Managing a Grid of Computer Laboratories for Educational Purposes

Luis C. E. De Bona Marcos Castilho Fabiano Silva
Daniel Weingartner Luis H. A. Lourenço Bruno Ribas

Center for Scientific Computing and Free Software (C3SL)
Federal University of Parana
Departament of Computer Science

Outline

Outline

Digital Inclusion Polices

- Computer laboratories are a necessary tool in a student's learning process
- The cost reduction of hardware led to the creation and expasion of digital inclusion policies
- The lack of specialized workforce is a huge problem!

Managing Computer Laboratories

- Having an expert to manage each laboratory in a huge public is impossible
- It demands a effort to define a new administration model
- All system management task should be performed either automatically or by experts
- But we need to minimize the experts' work

A New Administration Model

- This paper presents a model that allows the administration of thousands computing laboratories with minimum human intervention
- Based on autonomic computing concepts: self-configuration; self-optimization; self-healing; and self-protection
- Two kinds of human interventions:
  - Local execution of simple task through user-friendly interfaces
  - Remove execution of critical or unpredictable tasks by a team of experts

Parana Digital Project (PRD)

- The model was implemented by Parana Digital Project
- More than 2100 public school laboratories
- GNU/Linux software and GPL licence

Outline

A Model for Management of Computing Labs

- The goal:
  - To keep computing laboratories in working conditions
  - To dispense the use of specialized staff in each laboratory
- Geographically distributed computing laboratories are interconnected forming a computational grid
- Hardware, operating system and main application are considered fairly homogeneous
- Tasks needed to install and to manage the laboratories are classified into local and global management tasks

Management Tasks

- Local management tasks
  - Specific aspects of each laboratory
  - Translated into high level simple decisions offered by a user-friendly interface
  - Ordinary user called local manager
- Global management tasks
  - To guarantee that the laboratory offers the expected services
  - Determined in a global way and uniformly executed in the grid
  - Most of these task should be automated
  - In order to implement it the proposal is to apply concepts of autonomic computing

The Local Managers

- They are not computer experts
- They take high level and simple decisions
- They use a user-friendly interface to perform local tasks:
  - User accounts, disk quotas, etc.
  - Initial installation or re-installation procedures
  - Contact the call center in case of problems

Global managers: the Core

- A small group of specialized system managers
- They are able to manage hundreds of laboratories
- They specify what are the system's features, software to be installed, what should be restricted or allowed
- Their decisions are propagated through self-management
  - Self-configuration
  - Self-optimization
  - Self-recuperation
  - Self-protection

Self-Configuration

- Is a continuous process aiming to keep the system configured under varying time and environment conditions
- The configuration policies are defined for the entire grid by a core management team, no question should be asked to the local manager
- In order to reduce management complexity a computing model based on graphic terminals is employed
- After installation the system is continuously updated through the network
- The update system is based on modern software package managements systems

Self-Optimization

- Data provided by monitoring systems can be used by experts to optimize system parameters
- While laboratories have similar configurations global monitoring information is very usefull
- Historical and comparative analysis of performance metrics can be used to optimize the system
- Based on global monitoring information experts can enforce new configuration parameters

