Intrusion Detection

Dr. Demetrios Glinos
University of Central Florida

CIS3360 - Security in Computing

Readings

- "Computer Security: Principles and Practice", 3rd Edition, by William Stallings and Lawrie Brown
  - Chapter 8
  - Appendix J (Optional – on base rate fallacy)

Outline

- Intruders
- Intrusion Detection
- IDS Effectiveness
- Analysis Approaches
- Host-Based Intrusion Detection
- Network-Based Intrusion Detection
- Distributed or Hybrid Intrusion Detection
- Port Scanning
- Honeypots

Classes of Intruders

1. Cyber criminals:
   - individuals or organized crime groups
   - goal is financial reward
   - identity theft, theft of financial credentials, corporate espionage, data theft, or data ransoming
2. Activists (hactivists):
   - typically individual insiders, or outside groups
   - motivated by political or social causes
   - website defacement, DoS attacks, disclosure of embarassing information
3. State-sponsored groups:
   - sponsored by a government to conduct espionage or sabotage
4. Others:
   - classic hackers motivated by technical challenge and peer group esteem
   - also includes "hobby hackers" who use attack toolkits

Intruder Skill Levels

- Apprentice
  - minimal technical skill
  - primarily use existing attack toolkits
  - most attackers are of this type, including many criminal and activist attackers
- Journeyman
  - able to extend or modify existing toolkits to expoit newly discovered (or purchased) vulnerabilities, or to focus on different targets
  - may also be able to locate new vulnerabilities similar to some already known
- Master
  - high-level technical skills capable of discovering new categories of vulnerabilities, or writing new powerful attack toolkits
  - in this group: some classical hackers, plus some state-sponsored organizations

Typical Intruder Behavior

1. Target acquisition and information gathering
   - using publicly available information and network exploration tools
2. Initial access
   - via remote network access vulnerability, guessing weak username/passwords, or via malware using social engineering or drive-by download
3. Privilege escalation
   - once in, using various methods to acquire root (admin) privileges
4. Information gathering or system exploitation
   - finding and using information on the target, or navigating to other targets
5. Maintaining access
   - installing backdoors or other malicious software, adding hidden username/password, disabling OS updates or anti-virus
6. Covering tracks
   - disables or edits audit logs to remove evidence of attack
   - also using rootkits and other measures to hide covertly installed files or code

Intrusion and Intrusion Detection

From RFC 2828 (Internet Security Glossary):

Security Intrusion:

A security event, or a combination of multiple security events, that constitutes a security incident in which an intruer gains, or attempts to gain access to a system (or system resource) without having authorization to do so.

Intrusion Detection:

A security service that monitors and analyzes system events for the purpose of finding, and providing real-time or near real-time warning of, attempts to access system resources in an unauthorized manner.

Note: Intrusion detection services supplement, but do not replace firewalls, authentication facilities, and access control facilities

Intrusion Detection Systems

Intrusion detection systems (IDS)

