GreenLight Project

Towards an Energy-Aware Service-Oriented Architecture for GreenLight

July 1st, 2009

Ingolf Krueger,

Claudiu Farcas, Filippo Seracini

Outline

- GreenLight overview
- Challenges and motivation
- Our contribution
- Towards energy efficiency
- Service Oriented Architectural Blueprint
- Mapping to Deployment
- Future work

NSF Award to UCSD for $2M (equipment)

Objectives:

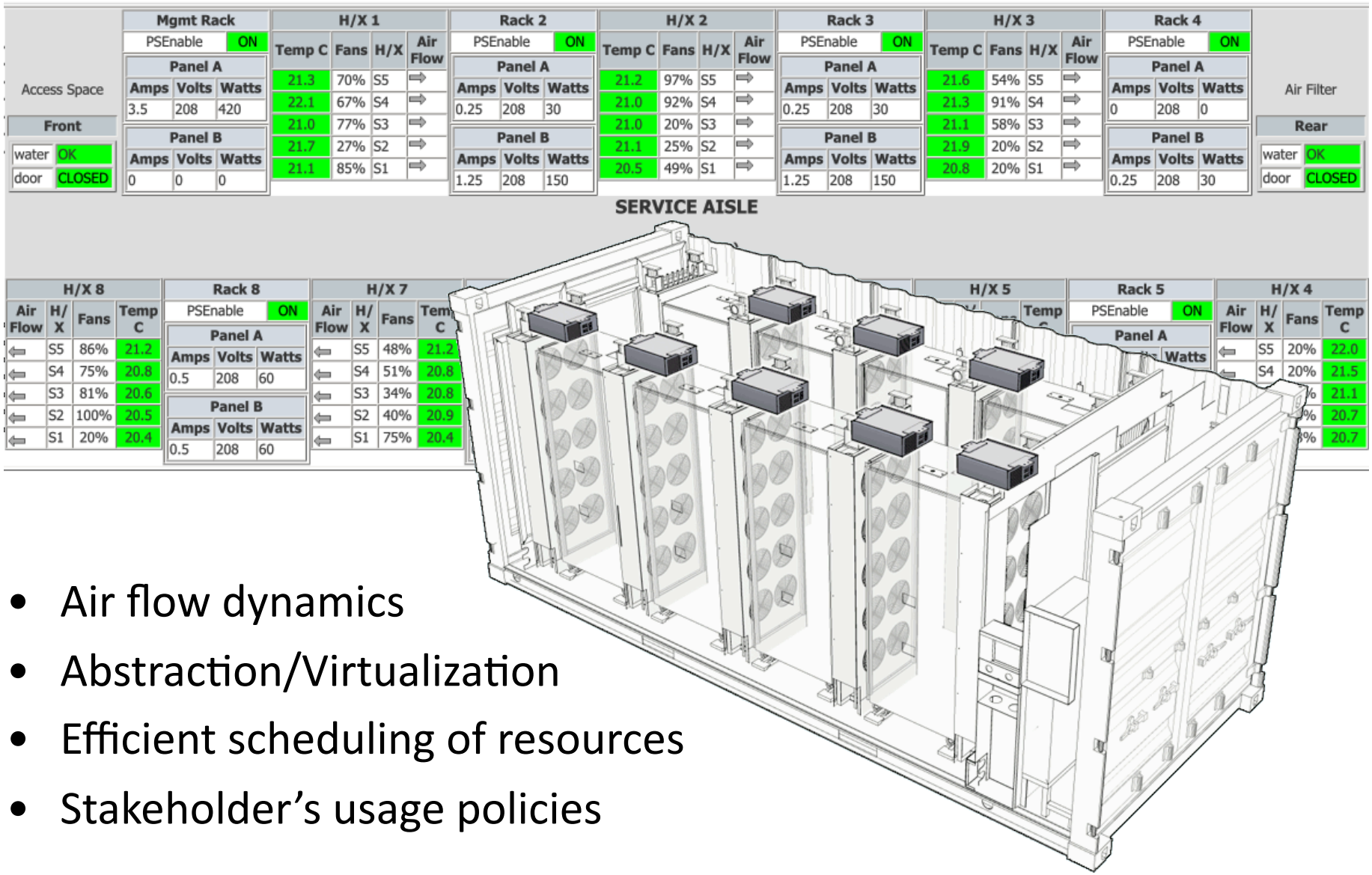
- Understand energy consumption related to task execution
- Create an infrastructure that allows to decrease the environmental impact of computation
- Provide to users different modes of computation (i.e. max performance, max energy saving, min computational cost, etc)

