HONG JIN, KANG

Postdoctoral Fellow
University of California, Los Angeles

hjkang@g.ucla.edu / kanghongjin@gmail.com
https://kanghj.github.io

RESEARCH INTERESTS

My research interest is to improve developer productivity by investigating active learning and other ways of better incorporating expert knowledge into automated systems. In my PhD, I focused on bugs and vulnerabilities from using APIs and third-party code, e.g., libraries.

EDUCATION AND POSTDOCTORAL EXPERIENCE

Postdoctoral Fellow   *Jan 2023 – present*
University of California, Los Angeles

Ph.D in Computer Science   *Jan 2019 – Jan 2023*
Singapore Management University

Bachelor of Computing   *Aug 2012 – May 2016*
National University of Singapore
First Class Honours/Highest Distinction

PEER-REVIEWED CONFERENCE/JOURNAL PUBLICATIONS

Chronos: Time-Aware Zero-Shot Identification of Libraries from Vulnerability Reports
Yunbo Lyu, Thanh Le-Cong, **Hong Jin Kang**, Ratnadira Widyasari, Zhao Zhipeng, Bach Le, Ming Li, David Lo
IEEE/ACM International Conference on Software Engineering (ICSE 2023)

Compressing Pre-trained Models of Code into 3 MB
Jieke Shi, Zhou Yang, Bowen Xu, **Hong Jin Kang,** David Lo
IEEE/ACM International Conference on Automated Software Engineering (ASE 2022)
(**Nominated** for a **Distinguished Paper Award**)

Detecting False Alarms from Automatic Static Analysis Tools: How Far are We?
Hong Jin Kang, Khai Loong Aw, and David Lo.
IEEE/ACM International Conference on Software Engineering (ICSE 2022)
(**Nominated** for a **Distinguished Paper Award**)

Test Mimicry to Assess the Exploitability of Library Vulnerabilities.
Hong Jin Kang, Truong Giang Nguyen, Xuan-Bach D. Le, Corina Pasareanu, David Lo.
ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2022).

How to Find Actionable Static Analysis Warnings: A Case Study with FindBugs
Rahul Yedida, **Hong Jin Kang,** Huy Tu, Xueqi Yang, David Lo, Tim Menzies
IEEE Transactions on Software Engineering (TSE). Already accepted (Expected to be published in 2023)

AutoPruner: Transformer-based Call Graph Pruning
Le-Cong Thanh, **Hong Jin Kang,** Truong Giang Nguyen, Stefanus Agus Haryono, David Lo, Xuan-Bach D. Le, Huynh Quyet Thang.
ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022)

HERMES: Using Commit-Issue Linking to Detect Vulnerability-Fixing Commits
Truong Giang Nguyen, **Hong Jin Kang,** David Lo, Sharma Abhishek, Santosa Andrew, Sharma Asankhaya, and Ming Yi Ang.
IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER 2022)

Active Learning of Discriminative Subgraph Patterns for API Misuse Detection
Hong Jin Kang, David Lo.
IEEE Transactions on Software Engineering (TSE 2022)

Automated Identification of Libraries from Vulnerability Data: Can We Do Better?
Stefanus A. Haryono, **Hong Jin Kang**, Abhishek Sharma, Asankhaya Sharma, Andrew Santosa, Ang Ming Yi, David Lo
IEEE/ACM International Conference on Program Comprehension (ICPC 2022)

AndroEvolve: Automated Android API Update with Data Flow Analysis and Variable Denormalization
Stefanus A Haryono, Ferdian Thung, David Lo, Lingxiao Jiang, Julia Lawall, **Hong Jin Kang,** Lucas Serrano, Gilles Muller
Empirical Software Engineering (EMSE 2022)

Adversarial Specification Mining
Hong Jin Kang, David Lo.
ACM Transactions on Software Engineering and Methodology (TOSEM 2021)

IoTBox: Sandbox Mining to Prevent Interaction Threats in IoT Systems
Hong Jin Kang, Sheng Qin Sim, David Lo
IEEE International Conference on Software Testing, Verification and Validation (ICST 2021)

BiasFinder: Metamorphic Test Generation to Uncover Bias for Sentiment Analysis Systems
Muhammad Hilmi Asyrofi, Zhou Yang, Imam Nur Bani Yusuf, **Hong Jin Kang**, Ferdian Thung, and David Lo
IEEE Transactions on Software Engineering (TSE 2021)

CC2Vec: Distributed representations of code changes,
Thong Hoang, **Hong Jin Kang**, Julia Lawall, David Lo.
ACM/IEEE International Conference on Software Engineering (ICSE 2020)

Towards Generating Transformation Rules without Examples for Android API Replacement
Ferdian Thung, **Hong Jin Kang**, Lingxiao Jiang, David Lo.
IEEE International Conference on Software Maintenance and Evolution (ICSME 2019)

Assessing the Generalizability of code2vec Token Embeddings
Hong Jin Kang, Tegawende F Bissyandé., David Lo.
ACM/IEEE International Conference on Automated Software Engineering (ASE 2019)

Semantic Patches for Java Program Transformation
Hong Jin Kang, Ferdian Thung, Julia Lawall, Gilles Muller, Lingxiao Jiang, David Lo.
European Conference on Object-Oriented Programming (ECOOP 2019)

SHORT/WORKSHOP/DEMO PAPERS

VulCurator: A Vulnerability-Fixing Commit Detector
by Truong Giang Nguyen, Thanh Le-Cong, **Hong Jin Kang,** and Bach Le, David Lo
ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2022) – Demonstrations

