Measuring Defense in Depth

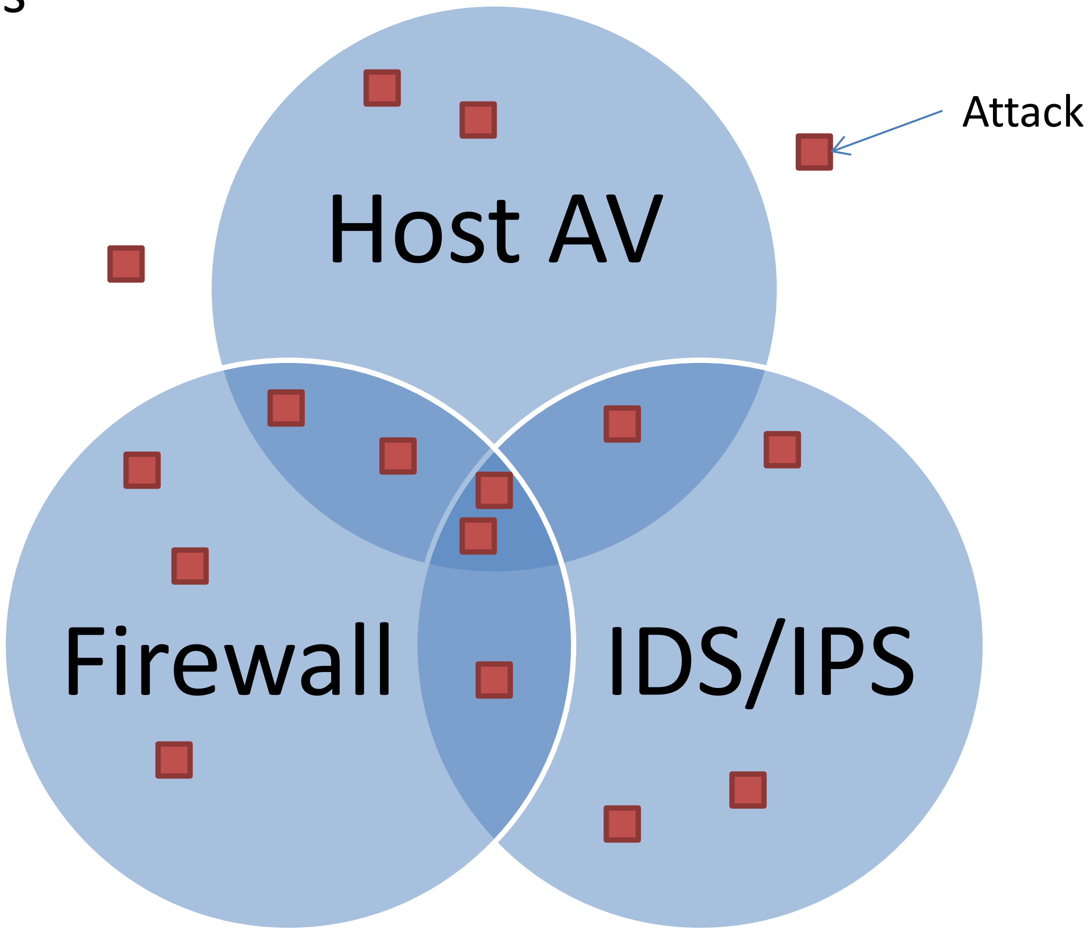
Nathaniel Boggs, Senyao Du,

Salvatore J. Stolfo

Columbia University

Defense in Depth

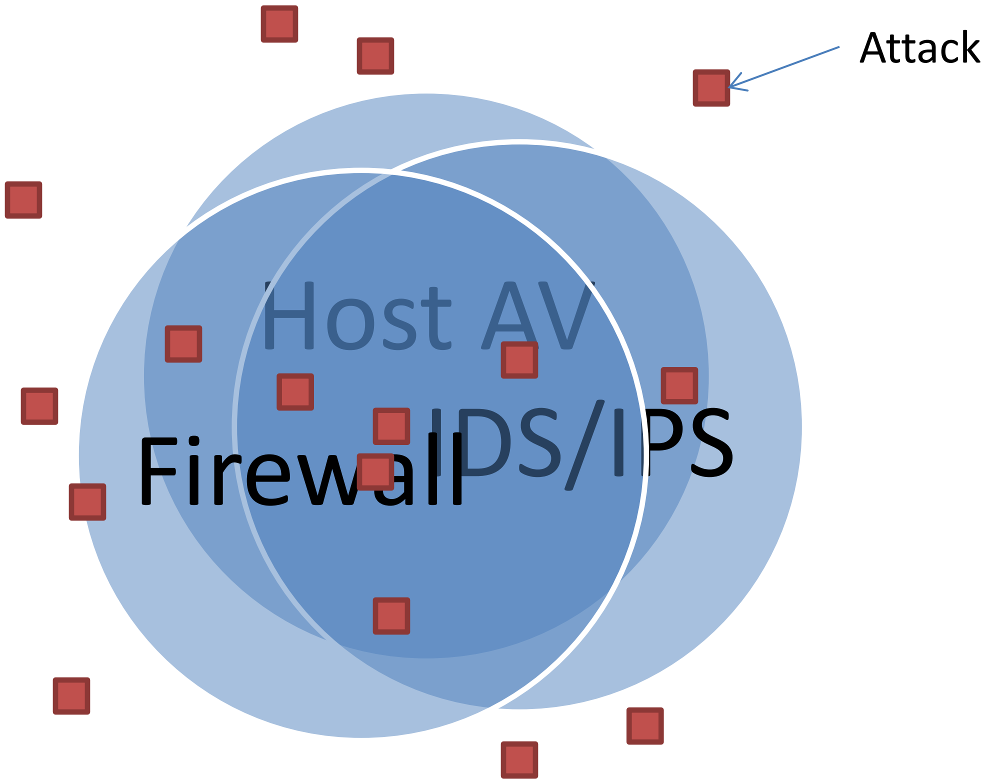
We assume layers provide broader coverage, better security.



