# Retro-Future Kick-off

**Cyber Security Division**
**2012 Principal Investigators' Meeting**
10 October 2012

PIs: Mike Fisk[1], *John Heidemann*[2],
Dan Massey[3], Christos Papadopoulos[3]

mfisk@lanl.gov, johnh@isi.edu, {massey,christos}@cs.colostate.edu

[1] Los Alamos National Laboratory
[2] *USC/Information Sciences Institute*
[3] Colorado State University

Copyright © 2012 by John Heidemann
Release terms: CC-BY-NC 3.0 unported

# the 0-day Challenge

in the future: all interesting security events
*will have already happened*

(interesting: like *0-day attacks* and *insider threats*)

pro-active security *will always fail* (eventually)
we know "interesting" *only after the fact*

# An unfortunate scenario…

- you're City Bank of Wells Chase
  - a list 1000 accounts from ATM *X* appeared on a hacker website two days ago
- you want to know
  - what computers talked to *X*? *(who got in?)*
  - what traffic was different recently? *(how did they get in?)*
  - was other information removed from *X*? *(what got out?)*
  - did *X* contact other internal computers? *(did it spread?)*
  - did any other ATMs see similar events? *(what else?)*
- process: gathering data, mitigation, understanding, prevention

# the Need: Post-Event Recovery and Understanding

- if security will fail (and it will)
  - 0-day attacks (by definition, not known in advance)
  - and insider threats (cannot be pre-emptively closed)
- we must support
  - forensics
  - recovery and mitigation
  - understanding what happened
- all *after-the-fact.* => we must unwind time
  - what happened? why? what was lost?
  - *understanding* will improve future prevention

# the Approach: an Internet "Tivo"

We plan an *Internet "Tivo"*: a new system to record and replay security events.

- remember all needed for analysis
  - traffic, naming, routing
  - from multiple perspectives
- archive for as long as possible
- and be deployable
  - policy and privacy controls *(acceptable)*
  - cost-effective *(affordable)*

# Retro-Future Project Goals

- goal: prototype an Internet Tivo *software and system*
- trial deployments to evaluate effectiveness
  - deployments subject to policy constraints
    - by default, no payloads and IP anonymization
    - or capture more where supported by local policy
    - expect to test federation and policy constraints
  - expect multiple, small trials
- non-goals:
  - new datasets, new detection methods
  - goal is to *develop new capability* to enable those

# Challenges

- maximize history
- cost-effective operation
- respect permissions and privacy

# Challenge: Maximize History

- challenge: make most of limited storage: *maximize utility of what is stored*
- approaches:
  - multi-resolution storage
    - recent history: full details *(packets)*
    - weeks: sparser *(flows)*
    - years: sparser still *(statistics)*
  - exploit application-specific knowledge
    - ex: don't save replies if one can regenerate them

# Challenge: Cost-Effective Operation

- challenge: make most of limited money: *avoid expensive hardware and big pipes*
- approaches:
  - exploit commodity hardware *(datacenter PCs)*
  - parallel search *(Map/Reduce-like compute)*
  - distributed data *(operate at observer)*

# Challenge: Permission and Privacy

- challenge: must respect polices and user privacy
  *one "size" will never fit all*
- approaches:
  - multi-organization federation *(you keep your data)*
  - distributed data *(…at your site)*
  - support varying policies *(…with your rules)*
  - separate storage from access control
    *(human and policy-based access controls)*
  - auditing of use *(accountability for actions)*

# Benefits

- post-facto understanding a compromise
  - what was lost? compromised? *(mitigate this event)*
  - what failed? *(prevent future events)*
- recovery from insider attacks
  - what was taken? seen? *(mitigate this event)*
  - signs of warning? *(prevent future events)*
- longitudinal studies of wide-area events
  - how do events propagate and grow? *(understanding)*
  - can we improve the emergent network? *(prevention)*