AndroEvolve: Automated Update for Android Deprecated-API Usages
by Stefanus Agus Haryono, Ferdian Thung, David Lo, Lingxiao Jiang, Julia Lawall, **Hong Jin Kang,** Lucas Serrano, Gilles Muller
ACM/IEEE International Conference on Software Engineering (ICSE 2021) – Demonstrations

BugsInPy: a Database of Existing Bugs in Python Programs to Enable Controlled Testing and Debugging Studies
Ratnadira Widyasari, Sheng Qin Sim, Camellia Lok, Haodi Qi, Jack Phan, Qijin Tay, Constance Tan, Fiona Wee, Jodie Ethelda Tan, Yuheng Yieh, Brian Goh, Ferdian Thung, **Hong Jin Kang**, Thong Hoang, David Li, Eng Lieh Ouh
ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2020) – Tool Demo

Automatic Android Deprecated-API Usage Update by Learning from Single Updated Example
Stefanus Agus Haryono, Ferdian Thung, **Hong Jin Kang**, Lucas Serrano, Gilles Muller, Julia Lawall, David Lo, Lingxiao Jiang.
International Conference on Program Comprehension (ICPC 2020) - Early Research Achievements

A Comparison of Word Embeddings for English and Cross-Lingual Chinese Word Sense Disambiguation
Hong Jin Kang, Tao Chen, Muthu Kumar Chandrasekaran, Min-Yen Kan.
The COLING Workshop on Natural Language Processing Techniques for Educational Applications Workshop (NLP-TEA-3 2016)

PAPERS UNDER REVIEW

Active Learning-based Input Selection for Fuzzing Deep Learning Libraries
by **Hong Jin Kang**, Pattarakrit Rattanukul, Stefanus Agus Haryono, Truong Giang
Nguyen, Chaiyong Ragkhitwetsagul, Corina Pasareanu, David Lo
(This work found vulnerabilities leading to the assignment of 23 CVEs)

SERVICE
JOURNAL REVIEWER: EMSE, TOSEM, IST
SUB-REVIEWER: ICSME 2020, APSEC 2020, MSR 2020, ASE 2021,
ICST 2022, ICSE 2022
PROGRAM COMMITTEE: SANER ERA 2023

MENTORING EXPERIENCE

I had the privilege of working with other students/junior researchers, including

- Aw Khai Loong   undergraduate student
- Sim Sheng Qin   undergraduate student
- Jonathan Ooi   masters student
- Le Cong Thanh   research engineer, starting a PhD in Jan 2023
- Lyu Yunbo   research engineer, starting a PhD in Aug 2023
- Truong Giang Nyugen   research engineer
- Stefanus Agus Haryono   research engineer

AWARDS

2022 SMU Dean's List
2020, 2021, 2022 SMU's Presidential Doctoral Fellowship

GRANT EXPERIENCE

- **"TrustedSEERs: Trusted Intelligent Work Bots for Engineering Better Sofware Faster"**
  Assisted in writing part of the proposal for funding for the NRF Investigatorship from National Research Foundation (NRF), Singapore. S$3.2M was awarded.

- **"Making Big Code Active: From Billions of Code Tokens to Automation"**
  Assisted in writing a joint proposal between Singapore Management University and Zhejiang University. S$600,000 was awarded by the Singapore Data Science Consortium

- **"Uncovering Vulnerabilities in Machine Learning Frameworks via Software Composition Analysis and Directed Grammar-Based Fuzzing"**
  For this grant (S$550,000) awarded by the National Satellite of Excellence in Trustworthy Software Systems, I coordinated the work of all three work packages of the grant. I had the pleasure of mentoring several students and researchers.

WORK EXPERIENCE

JAN 2023 – PRESENT
POSTDOCTORAL FELLOW, UCLA

- Investigating methods of improving interactions between humans and machine learning approaches, e.g., through active learning
- Developing an interface to better involve expert feedback from human users to mine program patterns for detecting possible bugs

AUG 2018 – DEC 2018
RESEARCH ENGINEER, SMU

- Developed Coccinelle4J, a port of Coccinelle for the Java programming language. Coccinelle is a widely adopted program matching and transformation tool for C systems software, including the Linux kernel. Over 6000 commits have been in the Linux kernel with the help of Coccinelle in the past 10 years. We apply Coccinelle4J to the automatic migration of deprecated Android API usages.

2018
BACKEND ENGINEER, GRAB

- Worked with cutting-edge tools and managed cloud-based infrastructure.
- Owned a microservice on the critical path of all bookings. Improved testing to identify single points of failure, improving the resilience of the service to ensure uptime and SLAs are met.

2016 –2018
SOFTWARE ENGINEER, WORKS APPLICATIONS CO. LTD

- Worked on enterprise-grade HR Enterprise Resource Management software.
- Mentored junior engineers.
- Full-stack development from the back-end using Spring, management of Cassandra, Elasticsearch, to implementing a new user experience on the front-end with modern frameworks.

AUG 2015 – MAY 2016
STUDENT SOFTWARE ENGINEER, TEAMMATES @ NUS

- Involved in both planning and development of the project. Planned and managed weekly releases.
- Part of the team that introduced static analysis tools and CI to the project.

MAY 2014 – AUG 2014
INTERN, TEAMIE

- Developed several new features, including a feature that summarizes users' activity logs to identify highlights of their history on the platform.