GENI Trace Collection for Security Studies

Yan Luo
Department of Electrical and Computer Engineering
University of Massachusetts Lowell

University of Massachusetts UMASS Lowell

Questions

1. What assets are you trying to protect?
2. What are the risks to these assets?
3. What are the security solutions?
4. How well does the security solution mitigate those risks?
5. What other risks does the security solution cause?
6. What cost and trade-offs does the security solution impose?

Bruce Schneier, *Beyond Fear*

What assets are you trying to protect?

- Computing nodes
- Programmable routers/switches
- Radar, sensors, ...
- Network bandwidth
- Application data
  - E-commerce data?
  - Healthcare app data?
- Experiments
- Users

What are the risks to the assets?

- Shutdown/disable GENI hardware
- Breach of privacy data
- Misuse of allocated GENI resources
- Unauthorized usage of GENI resources
- Interrupting user experiments
- Users losing interest due unavailability

What are the security solutions?

- Our proposed solution:
  - providing a mechanism of capturing and analyzing packet traces on GENI.

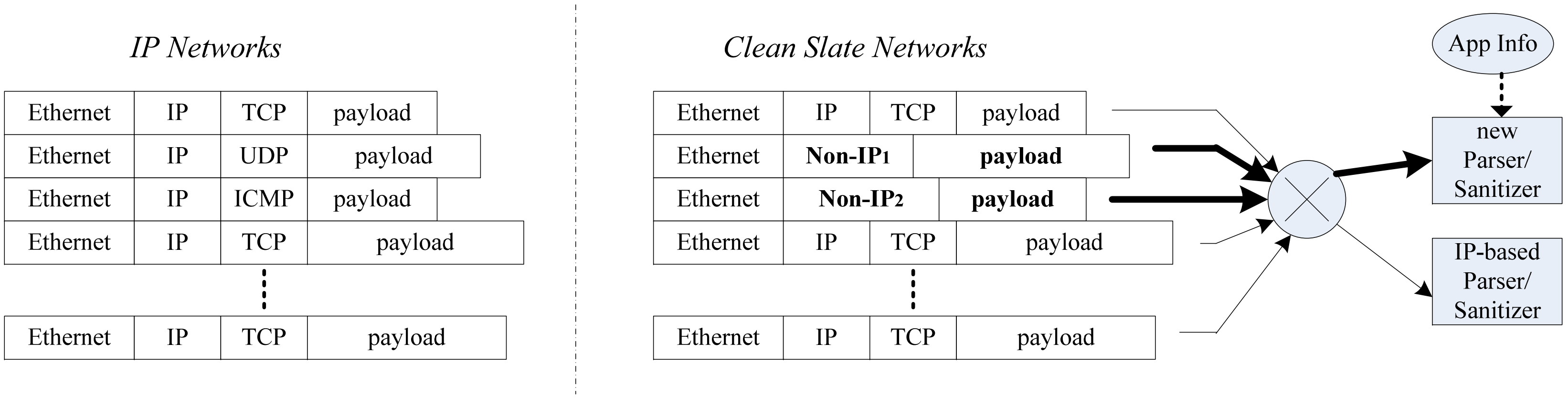
Trace Collection and Analysis in Network Research

- Long history
  - 1994: DAG card developed by University of Waikato networking research group
  - 1995: NLANR established the NLANR/Fix-West real time flow data web site
- Popular trace archives
  - NLANR
  - Internet Measurement Data Catalog http://imdc.datcat.org/Home
  - WITS: Waikato Internet Traffic Storage http://www.wand.net.nz/wits/
- Proved to be beneficial
  - Hundreds of papers

Challenges in Network Trace Studies

- High speed trace capture and archive
  - Specialized hardware
  - Enormous storage space
- Trace anonymization
  - Protect privacy
  - Facilitate trace sharing
- New challenges for GENI
  - Packet formats
  - Experimental applications

