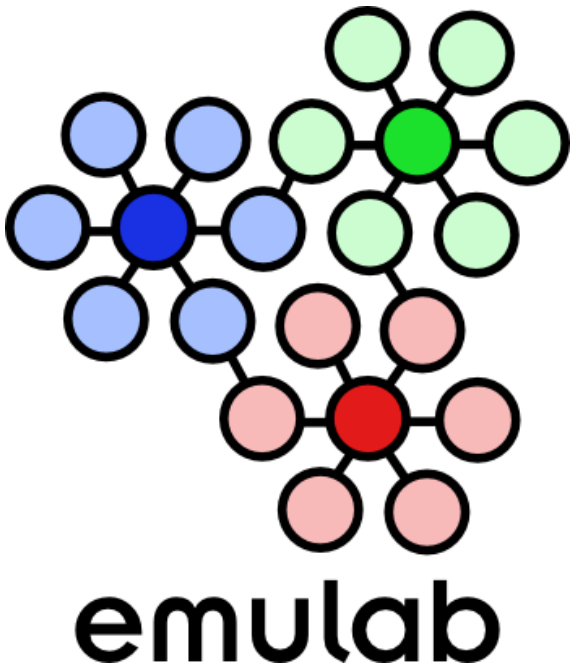
Integrated Scientific Workflow Management for the Emulab Network Testbed



Eric Eide, **Leigh Stoller, Tim Stack, Juliana Freire, and Jay Lepreau**

University of Utah, School of Computing

USENIX 2006 / June 3, 2006

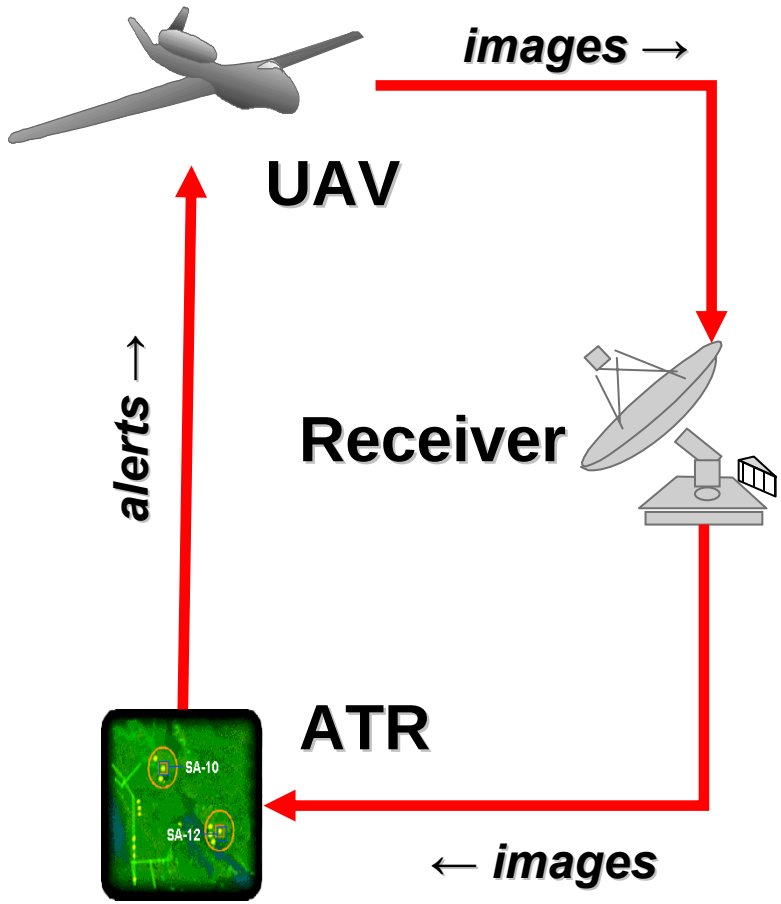
This Talk in One Slide

- Current network testbeds
  - *…manage the "laboratory"*
  - *…not the experimentation process.*
- → A big problem for large-scale activities!