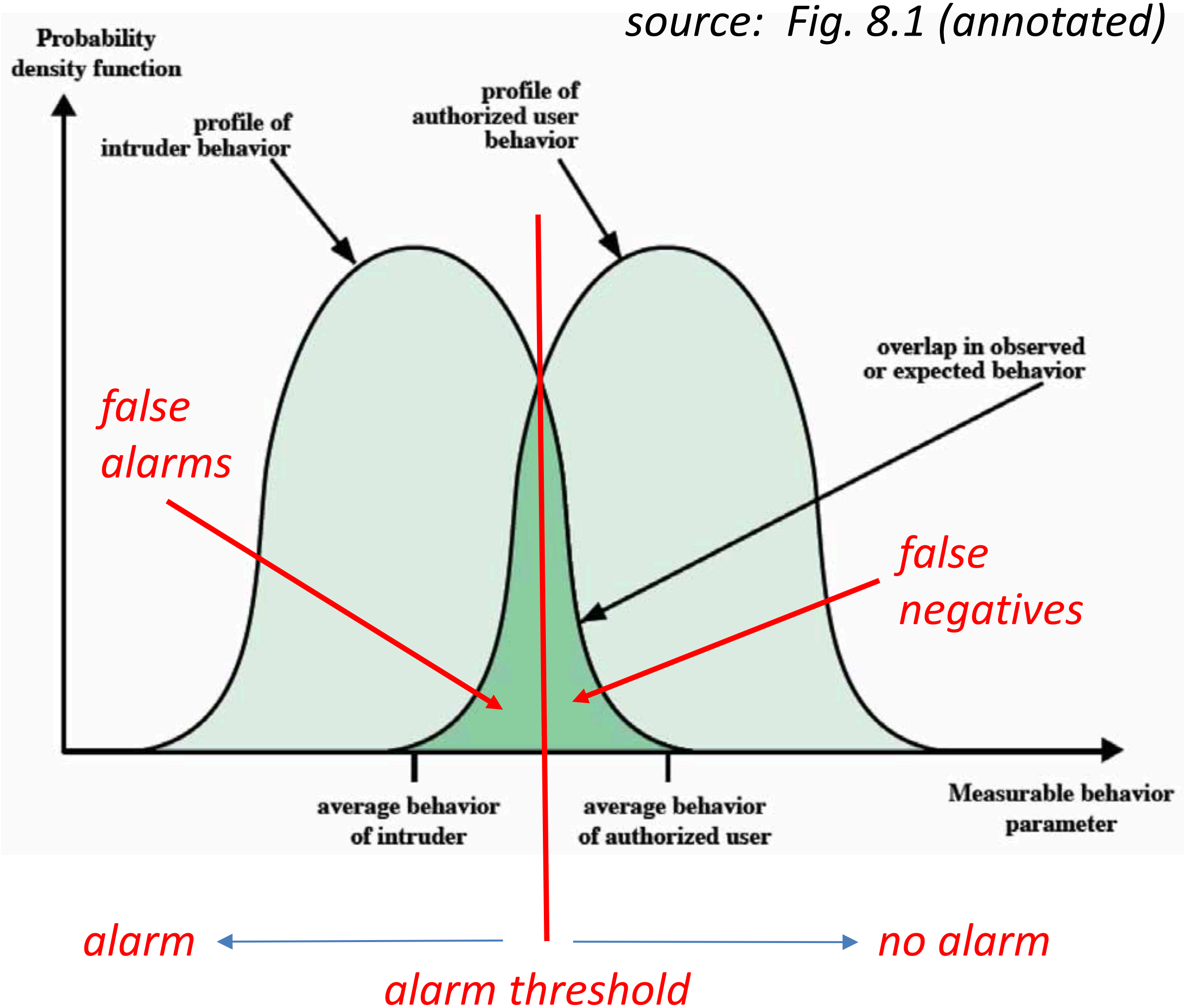
- Hardware and/or software system used to detect malicious activity on a network or on an individual computer

- **IDS Components**
  - **Sensors**
    - used to collect data that may contain evidence of an intrusion or intrusion attempt
    - examples: network packets, log files, system call traces
  - **Analyzers**
    - determine whether an intrusion has occurred, based on the data collected
    - issues an "alarm" when an intrusion is detected
  - **User Interface**
    - display and management component that enables the user to view the output of the IDS and to control the behavior of the system

Basis for Intrusion Detection

- Basic idea: Intruders act differently from legitimate users in quantifiable ways
- However, there is overlap
- As a result, for every IDS, there are
- **false alarms (positives)**
    - no intrusion, but alarm issued
- **false negatives**
    - attack is real, but no alarm

source: Fig. 8.1 (annotated)



IDS Effectiveness

- **Ideal IDS** produces only true positives and true negatives
- **Actual IDSs** produce some false positives and false negatives

- *Effectiveness of IDS is often counter-intuitive due to the effect of the "base rate fallacy"*

- **Base rate fallacy**
  - An error in thinking
  - If presented with related *base rate information* and *specific information*, the mind tends to focus on the specific

  - **Example:**
    - Suppose the likelihood of a malicious attack is 0.01% and an IDS will correctly recognize a malicious event 99% of the time (i.e., if the situation is malicious, the IDS sounds an alarm 99% of the time)
    - Suppose also that the IDS issues an alarm on a benign event only 0.1% of the time

    → What is the false alarm rate? (i.e., *the likelihood that if an alarm is raised, the situation is benign -- In other words: what is the false positive rate?*)

False Alarm Rate Calculation (1)

- One is tempted to think the false alarm rate is 0.1%, but this is not so
- Analysis uses Bayes' theorem from statistics
- See http://yudkowsky.net/rational/bayes for an interesting explanation

