Intrusion Detection Systems

CSE497b - Spring 2007
Introduction Computer and Network Security
Professor Jaeger
www.cse.psu.edu/~tjaeger/cse497b-s07/
Intrusion Detection

• An IDS system find anomalies

• “The IDS approach to security is based on the assumption that a system will not be secure, but that violations of security policy (intrusions) can be detected by monitoring and analyzing system behavior.” [Forrest 98]

• However you do it, it requires

• Training the IDS (training)

• Looking for anomalies (detection)

• This is an explosive area in computer security, that has led to lots of new tools, applications, industry
Intrusion Detection Systems

- IDS systems claim to detect adversary when they are in the act of attack
  - Monitor operation
  - Trigger mitigation technique on detection
  - Monitor: Network, Host, or Application events
- A tool that discovers intrusions “after the fact” are called *forensic analysis* tools
  - E.g., from system logfiles
- IDS systems really refer to two kinds of detection technologies
  - Anomaly Detection
  - Misuse Detection
Anomaly Detection

- Compares profile of normal systems operation to monitored state
  - Hypothesis: any attack causes enough deviation from profile (generally true?)

- Q: How do you derive normal operation?
  - AI: learn operational behavior from training data
  - Expert: construct profile from domain knowledge

- Black-box analysis (vs. white or grey?)

- Q: Will a profile from one environment be good for others?

- Pitfall: false learning
Misuse Detection

- Profile signatures of known attacks
  - Monitor operational state for signature
  - Hypothesis: attacks of the same kind have enough similarity to distinguish from normal behavior

- Q: Where do these signatures come from?
  - Record: recorded progression of known attacks
  - Expert: domain knowledge

- AI: Learn by negative and positive feedback

- Pitfall: *too specific*
Network Intrusion Detection

- Intrusion Detection in the network
  - On a switch, router, gateway
  - End-point would be host IDS

- Why do network IDS?
  - Single point of mediation
  - Systems protections are harder to update

- Inspect packets -- What are you looking for?
  - Port scans (or specific service ports)
  - Expected or malformed payloads (signatures)
  - Insider attacks
Snort

- Lots of Network IDS products
  - Firewalls on steroids
- Snort
  - Open source IDS
  - Started by Martin Roesch in 1998 as a lightweight IDS
- Snort rules
  - Sample: alert tcp any any -> 192.168.1.0/24 111 (content:"l00 01 86 a5l"; msg: "mountd access");
  - Rule Header: Action, Protocol, Src+Port -> Dest+Port
  - Rule Options: Alert messages and Packet Content
Sequences of System Calls

- Forrest et al. in early-mid 90s, understand the characteristics of an intrusion

**Event Stream**

WRITE → READ → WRITE → SEND → SEND

**Attack Profile**

READ → WRITE → SEND

- **Idea:** match sequence of system calls with profiles
  - *n-grams* of system call sequences (learned)
    - Match sliding windows of sequences
    - If not found, then trigger anomaly
    - Use *n*-grams of length 6, and later studies of 10.

- If found, then it is normal (w.r.t. learned sequences)
Analyzing IDS Effectiveness

• What constitutes a intrusion/anomaly is really just a matter of definition
  – A system can exhibit all sorts of behavior

Detection Result

<table>
<thead>
<tr>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Reality

F T
T F

• Quality determined by consistency with a given definition
  – context sensitive
Intrusion Detection

- Monitor for illegal or inappropriate access or use of resources
  - Reading, writing, or forwarding of data
  - DOS
- Hypothesis: resources are not adequately protected by infrastructure
- Often less effective at detecting attacks
  - Buttress existing infrastructure with checks
  - Validating/debugging policy
  - Detects inadvertent, often catastrophic, human errors
    - “rm -rf /” issue
- Q: Who is the intruder?
IDS vs Access Control

- IDS rules describe
  - subjects (sources), objects (addresses and ports), operations (send/receive)
    - Like access control