- Evolve Emulab for experiments based on scientific workflows
  - *Big mutual benefits: testbed ↔ workflow*
  - *Work in progress*

Example: UAV Simulation

- A distributed, real-time application
- Evaluate improvements to real-time middleware
  - *vs. CPU load*
  - *vs. network load*
- ***4 research groups***
- ***x 19 experiments***
- ***x 56 metrics***

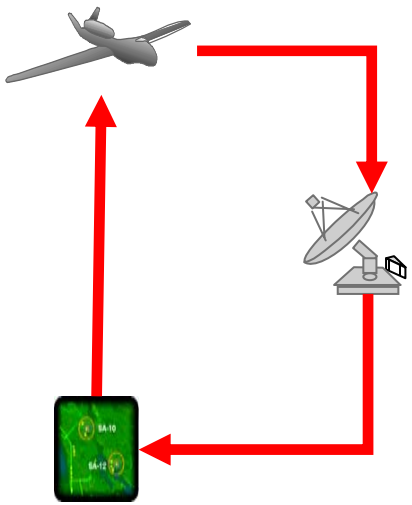


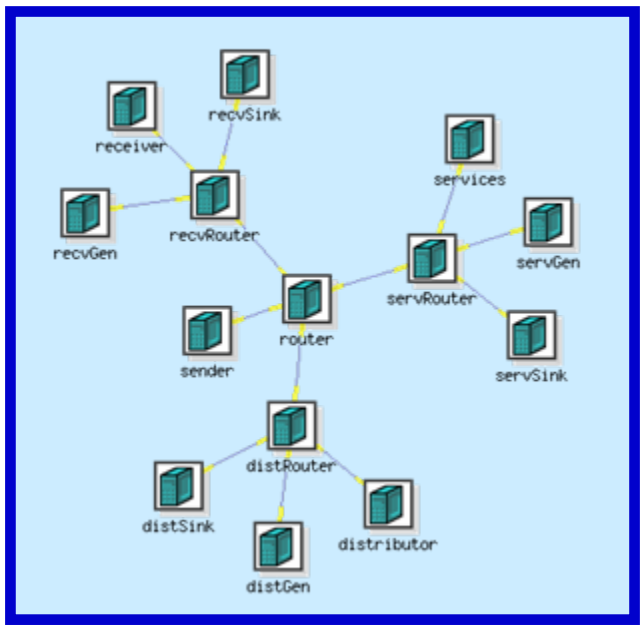
Use Emulab

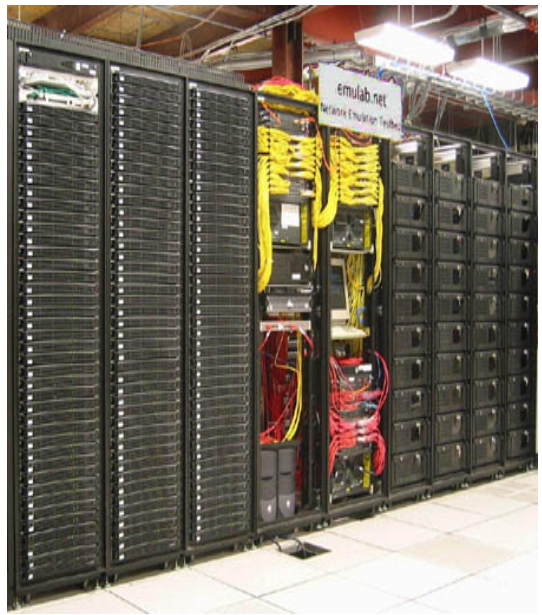
write "ns" file

"swap in"

