

Towards Collaborative Data Sharing “U.S. Perspective”

Karl Levitt

Department of Computer Science

UC Davis

Previous Affiliations: NSF, SRI International)

levitt@cs.ucdavis.edu

July 6, 2011

U.S. Perspective is Impossible

U.S. Government Agencies with Research Interest in Security

NSF, DHS, DARPA, IARPA, DoD, DOE, Treasury, NIH, NASA, FTC, NSA, CIA, FBI, Census, State, ...

150 Research Universities

Many Research Institutions

Many Companies

Many Consultants

Many Cyber Criminals

Many Victims

Outline

- Motivating examples
 - International data sharing is essential in ongoing incidents
 - Need an architecture, mechanized privacy policies, etc.
 - GENI: NSF's network testbed gets bigger and goes International
- Jim's provocative questions

Opportunities/Needs for International Cooperation (thanks to Vern Paxson)

- Much attack activity is **indiscriminant** \Rightarrow significant utility in sharing information via distributed sensors
 - With caveat that even so, perspectives are *not* homogeneous
- \Rightarrow Non-local defenses require international coordination
 - Whether proactive (e.g., anti-spoofing) or reactive
- \Rightarrow Incident response & forensics require international coordination
- Some facets of organized cybercrime appear to have national components (e.g., Mafia for country x)

Envisioning a Rich Inter-site Analysis for Cooperative Attack Mitigation

- Sites deploy *activity repositories* using common data format
- Site A can send request for analysis against activity seen by Site B
 - E.g. “have you seen the following access sequence?”
 - Done by sending an *analysis program*
 - Note: due to co-aligned threat models, it's often in B's interest to investigate
- B runs query against their repository ...
 - ... can also install **same** query against **future activity**
- B decides what (**sanitized**) results to return to A
 - If request was unreasonable, B can **smack** requestor

Fundamental Premise

- Modern *serious* attacks often manifest
 - Over a range of time scales
 - Involving numerous system components
- Serious =
 - E.g. stolen credentials
 - E.g. insiders
- Detecting these requires broad **visibility**
 - Across time (into the past; *looking to the future*)
 - Across space (different forms of sensing; *inter-site*)