Self-Healing and Self-Protection

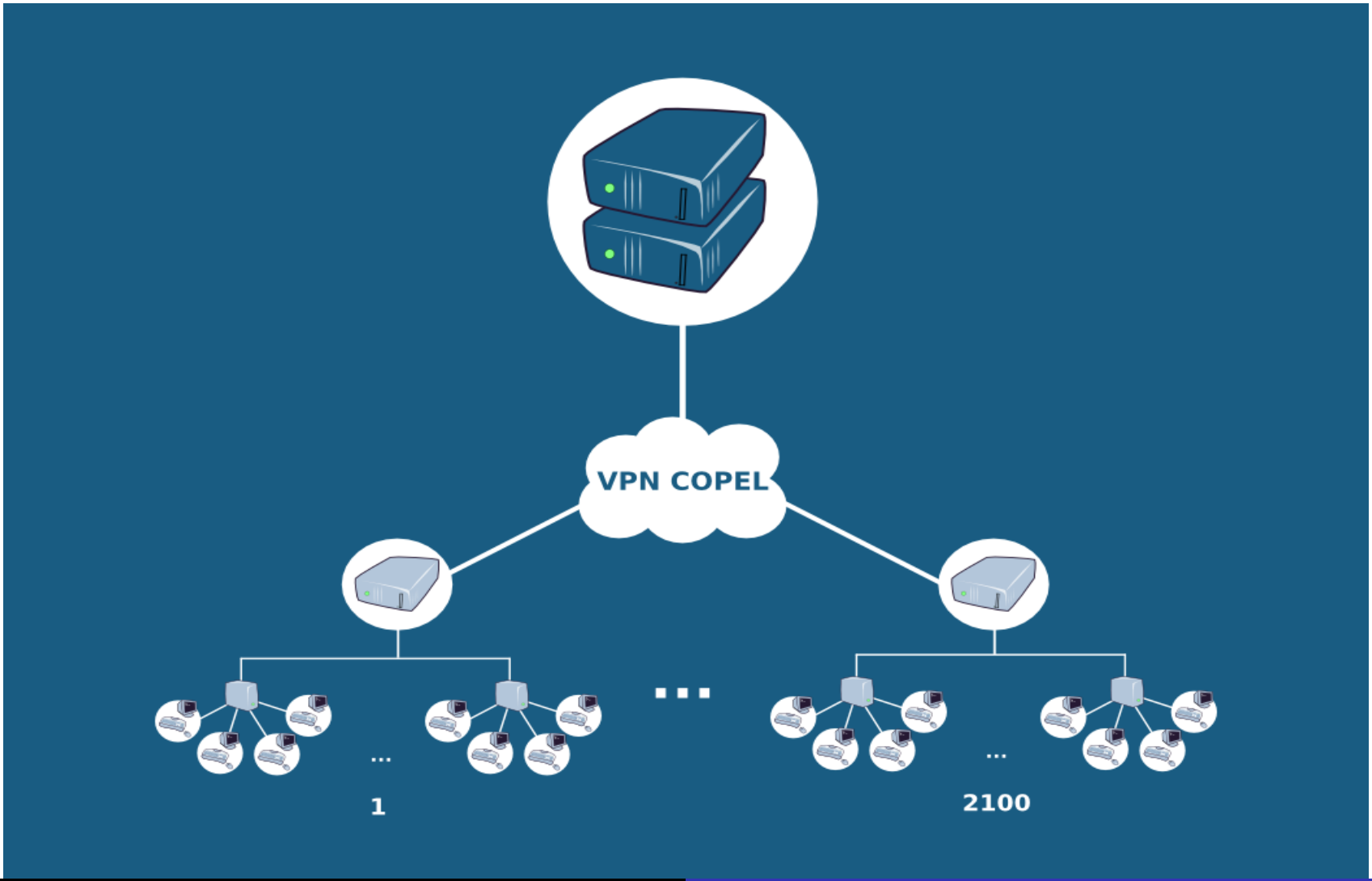
- The system must detect, diagnose, treat and prevent problems due to bugs or hardware failures, leaving minimal decisions to the local manager
- The downtime of computing laboratories can be drastically reduced when certain aspects of the hardware and the operating system are tracked
  - Maintaining the system updated is the first step
  - Hard disks and filesystem are the most failure-prone components
  - Hard disk's self monitoring facility must be monitored
  - Redundant Arrays (RAIDS) are really usefull
  - But is also necessary to perform periodic filesystem integrity verification
- When complete re-installation is need a local procedure must tries automatically reinstall the system without loss of user data

Outline

Paraná Digital Project

- The aim of the project is to provide every school of the Parana Sate with a computing laboratory
- Parana sate has about 1.500.000 students, 57.000 teachers, 2.100 schools distributed over 399 cities over 199,314 Km^2
- A huge testbed for the proposed model
- The first laboratory was installed in June/2006 and, as of August 2008, 2.126 schools were operational. The management team is composed of 12 highly trained Unix managers, taking care of the entire network (approximately 44 thousands stations)

PRD's architecture



PRD's architecture

- In each school:
  - a laboratory composed of 20 X-terminals
  - one processing server called *the school server*
- The school server acts simultaneously as
  - processing and storage unit, gateway to the network, firewall and access point to the Core
  - it runs a Debian-based GNU/Linux distribution,
  - all servers have the same software packages installed.
- At the Core, a proxy-controlled connection to the Internet is provided

Initial installation

- A CD-ROM containing a standard system image
- The server checks periodically for updates at the central mirror
- X-terminals must automatically recognize and configure hardware
- It allows automated installation and configuration by the local manager
- This automated procedure minimizes service disruption and data loss
- A regular user must have an account created by the local manager.
- The local manager does not have root powers
- This ensures that critical tasks are globally defined on the grid.