Concept → **Experiment** → **Emulate**

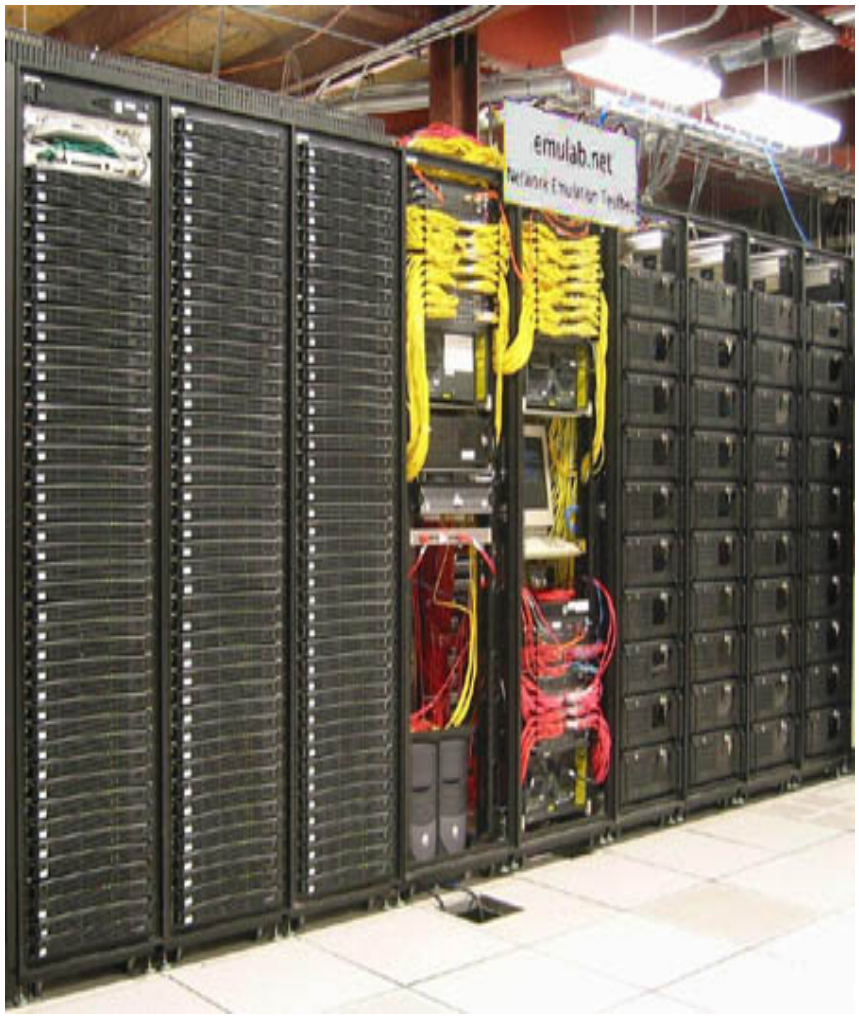




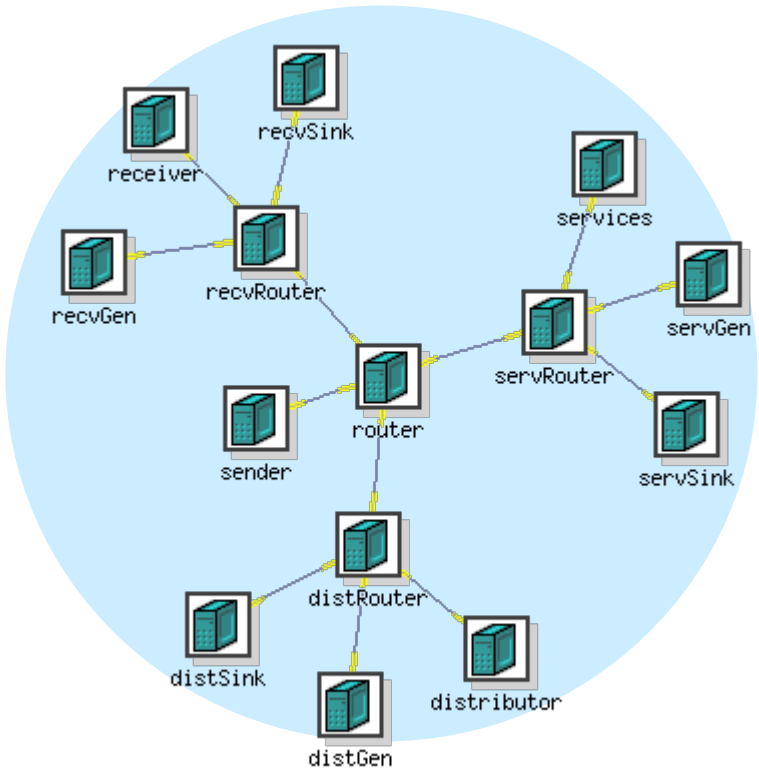


Problems Solved

- I get machines!
  - *328 PCs, and more*
  - *Time- & space-shared*
  - *Loads OS and software*
- I get network!
  - *Config. topology & quality*
- I get to collaborate!
  - *Available to researchers and educators worldwide*
  - *File storage, email, …*

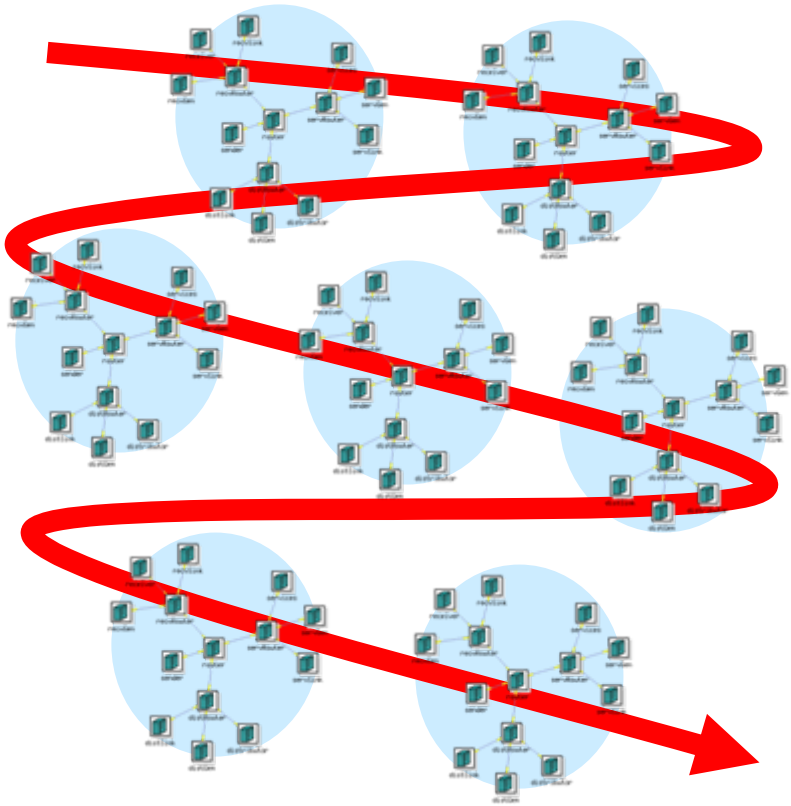


Problems *Not* Solved



- **"Now what?"**
- Getting off the ground
  - *Run all my software*
  - *Add instrumentation*
  - *Collect all my data*
  - *Analyze it*
- Scaling up
  - *19 configurations*
  - *Automation*

More Problems *Not* Solved



- **"How did I get here?"**
- Over the short term…
  - *"Where are the results I got last week?"*
  - *"How did I get those results anyway?"*
  - *"What if…?"*
- …and the long term
  - *Reproducing results*
  - *Reusing artifacts*

Idea: Scientific Workflow

- Managing activities, inputs, and outputs is the job of a ***scientific workflow system***

- ***Our approach:*** evolve Emulab with integrated support for scientific workflows
  - *Build on existing abstractions & mechanisms*
  - *Resource focus → user & task focus*
  - *Users work "within" and "across" experiments*