We are given

The likelihood of a malicious attack is **0.01%** (this is our given *base rate*)

(*So, out of 10,000 events, 1 is malicious -- that's what 0.01% means*)

99% of the time, the IDS recognizes a malicious event as malicious

0.1% of the time, the IDS issues an alarm when the event is benign

Let

M represent a malicious event

B represent a benign event

A represent an alarm being issued by the IDS

Then our given information is

$p(B) = .9999$	Out of 10,000 events, 9,999 of them are benign
$p(M) = .0001$	Out of 10,000 events, 1 of them is malicious
$p(A\|M) = .99$	99% of the time recognizes an M as an M
$p(A\|B) = .001$	this is the 0.1% of the time IDS thinks a B is an M

False Alarm Rate Calculation (2)

Our given information, once again, is

p(B) = .9999	Out of 10,000 events, 9,999 of them are benign
p(M) = .0001	Out of 10,000 events, 1 of them is malicious
p(A|M) = .99	99% of the time recognizes an M as an M
p(A|B) = .001	.1% of the time IDS thinks a B is an M

Now, the **false alarm rate** is **p(B|A)**, which, by Bayes' Theorem, is computed as:

$$\boxed{p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}}$$

Now, p(A) = p(A|M)p(M) + p(A|B)p(B)

= (.99)(.0001) + (.001)(.9999)

= .000099 + .0009999

= .0010989

So, p(B|A) = [(.001)(.9999)] / .0010989 = .0009999 / .0010989 = .9099

→ **False alarm rate is 90.99 %**

False Alarm Rate Calculation (3)

- Q: Does this make sense?
- A: Yes!

- Think of it this way
  - Out of 10,000 events, 1 of them is malicious ***and*** it is probably detected (99%)

    .99 alarms issued ← *true positives*
  - But also, out of the remaining 9,999 events, which are all benign, 0.1% of them result in alarms

    9.999 (false) alarms issued ← *false positives*

  - So, out of the approximately 11 events thought to be malicious by the IDS, about 10 out of 11 of them are false alarms

false alarm rate = (# false alarms) / (# total alarms)

= 9.999 / (9.999 + .99)

= 9.999 / 10.989

= **90.99%** *(same result as before)*

Analysis Approach: Anomaly Detection

- IDSs typically use one of these analysis approaches

Anomaly detection

- collect lots of behavior data for legitimate users over a period of time
- determine the ***statistical profile*** of characteristics of "normal" behavior
- analyze current observed behavior and compare against expected behavior

- Typical statistical information:
  - *Count, Average, Percentage, Metering, Time-interval length*

- Thresholds used to trigger alarms based on significant deviations

Analysis Approach: Signatures or Heuristics

Signature or Heuristic detection

- use a set of known malicious data patterns (signatures) or attack heuristics encoded as ***rules***
- compare current observed behavior against these patterns or rules to determine if a known attack is being conducted
- this approach can only detect known attacks for which it has patterns or rules

- Typical data in event records:
  - *Subject, Object, Action, Exception-handling, Resource usage, Time stamp*

- Example rules:
  - Desktop computers may not be used as HTTP servers
  - HTTP servers may not acept unencrypted telnet or FTP sessions

