Automatic Performance Analysis for Parallel Applications: Initial Ideas

Flavio Alles Rodrigues
Lucas Mello Schnorr
{farodrigues, schnorr}@inf.ufrgs.br
Motivation

Parallel computers are increasing in size...
Motivation

Parallel computers are increasing in size...

...becoming more heterogeneous...
Motivation

Parallel computers are increasing in size...
...becoming more heterogeneous...
and consuming more power.
Motivation

• Programming such a system is challenging
Motivation

- Programming such a system is challenging
- Performance analysis tools are essential
Outline

[1] Automatic Performance Analysis
Outline

[1] Automatic Performance Analysis
[2] Survey of Tools
Outline

[1] Automatic Performance Analysis
[2] Survey of Tools
[3] Issues
Outline

[1] Automatic Performance Analysis
[2] Survey of Tools
[3] Issues
Automatic Performance Analysis

• Detect behavioral patterns and performance issues automatically
Automatic Performance Analysis

• Detect behavioral patterns and performance issues automatically

• Increases the amount of information that can be analysed
Automatic Performance Analysis

• Detect behavioral patterns and performance issues automatically
• Increases the amount of information that can be analysed
• Large-scale parallel applications generate huge amounts of data
Automatic Performance Analysis

• Detect behavioral patterns and performance issues automatically
• Increases the amount of information that can be analysed
• Large-scale parallel applications generate huge amounts of data
• Path to scale analysis to the required levels
Survey of Tools

- Automate some aspect of analysis
Survey of Tools

• Automate some aspect of analysis

• Focus is on parallel applications
Survey of Tools

- Automate some aspect of analysis
- Focus is on parallel applications
- Presented to public in peer-reviewed articles
Survey of Tools

• Paradyn [1]
• Periscope [2]
• Scalasca [3]
Paradyn

- Online, profile-based analysis
Paradyn

- Online, profile-based analysis
- \(W^3\) Search Model
Paradyn

- Online, profile-based analysis
- $W^3$ Search Model
- *Why* is the application performing poorly?
Paradyn

- Online, profile-based analysis
- $W^3$ Search Model
  - *Why* is the application performing poorly?
  - *Where* is the bottleneck?
Paradyn

- Online, profile-based analysis
- W³ Search Model
  - Why is the application performing poorly?
  - Where is the bottleneck?
  - When does the problem occur?
Periscope

• Online, profile-based analysis
Periscope

- Online, profile-based analysis
- Distributed autonomous search for pre-defined bottlenecks (ASL [4])
Periscope

- Online, profile-based analysis
- Distributed autonomous search for pre-defined bottlenecks (ASL [4])
- Two distinct search strategies
Periscope

- Online, profile-based analysis
- Distributed autonomous search for pre-defined bottlenecks (ASL [4])
- Two distinct search strategies
  - Single-node Performance (i.e. stalled cycles)
Periscope

- Online, profile-based analysis
- Distributed autonomous search for pre-defined bottlenecks (ASL [4])
- Two distinct search strategies
  - Single-node Performance (*i.e.* stalled cycles)
  - MPI Performance (*i.e.* load imbalances)
Scalasca

• Offline, trace-based analysis
Scalasca

- Offline, trace-based analysis
- Parallel trace analysis: communications replay to search for pre-defined performance properties
Scalasca

- Offline, trace-based analysis
- Parallel trace analysis: communications replay to search for pre-defined performance properties
- Wait states (forward replay) and its root causes (backwards replay)
Issues

• Static expectations of automatic analysis
Issues

• Static expectations of automatic analysis
• Problems that automatic analysis tools diagnose
• Static expectations of automatic analysis
• Problems that automatic analysis tools diagnose
• Assumption that these tools make regarding the machines where the applications are executed
Potential Approaches

• Machine Learning
Potential Approaches

• Machine Learning
  • Unsupervised learning techniques
Potential Approaches

- Machine Learning
  - Unsupervised learning techniques
    - Clustering
    - Hidden Markov Models
    - Artificial Neural Networks
      - Self Organizing Maps
      - Adaptive Resonance Theory
Upcoming Work

- Python
- scikit-learn [5]
- Pajé Traces [6]
References


[5] [http://scikit-learn.org](http://scikit-learn.org)

Automatic Performance Analysis for Parallel Applications: Initial Ideas

Flavio Alles Rodrigues
Lucas Mello Schnorr
{farodrigues, schnorr}@inf.ufrgs.br