Tracing and Monitoring Framework Impact Prediction

Douglas Santos
Michel Dagenais

December 8, 2010
École Polytechnique, Montreal

Summary

- **Introduction**
- **State of the Art**
- **Monitoring Framework**
- **Future Work**

Introduction

- Large multi-core distributed systems
- Production environment
- Impact of Tracing ?
- System Health

State of the Art

- Resource consumption aware [1]
  - Video on Demand (client / server)
  - Objective: Guarantee QoS for network streams
  - Admission control
    - Monitor: network and memory

      Results: better resource utilization

State of the Art

- Application performance prediction [2]
  - Web applications
  - Objective: Predicting application performance
  - System that predict response times
    - Monitor: request delay, CPU, memory

      Results: predict response time within 15%

State of the Art

- QoS Prediction [3]
  - Real time systems
  - Objective: Guarantee QoS for messages
  - Resource monitoring (profiler)
    - Monitor: execution time, CPU, memory

      Results: good CPU utilization

State of the Art

[1] Dongmahn Seo, Joahyoung Lee, Yoon Kim, Chang Yeol Choi, Manbae Kim, and Inbum Jung. Resource consumption-aware qos in cluster-based vod servers. J. Syst. Archit., 53(1):39–52, 2007.

[2] Shobhana Kirtane and Jim Martin. Application performance prediction in autonomic systems. In ACM-SE 44: Proceedings of the 44th annual Southeast regional conference, pages 566–572, New York, NY, USA, 2006. ACM.

[3] Eui-nam Huh, Lonnie R. Welch, Behrooz Shirazi, Brett C. Tjaden, and Charles Cavanaugh. Accommodating qos prediction in an adaptive resource management framework. In IPDPS '00: Proceedings of the 15 IPDPS 2000 Workshops on Parallel and Distributed Processing, pages 792–799, London, UK, 2000. Springer-Verlag.

[4] Stephen A. Jarvis, Daniel P. Spooner, Helene N. Lim Choi Keung, Graham R. Nudd, Junwei Cao, and Subhash Saini. Performance prediction and its use in parallel and distributed computing systems. In Proceedings of the 17th International Symposium on Parallel and Distributed Processing, IPDPS '03, pages 276.1, Washington, DC, USA, 2003. IEEE Computer Society.

[5] A. J. Oliner and J. E. Moreira. Probabilistic qos guarantees for supercomputing systems. In DSN '05: Proceedings of the 2005 International Conference on Dependable Systems and Networks, pages 634–643, Washington, DC, USA, 2005. IEEE Computer Society.

State of the Art

[6] Philipp Leitner, Branimir Wetzstein, Florian Rosenberg, Anton Michlmayr, Schahram Dustdar, and Frank Leymann. Runtime prediction of service level agreement violations for composite services. In Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops, volume 6275 of Lecture Notes in Computer Science, pages 176–186. Springer Berlin / Heidelberg, 2010.

[7] Stephen A. Jarvis, Daniel P. Spooner, Helene N. Lim Choi Keung, Junwei Cao, Subhash Saini, and Graham R. Nudd. Performance prediction and its use in parallel and distributed computing systems. Future Gener. Comput. Syst., 22(7):745–754, 2006.

[8] Dan G. Waddington and David Hutchison. End-to-end qos provisioning through resource adaptation. In HPN ’98: Proceedings of the IFIP TC-6 Eigth International Conference on High Performance Networking, pages 309–326, Deventer, The Netherlands, The Netherlands, 1998. Kluwer, B.V.