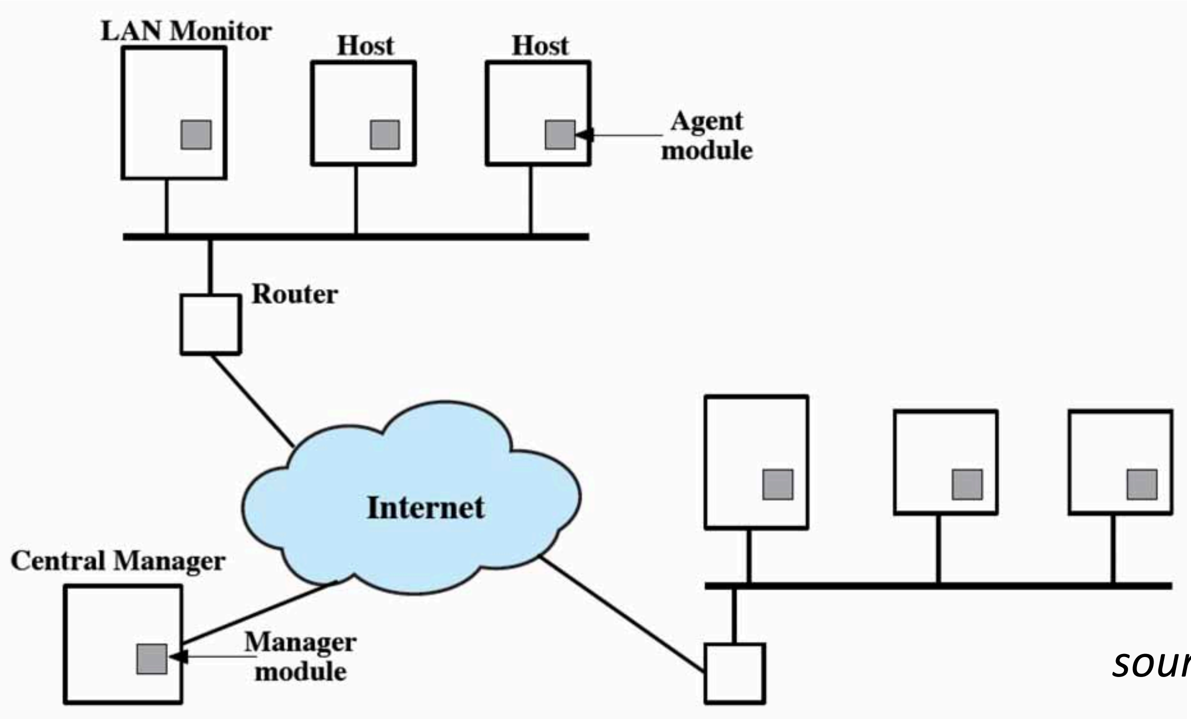
Types of IDSs

- **Host-based IDS (HIDS)**
    - monitors the characteristics of a ***single host*** and the events occurring within it, for evidence of suspicious activity
    - examples: system call traces, log file records, registry access, file integrity checksums

- **Network-based IDS (NIDS)**
    - monitors ***network traffic*** on particular network segments or devices
    - analyzes network, transport and application protocols to identify suspicious activity

- **Distributed or hybrid IDS**
    - combines information from a number of sensors, often both host and network-based
    - uses a ***central analyzer***