System Upgrade

- Frequent system upgrades are necessary to provide new functionalities, address security problems and propagate new software, tools, or policies from the Core
  - automatic daily upgrades and
  - triggered upgrades

The daily automatic upgrade

- Based on Debian's *apt-get* tools
- Every night, each school server looks for new software packages
- The single mirror ensures that all servers will install exactly the same software
- It is quite simple to propagate a new tool or configuration over the entire network

The triggered upgrade

- The Core can force all school servers to upgrade
- This happens as soon as the network link turns on
- Use cases:
  - the Kernel exploit that allowed an ordinary user to become root
  - the upgrade from *sarge* to *etch* (in February 2008)

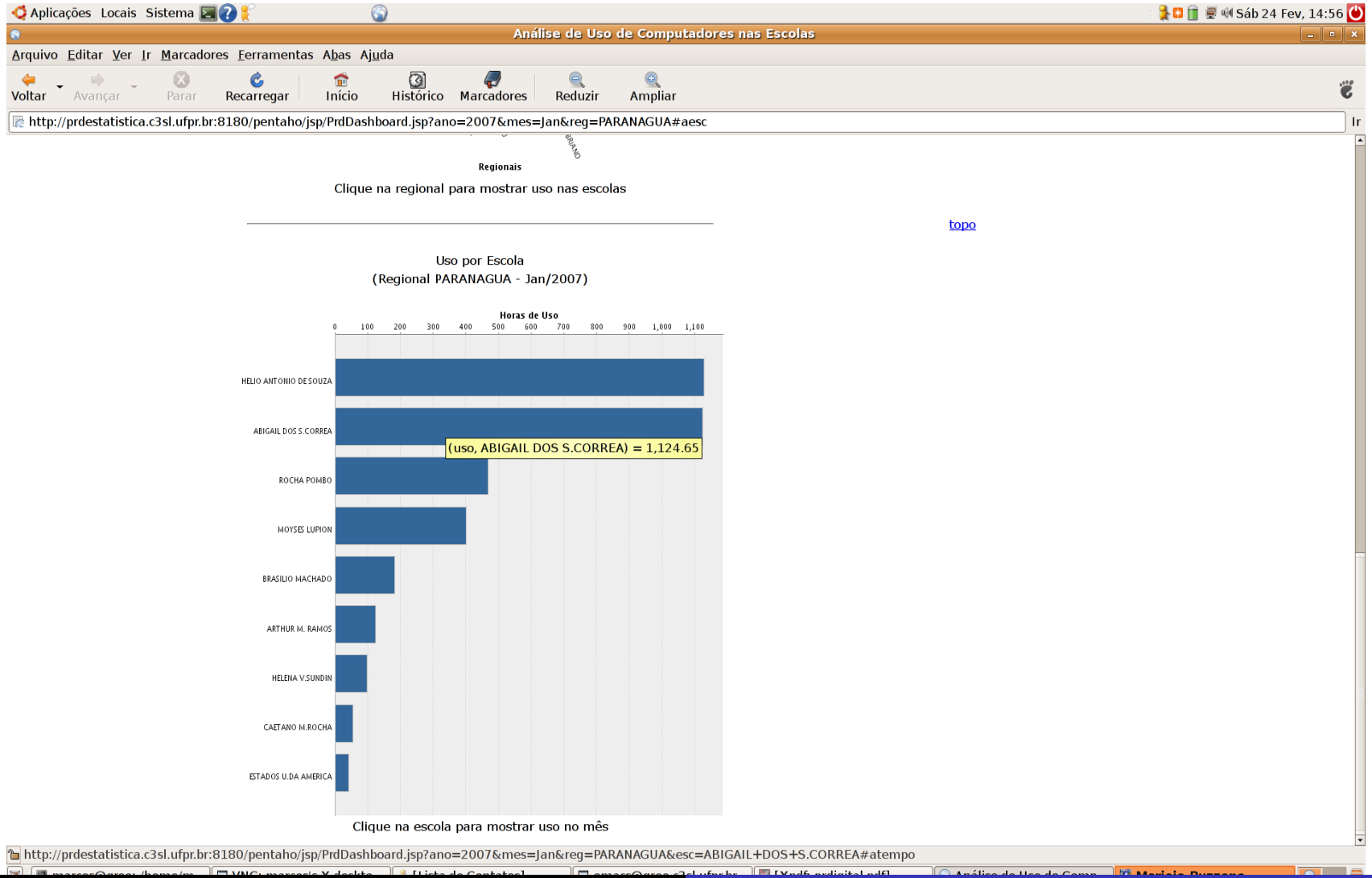
System Monitoring

- Is an essential feature of autonomic systems, and provides information to allow the system's self-optimization and self-recuperation
- It is at the heart of the PRD network, revealing the real state of the whole grid
- In the PRD model, there are two different monitoring systems:
  - the statistics center
  - diagnosis system

