Carnegie Mellon, Microsoft Research Automate Privacy Compliance for Big Data Systems-Carnegie Mellon News - Carnegie Mellon University

Carnegie Mellon University

Search Only Carnegie Mellon News SEARCH
Submit Carnegie Mellon University Search

Skip navigation and jump directly to page content

Carnegie Mellon
News

- News Stories
  - All News
    - 2014
      - July
      - June
      - May
        - Press Release: Chemists Identify Key Step in Biosynthesis of Carbapenem Antibiotics
        - Press Release: Negative Social Interactions Increase Hypertension Risk In Older Adults, Carnegie Mellon Researchers Find
        - Press Release: Carnegie Mellon Researchers Find Spontaneous Thoughts Are Perceived To Reveal Meaningful Self-Insight
        - Press Release: Heavily Decorated Classrooms Disrupt Attention and Learning In Young Children, According To New Carnegie Mellon Research
        - News Brief: Jared L. Cohon Appointed as Co-Chair on the Commission to Review the Effectiveness of the National Energy Laboratories
        - Media Advisory: LaunchCMU To Feature Startups, New Technology and Research Innovations from Carnegie Mellon's

Entrepreneurship and Technology Community
- Carnegie Mellon Researchers Use Autonomous Airboats To Monitor Hippo Dung in Kenya's Mara River Basin
- Carnegie Mellon, Microsoft Research Automate Privacy Compliance for Big Data Systems
- Carnegie Mellon's Pittsburgh Queer History Project Presents its First Exhibit, "Lucky After Dark"
- Media Advisory: Shutterfly President and CEO Jeffrey Housenbold To Deliver Keynote Address at Carnegie Mellon Commencement, May 18
- Press Release: International Group of Researchers Shows Emissions From Forests Influence Very First Stage of Cloud Formation
- Press Release: Carnegie Mellon Will Test New Internet Architecture in Vehicular Network and for Delivering Online Video
- Press Release: Carnegie Mellon's Integrated Innovation Institute First To Cross-Train in Engineering, Design and Business
- Press Release: Carnegie Mellon Opens Science of Security Lablet with Support from National Security Agency
- Press Release: Carnegie Mellon's David S. Kaufer Selected as Rhetoric Society of America Fellow
- Press Release: Videos by Local High School Students Address Technology, Depression, Adversity
- News Brief: Carnegie Mellon Researchers Find Mandates for Renewable Energy Are Driving Wind-Related Patents
- Press Release: Carnegie Mellon Creates Interactive Map Showing Income Inequality Using Data Analyzed in Piketty's "Capital in the 21st Century"
- Press Release: Carnegie Mellon Partners With IBM To Offer Cognitive Computing Course Featuring Watson
- Press Release: “ Birth Marks ” by Carnegie Mellon ’ s Jim Daniels Wins Two More Prominent Awards
- News Brief: Cybersecurity Author and Alumnus Visits Campus for Book Signing
- Media Advisory: Carnegie Mellon Students Present Prototypes To Solve Medical Problems
- News Brief: Three Recent Hilary Masters' Pieces Land on Best American Essays' Notable Lists
- Press Release: Leaders in Business, Drama, Music and Technology Highlight Carnegie Mellon's Commencement, May 18

- April
- March
- February

Wednesday, May 21, 2014

Carnegie Mellon, Microsoft Research Automate Privacy Compliance for Big Data Systems

Search Engine Code Is Moving Target That Eludes Manual Audits

Contact: Byron Spice / 412-268-9068 / bspice@cs.cmu.edu

PITTSBURGH—Web services companies, such as Facebook, Google and Microsoft, all make promises about how they will use personal information they gather. But ensuring that millions of lines of code in their systems operate in ways consistent with privacy promises is labor-intensive and difficult. A team from Carnegie Mellon University and Microsoft Research, however, has shown these compliance checks can be automated.

The researchers developed a prototype automated system that is now running on the data analytics pipeline of Bing, Microsoft's search engine. According to Saikat Guha, researcher at Microsoft, it's the first time automated privacy compliance analysis has been applied to the production code of an Internet-scale system and is a reflection of Microsoft's commitment to creating the technology necessary to further safeguard the privacy of customers.

Employing a new, lawyer-friendly language to specify privacy policies and using a data inventory to annotate existing programs, the researchers showed that a team of just five people could manage a daily compliance check on millions of lines of code written by several thousand developers.

They presented their research findings at the 35th IEEE Symposium on Security & Privacy, May 18-21, in San Jose, Calif.

"Companies in the United States have a legal obligation to declare how they use personal information they gather and it's also good business to establish a bond of trust with customers," said Anupam Datta, associate professor of computer science and electrical and computer engineering. "But these systems are constantly evolving and their scale can be daunting. The manual methods typically used for checking compliance are labor intensive, yet too often fail to catch all violations of policy."

"Tens of millions of lines of code are already in the pipeline," noted Shayak Sen, a Ph.D. student in computer science who interned at Microsoft Research India and the lead student author on the study. "And during our implementation on Bing, we found that more than 20 percent of the code was changing on a daily basis." At these large scales, automated methods offer the best hope of verifying compliance.

"One reason that gaps exist between policies set by a company's privacy team and the code written by software developers is that the two groups don't speak the same language," Datta said. Lawyers and privacy champions typically have little experience in programming and developers attempting to translate policies into code can get tripped up by ambiguities in the language of the privacy policies.

So the researchers developed a language — Legalease — that could be easily learned and used by privacy advocates. It employs allow-deny rules with exceptions, a structure that is

found in many privacy policies and laws, such as the Health Insurance Portability and Accountability Act (HIPAA), and is expressive enough to capture the real policies of an industrial-scale system such as Bing.

In preliminary usability testing, a dozen Microsoft employees were given a one-page document explaining Legalease and spent an average of under 5 minutes studying it. They then took an average of less than 15 minutes to encode nine Bing policy clauses regarding how user information can be used. "They were able to perform this task with a high degree of accuracy, which is encouraging," Sen said.

But encoding privacy policies correctly means little if it cannot be applied to large codebases written by large teams of programmers. To solve this dilemma, the researchers leveraged Grok — a data inventory that annotates existing programs written in languages typically employed by MapReduce-like systems, such as those used by Bing and Google — for their backend data analytics over user data.

Grok performs this automated annotation by combining information from different sources with varying levels of confidence. For instance, automated pattern-matching to column names can be performed across an entire database, but with low confidence, while annotations by developers have high confidence, but low coverage.

Grok had been developed by Microsoft Research and deployed by Bing for the express purpose of automating privacy compliance checking the previous year, but writing policies for Grok was cumbersome.

"Legalease was the final piece of the automated privacy compliance jigsaw puzzle," Guha said. "Developed over Sen's internship and subsequent collaboration with CMU, Legalease bridged privacy teams with Grok, and through Grok, with the developers."

Datta said automating the process of compliance checks could push the industry to adopt stronger privacy protection policies.

"Sometimes, companies want to make their policies stronger, but hesitate because they are not sure they can ensure compliance in these large systems," he explained, noting that online privacy policy compliance is enforced in the United States by the Federal Trade Commission.

The research team included Sriram K. Rajamani of Microsoft Research in Bangalore, India; Janice Tsai of Microsoft Research, Redmond, and Jeannette Wing, corporate vice president of Microsoft Research and former head of CMU's Computer Science Department.

This research was supported, in part, by the Air Force Office of Scientific Research and the National Science Foundation.

###

- Carnegie Mellon News
- 5000 Forbes Avenue Pittsburgh, PA 15213
- (412) 268-2900

- Carnegie Mellon Home