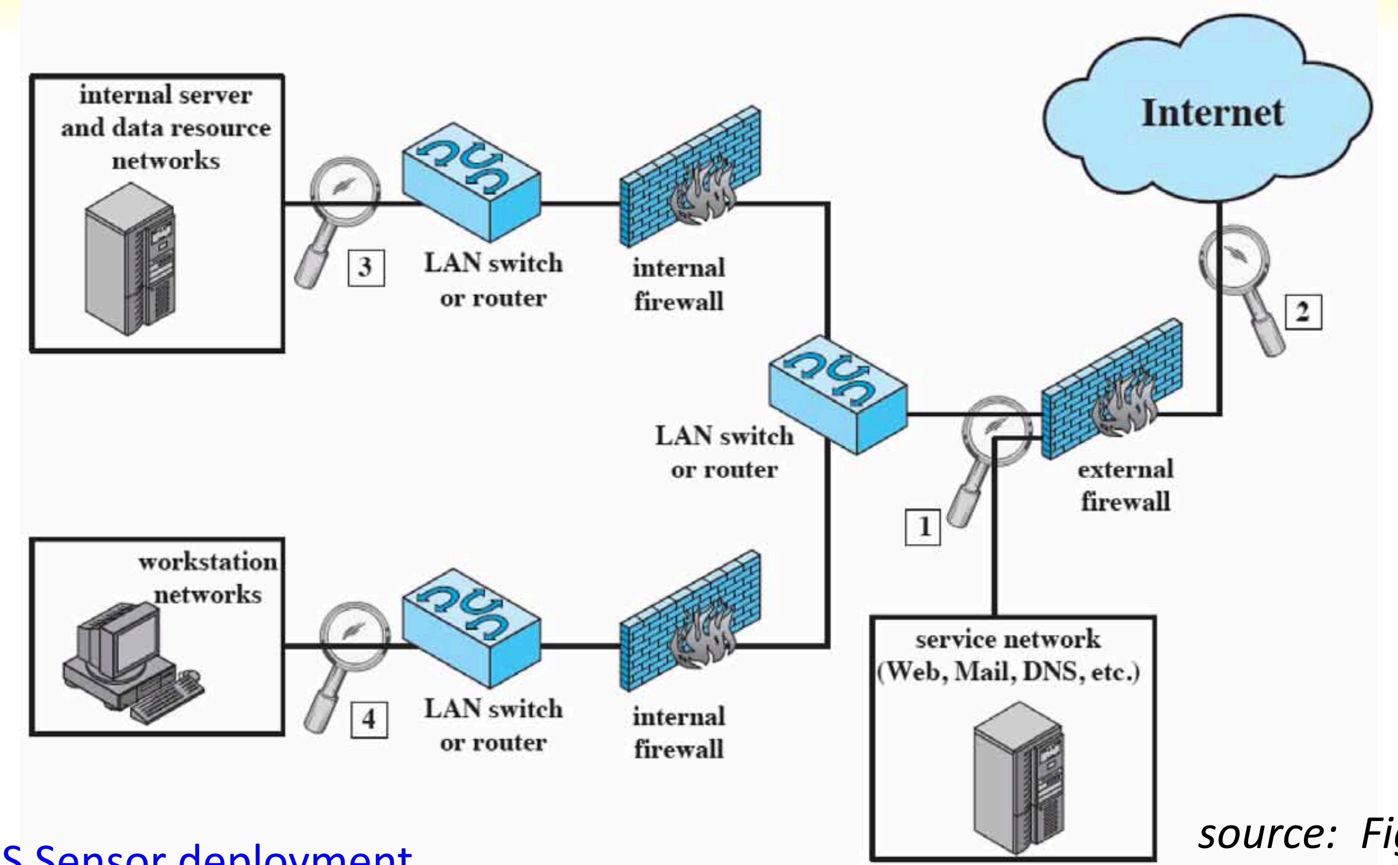
Distributed Host-Based IDS Architecture



source: Fig. 8.2

- Host agent modules – background process on hosts, collects log data
- LAN monitor agent modules - monitors traffic on LANs
- Central manager module – correlates all data for comprehensive intrusion detection analysis

Network-Based Intrusion Detection Systems



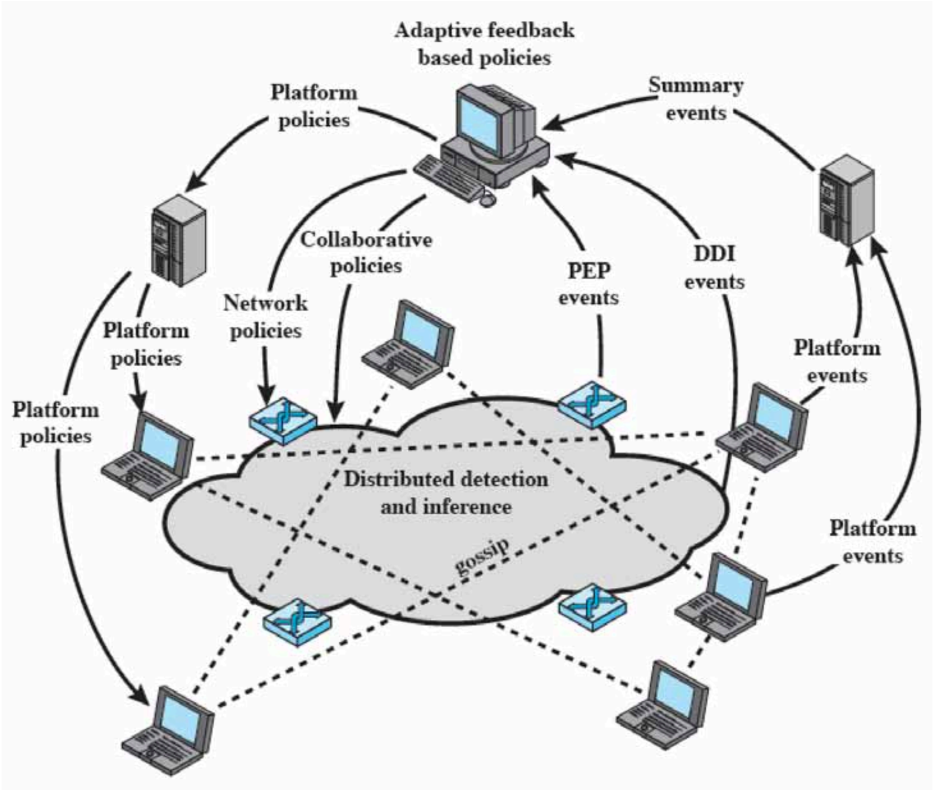
source: Fig. 8.5

NIDS Sensor deployment

1. just inside firewall (typical)
2. outside firewall (also possible)
3. & 4. sensors to protect backbone networks

