Kihong Heo

Post-doctoral Researcher
Department of Computer and Information Science
University of Pennsylvania
3330 Walnut St, Philadelphia, PA 19104, USA
✉ : kheo@cis.upenn.edu : http://www.cis.upenn.edu/~kheo

Research Interests

I am interested in semantic-based static analysis for safe and reliable softwares. In particular,

- **Sound, Scalable & Precise Static Analysis**
  I have been developing techniques to achieve sound, scalable, yet precise static analysis in a single analyzer. The challenge has been handled by making the analysis compute only right part at right moment (sparse analysis) and apply expensive sensitivities only when they benefit (selective X-sensitivity). The key part of these approaches is to estimate the sparsity and the impact of the sensitivities. To this end, I have been working on statistical approaches (e.g., machine learning) as well as logical techniques (e.g., pre-analysis).

- **Data-driven Adaptive Program Analysis guided by Machine Learning**
  Thanks to the abundance of program code and analysis result, now it is possible to exploit machine learning techniques for improving the performance of program analyses. In particular, I am working on inferring effective abstractions using machine learning. For example, I try to infer an effective variable clustering strategy for selectively relational analysis and harmless unsoundness for selectively unsound analysis.

Education

Computer Science and Engineering, Seoul National University Mar 2009 – Aug 2017
Doctor of Philosophy in Computer Science and Engineering
Thesis: Selectively Sensitive Static Analysis by Impact Pre-analysis and Machine learning
Outstanding Dissertation Award
Advisor: Prof. Kwangkeun Yi

Computer Science and Engineering, Seoul National University Mar 2005 – Feb 2009
Bachelor of Science in Computer Science

Experience

University of Pennsylvania Jul 2017 – present
Post-doctoral Researcher
Advisor: Prof. Mayur Naik

Facebook Apr 2017 – Jun 2017
Research Scientist (contingent)

The Hong Kong University of Science & Technology Sep 2011 – Feb 2012
Visiting Student

Advisor: Prof. Sunghun Kim

Research Project

- **Inferbo: Infer-based buffer overrun analyzer** 2016 – present
  I have been a core developer of Inferbo that is a precise and scalable buffer-overrun analyzer based on the Facebook Infer analyzer. Inferbo scales to a large and fast-moving codebase like Facebook thanks to Infer's modular analysis engine. Modular analysis separately analyzes each sub-part (e.g., procedure) of a large program and composes the whole analysis result using the partial information. Despite its efficiency, modular analysis has previously been used for checking relatively simple or inductively-defined properties. To achieve a modular buffer-overrun analyzer which requires sophisticated numerical reasoning, we designed a symbolic abstract domain and procedure summary by observing some common buffer-overrun issues in Facebook's codebase. Inferbo is now merged in Facebook Infer and available via the Infer GitHub. (`https://github.com/facebook/infer`)

- **Sparrow: a static analyzer for C program** 2011 – present
  I have been a core developer of Sparrow and mainly involved in its relational analysis engine. Sparrow is a state-of-the-art static analyzer that aims to verify the absence of fatal bugs in C source. In particular, I have been developing techniques for cost-effective relational analysis, context-sensitive analysis, unsound analysis enabled by semantic-based pre-analysis and machine learning. Sparrow is now open-source and available via GitHub. (`http://www.github.com/ropas/sparrow`)

- **Selective X-sensitive Analysis** 2013 – present
  I have been working on the selective X-sensitive analysis framework and mainly developing selectively relational analysis. Selective X-sensitive analysis applies certain sensitivity X (e.g. context, flow, or relational analysis) only when and where doing so is likely to improve the precision of the main analysis. The challenge is to estimates the impact of X on the main analysis's precision. To this end, we have developed 1) impact pre-analyses that are based on the abstract interpretation framework, and 2) machine learning techniques that learn the behaviors of the impact pre-analyses. In the project, I mainly designed an impact pre-analysis and machine learning techniques for the octagon relational analysis.

- **Global Sparse Analysis Framework** 2011 – 2012
  I joined the sparse analysis project and designed the sparse interval analysis engine part. Our sparse analysis framework provides a general method for achieving global static analyzers that are precise, sound, yet also scalable. Based on the framework, we have derived a sparse version of Sparrow which is 175x more scalable than the baseline in terms of lines of code and scales to a million lines of C programs. In the project, I participated in designing a pre-analysis for the interval analysis and implementing the sparse interval analysis. (`http://ropas.snu.ac.kr/sparseanalysis`)

- **SNEC: Semantic-based Non-Essential Change Detection** 2011 – 2012
  I have developed SNEC, a semantic-based non-essential change detector. Non-essential change is a code change that does not alter the semantics such as refactoring. Abundant non-essential changes in software history have negative impacts on software mining tasks. SNEC identifies non-essential changes by observing semantic equivalence using a semantic-aware static analyzer. I designed and implemented SNEC based on a commercial static analysis engine for JAVA programs. (`http://ropas.snu.ac.kr/snec`)