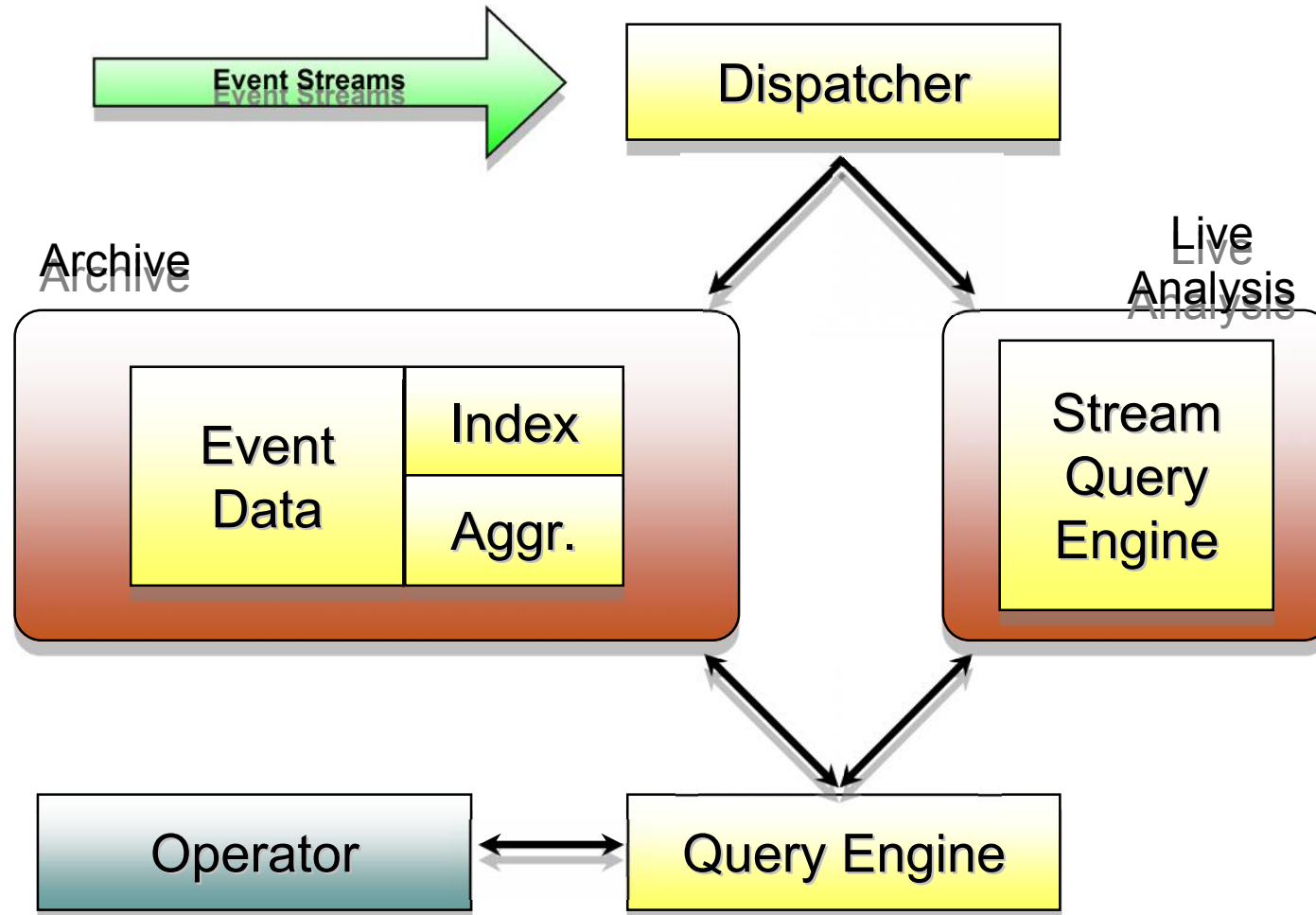
Realizing Visibility: Data Guidelines

- Data breadth:
 - Application logs, IDS, routers, firewalls, syslog
 - External information
 - **Policy-neutral** data
 - Do **not** pre-suppose good/bad judgment
- Uniform data model:
 - Asynchronous, typed events
 - Encompasses different detail/semantic level
 - E.g., “packet seen”, “TCP connection begun”,
“URL U fetched from server S by client C”
 - Can aggregate group of events into new event

Data Guidelines, con't

- Maintain extensive history:
 - Initial capture triage (e.g., heavy-tail cutoff)
 - Aging mechanisms distill older data into coarser info (e.g., packets → flow records) rather than discard
 - Via event aggregation
- Sanitization:
 - Presentation: keep operators from inadvertently tripping over sensitive material
 - Underlying: prevent leaks (e.g., subpoena)
 - Must consider both indiv. data & in aggregate