Mixed Network Traffic in Clean Slate Networks



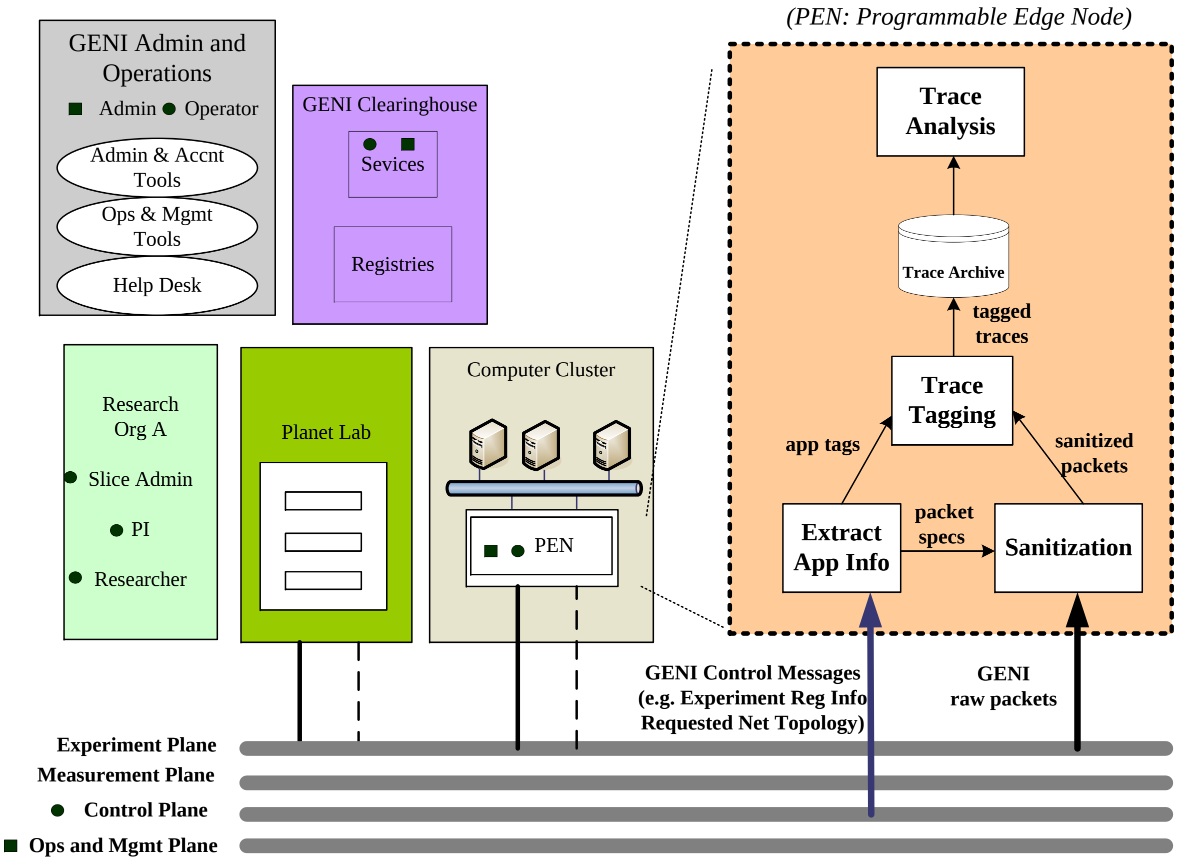
IP Networks

Ethernet	IP	TCP	payload
Ethernet	IP	UDP	payload
Ethernet	IP	ICMP	payload
Ethernet	IP	TCP	payload
		⋮	
Ethernet	IP	TCP	payload

