

Online System Monitoring, Diagnosing, and Predicting for Large-Scale Computing

This project is dedicated to designing and developing an extensive software infrastructure to provide online system monitoring, diagnosing, and predicting for large-scale systems. The goal is to enhance existing resilience technologies as well as to facilitate new resilience techniques. Our major research tasks include effective log preprocessing, online failure prediction, automated anomaly diagnosis, and reliability-aware modeling. The work was selected for the Top-10 Data Mining Case Studies at the 10th ICDM [\[Website\]](#). The team also won "The Cray Log Analysis Contest" at the First USENIX Workshop on the Analysis of System Logs (WASL'08), which was co-located with the 8th USENIX Symposium on Operating Systems Design and Implementation (OSDI'08), December 8-10, 2008. A prototype software tool named SysDP is available at [website](#).

Members:

- Zhiling Lan (faculty)
- Li Yu (Ph.D. student)
- Sean Wallace (Ph.D. student)
- Eduardo Berrocal (Ph.D. student)
- Ziming Zheng (Ph.D. graduated in 2012)

Collaborators:

- Ann Gentile (Sandia Lab)
- Jim Brandt (Sandia Lab)
- Susan Coghlan (Argonne National Lab)
- Rajeev Thakur (Argonne National Lab)
- Terry Jones (Oak Ridge National Lab)

Publications:

- L. Yu and Z. Lan, "A Scalable, Non-Parametric Anomaly Detection Framework for Hadoop", *Proc. of the ACM Cloud and Autonomic Computing Conference (CAC'13)*, 2013.
- Z. Zheng, L. Yu, Z. Lan, and T. Jones, "3-Dimensional Root Cause Diagnosis via Co-Analysis", *Proc. of ICAC'12*, 2012.
- L. Yu, Z. Zheng, Z. Lan, T. Jones, J. Brandt, and A. gentile, "Filtering Log Data: Finding the needles in the Haystack", *Proc. of DSN'12*, 2012. [\[PDF\]](#)
- L. Yu, Z. Zheng, Z. Lan, and S. Coghlan, "Practical Online Failure Prediction for Blue Gene/P: Period-based vs Event-Driven", *Proc. of Proactive Failure Avoidance, Recovery, and Maintenance Workshop (PFARM) (in conjunction with DSN'11)*, 2011. [\[PDF\]](#)
- Z. Zheng, L. Yu, W. Tang, Z. Lan, R. Gupta, N. Desai, S. Coghlan, and D. Buettner, "Co-Analysis of RAS Log and Job Log on Blue Gene/P", *Proc. of IPDPS'11*, 2011. [\[PDF\]](#)
- Z. Zheng, Z. Lan, R. Gupta, S. Coghlan, and P. beckman, "A Practical Failure Prediction with Location and Lead Time for Blue Gene/P", *Proc. of the 1st Workshop on Fault-Tolerance for HPC at Extreme Scale (FTXS), in conjunction with DSN'10*, 2010. [\[PDF\]](#)
- Z. Lan, J. Gu, Z. Zheng, R. Thakur, and S. Coghlan, "A Study of Dynamic Meta-Learning for Failure Prediction in Large-Scale Systems" *Journal of Parallel and Distributed Computing (JPDC)*, 2010. [\[PDF\]](#)

Contact:

Dr. Zhiling Lan (lan AT iit DOT edu)

This work is partially supported by DOE/Sandia.