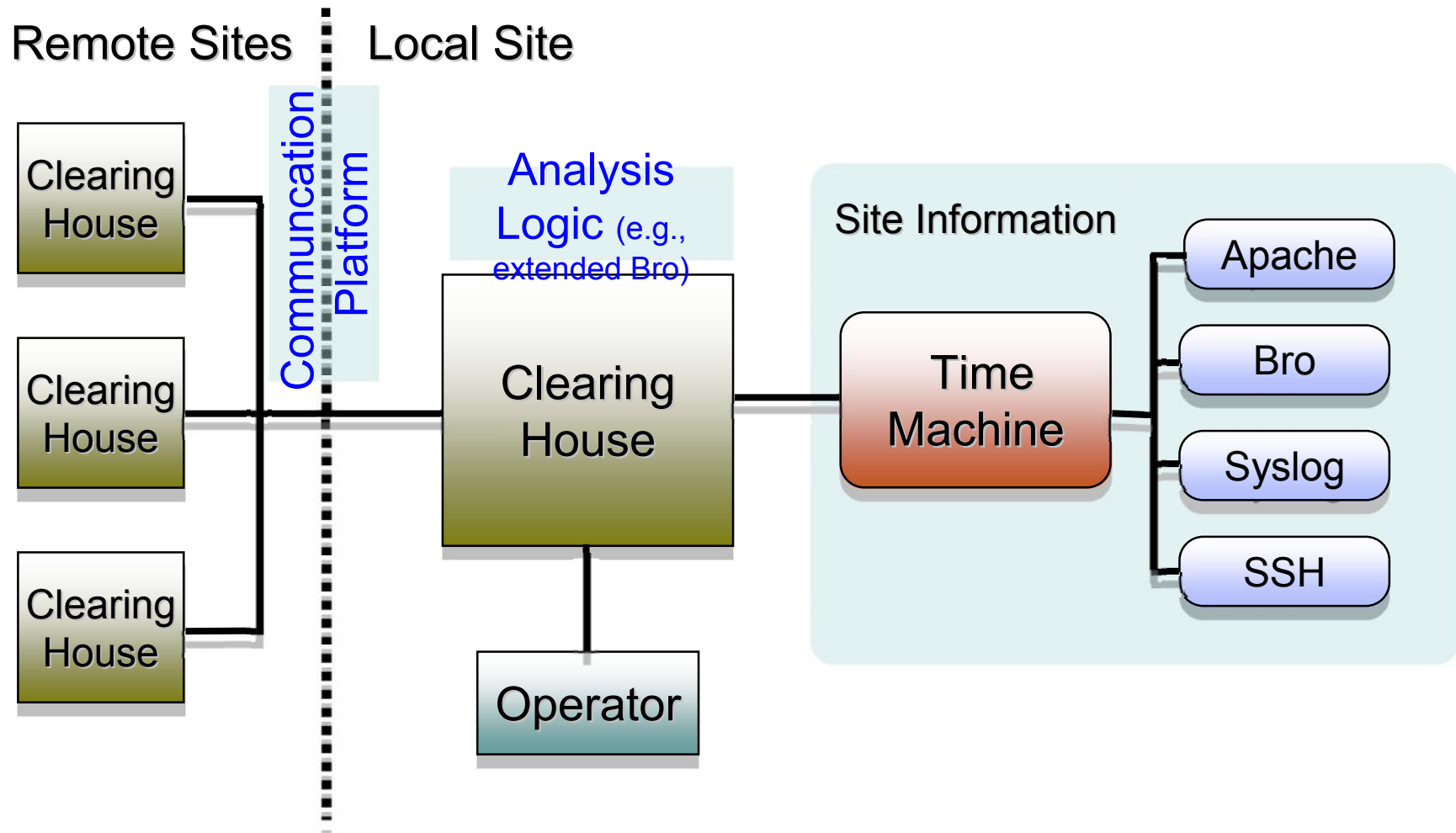
Envisioned System Architecture



Going Beyond A Single Site

- Premise: global log-sharing (e.g. DShield) fundamentally limited by issues of **trust**
 - Leakage of sensitive information
 - Poisoning by bad actors
- Underdeveloped **sweet spot**: sharing between sites with
 - Co-aligned threat models for **high-cost events**
 - E.g. credential-sharing
 - Functional administrative ties
 - If remote site misbehaves, there's someone to complain to
- Such sites *already work together today*
 - But it's **crude**: telephone calls, email, informal sketches

Clearing House Architecture





GENI

Exploring Networks of the Future

Now going live across the US!
Thanks to Chip Elliott -- BBN

GENI Project Office
July 2011
www.geni.net

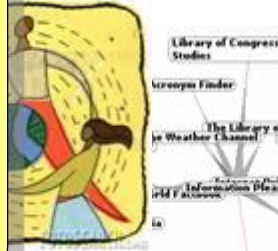
Outline

- GENI – Exploring future internets at scale
- Introducing GENI: an example
- GENI's growing suite of infrastructure
- Experiments going live across the US!
- What's next for GENI?
- GENI and US Ignite
- How can you participate?

GLOBAL NETWORKS ARE CREATING extremely important new challenges

Science Issues

We cannot currently understand or predict the behavior of complex, large-scale networks