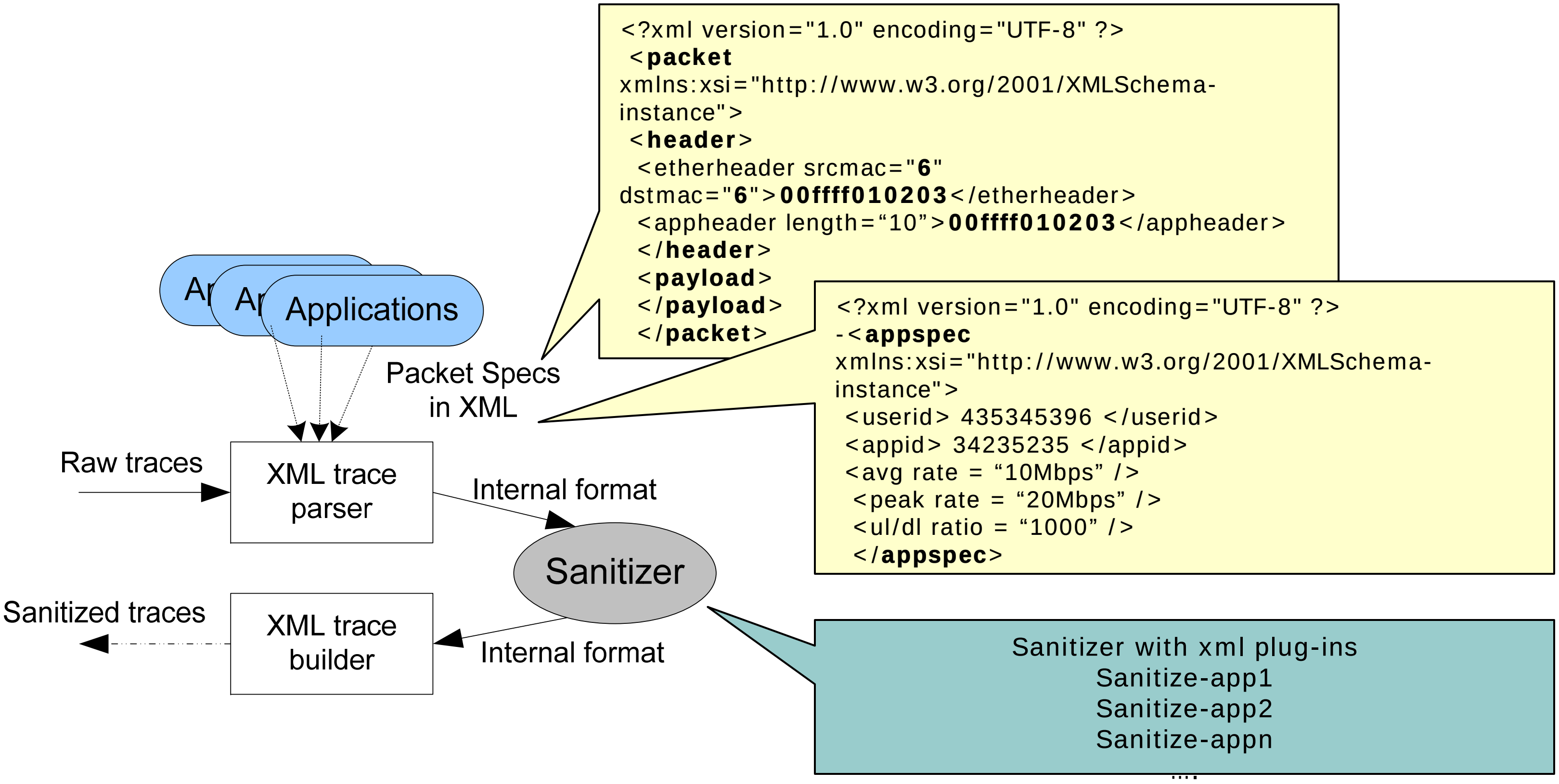
Clean Slate Networks

Ethernet	IP	TCP	payload
Ethernet	**Non-IP$_1$**	**payload**	
Ethernet	**Non-IP$_2$**	**payload**	
Ethernet	IP	TCP	payload
		⋮	
Ethernet	IP	TCP	payload

Proposed Trace Collection Architecture



Trace Specification and Anonymization



```
<?xml version="1.0" encoding="UTF-8" ?>
 <packet
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
 <header>
  <etherheader srcmac="6"
dstmac="6">00ffff010203</etherheader>
  <appheader length=“10”>00ffff010203</appheader>
  </header>
  <payload>
  </payload>
  </packet>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
-<appspec
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
 <userid> 435345396 </userid>
 <appid> 34235235 </appid>
 <avg rate = “10Mbps” />
  <peak rate = “20Mbps” />
  <ul/dl ratio = “1000” />
  </appspec>
```

Sanitizer with xml plug-ins
Sanitize-app1
Sanitize-app2
Sanitize-appn
....

How well does the security solution mitigate those risks?

- Online capture and anonymization of packet traces for post-analysis
- Facilitate trace sharing and publication
- Audit experiments and their packets
- Detect abnormal behavior
  - Invalid packet formats
  - Misuse of GENI resource
  - unauthorized usage
  - Unexpected experiment behavior (duration, burst, ul/dl, etc)

What other risks does the security solution cause?

- Additional design complexity of GENI infra.
- Users lose interest because of cumbersome application specs
- Weakest link targeted by phony/malicious application/packet specs
- Performance distortion/degradation of experiments
  - Additional packet processing

What cost and trade-offs does the security solution impose?

- Packet and application specs expected from GENI user
  - Detailed specs
  - better understanding of the apps and network activities
  - Cumbersome to user

  vs.
  - Simplified specs
  - minimal knowledge of apps and activities
  - simple to user