Seongmin Lee

Max Planck Institute for Security and Privacy (MPI-SP)
Universitätsstraße 140
44799 Bochum
Germany

☎ +49 177 783 5480   ✉ seongmin.lee@mpi-sp.org   🏛 Google Scholar

Research Interests

My research interest lies in software testing and program analysis. The overarching objective of my research is **achieving practical software testing in real-world scenarios** by addressing the challenges associated with the scale and complexity of software systems, overcoming the limitations of formal methods, and taking account of the operational behavior of software systems. I **utilize statistical methods to analyze dynamic information from program execution**, facilitating the reasoning of a program's semantic properties and addressing empirical challenges in software testing.

Education and Employment

Max Planck Institute for Security and Privacy	Germany
Postdoctoral Researcher, Software Security Research group	Sep. 2022 – Present
Group head: Dr. Marcel Böhme	
Korea Advanced Institute of Science and Technology	Republic of Korea
Doctor of Philosophy, School of Computing	Sep. 2016 – Aug. 2022
Advisor: Dr. Shin Yoo	
Bachelor of Science, School of Computing	Feb. 2012 – Aug. 2016
Bachelor of Science, Department of Mathematical Sciences	

Publications

Preprints

- Seongmin Lee, Shreyas Minocha, and Marcel Böhme. Accounting for missing events in statistical information leakage analysis – *Under review with ICSE'25*

- Seongmin Lee and Marcel Böhme. How much is unseen depends chiefly on information about the seen, 2024 – *Under review with ICML'24*

- Seongmin Lee and Marcel Böhme. Predicting the number of new discoveries, 2024 – *Under review with ICML'24*

- Seongmin Lee, Dave W. Binkley, Robert Feldt, Nicolas Gold, and Shin Yoo. Causal program dependence analysis. 2021 – *Under review with Science of Computer Programming*

Peer-Reviewed Conference and Workshop Papers

- *Danushka Liyanage, *Seongmin Lee, Chakkrit Tantithamthavorn, and Marcel Böhme. Extrapolating coverage rate in greybox fuzzing. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, ICSE '24, New York, NY, USA, 2024. Association for Computing Machinery
  (*Co-first authors with equal contribution)

- Seongmin Lee and Marcel Böhme. Statistical reachability analysis. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ESEC/FSE 2023, pages 326–337, New York, NY, USA, 2023. Association for Computing Machinery

- Saeyoon Oh, Seongmin Lee, and Shin Yoo. Effectively sampling higher order mutants using causal effect effectively sampling higher order mutants using causal effect. In *The 16th International Workshop on Mutation Analysis*, 2021

- Seongmin Lee. Scalable and approximate program dependence analysis. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Companion Proceedings*, ICSE '20, pages 162–165, New York, NY, USA, 2020. Association for Computing Machinery

- S. Lee, S. Hong, J. Yi, T. Kim, C. Kim, and S. Yoo. Classifying false positive static checker alarms in continuous integration using convolutional neural networks. In *2019 12th IEEE Conference on Software Testing, Validation and Verification (ICST)*, pages 391–401, 2019

- S. Lee, D. Binkley, R. Feldt, N. Gold, and S. Yoo. Moad: Modeling observation-based approximate dependency. In *2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pages 12–22, 2019

- Seongmin Lee and Shin Yoo. Hyperheuristic observation based slicing of guava. In Tim Menzies and Justyna Petke, editors, *Search Based Software Engineering*, pages 175–180, Cham, 2017. Springer International Publishing

- Gabin An, Jinhan Kim, Seongmin Lee, and Shin Yoo. PyGGI: Python General framework for Genetic Improvement. In *Proceedings of Korea Software Congress*, KSC 2017, December 2017

- Jeongju Sohn, Seongmin Lee, and Shin Yoo. *Amortised Deep Parameter Optimisation of GPGPU Work Group Size for OpenCV*, pages 211–217. Springer International Publishing, Cham, 2016

Peer-Reviewed Journal Articles

- Seongmin Lee, David Binkley, Robert Feldt, Nicolas Gold, and Shin Yoo. Observation-based approximate dependency modeling and its use for program slicing. *Journal of Systems and Software*, page 110988, 2021

- Seongmin Lee, David Binkley, Nicolas Gold, Syed Islam, Jens Krinke, and Shin Yoo. Evaluating lexical approximation of program dependence. *Journal of Systems and Software*, 160:110459, 2020

Posters & Technical Reports

- Seongmin Lee, Dave Binkley, Robert Feldt, Nicolas Gold, and Shin Yoo. Causal program dependence analysis and causal fault localization. Technical Report CS-TR-2021-423, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, Korea 34141, January 2021

- Seongmin Lee, David Binkley, Nicolas Gold, Syed Islam, Jens Krinke, and Shin Yoo. Mobs: Multi-operator observation-based slicing using lexical approximation of program dependence. In *Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings*, ICSE '18, pages 302–303, New York, NY, USA, 2018. Association for Computing Machinery

- Seongmin Lee and Shin Yoo. Using source code lexical similarity to improve efficiency of Observation Based Slicing. Technical Report CS-TR-2017-412, School of Computing, KAIST, May 2017