Contributions

- Address demand + opportunity
  - *Users need to manage large-scale complexity*
  - *A symbiotic combination: leverage and impact*
- Advance the applicability of testbeds
  - *Not just Emulab — e.g., PlanetLab and DETER*
- Advance scientific workflow systems
  - *Exploit testbed capabilities — e.g., "total control"*
  - *Address testbed requirements — e.g., flexible use*

Issue: Encapsulation

- Current "experiment" model is not fully encapsulating
  - *Topology + static events*
  - *Need everything else!*
- **Challenge: specification**
  - *Complete and precise…*
  - *…w/o huge user burden*
- **Approach: be automatic**
  - *E.g., track files used*
  - *Snapshot, archive, restore*
  - *User can refine "extent"*

ns file

OSes

packages

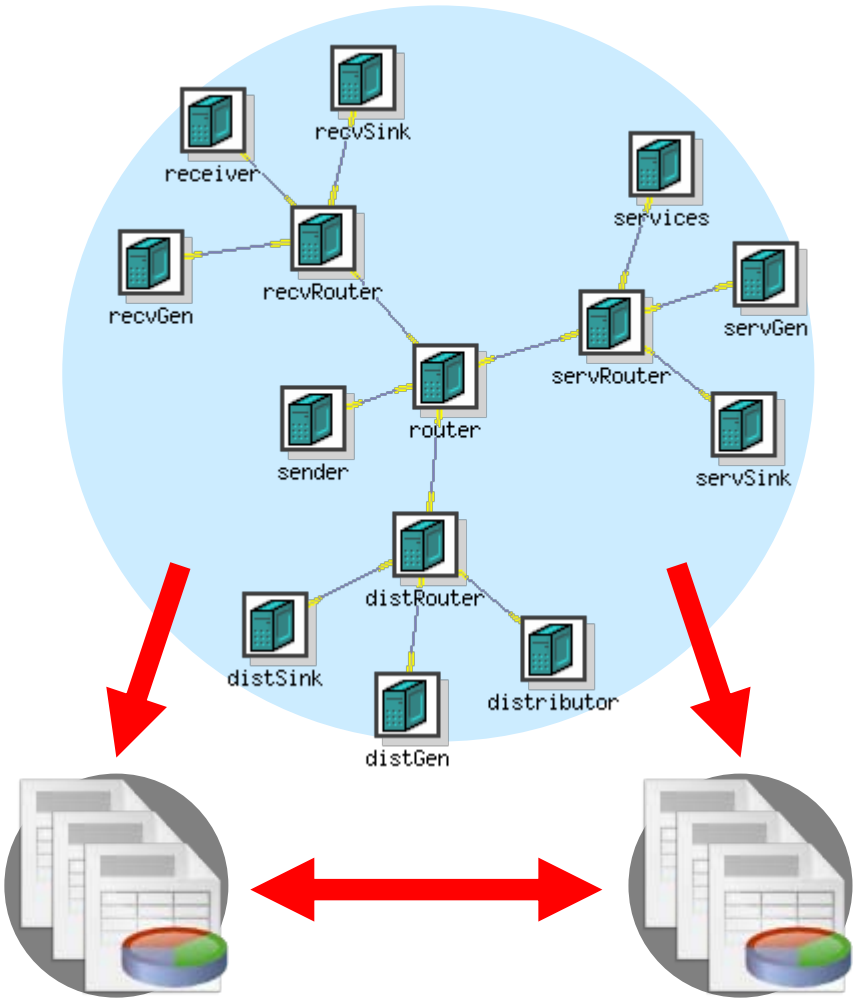
my software

inputs

outputs

NFS monitors	***Subversion repo.***
packet monitors	***datapository (DB)***
AJAX GUI	***research filesystems***