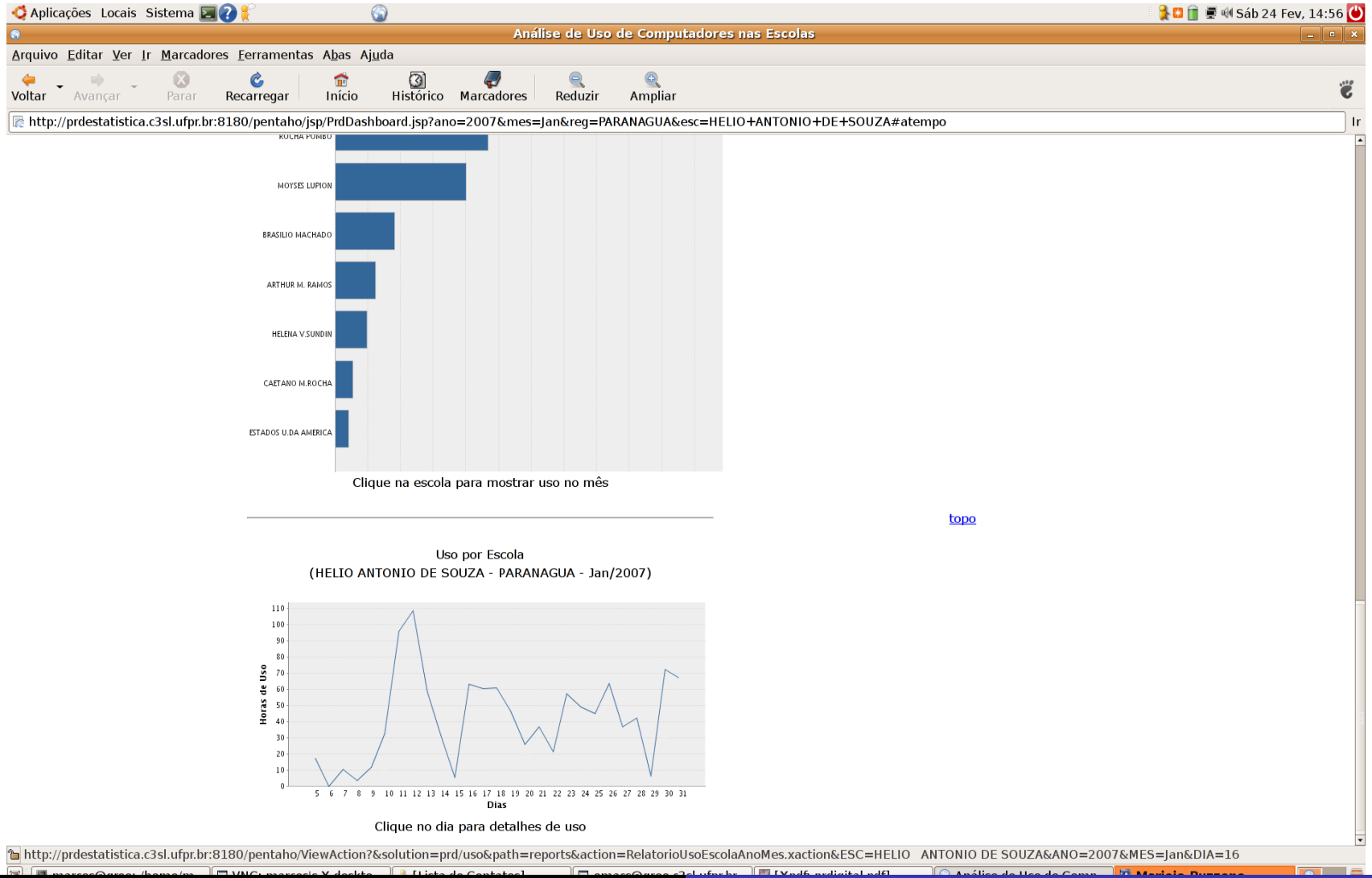
The Statistics Center

- A web site with strategic information
- It allows an overview of the network's growth and provide data concerning the laboratories usage
- This is automatically stored in the central database
- This provides strategic information for decision support