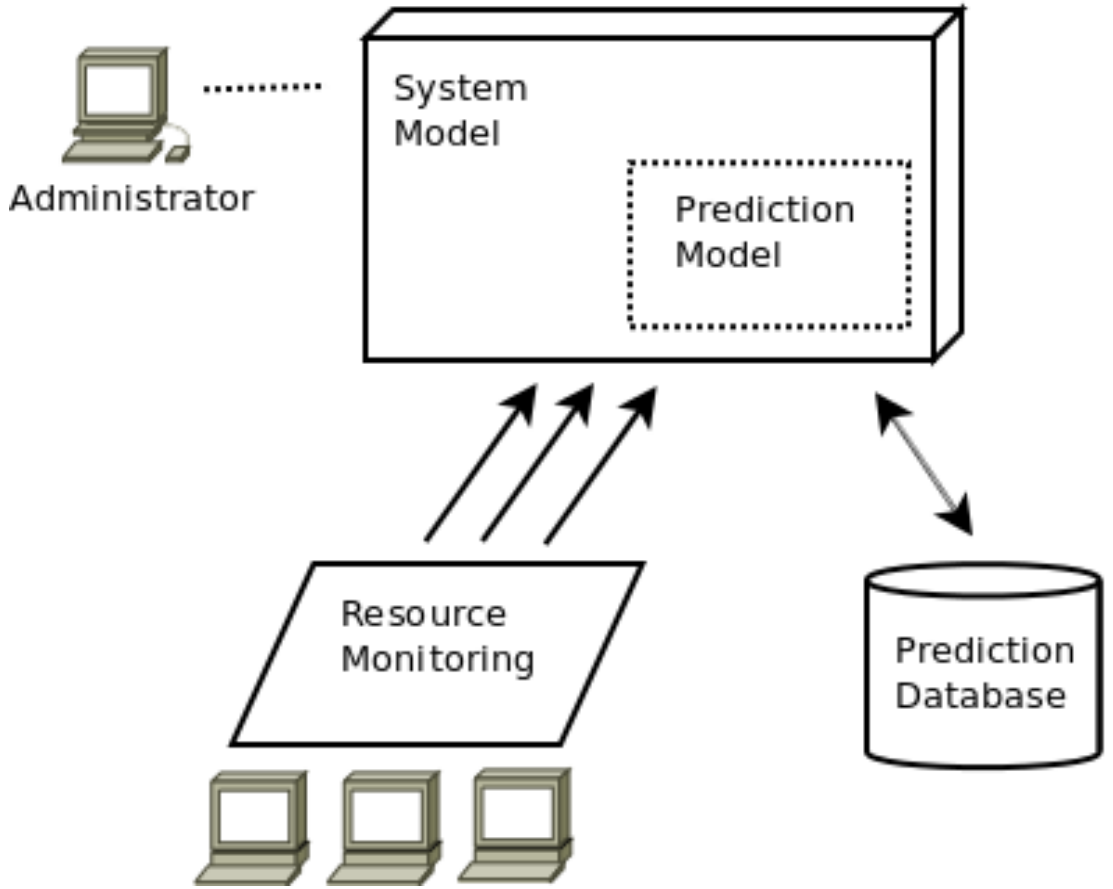
[9] David A. Bacigalupo, Stephen A. Jarvis, Ligang He, Daniel P. Spooner, Donna N. Dillenberger, and Graham R. Nudd. 2005. An Investigation into the Application of Different Performance Prediction Methods to Distributed Enterprise Applications. J. Supercomput. 34, 2 (November 2005), 93-111.

Monitoring Framework

- Proposal
  - Resource monitoring and modeling
    (CPU, memory, disc, net)
  - System Health view
  - Based on the model, predict the impact of Tracing
  - QoS

