Time-Series analysis aids performance monitoring and anomaly detection in computer networks

Marcos Portnoi, Priscilla Moraes, Martin Swany
Computer and Information Sciences, University of Delaware,
Newark, Delaware

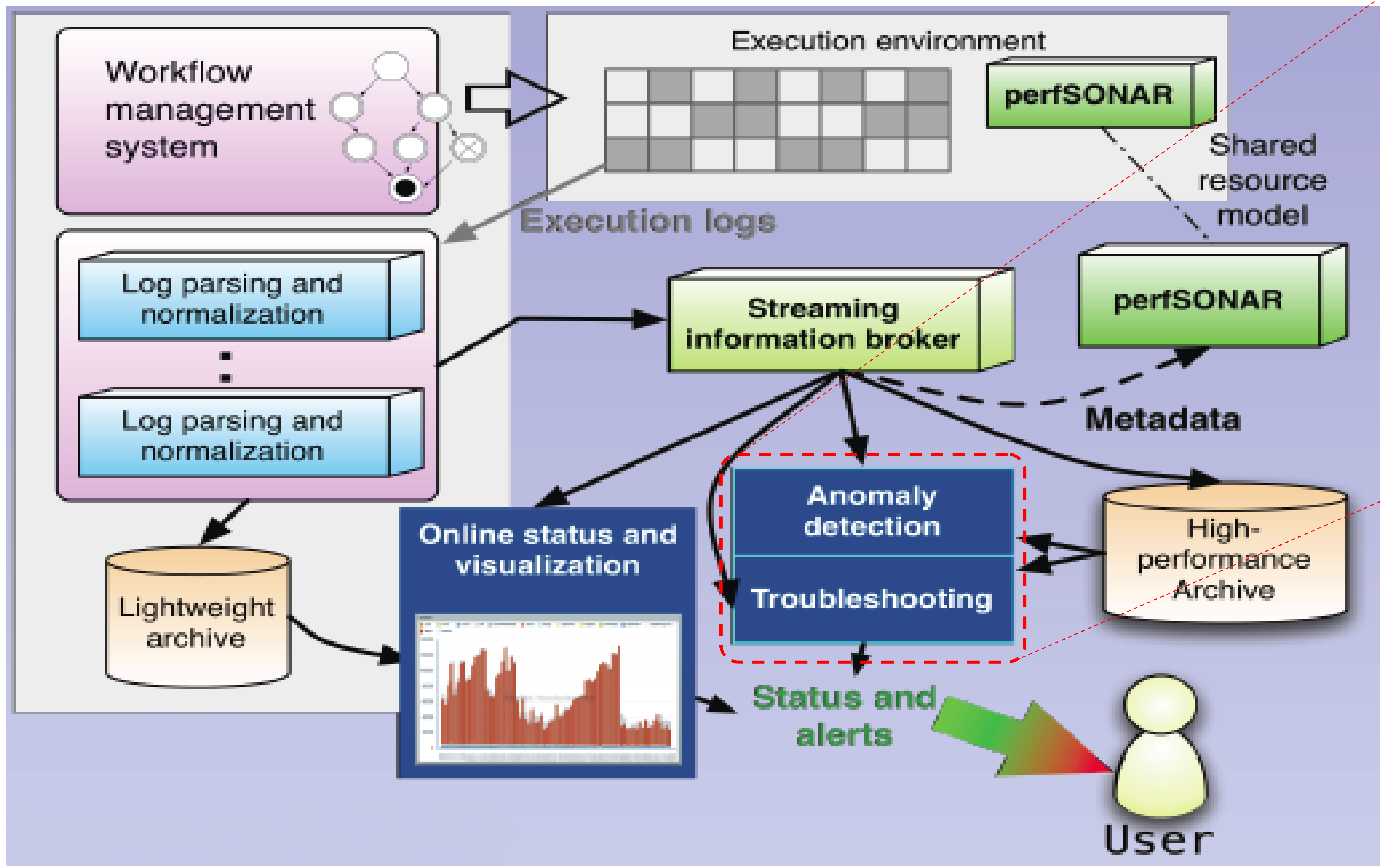
Problems:

- **Scenario 1:**
  - Scientific research requires very large-scale calculations.
  - Analyses are composed of thousands to millions of coordinated tasks.
  - Need to be executed efficiently and reliably.
  - Hard to analyze what happened, what performance was achieved, how well the computation progressed, what were the problems.
- **Scenario 2:**
  - In computer networks, applications usually compete for network resources.
  - This can often result in applications getting a "fair" share of the resources… but fairness in the point of view of the network, not necessarily the applications.
  - Tasks which composes workflow runs may take too long or even not finish. This is then considered an anomalous behavior which needs to be detected.