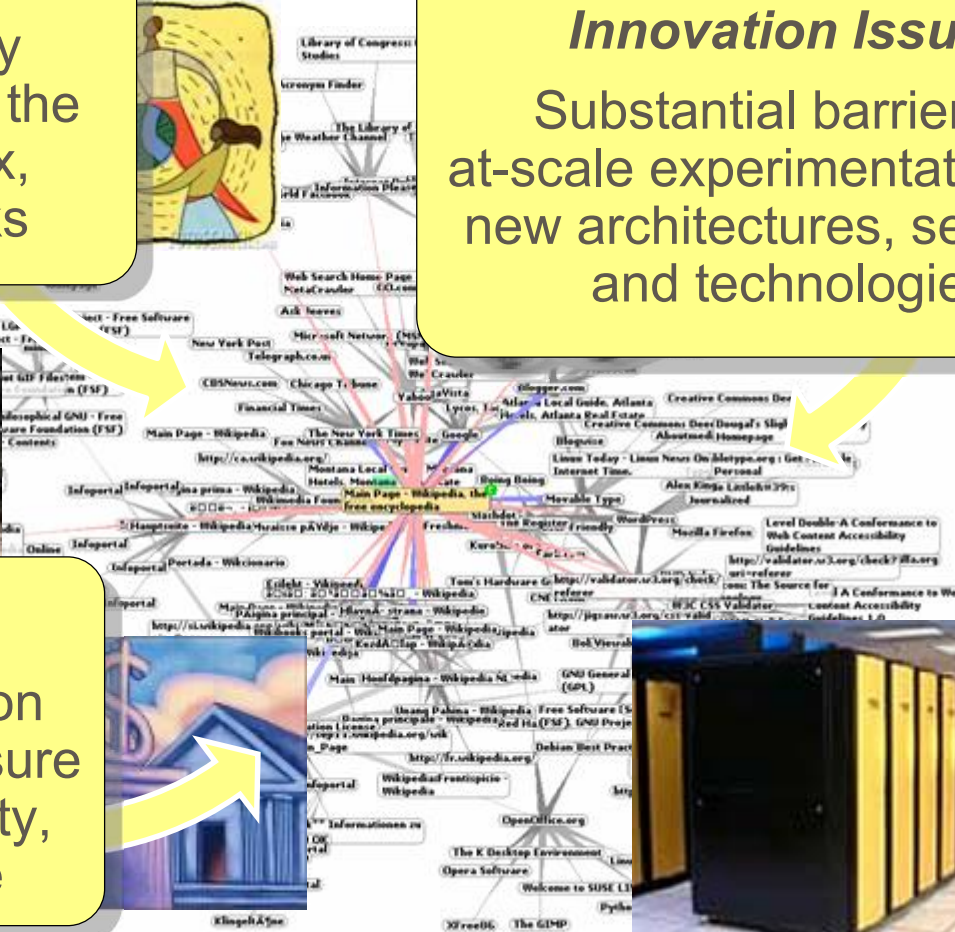


Innovation Issues

Substantial barriers to at-scale experimentation with new architectures, services, and technologies

Society Issues

We increasingly rely on the Internet but are unsure we can trust its security, privacy or resilience



