Cyber Security Division
2013 Principle Investigators’ Meeting

Retro-Future

PIs: Mike Fisk\(^1\), John Heidemann\(^2\), Christos Papadopoulos\(^3\)

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

\(^1\) Los Alamos National Laboratory
\(^2\) USC/Information Sciences Institute
\(^3\) Colorado State University

Copyright © 2013 by John Heidemann
Release terms: CC-BY-NC 3.0 unported
The Retro-Future Team: PIs, staff, and students

Mike Fisk, LANL

John Heidemann, USC/ISI

Christos Papadopoulos, CSU

Yuri Pradkin, USC/ISI

Xun Fan, USC/ISI

Zi Hu, USC/ISI

Han Zhang, CSU
the 0-day Challenge

in the future: all interesting security events involve multiple parties and will have already happened

interesting: like 0-day attacks and insider threats

networking is many organizations (=> many policies)

pro-active security always fails (eventually) we know “interesting” only after the fact
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
- constrained by:
  - after-the-fact => we must unwind time
    - what happened? why? what was lost?
    - understanding will improve future prevention
  - in a multi-party, multi-policy world
the Retro-Future Goal: an Internet “Tivo”

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
• is deployable:
  – acceptable: policy and privacy controls
  – affordable: cost-effective
Retro-Future Project Approach

- prototype an Internet “Tivo” *software and system*
- *evaluate effectiveness* through target applications
  - emphasize key technologies
  - real-world policy constraints: federation and collaboration
    - default for safety (no payload and IP anonymization)
    - or more where supported by local policy
- non-goals:
  - new datasets, new detection methods
  - goal is to *develop new capability* to enable those
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)*

inputs *(subject to local policy)*

multi-organization collaboration *(subject to sharing policies)*

other net. info

routing

packet headers

inputs *(subject to local policy)*

multi-organization collaboration *(subject to sharing policies)*

other net. info

routing

packet headers

inputs *(subject to local policy)*

multi-organization collaboration *(subject to sharing policies)*

other net. info

routing

packet headers
Applications to Prioritize Challenges

**Pathscan**—LANL-developed approach to detect network traversals (internal attack behavior)
- *their goal*: efficient, federated (decentralized) observation
- *we bring*: packet and flow observation with time travel

**Gloriad.org**—a research and academic network
- *their goal*: understand heavy hitters; improve security
- *we bring*: retrospective packet and flow analysis that crosses organizations

**Multi-View IPv4**
- *our goal*: integrate routing, allocation, use of IPv4 address space
- *we bring*: multiple data sources, time travel

<table>
<thead>
<tr>
<th>Probability</th>
<th>Source</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.57E-08</td>
<td>10.0.10.11</td>
<td>10.0.30.13</td>
</tr>
<tr>
<td>3.67E-08</td>
<td>10.0.20.12</td>
<td>10.0.30.15</td>
</tr>
<tr>
<td>3.58E-08</td>
<td>10.0.20.12</td>
<td>10.0.30.17</td>
</tr>
<tr>
<td>3.59E-08</td>
<td>10.0.20.12</td>
<td>10.0.30.19</td>
</tr>
<tr>
<td>4.03E-08</td>
<td>10.0.20.12</td>
<td>10.0.30.23</td>
</tr>
<tr>
<td>4.10E-08</td>
<td>10.0.30.13</td>
<td>10.0.20.12</td>
</tr>
<tr>
<td>5.08E-08</td>
<td>10.0.20.12</td>
<td>10.0.30.14</td>
</tr>
<tr>
<td>5.53E-08</td>
<td>10.0.20.12</td>
<td>10.0.70.18</td>
</tr>
<tr>
<td>5.58E-08</td>
<td>10.0.20.12</td>
<td>10.0.60.19</td>
</tr>
<tr>
<td>5.73E-08</td>
<td>10.0.20.12</td>
<td>10.0.70.20</td>
</tr>
<tr>
<td>5.75E-08</td>
<td>10.0.20.12</td>
<td>10.0.60.23</td>
</tr>
</tbody>
</table>
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)*

*and deployable:* given budget and policy constraints
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

• commercial systems exist: NetWitness, Solara
  ⇒ we aim to manage policy, privacy and federation,
    and leverage open-source for lower deployment cost

• commodity systems move fast
  ⇒ we will leverage open source, evolving with it
Status and Next Steps (as of Fall 2013)

- status
  - identified driving applications and initial partners (LANL and Gloriad)
  - prototyping data streams
  - initial search API and evaluation of federation
- next steps
  - from components to prototype applications
  - experience with federated search and data integration
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

• contact us:
  – retrofuture@isi.edu
  – http://www.isi.edu/ant/retrofuture/