# Alternatives

- many siloed archives exist
  - routing (RouteViews)
  - custom packet- and flow-storage
  - application-level systems

  ⇒ we aim to span multiple levels
  and manage policy and privacy up front
- commodity systems move fast

  ⇒ we will leverage open source, evolving with it

# The Principle Investigators

Mike Fisk, LANL

background: network security, attack detection, large scale data processing

Dan Massey, CSU

background: BGP, DNS and routing security

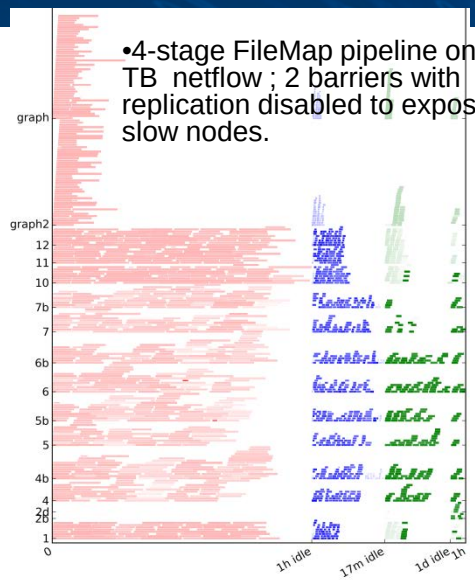
John Heidemann, USC/ISI

background: network traffic analysis, network topology

Christos Papadopoulos, CSU

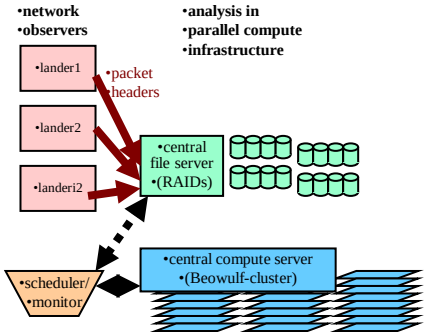
background: network security, attack detection

# Some of Our Prior Results




## SMACQ and FileMap (LANL)

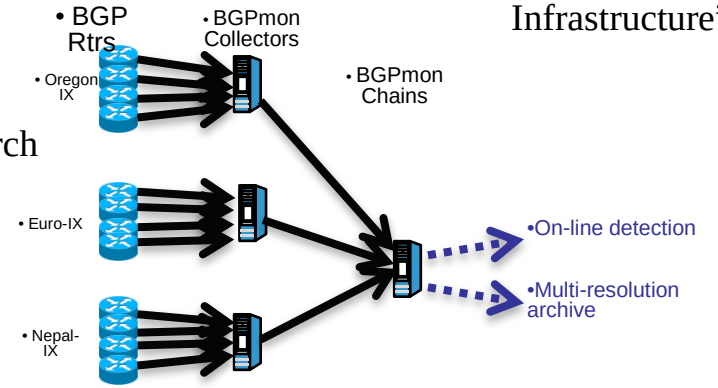
packet and flow archive and parallel search
*http://smacq.sourceforge.net/*




LANDER scales
up to 2000-node cluster
and down to 4-core stand-alone observer

## LANDER (USC & CSU)

packet capture, parallel processing, and anonymization, see: Hussain et al, "Experiences with a Continuous Network Tracing Infrastructure", ACM MineNet 2005




## BGPMon (CSU)

routing observation and archive. See Yan et al, "BGPmon: A real-time, scalable, extensible monitoring system", IEEE CATCH, 2009 *http://bgpmon.netsec.colostate.edu*

# Conclusions

- Retro-future: an Internet Tivo for security events
  - multi-resolution storage to maximize lifetime
  - cost-effective, commodity, parallel hw & sw
  - federated policy and privacy
- important applications
  - understanding and recovering from…
  - 0-day attacks, insider-threat, wide-area events
  - …understand the past to protect the future
- just getting started, but building on experience