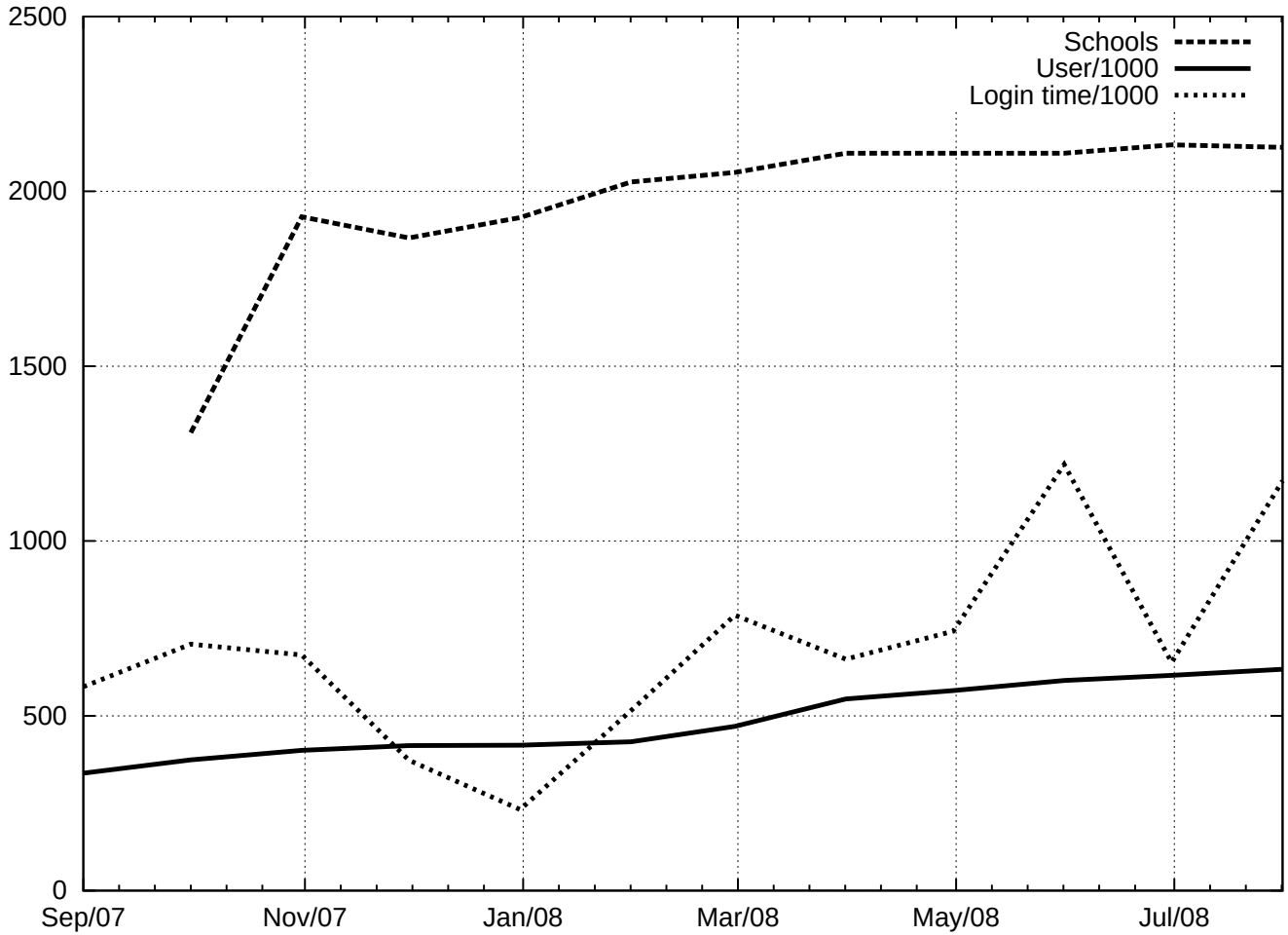
Snapshot of the statistics center web page



Snapshot of the statistics center web page



The Instant Diagnosis System



Manual System Inspection

- Manual interventions are undesired
- The core can remotely log in a server via SSH
- But just to inspect the server's behavior
- All changes should happen on a global manner in the grid
- This means: global management!

Outline

Conclusion

- How to manage computing laboratories without a local system manager
- We proposed a model that allows the administration of thousands of computing laboratories with minimum humam intervention
- The system is managed as a whole
- Paraná Digital has more than 2,100 schools and only 12 managers