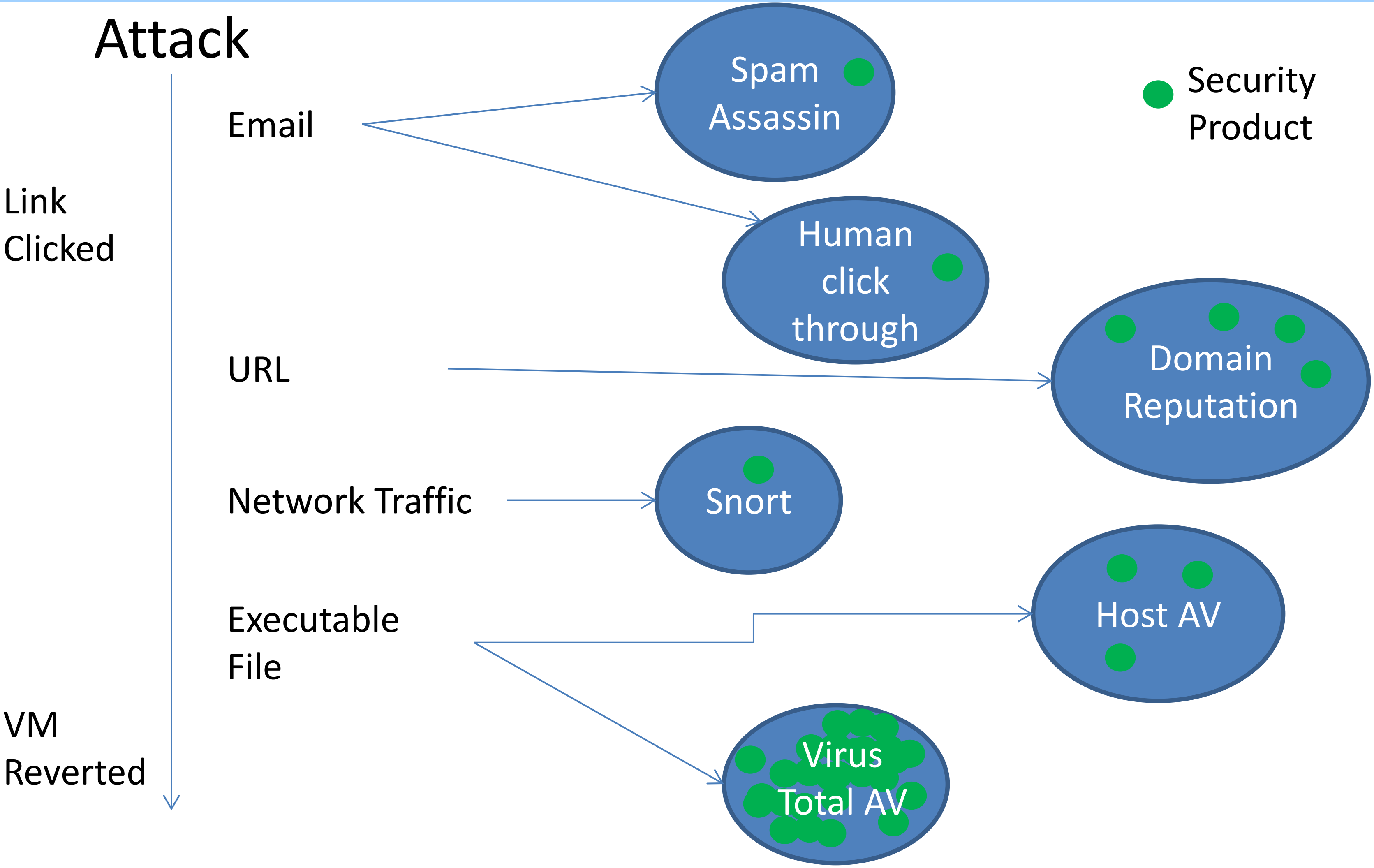
Defense in Depth

What if they look more like this?

We measure overlap between products and total coverage!