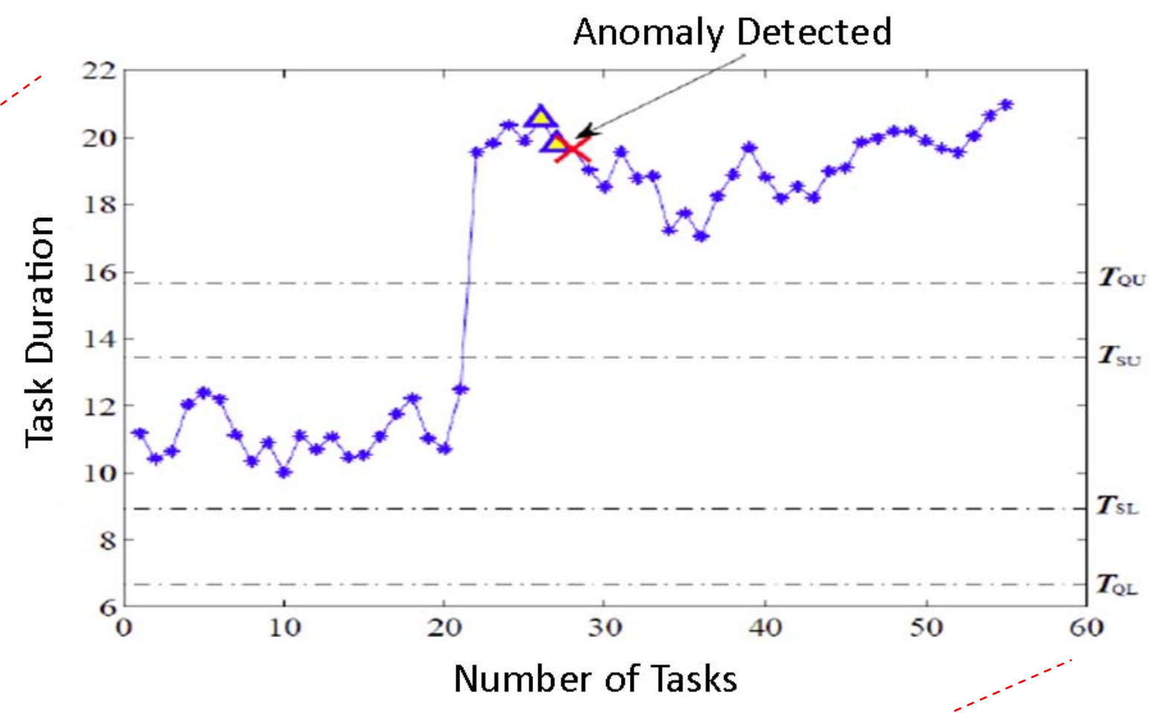
Proposals for solving the problems:

- Employ *Time-Series Analysis* on previous history of applications' behavior.
- A *probe* monitors network flows or scientific workflows.
- A *Statistical Engine* inside the probe uses time-series analysis techniques to detect flow or workflow statistical behavior.
- Time-series analysis performed by the Statistical Engine may be employed in *anomaly detection:*
  - **Scenario 1:** Data analysis and troubleshooting of anomalous tasks in scientific workflow applications environment.
  - **Scenario 2:** A *network performance monitoring architecture*, such as perfSONAR, to provide services for *event triggering*, *alarming*, and *statistical auditing*..
- Can also be used in *forecasting*, where the history of the application or network behavior and usage is exploited to predict future performance.