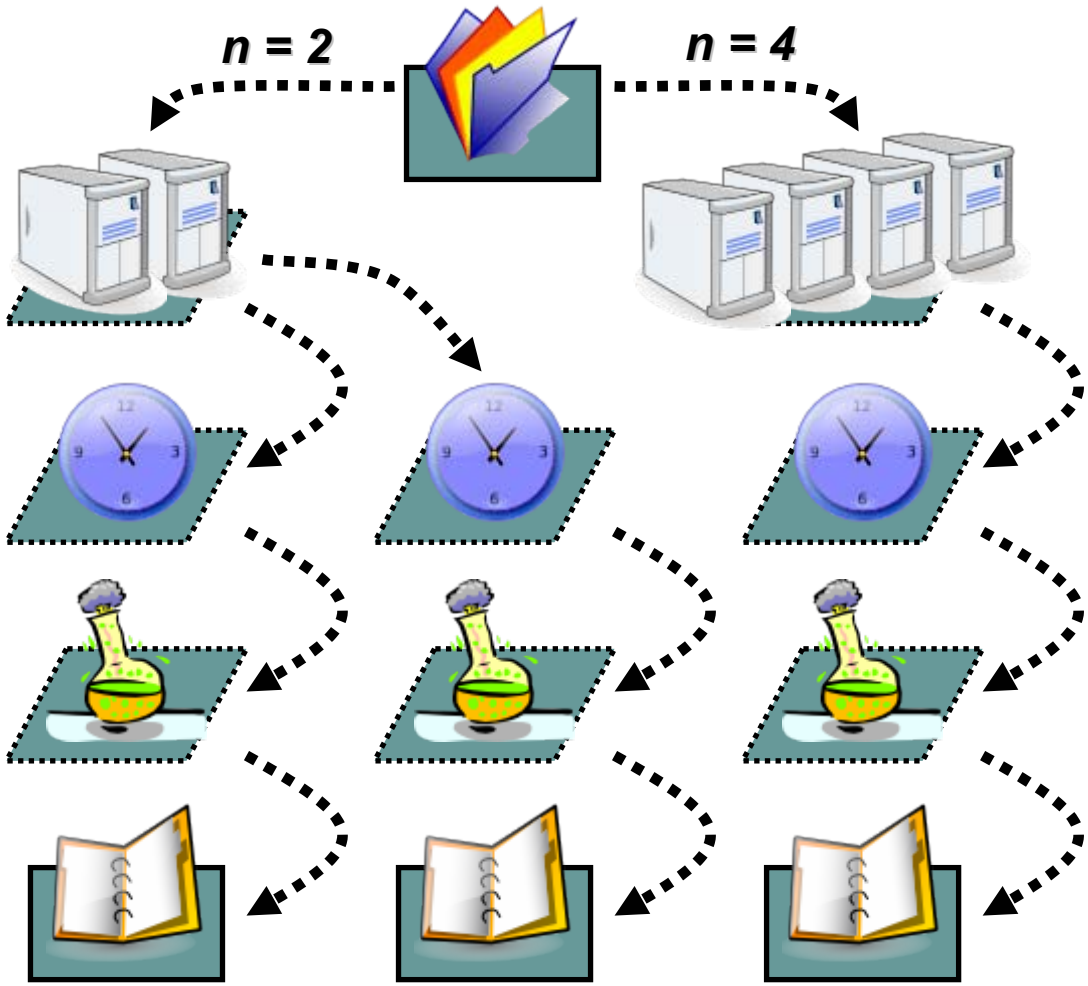
Issue: Definition vs. Execution



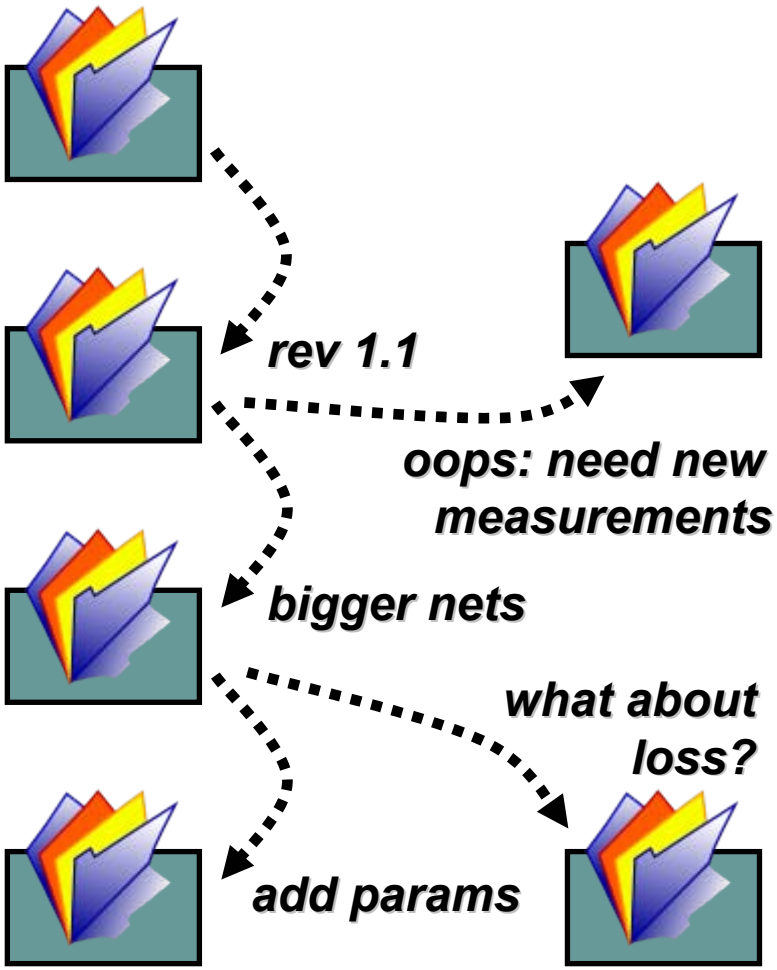
- Current “experiment” has multiple roles
  - *Definition*
  - *The thing that you run*
- **Challenge: representing relationships**
  - *Multiple runs of one setup*
  - *Similar configurations*
- **Approach: a new model of experimentation**
  - *Separate the roles*
  - *Evolve the new abstractions*

