Performance Analysis of Volunteer Computing Traces

Trlice Estrada and Michela Taufer

University of Delaware

Kevin Reed

IBM

**Volunteer Computing (VC)**

VC projects deploy resources connected to the Internet and owned by the public. VC projects are throughput-based: performance is measured in terms of number of valid results that are delivered to the scientists within a given interval of time.

BOINC projects are VC projects that provide tasks, such as longer or shorter periods of time, to different groups of people. Projects with different degrees of availability and reliability can support a VC project. This heterogeneity across VC projects makes it challenging to find performance bottlenecks, as well as common scheduling strategies across projects that are able to tackle these challenges.

---

**Research Objective**

Implement a statistical methodology that provides project designers and administrators with a quantitative approach to predict times in the life cycle of WUIs.

---

**Methodology**

**Statistical Distributions**

- **Density function**: the distribution function of a random variable
- **Cumulative distribution function**: the probability that a random variable is less than or equal to a given value

**Adjustment Results:**

- **Adjust Traces with a Single Distribution**
  - **Distributions**: Weibull, Levy, and Gamma
  - **Adjustment method**: Least Squares

- **Adjust Traces with Mixture of Gaussians**
  - **Adjustment algorithm**: Obtain a probability model mixture by using the Expectation Maximization (EM) algorithm.
  - **EM** performs an expectation step (to calculate the likelihood of the model) and a maximization step (to maximize the likelihood found in the previous step).

**Challenges**

- The times WUIs wait for distribution, are in progress (waiting on the worker for execution, in execution, or waiting for being returned), or are waiting for validation are difficult to estimate because of the unpredictable behavior of volunteer resources. If project designers and administrators cannot estimate these times, VC projects can be affected by significant delay in delivering results to the scientists.

---

**Future Work**

- **Tuning the training and testing intervals**
- **Use heuristics and machine learning techniques**
- **Comparison with Binomial Method Batch Prediction (R. Wolski et al., 2006)**
- **Integration the method in SimBA**
- **Integration in real VC projects (Docking/Home)**

---

**Conclusions**

- **Accuracy Study**
  - **Accuracy in Predicting Distribution Delay**
  - **Accuracy in Predicting Time in Progress**
  - **Accuracy in Predicting Validation Delay**

**Sponsors**

- InnoDB
- Contacts: (estrada, taufer)@udel.edu

**References**

- [R. Wolski et al., 2006]
- [SimBA]
- [Integration in real VC projects (Docking/Home)]