Motivation

- Industry
  - Microsoft's $500 M new data center in Chicago
    - 220 containers
    - Up to 550,000 servers
  - Google
    - patented datacenter-in-a-shipping-container
    - Water-based data center (waves powered)
- Government
  - Data center energy efficiency research part of the US stimulus package
- Community
  - GreenLight got CENIC Experimental/Developmental Applications 2009 Award

Inside the Blackbox

Challenges



- Air flow dynamics
- Abstraction/Virtualization
- Efficient scheduling of resources
- Stakeholder's usage policies

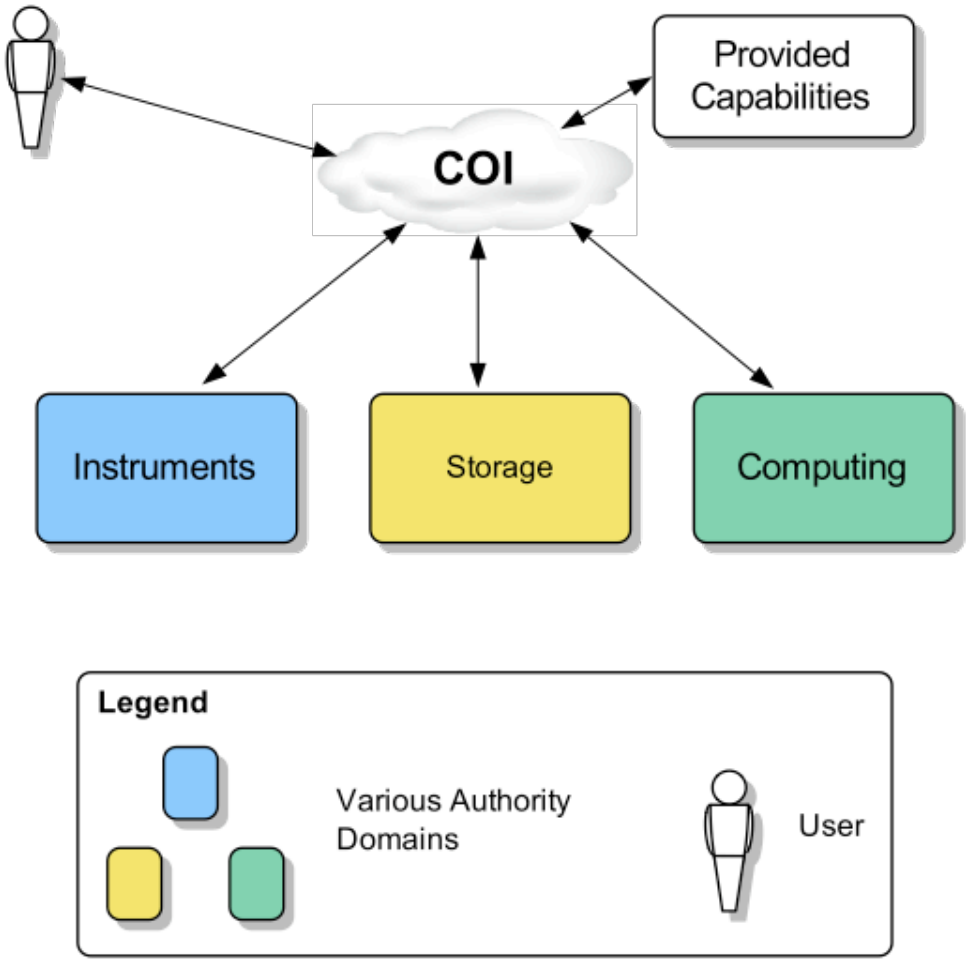
Our Team's contributions

- Create a SOA-based cyberinfrastructure to:
  - Manage and control the *Blackbox*
  - Run scientific experiments (various tasks)
  - Provide green data related to task execution
  - Apply strategies to improve energy consumption, minimize thermal footprint, reduce noise, etc