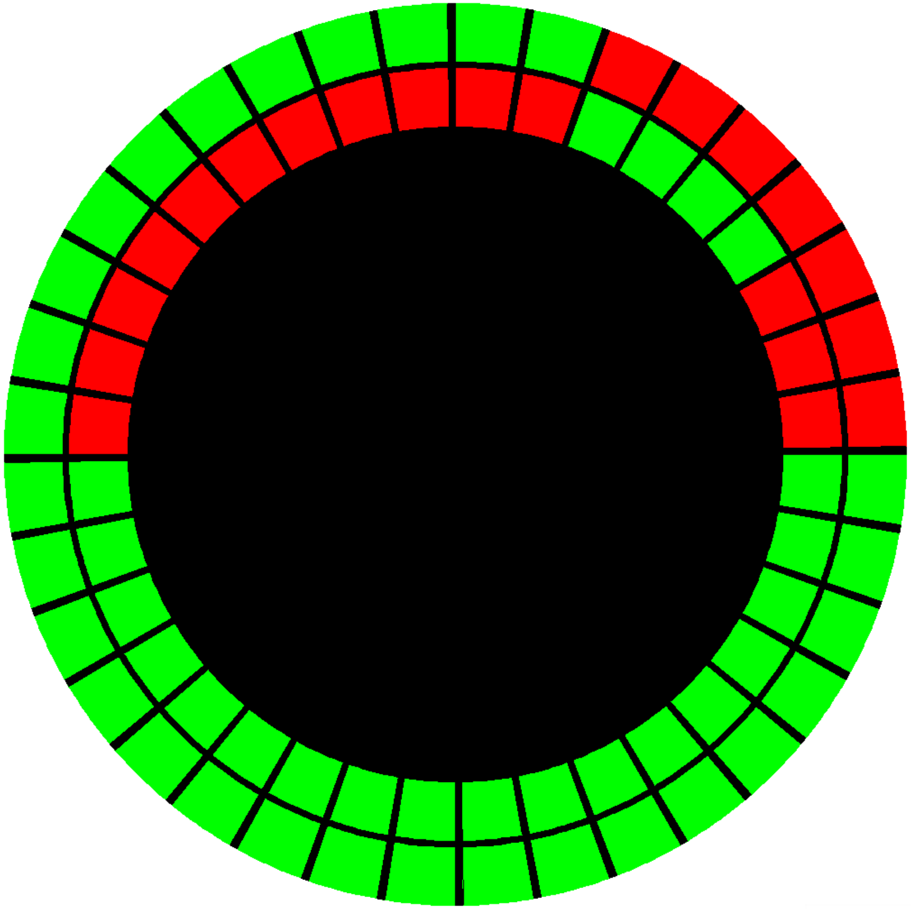


Attack Data Scanned by Real Security Products at Different Layers



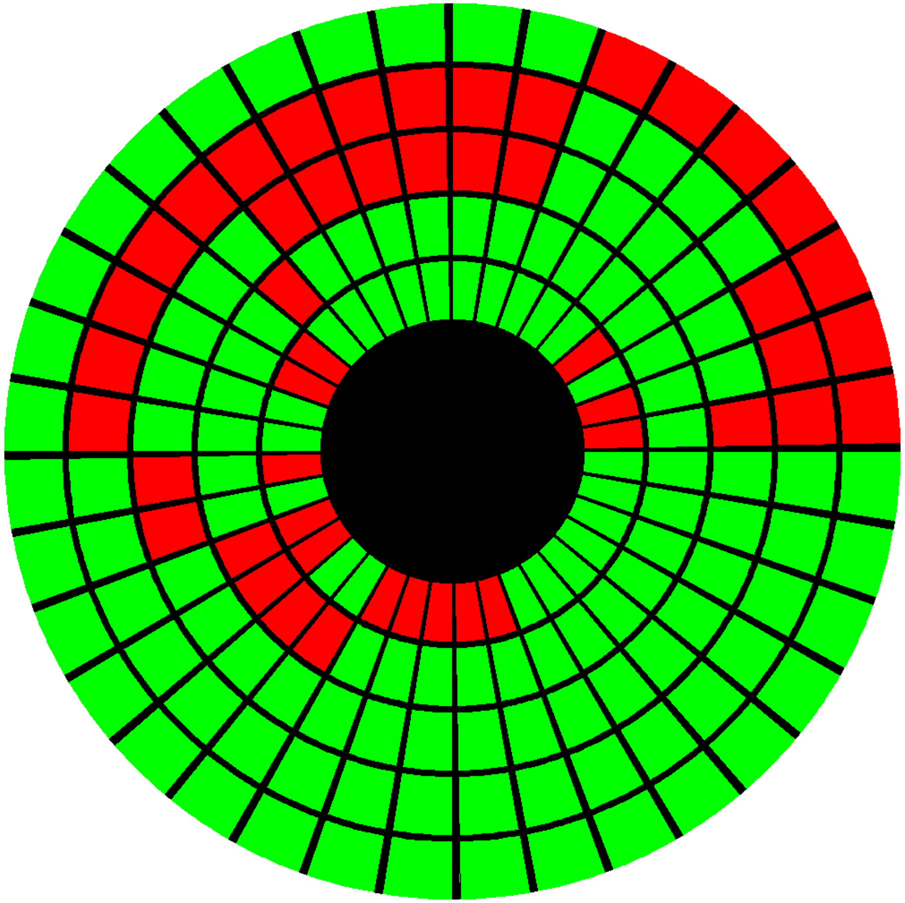
Example Using Real Data

- Assume a small organization with the best AV and best domain reputation seen in our experiment
- AV: detects 29/36 attack clusters
- Domain reputation detects 22/36
- Current state of the art



Example Using Real Data

- AV + Domain Reputation detect: 33/36
- Snort detects 27/36 (2/3)
- Spam Assassin detects 31/36 (3/3)
- Humans not clicking detect 23/36
- Imagine zero day attacks, more layers, more security products tested



Our Approach's Key Attributes

- Products tested individually
- Expandable framework
  - Break down attack vectors into distinct types of linked data
  - Any 'attack' representable
- Evaluate products in the context of existing layers of security rather than in absolute/isolated terms

Future Work - Additional Metrics

- Web application attack vector (i.e. SQL injection)
- False positive rate per set of security products
- Redundancy
  - Good redundancy (detection methods differ)
  - Bad redundancy (Attacker can bypass both security products with one change)
  - Classify detection method

Questions?