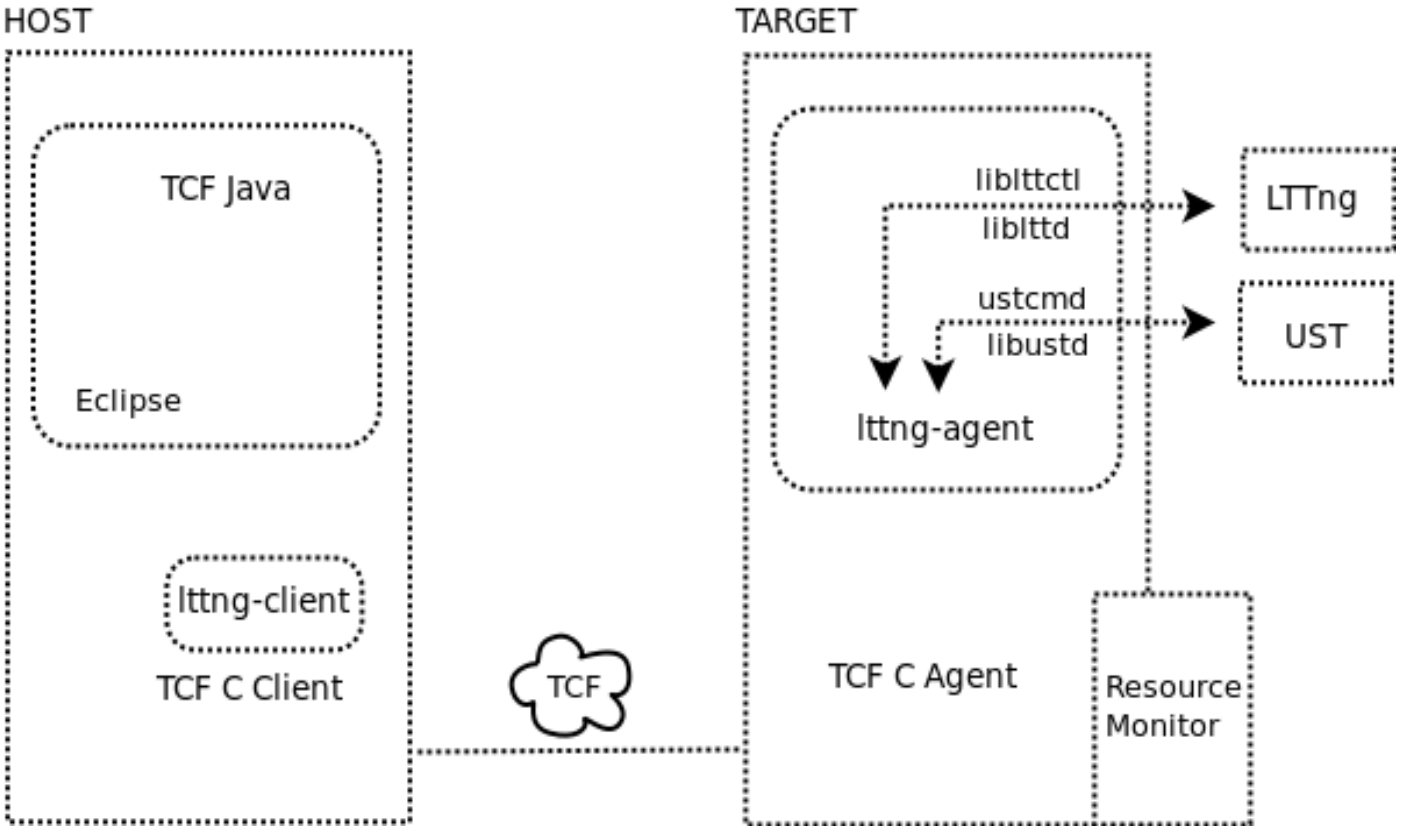
Monitoring Framework

- Model

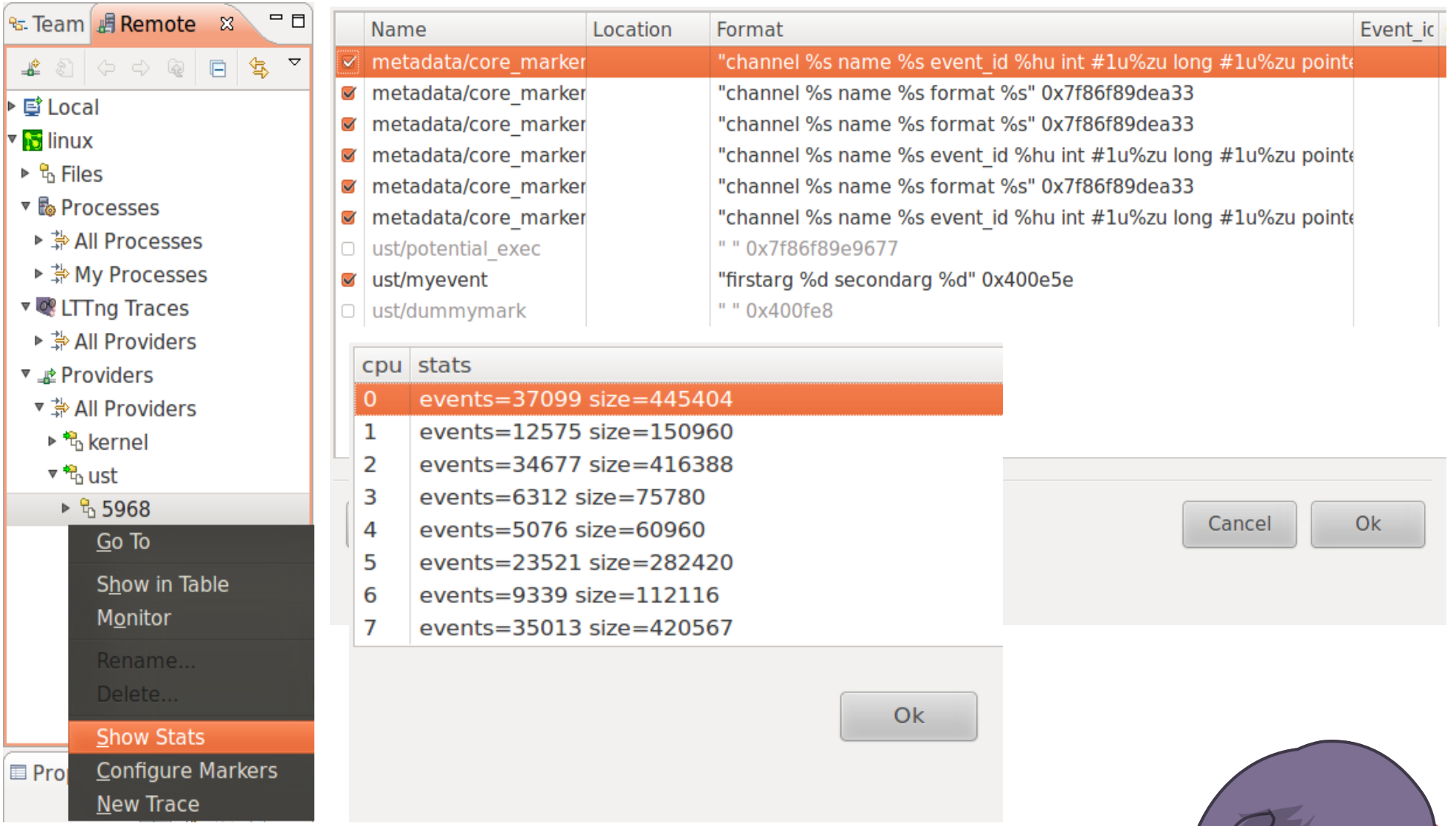


Monitoring Framework

- Architecture



Monitoring Framework



Future Work

- Integrate a complete resource monitoring system into the TCF framework
- Work on the system model and health monitoring impact prediction
- Work on high level QoS and admission control

Questions ?