Cyberinfrastructure

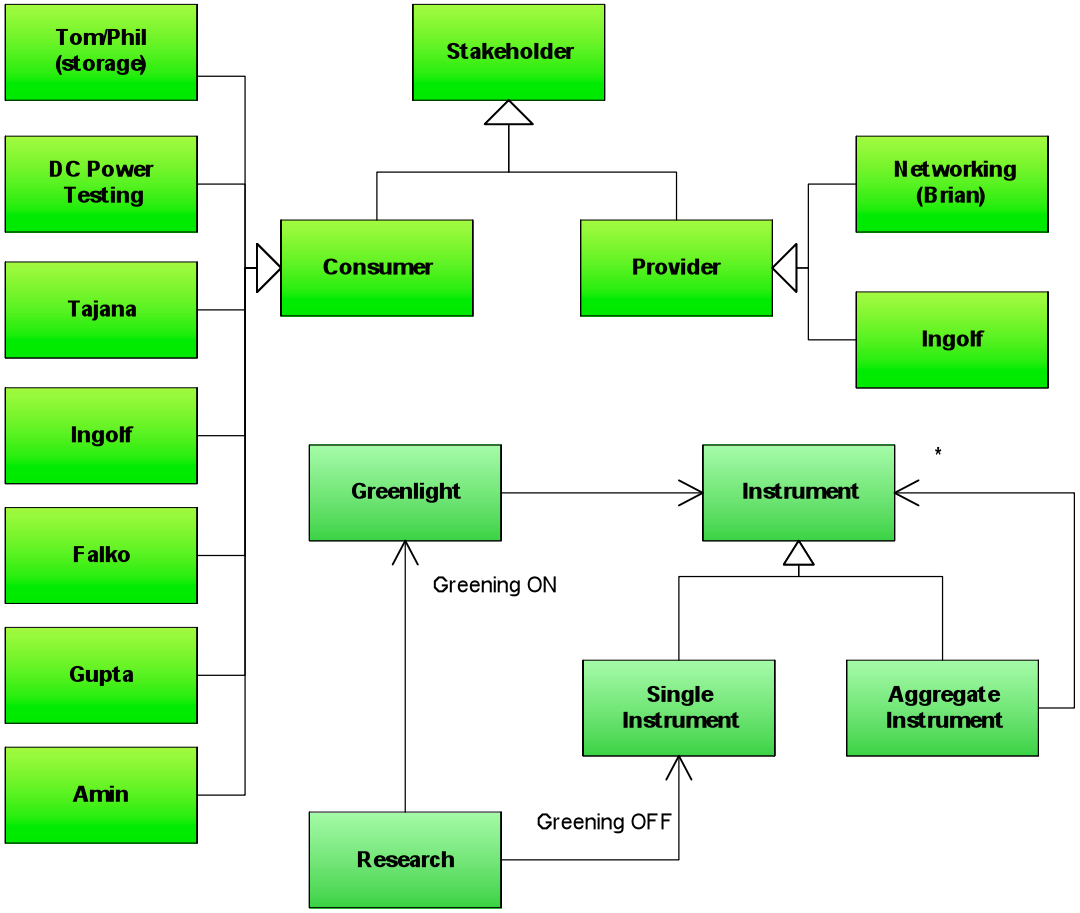
- **Scenario**: A scientist is trying to setup up a facility out of resources (instruments, computing capabilities, storage) spread out over a variety of authority domains

- **Challenges**
  - Resource discovery (instruments, storage, computation)
  - Resource access (seamlessly across infrastructure)
  - Resource Model (adding/ removing an instrument, ...)
  - Authentication, authorization, and other policies,
  - Governance
  - Capability presentation



Requirements Engineering

- GreenLight Researchers are interested in both producing and consuming greening data such as temperature and power measurements