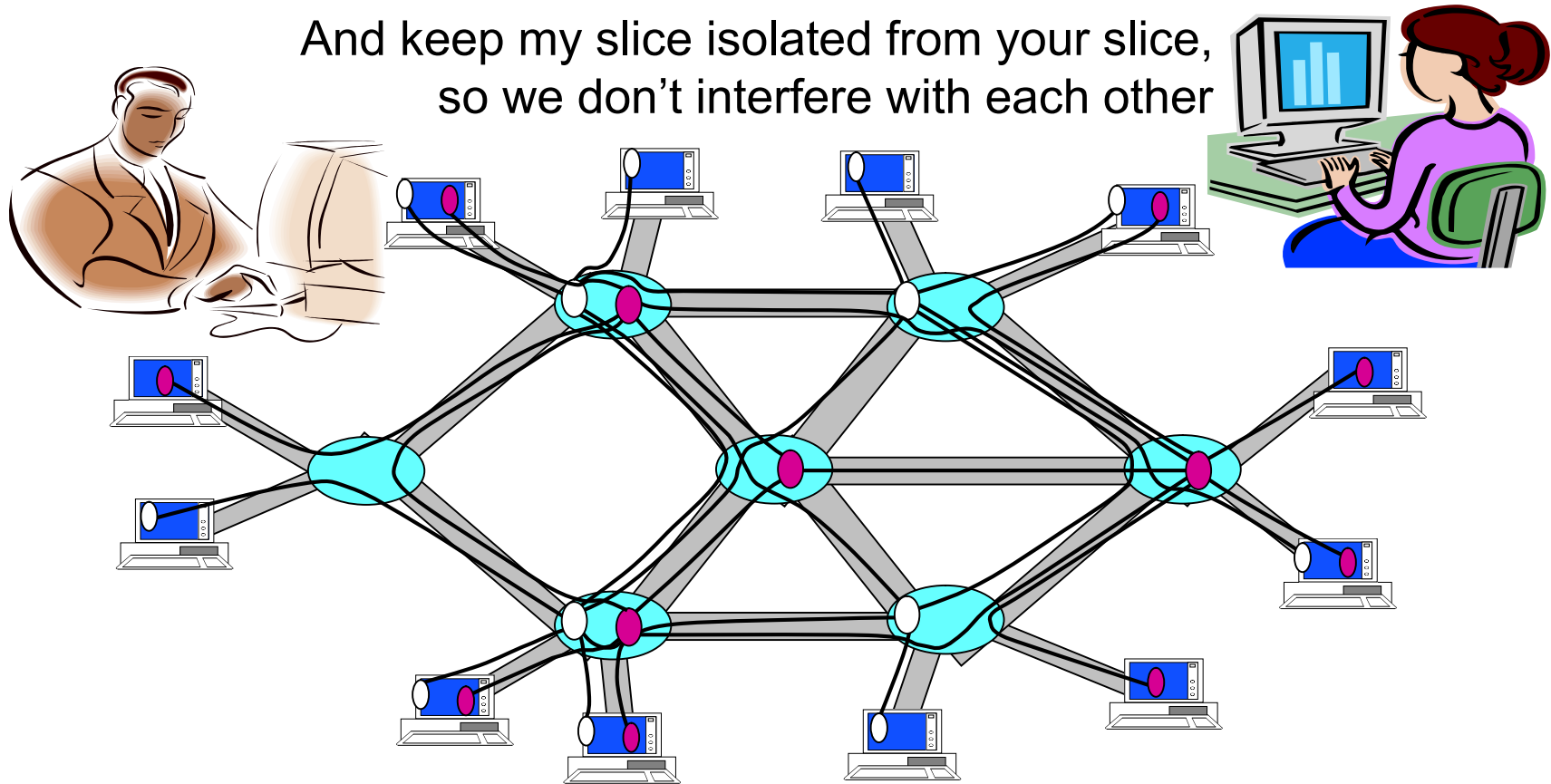
What is GENI?

- GENI is a virtual laboratory for **exploring future internets at scale**, now rapidly taking shape in prototype form across the United States
- GENI opens up huge new opportunities
 - **Leading-edge research** in next-generation internets
 - **Rapid innovation** in novel, large-scale applications
- Key GENI concept: slices & deep programmability

Revolutionary GENI Idea Slices and Deep Programmability

Install the software I want *throughout* my network slice
(into firewalls, routers, clouds, ...)

And keep my slice isolated from your slice,
so we don't interfere with each other



