source Contributor — Oct. 2014 – Aug. 2015

- Optimized the Just-in-Time compiler of SpiderMonkey JavaScript engine via data flow analysis to eliminate redundant computation, contributing to features deployed in Firefox 35 to enhance web browsing experience.

National Chiao Tung University — **Hsinchu, Taiwan**
Research Assistant cooperating with Industrial Technology Research Institute — May. 2012 – Jun. 2013

- Developed a real-time record-and-replay system on Android to monitor high-risk apps by tracking critical app events and decoupling analysis modules in the cloud, leading to merely 8% of CPU overhead on real devices.

OPEN SOURCE PROJECTS

- **ProbeDroid** (C++, Java | Android Runtime | Debugger) – Dynamic binary instrumentation kit to profile runtime performance, explore code coverage, and track high-risk behaviors of Android apps on Android 5.0 and above.
- **YADD** (C++ | Android Reverse Engineering | Compiler Technique) – Fast static Android app disassembler for DEX class and method signature extraction and code structure and information flow visualization.
- **MeltingPot** (C | Feature Extraction | Malware Clustering) – Tool to slice and cluster similar executables, such as PEs and ELFs, extract common binary signatures, and generate Yara patterns for malware family detection.
- **LibCDS** (C | Language Utility) – Fast and memory efficient cross-platform C library to manipulate data structures, such as vector, hash map, tree map, and trie with complete unit tests and documents.

TECHNICAL SKILLS

Languages: C++ 14/17 (advanced), C (advanced), Python, Java, JavaScript, SQL, Intel x86/64 assembly
Platform/Toolkits: Cassandra, PostgreSQL, Prometheus, Kafka, Grafana Ansible, Docker, Kubernetes, CMake, GDB