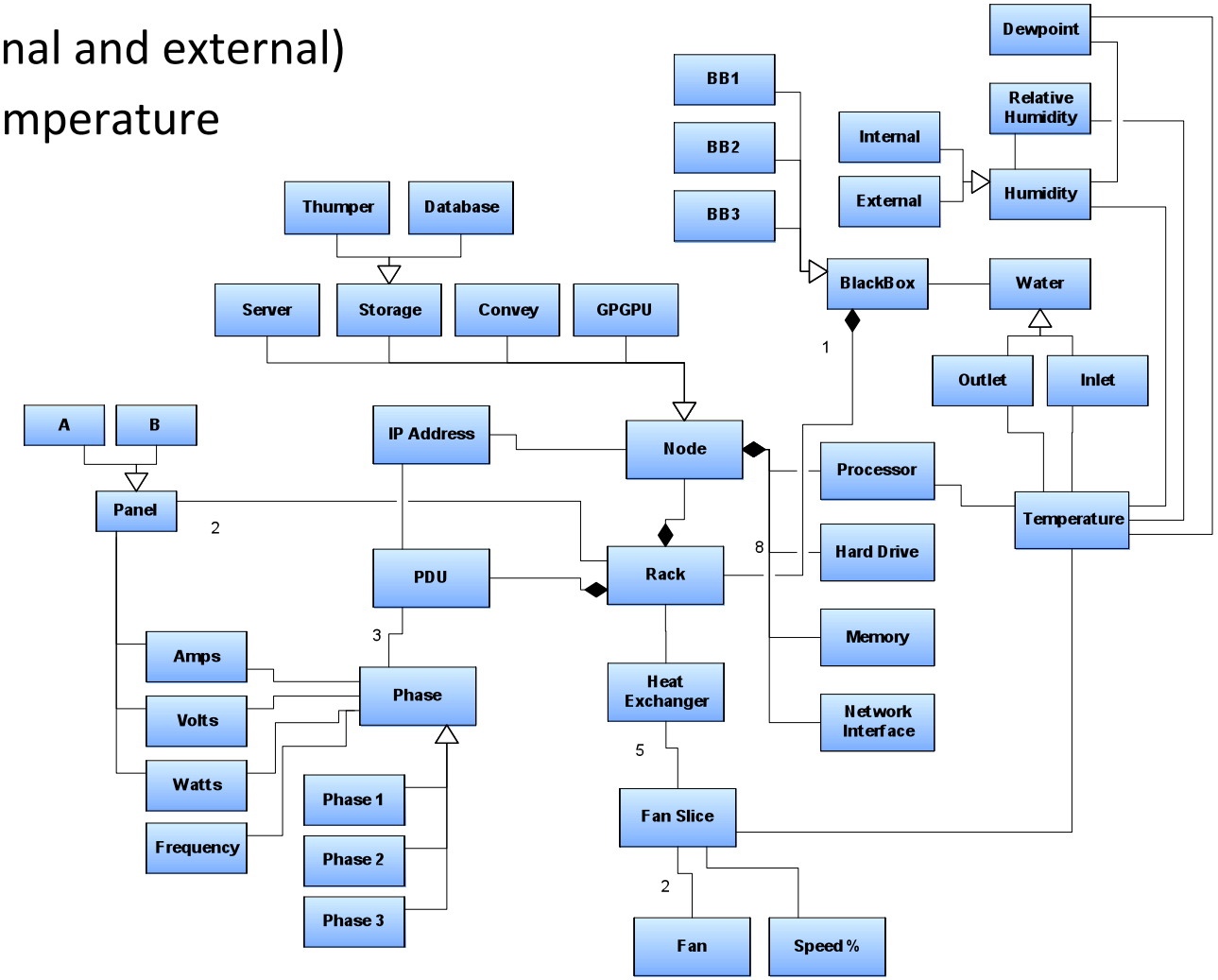


A few important questions:

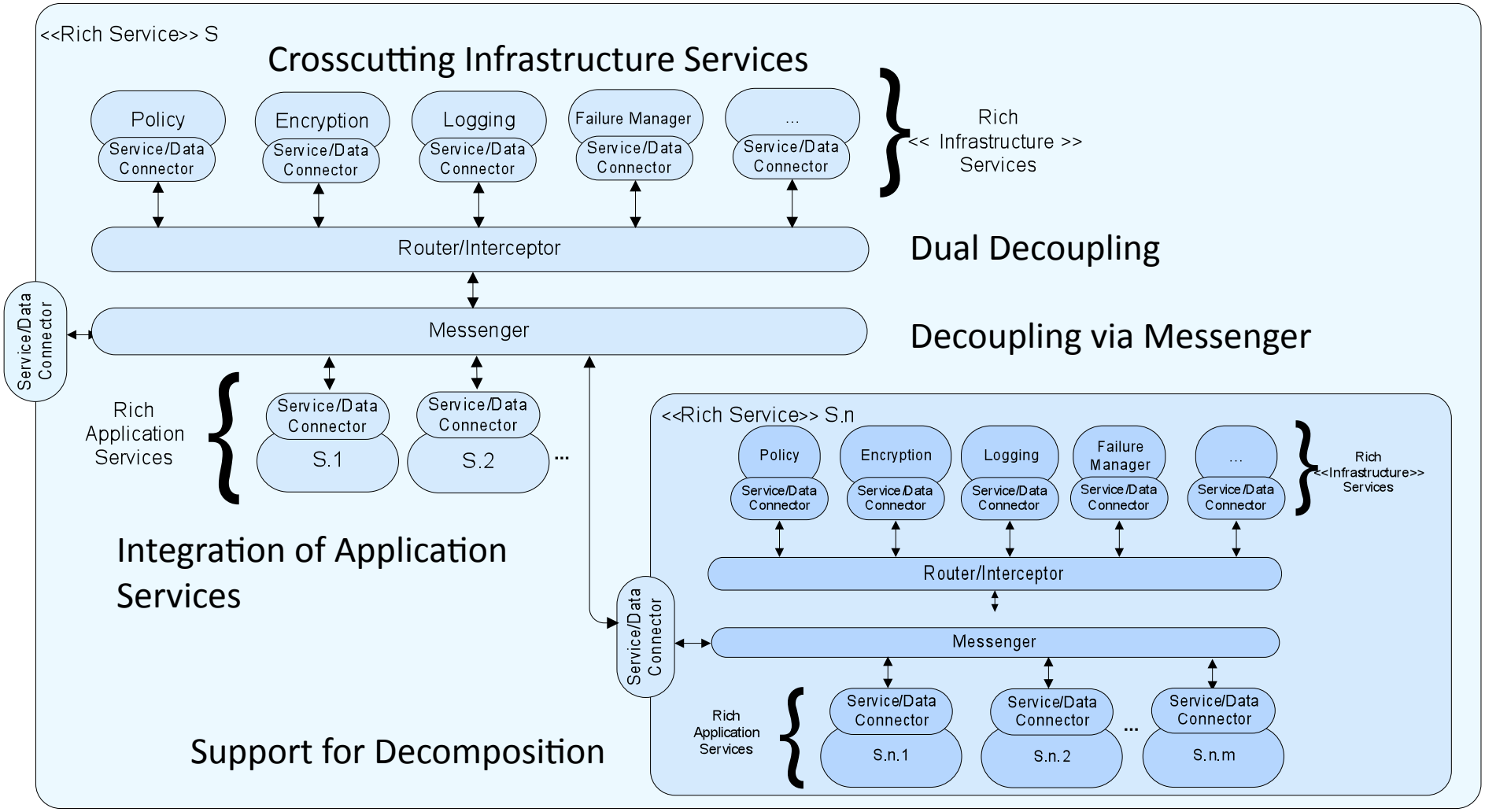
- What are the data sources?
- What can be measured?
- How is data stored?
- How is data represented?
- Who wants what?
- How to share data?
- How to best use data?
- Strategies to optimize power consumption?

Domain Modeling

- Multiple data collection points
  - Air temperature (40 sensors for rack-level air, hundreds internal)
  - Humidity (internal and external)
  - Intake water temperature
  - Power usage
  - Fan speeds



Architectural Blueprint



Recursive pattern as integration strategy for GreenLight components

Rich Services – Core

- Main entities of the architectural blueprint
  - Service/Data Connector - interaction between the Rich Service and its environment
  - the Messenger and the Router/Interceptor - communication infrastructure
  - Rich Services - encapsulate various application and infrastructure functionalities
- Rich Application Services
  - interface directly with the Messenger
  - provide core application functionality
- Rich Infrastructure Services
  - interface directly with the Router/Interceptor
  - provide infrastructure and crosscutting functionality
  - Examples: policy monitoring/enforcement, encryption, authentication

Mapping to Deployment (1)

- Provisioning of computational resources
  - Choosing an appropriate infrastructure resource management platform: Rocks, Perceus, OpenQrm, OpenNebula, Eucalyptus (EC2)
  - Job dispatcher: SGE, Condor
  - Execution of scientific workflows: Pegasus
- Provisioning of storage resources
  - Localized vs. Distributed file system (e.g., Thumpers vs. local hard drives)
  - Analyze tradeoffs between bandwidth, performance, power consumption etc.

Mapping to Deployment (2)

- Data
  - Collection: Intermapper
  - Data storage: postgreSQL
  - Data model to store/organize power related data: XDR, SDXF
  - Data model to communicate such data: DAP, XML/SOAP
- Control Models
  - What can be controlled and how (i.e. fan, cpu speed)
  - Algorithms under development by Tajana's group
- Application integration
  - ESB strategy: Mule, ServiceMix
  - Message Oriented Middleware: AMQP, Jabber/XMPP

Next Steps

- Models for "green data" for various applications: Proteomics, Ocean Observatories, Software Engineering, …
- Resource model for GreenLight resources (e.g., CPUs, VMs, nodes, etc)
- Usage policies and adequate scheduling algorithms to improve efficiency
- Deployment of SOA-based infrastructure to use and manage the *Blackbox*
- Expose the "green data" as web services & portal

The End

Thank you!