We can run many different “future internets” in parallel

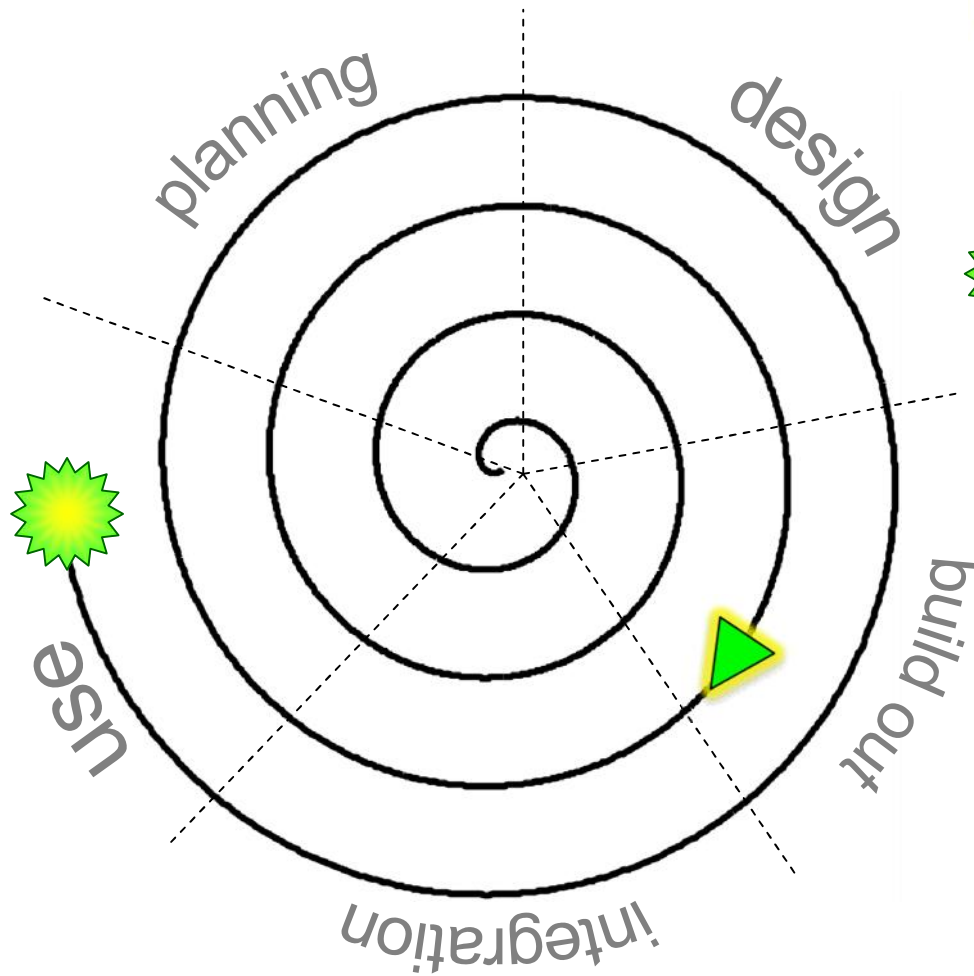
GENI IS NOW GOING LIVE ACROSS the US ; soon Internationally

GENI enabling testbeds, campuses, and backbones





Spiral Development

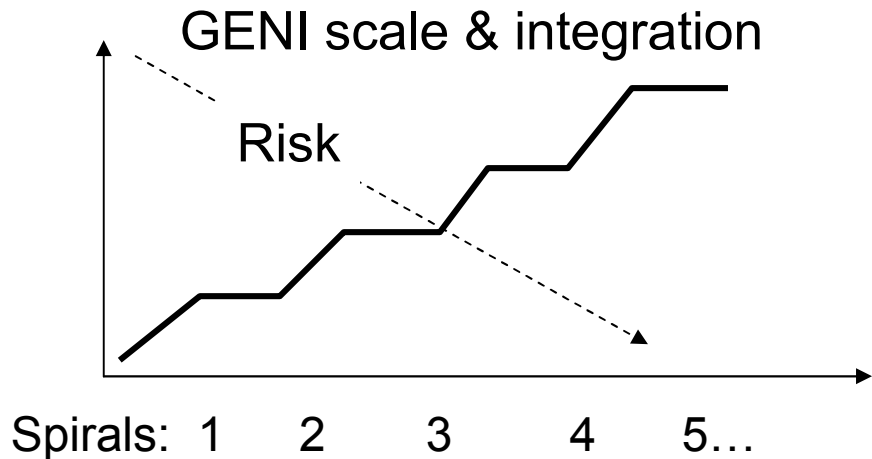
GENI grows through a well-structured, adaptive process



GENI Prototyping Plan

 **GENI Spiral 3**
Early experiments, meso-scale build, interoperable control frameworks, ongoing integration, system designs for security and instrumentation, starting up operations.

 **Envisioned ultimate goal**
Large-scale distributed computing resources, high-speed backbone nodes, nationwide optical networks, wireless & sensor nets, etc.



The organiser's BIG Questions

What are we doing and why?

Developing a framework for data sharing

To support productive and relevant research

What are expected impacts

Faster and more accurate response to attacks

Better security products and services

What kind of data should we share?

Attack data

Normal activity

What kind of collaboration do we need?

Different data sources

Reverse engineering protocols used by underground cyber world

Agreement on overall architecture

The organiser's BIG Questions (cont.)

What kinds of analysis do we need?

Data sanitization policies, implementations

What are the incentives to participate?

An effective global defense posture

Economic: Build a new industry based on collaborative defense

What are the risks?

Data sharing substrate is a vulnerability attackers can exploit

Major privacy breaches – more than today?

We might actually succeed!