PEIPEI WANG

◇ *Phone:* (919) 592-2485 ◇ *E-mail:* wangpeipei.90@gmail.com
◇ *Homepage:* `https://wangpeipei90.github.io`
◇ *Google Scholar:* `http://tiny.cc/PeipeiWangScholarProfile`

EDUCATION

North Carolina State University (NC State) — Raleigh, NC
Ph.D. in Computer Science — 2020(expected)
Advisor: Dr. Kathryn T. Stolee

Xi'an Jiaotong University (XJTU) — Xi'an, China
Master of Science in Computer Science — 2013
Bachelor of Engineering in Software Engineering — 2010

SKILLS AND QUALIFICATIONS

Language: proficient in Java, Python, R scripts, familiar with C/C++, Shell script, JavaScript, and JSP; (Passed LinkedIn skill assessment of Python and Java.)
OS: Linux, Windows;
Data Processing: Aapche Hadoop, Apache Spark, GraphQL;
Cloud Systems: Docker, Cassandra, MySQL, Apache HTTP server, Squid, Memcached;
Build & SCM Tools: Make, Git, Maven, SVN.

INTERNSHIP AND RESEARCH EXPERIENCE

NSF founded research project, NC State — Research Assistant
Java Microbenchmark Harness (JMH) — *Jan 2020 - Present*

- **Performance Evaluation of Regular Expression Usage.** Measure the performance of different ways of data processing with regular expressions, and also compare them with alternative solutions implemented by string operations.

NSF founded research project, NC State — Research Assistant
Regular Expression Bugs, Fix Complexity, API, Bad Smells, Pull Requests, GraphQL — *Jan 2019 - Jan 2020*

- **An Empirical Study on Regular Expression Bug Characteristics.** Studied GitHub pull requests from Java, JavaScript and Python repositories to understand why changes are made to regular expressions. Measured the complexity of regular expression bug fixes. Found that regular expression bugs are caused by not only bad regexes but also bad smells in composing regular expressions and the usage of regex APIs.

NSF founded research project, NC State — Research Assistant
Regular Expression, Maven, Java Instrumentation, Testing Coverage, DFA, re2 — *Jan 2017 - Jan 2018*

- **Measuring regular expression testing coverage.** Instrumented GitHub Java maven projects and collected regular expressions and its inputs in test suites which are used to measure regular expression testing coverage via transformed DFA metrics. Found that the testing coverage of regular expression is low and lack of negative inputs.

Performance Regression, Facebook, Inc. — Software Development Intern
Cherry-Pick, Fblearner Workflow, Breakage — *June, 2020 - Aug, 2020*

- **Automated cherry-pick manager.** Automated the process of discovering pairs of commits that introduces breakages and fix breakages, in other words, cherry-picks, which will be applied to run experiments. This project enables not only cherry-picks that developers manually put into the performance regression experiment pipelines but also automatically manage them in databases through analyzing the history information of successful and failed experiments in the incident tracker. The automation is extended to handle not just one single breakage, but also consider multiple breakage scenarios.

DevOps Insights, IBM Inc. — Data Scientist Intern
GitHub, Mining Software Repositories, Rosie Pattern Language, Spark — *June, 2018 - Aug, 2018*

- **The importance of mining domain-specific knowledge.** Identified string literals in thousands of selected GitHub projects by leveraging Rosie pattern language, which is an alternative of regular expression language, and Apache Spark for data processing. The analysis of those string literals demonstrated the difficulty of avoiding hard-coding and the security vulnerabilities brought in by hard-coding string literals.

Morgan Stanley Inc. (Shanghai, China) Technology Intern
Library Dependency, d3.js, Graph Cycle Detection *Jul, 2012 - Sep, 2012*

- **Visualized library dependencies among projects in JavaScript.** Developed a Web UI to plot dependency paths/cycles and library conflicts between projects with d3.js, a JavaScript library for producing dynamic, interactive data visualizations in web browsers.

