

Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (SLAML 2011)

At the ACM Symposium on Operating Systems Principles
October 23-26, 2011
Carcais, Portugal

Important Dates

- Full paper submission due: **Friday, June 17th, 2011**
- Notification of acceptance: **Friday, July 15th, 2011**
- Final papers due: **Friday, August 12th, 2011**

Overview

Modern large-scale systems are challenging to manage. Fortunately, as these systems generate massive amounts of performance and diagnostic data, there is an opportunity to make system administration and development simpler via automated techniques to extract actionable information from the data. This workshop addresses this problem in two thrusts: (i) the analysis of raw system data logs, and (ii) the application of machine learning to systems problems. We expect the large overlap in these topics to promote a rich interchange of ideas between the areas.

Log Analysis: It is well known that raw system logs are an abundant source of information for the analysis and diagnosis of system problems and prediction of future system events. However, a lack of organization and semantic consistency between system data from various software and hardware vendors means that most of this information content is wasted. Current approaches of extracting information from the raw system data capture only a fraction of the information available and do not scale to the large systems common in business and supercomputing environments. It is thus a significant research challenge to determine how to better process and combine information from these data sources.

Machine Learning: The large scale of available data requires automated and machine-assisted analysis. Statistical machine learning techniques have recently shown great promise in meeting the challenges of scale and complexity in datacenter-scale and Internet-scale computing systems. However, applying these techniques to real systems scenarios requires careful analysis and engineering of the techniques to fit them to specific scenarios; there is also sometimes the opportunity to develop new algorithms specific to systems scenarios. This workshop thrust thus also presents a substantial research area: the exploration of new approaches to using machine learning to help us understand, measure, and diagnose complex systems.

Topics

Topics include but are not limited to:

- Reports on publicly available sources of sample system logs
- Prediction of malfunction or misuse based on system data
- Statistical analysis of system logs
- Applications of Natural-Language Processing (NLP) to system data
- Techniques for system log analysis, comparison, standardization, compression, anonymization, and visualization
- Applications of log analysis to system administration problems
- Use of machine learning techniques to address reliability, performance, power management, security, fault diagnosis, scheduling, or manageability issues
- Challenges of scale in applying machine learning to large systems
- Integration of machine learning into real-world systems and processes
- Evaluating the quality of learned models, including assessing the confidence/reliability of models and comparisons between different methods

Workshop Organizers

Program Co-Chairs

- Peter Bodik, *Microsoft Research*
- Marc Casas, *Lawrence Livermore National Laboratory*
- Greg Bronevetsky, *Lawrence Livermore National Laboratory*

Submission Guidelines

Submitted papers must be no longer than 8 (8) 8.5"x11" or A4 pages, using a 10 point font on 12 point (single spaced) leading, with a maximum text block of 6.5 inches wide by 9 inches deep. The page limit includes everything except for references, for which there is no limit. The use of color is acceptable, but the paper should be easily readable if viewed or printed in gray scale. Authors must make a good faith effort to anonymize their submissions, and they should not identify themselves either explicitly or by implication (e.g., through the references or acknowledgments). Submissions violating the detailed formatting and anonymization rules on the Web site will not be considered for publication. There will be no extensions for reformatting.

Blind reviewing of full papers will be done by the program committee, with limited use of outside referees. Papers will be provisionally accepted subject to revision and approval by a program committee member acting as a shepherd. On acceptance, authors will be required to sign an ACM copyright release form. Your submission indicates that you agree to this. Papers will be held in full confidence during the reviewing process, but papers accompanied by nondisclosure agreement forms are not acceptable and will be rejected without review. Authors of accepted papers will be expected to supply electronic versions of their papers and encouraged to supply source code and raw data to help others replicate and better understand their results.