New Model

- Template
- Swapin
- Experiment
- Activity
- Record



Issue: History

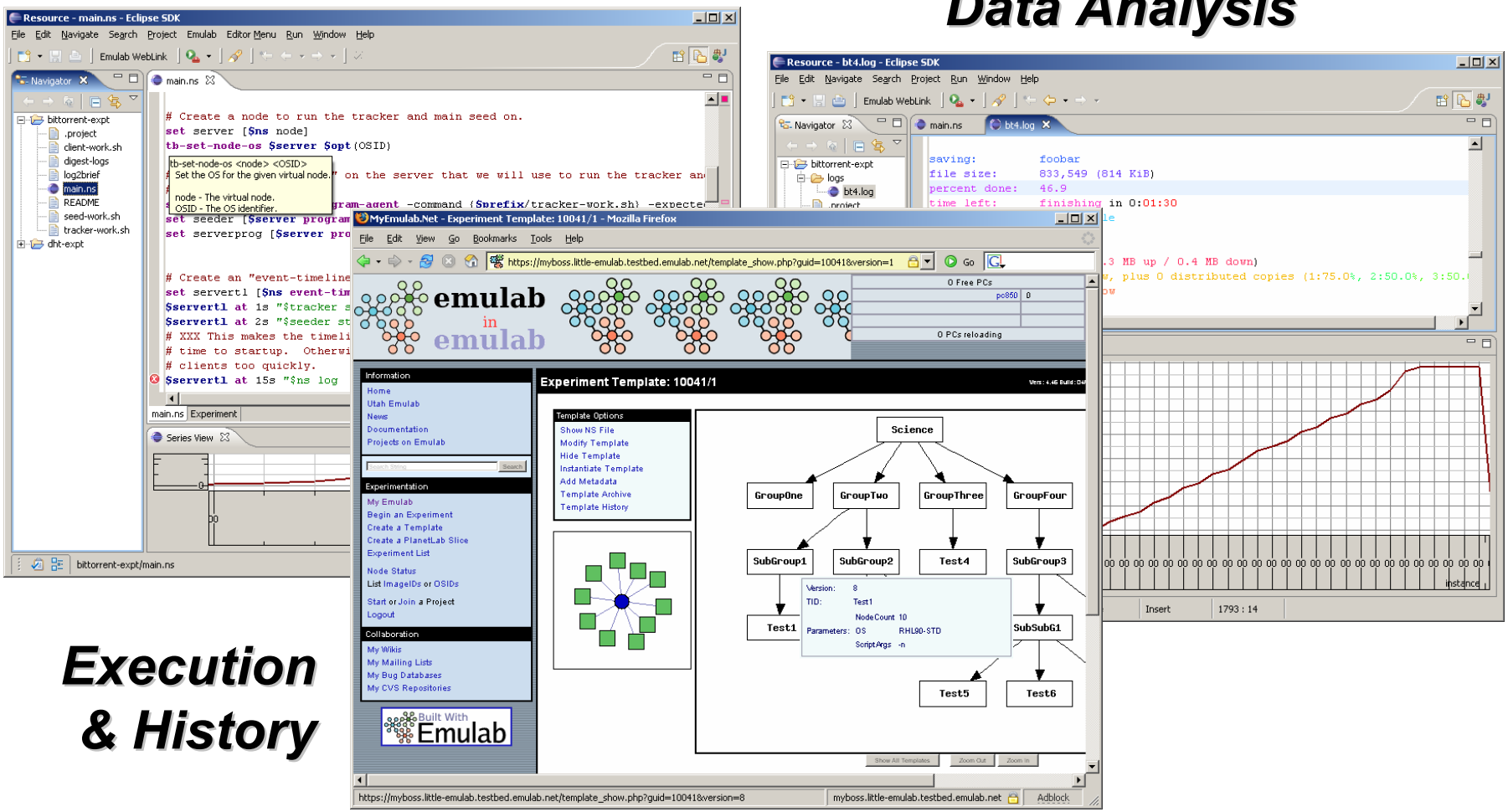


- Research and educational plans are dynamic
  - *By design & by discovery*
- **Challenge: safe exploration**
  - *Fork*
  - *Back up*
- **Approach: keep history & support temporal navigation**
  - *Keep template revisions*
  - *Track provenance*
  - *Locate, repeat, and reuse*

Implementation in Progress

Definition

Data Analysis

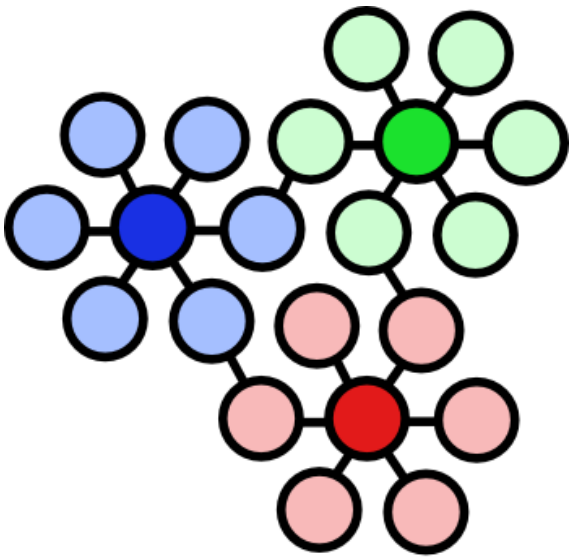


Execution & History

Conclusion

- Large and powerful testbeds
  - *…enable complex and large-scale activities*
  - *…lead to complex and large-scale workflow management problems*
- Integrated workflow management can leverage the strengths of testbeds
  - *Systems approach — and systems challenges*
- → Better testbeds and workflow systems

http://www.emulab.net/



Thanks!

Extra Slides After This Point