National Fundamental Software Project (equivalent to NSF), Xi'an Jiaotong University Research Assistant
Java EE, EJB, Middleware, AOP, JVM, Remote Method Invocation, Classloader *Jul, 2010 - May, 2012*

- **Java EE5 Application Server Development and Optimization.** Developed a JAVA EE Application Server to run EJB3 web applications according to the JAVA EE 5 and EJB 3 specifications. Specifically implemented the AOP(Aspect-oriented programming) feature with Java method reflection and conducted code optimization, code refactoring and software test of this product.

PUBLICATIONS

Regular Expression Analysis in Software Engineering Testing, Repair, Comprehension, and Maintenance

- "An Empirical Study on Regular Expression Bugs". Peipei Wang, Chris Brown, Jamie Jennings, Kathryn T. Stolee. *International Conference on Mining Software Repositories (MSR), 2020*.
- "Exploring Regular Expression Evolution". Peipei Wang, Rui Bai, Kathryn T. Stolee. *IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), 2019*. ☞
- "How Well Are Regular Expressions Tested in the Wild?" Peipei Wang, Kathryn T. Stolee. *Symposium on the Foundations of Software Engineering (FSE), 2018*. ☞
- "Exploring Regular Expression Comprehension". Carl Chapman, Peipei Wang, and Kathryn T. Stolee. *International Conference on Automated Software Engineering (ASE), 2017*. ☞

Distributed System Failure Diagnosis and Security Isolation, Performance, Bug Reproduction

- "DScope: Detecting Real-World Data Corruption Hang Bugs in Cloud Server Systems". Ting Dai, Jingzhu He, Xiaohui Gu, Shan Lu, and Peipei Wang. *ACM Symposium on Cloud Computing (SOCC), 2018*. ☞
- "Hytrace: A Hybrid Approach to Performance Bug Diagnosis in Production Cloud Infrastructures". Ting Dai, Daniel Dean, Peipei Wang, Xiaohui Gu, Shan Lu. *IEEE Transactions on Parallel and Distributed Systems (TPDS), 2018*. ☞
- "RDE: Replay DEbuggging for Diagnosing Production Site Failures". Peipei Wang, Hiep Nguyen, Xiaohui Gu, Shan Lu. *IEEE International Symposium on Reliable Distributed Systems (SRDS), 2016*. ☞
- "A Study of Security Isolation". *ACM Computing Surveys (CSUR), 2016*. Rui Shu, Peipei Wang, Sigmund A. Gorski III, Benjamin Andow, Adwait Nadkarni, Luke Deshotels, Jason Gionta, William Enck and Xiaohui Gu. ☞
- "PerfCompass: Online Performance Anomaly Fault Localization and Inference in Infrastructure-as-a-Service Clouds". Daniel Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu, Anca Sailer, Andrzej Kochut. *IEEE Transactions on Parallel and Distributed Systems (TPDS), 2015*. ☞
- "Automatic Server Hang Bug Diagnosis: Feasible Reality or Pipe Dream?". Daniel Dean, Peipei Wang, Xiaohui Gu, William Enck, Guoliang Jin. *IEEE International Conference on Autonomic Computing (ICAC), 2015*. ☞
- "Understanding Real World Data corruption Bugs in Cloud Systems". Peipei Wang, Daniel Dean, Xiaohui Gu. *IEEE International Conference on Cloud Engineering (IC2E), 2015*. ☞
- "PerfCompass: Toward Runtime Performance Anomaly Fault Localization for Infrastructure-as-a-Service Clouds". Daniel Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu. *USENIX Workshop on Hot Topics in Cloud Computing (HotCloud), 2014*. ☞

PATENT

[1] "An implement of EJB Container for AOP Based on dynamic stack". Yong Qi, Peipei Wang, Tao Yang, Yingyao Hao. China Patent, Publication Number: CN102508668 A, Application Number: CN201110357781.6. ☞