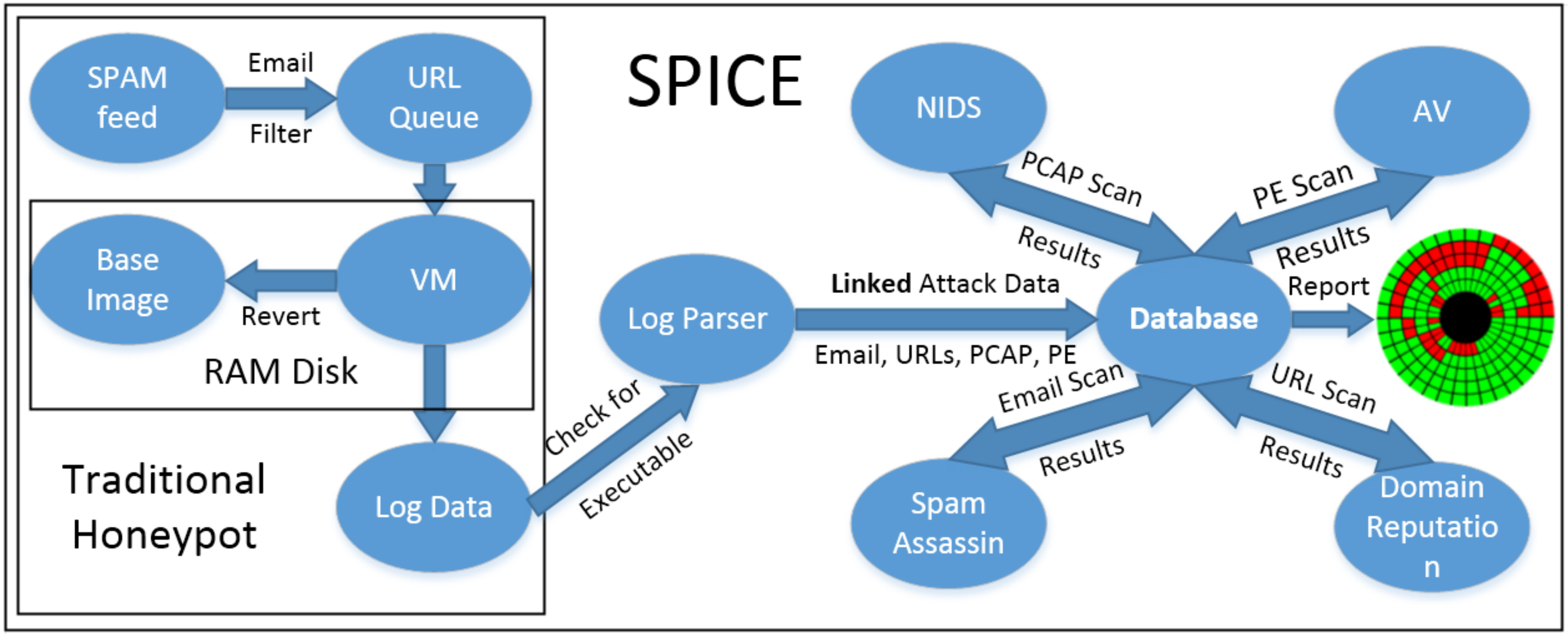
- ids.cs.columbia.edu
- boggs@cs.columbia.edu
- sal@cs.columbia.edu
- Under submission to USENIX Security

Backup Slides

Measure Different Classes of Attackers/Attack Vectors Separately

	Drive-by Download	Server Exploits	…	Exfiltration
Exploit Kit User	Our Experiment			
Targeted Attacker				
…				
Nation State				

System Architecture



Attack Data Collected

- 1463 malicious site visits by VMs ending in compromise
- 730 unique malicious emails
- 576 unique executables
- 36 clusters of distinct email content

Inline AV

- Install AV in VM
- Harder to measure
- If not infected, blocked by AV or other failure?
- Sent VMs to about 2 hundred known infected sites
- 2 of the 3 AVs compromised
- Future work

Human Factor

- Measure spam click through rate
- Sent sanitized versions of spam email
- Columbia University students/faculty/staff (IRB Approved)
- 360 chosen randomly
- 10 emails sent per attack cluster
- 17 click throughs
- At most 2 in same cluster

Results –Findings

- Most security products are horrible
  - Mean detections: 11.3/36 clusters
- No security product is perfect
  - No single product detected all clusters
- With time most products can detect attacks
  - Eventually detected mean: 27.3/36 clusters

Challenges – Data Sets

- Some attack vectors are harder
  - Insider
  - Data exfiltration
- How to link ‘attacks’
- Define ‘attacks’
- Future attacks differ?