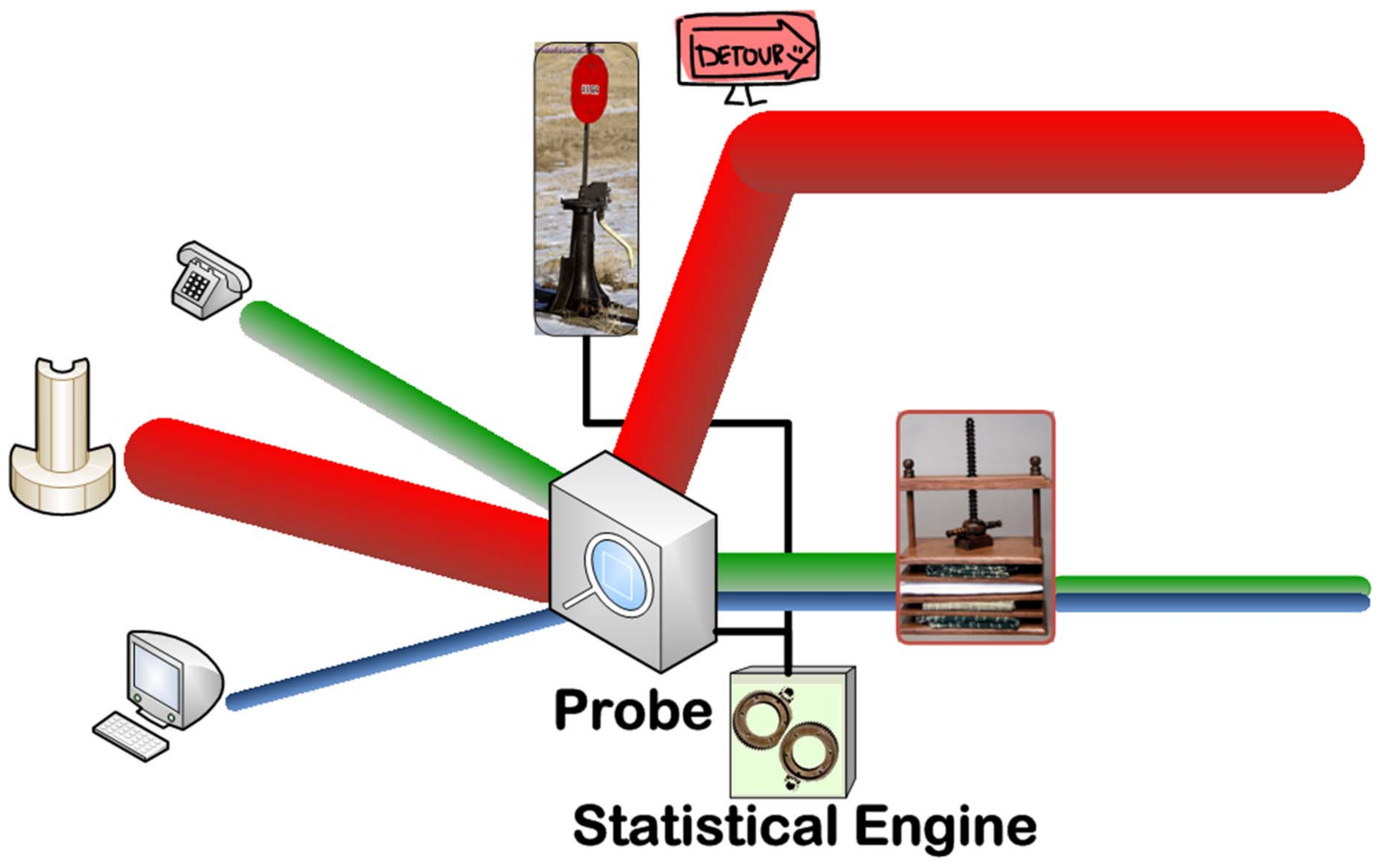
Scenario 1: STAMPEDE



Middleware for Monitoring and Troubleshooting of Large-Scale Applications on National Cyberinfrastructure