Software Engineering Notes

- William B. Langdon, Westley Weimer, Justyna Petke, Erik Fredericks, Seongmin Lee, Emily Winter, Michail Basios, Myra B. Cohen, Aymeric Blot, Markus Wagner, Bobby R. Bruce, Shin Yoo, Simos Gerasimou, Oliver Krauss, Yu Huang, and Michael Gerten. Genetic improvement @ icse 2020. *SIGSOFT Softw. Eng. Notes*, 45(4):24–30, October 2020

Academic Services

- Program committee: ISSTA'24, ASE'23 / (Artifact Evaluation) ISSTA'24, ECOOP'24, USENIX'24, ICSE'24, ISSTA'23, ICSME'22, ICSME'21 / (Student Research Competition) FSE'24

- Reviewer: TOSEM'24, TSE'24, IST'24, ASE'24, TOSEM'22, JSS'21, JSS'20 / (External) ICSE'24, FSE'24, ECOOP'24, ICSE'23, ISSTA'23

Grants and Fellowships

- Title: *Statistical Security Analysis for Large, Evolving Software*
  Funding Agency: CASA - Cyber Security in the Age of Large-Scale Adversaries
  Grant ID: DFG under Germany's Excellence Strategy - **EXC 2092 CASA - 390781972**
  Amount: Salary according to the remuneration group E 14 TV-L (full time, $\sim$ €65,000)
  Duration: 2024.01.01 – 2025.12.31

Invited Talks

- *Statistical Program Analysis*
  Korea Advanced Institute of Science and Technology (KAIST), 2024

- *Statistical Program Analysis*
  Ulsan National Institute of Science and Technology (UNIST), 2024

- *Causal Program Dependence Analysis*
  Sheffield Causality and Testing Workshop, 2023

- *Statistical program dependence analysis*
  Handong Global University, 2022

- *Observation-based approximate dependency modeling and its use for program slicing*
  Korea Conference on Software Engineering, 2022

- *MOBS: Multi-Operator Observation-Based Slicing using Lexical Approximation of Program Dependence*
  59th CREST Open Workshop - Multi-language Software Analysis, 2018

Awards and honors

- **PhD Dissertation Award**, School of Computing, KAIST, 2022
  - *Title of Dissertation: Statistical Program Dependence Approximation*

- **2021 Naver Ph.D. Fellowship Award**: Awarded by NAVER Corp. to Ph.D. candidates who have published an outstanding research paper or have excellent publication performance, 2021

- Government-sponsored Scholarship, Ministry of Science and ICT of Korea, 2016 - 2022

- Government-sponsored Scholarship, Ministry of Science and ICT of Korea, 2012 - 2016

Research Experience

Software Security Group, MPI-SP   **Sep. 2022 – Present**
Postdoc   *Bochum, Germany*

- Researching on unbiased estimation of the missing mass/probability/expected number of discovering new classes of unknown multinomial distribution
- Researched applying biostatistics for information leakage analysis to suggest the accurate and safe information leakage estimator – Statistical Information Leakage Analysis
- Researched on extrapolating the coverage rate of the Greybox Fuzzing using the statistical model – Greybox Fuzzing Extrapolation
- Researched applying statisical methods for program analysis to overcome the scalability issue of the static analysis – Statistical Reachability Analysis

Computational Intelligence for Software Engineering Laboratory (COINSE), KAIST   **Sep. 2016 – Aug. 2022**
Ph.D. Student   *Daejeon, Republic of Korea*

- Researched approximating the degree of dependence between program element using causal inference – CPDA
- Researched applying statistical models on the observation data to approximate the program dependence – MOAD
- Researched inferencing the type information in the binary executables using RNN with National Security Research Institute
- Researched classifying the false positive alarms from static checker in continuous integration pipeline using CNN with Samsung Research
- Researched program dependence approximation using the lexical model on the source code – MOBS

Computational Intelligence for Software Engineering Laboratory (COINSE), KAIST   **Mar. 2016 – Aug. 2016**
Undergraduate Research Intern   *Daejeon, Republic of Korea*

- Researched on the amortised deep parameter optimisation of GPGPU work group size for OpenCV.
- Accelerated the scalablility of Observation based slicing (ORBS) by applying a code distance metric during the slicing.

Programming Language Research Group (PLRG) Lab, KAIST   **Jul. 2015 – Feb. 2016**
Undergraduate Research Intern   *Daejeon, Republic of Korea*

- Developed a source code translator from C# to C++ with F#.
- Developed a frontend of Scalable Analysis Framework for ECMAScript (SAFE), a Javascript static analysis tool.

Teaching Experience

- Teaching Assistant, Automated Software Testing (CS453), School of Computing, KAIST, Spring 2019

- Teaching Assistant, Artificial Intelligence Based Software Engineering (CS454), School of Computing, KAIST, Fall 2018

- Teaching Assistant, Introduction to Logic for Computer Science (CS402), School of Computing, KAIST, Spring 2018

- Teaching Assistant, Artificial Intelligence Based Software Engineering (CS454), School of Computing, KAIST, Fall 2017

- Teaching Assistant, Introduction to Logic for Computer Science (CS402), School of Computing, KAIST, Spring 2017

- Teaching Assistant, Special Topics in Computer Science ⟨Search Based Software Engineering⟩ (CS492), School of Computing, KAIST, Fall 2016