Publications

1. Adapting Static Analysis via Learning with Bayesian Optimization
   Kihong Heo, Hakjoo Oh, Hongseok Yang and Kwangkeun Yi
   TOPLAS: *ACM Transactions on Programming Languages and Systems*, 2017 (to appear)

2. Automatically Generating Features for Learning Program Analysis Heuristics
   Kwonsoo Chae, Hakjoo Oh, **Kihong Heo**, and Hongseok Yang
   OOPSLA 2017: *ACM Conference on Object-Oriented Programming, Systems, Languages, and Applications*, 2017 (to appear)

3. Machine-Learning-Guided Selectively Unsound Static Analysis
   Kihong Heo, Hakjoo Oh, and Kwangkeun Yi
   ICSE 2017: *The 39th International Conference on Software Engineering*, 2017

4. Selective Conjunction of Context-sensitivity and Octagon Domain toward Scalable and Precise Global Static Analysis
   Kihong Heo, Hakjoo Oh, and Kwangkeun Yi
   SP&E: *Software-Practice and Experience*, 2017 (to appear)

5. Sound Non-Statistical Clustering of Static Analysis Alarms
   Woosuk Lee, Wonchan Lee, Dongok Kang, **Kihong Heo**, Hakjoo Oh, and Kwangkeun Yi
   TOPLAS: *ACM Transactions on Programming Languages and Systems*, 2017 (to appear)

6. Learning a Variable-Clustering Strategy for Octagon from Labeled Data Generated by a Static Analysis
   Kihong Heo, Hakjoo Oh, and Hongseok Yang
   SAS 2016: *The 23rd International Static Analysis Symposium*, 2016

7. Selective X-Sensitive Analysis Guided by Impact Pre-Analysis
   Hakjoo Oh, Wonchan Lee, **Kihong Heo**, Hongseok Yang, and Kwangkeun Yi
   TOPLAS: *ACM Transactions on Programming Languages and Systems*, Vol. 38, Issue 2, Jan. 2016

8. Widening with Thresholds via Binary Search
   Sol Kim, **Kihong Heo**, Hakjoo Oh, Kwangkeun Yi
   SP&E: *Software-Practice and Experience*, 46(10), 2016

9. Selective Context-Sensitivity Guided by Impact Pre-Analysis
   Hakjoo Oh, Wonchan Lee, **Kihong Heo**, Hongseok Yang, and Kwangkeun Yi
   PLDI 2014: *The 35th ACM SIGPLAN Conference of Programming Language Design and Implementation*, 2014

10. General Sparse Analysis Framework
    Hakjoo Oh, **Kihong Heo**, Wonchan Lee, Woosuk Lee, Daejun Park, Jeehoon Kang, and Kwangkeun Yi
    TOPLAS: *ACM Transactions on Programming Languages and Systems*, Vol. 36, Issue 3, Sept. 2014

11. A Sparse Evaluation Technique for Detailed Semantic Analyses
    Yoonseok Ko, **Kihong Heo**, and Hakjoo Oh
    Computer Languages, Systems, & Structures, Vol. 40, Issues 3–4, October–December 2014

12. Design and Implementation of Sparse Global Analyses for C-like Languages
Hakjoo Oh, **Kihong Heo**, Wonchan Lee, Woosuk Lee, and Kwangkeun Yi
PLDI 2012: *The 33rd ACM SIGPLAN Conference of Programming Language Design and Implementation*, 2012

Software

I have contributed to the following open-source software:

- Sparrow: a static analyzer for C program
`http://www.github.com/ropas/sparrow`
- Infer: a static analyzer for Java, C, C++, and Objective-C
`http://www.github.com/facebook/infer`

Talks

- Machine-Learning-Guided Selectively Unsound Static Analysis
Paper presentation. ICSE 2017. 05/26/2017
- Inferbo: Infer-based buffer-overrun analyzer
Invited talk. Korea University. 04/14/2017
- Inferbo: Infer-based buffer-overrun analyzer
Invited talk. KAIST. 03/24/2017
- Selectively Sensitive Static Analysis by Impact Pre-analysis and Machine Learning
Invited talk. Codemind. 02/20/2017
- Learning a Variable-Clustering Strategy for Octagon from Labeled Data Generated by a Static Analysis
Paper presentation. SAS 2016. 09/08/2016

Teaching Experience

- **Teaching Assistant** : SNU 4541.664 Program Analysis (grad) Spring 2010
- **Teaching Assistant** : SNU 4190.210 Programming Languages Spring 2009

References

Kwangkeun Yi
Professor
Dept. of Computer Science and Engineering
Seoul National University
Email: kwang@ropas.snu.ac.kr

Hongseok Yang
Professor
Dept. of Computer Science
University of Oxford
Email: hongseok.yang@cs.ox.ac.uk

Hakjoo Oh
Assistant Professor
Dept. of Computer Science and Engineering
Korea University
Email: hakjoo_oh@korea.ac.kr

Peter O'Hearn
Engineering Manager
Facebook, Inc.
Email: peteroh@fb.com

Last updated: August 23, 2017
http://www.cis.upenn.edu/~kheo