- But, also
  - Argument values
  - Order of messages
  - Protocols

- Claim: IDS is more complex than access control
  - IDS allows access, but tries to determine intent
  - Allow a move in chess, but predict impact
"gedanken experiment"

- Assume a very good anomaly detector (99%)
- And a pretty constant attack rate, where you can observe 1 out of 10000 events are malicious

- Are you going to detect the adversary well?
Bayes’ Rule

- \( \Pr(x) \) function, probability of event \( x \)
  - \( \Pr(\text{sunny}) = .8 \) (80% of sunny day)

- \( \Pr(x|y) \), probability of \( x \) given \( y \)
  - Conditional probability
  - \( \Pr(\text{cavity}|\text{toothache}) = .6 \)
    - 60% chance of cavity given you have a toothache

- Bayes’ Rule (of conditional probability)

\[
\Pr(B|A) = \frac{\Pr(A|B) \Pr(B)}{\Pr(A)}
\]

- Now: \( \Pr(\text{cavity}) = .5 \), \( \Pr(\text{toothache}) = .1 \)
The (base-rate) Bayesian Fallacy

• Setup
  • $\Pr(T)$ is attack probability, 1/10,000
    • $\Pr(T) = .0001$
  • $\Pr(F)$ is probability of event flagging, unknown
  • $\Pr(F|T)$ is 99% accurate (much higher than most known techniques)
    • $\Pr(F|T) = .99$

• Deriving $\Pr(F)$
  • $\Pr(F) = \Pr(F|T)*\Pr(T) + \Pr(F|!T)*\Pr(!T)$
  • $\Pr(F) = (.99)(.0001) + (.01)(.9999) = .010098$

• Now, what’s $\Pr(T|F)$?
The Bayesian Fallacy (cont.)

• Now plug it in to Bayes Rule

\[
Pr(T|F) = \frac{Pr(F|T) \cdot Pr(T)}{Pr(F)} = \frac{Pr(.99) \cdot Pr(.0001)}{Pr(.010098)} = .0098
\]

• So, a 99% accurate detector leads to …
  • 1% accurate detection.
  • With 99 false positives per true positive
  • This is a central problem with ID

• Suppression of false positives real issue
  • Open question, makes some systems unusable
Where is Anomaly Detection Useful?

| System | Attack Density $P(T)$ | Detector Flagging $Pr(F)$ | Detector Accuracy $Pr(F|T)$ | True Positives $P(T|F)$ |
|--------|-----------------------|---------------------------|-----------------------------|------------------------|
| A      | 0.1                   |                           | 0.65                        |                        |
| B      | 0.001                 |                           | 0.99                        |                        |
| C      | 0.1                   |                           | 0.99                        |                        |
| D      | 0.00001               |                           | 0.99999                     |                        |

$$Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$$
## Where is Anomaly Detection Useful?

| System | Attack Density $P(T)$ | Detector Flagging $Pr(F)$ | Detector Accuracy $Pr(F|T)$ | True Positives $P(T|F)$ |
|--------|-----------------------|---------------------------|-----------------------------|------------------------|
| A      | 0.1                   | 0.38                      | 0.65                        | 0.171                  |
| B      | 0.001                 | 0.01098                   | 0.99                        | 0.090164               |
| C      | 0.1                   | 0.108                     | 0.99                        | 0.911667               |
| D      | 0.00001               | 0.00002                   | 0.999999                    | 0.5                    |

$$Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$$
The reality …

• Intrusion detections systems are good at catching demonstrably bad behavior (and some subtle)

• Alarms are the problem

  • How do you suppress them?
  • and not suppress the true positives?
  • This is a limitation of probabilistic pattern matching, and nothing to do with bad science

• Beware: the fact that an IDS system is not alarming does not mean the network is safe

• All too often: used as a tool to demonstrate all safe, but is not really appropriate for that.