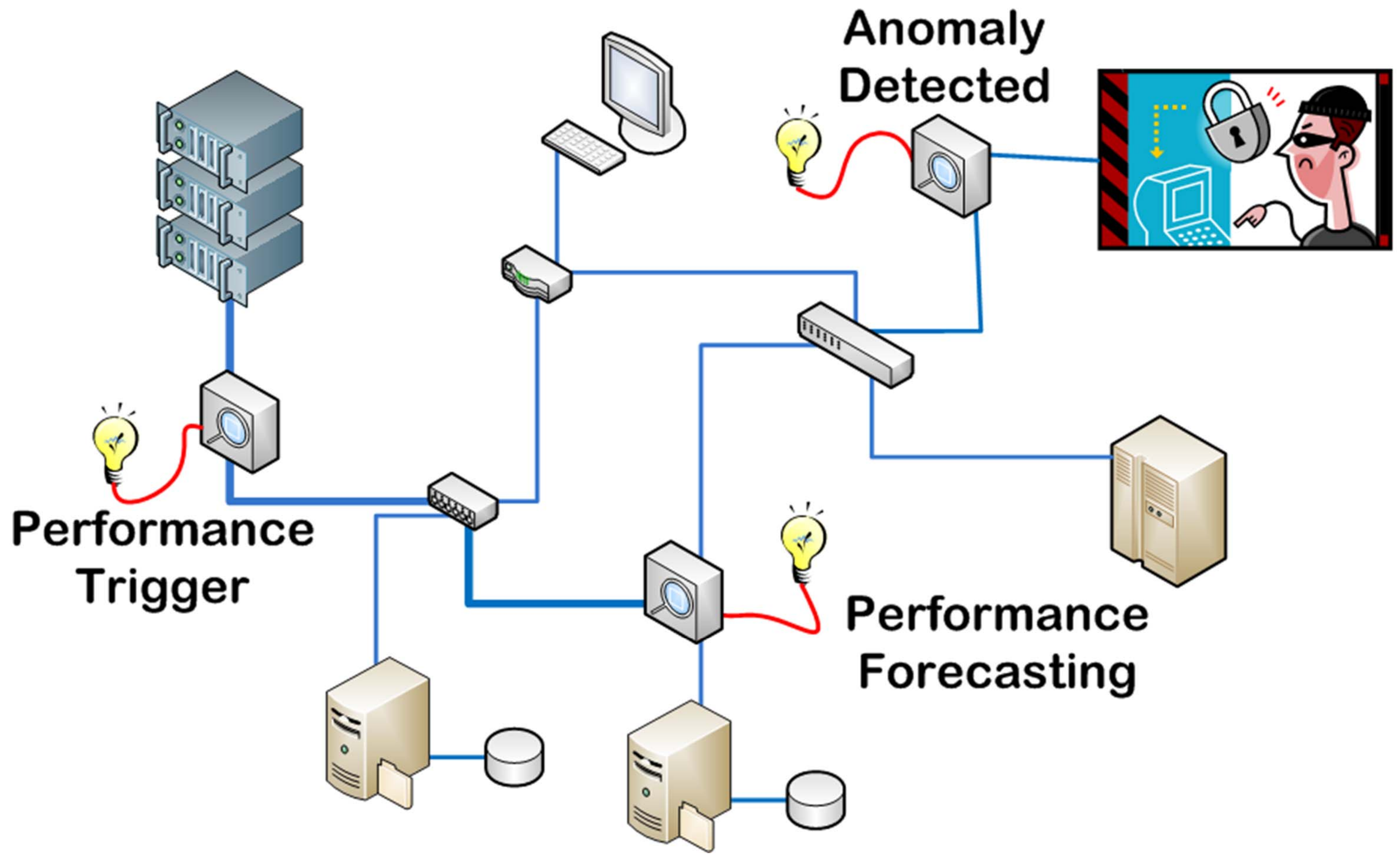


Scenario 2: Computer Network



- **Scenario 1:** *Pegasus Workflow Management System*, through the project *Synthesized Tools for Archiving, Monitoring Performance and Enhanced DEbugging* (STAMPEDE), where the applications generate scientific workflows.
  - Proven to provide a reliable and efficient platform for the execution of complex scientific workflows.
  - Executes analyses on the the TeraGrid and the Open Science Grid.
  - Supported applications: astronomy, bioinformatics, earthquake sciences, gravitational-wave physics, and others.
  - Uses Netlogger toolkit to normalize and correlate the flood of log information.
  - A reliable, efficient, general-purpose infrastructure for collecting end-to-end monitoring information (application, middleware, and network) is under development: STAMPEDE.
- A version of the Statistical Engine is being constructed for anomaly detection.
- Here, the Statistical Engine uses techniques such as *mean standard deviations* (MSD) and *cumulative distribution function* (CDF).
- The detector may trigger alarms of anomalies detected among tasks durations.
- When a historical high demanding flow appears, the engine recognizes it and takes action, for example, triggering a Virtual Circuit.

- **Scenario 2:**
  - The probe monitors flows network flows through CISCO Netflow data.
  - The Statistical Engine currently uses Finite State Automata techniques to detect flows by their duration.
  - When a historical long duration flow appears, the engine recognizes it and takes action, for example, triggering a Virtual Circuit.


Anomaly Detected
Performance Trigger
Performance Forecasting