Distributed or Hybrid Intrusion Detection

- Basic idea: gathering more comprehensive data allows more subtle intrusion detection and quicker adaptation
  - each host and network device (router) contains sensor
  - e.g., Intel's "autonomic enterprise security" architecture



PEP = policy enforcement point
DDI = distributed detection and inference

source: Fig. 8.6

Port Scanning – Part 1

- **Port Scanning**
  - a technique for finding out which ports on a machine are accepting connections
- Can be legitimate use (e.g., to evaluate security of own network)
- Commonly used to perform pre-attack reconnaissance
  - Port scanners can sometimes determine remote service or OS features (this is called *fingerprinting*)

→ *Detecting port scanning is an important form of preliminary intrusion detection*

Port Scanning – Part 2

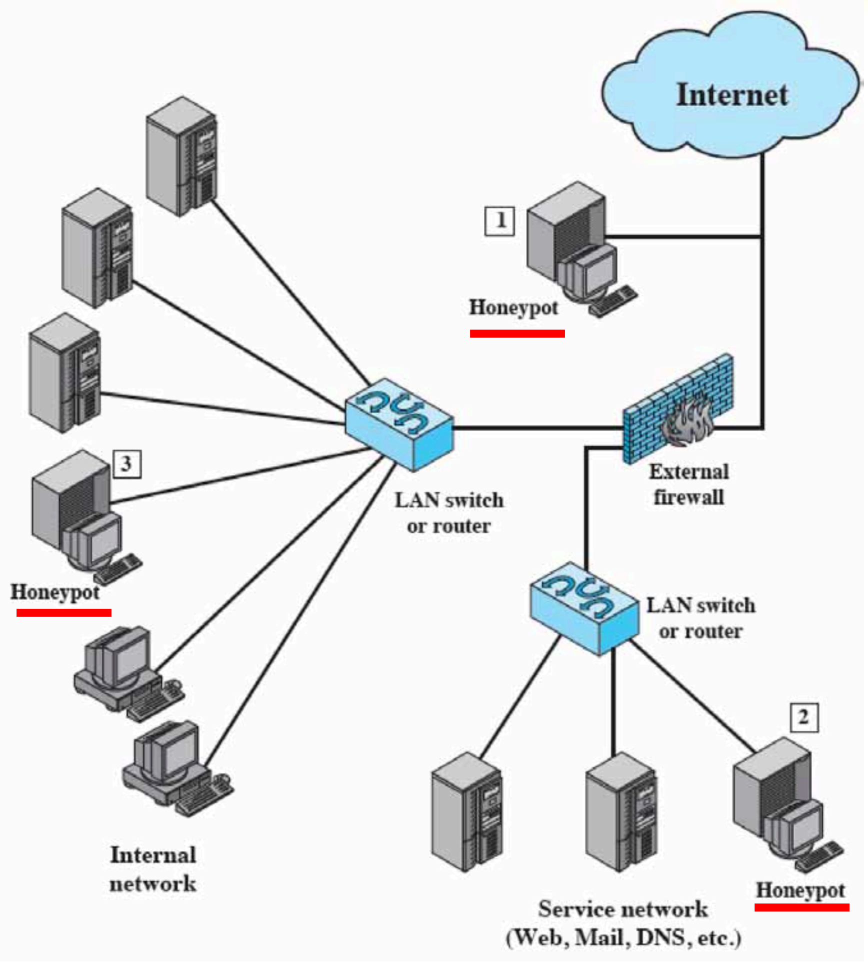
- **Port scanning methods**
  - **TCP scans** – attempt to connect, if successful, then port is open.
  - **SYN scan** – send only the SYN packet; if receive SYN-ACK, send RST (reset)
  - **Idle scanning** – Use a zombie to scan a target machine and hide attacker ID
    - Attacker sends zombie unsolicited SYN-ACK; zombie sends RST with seq ID
    - Attacker sends SYN to target, spoofing source as zombie
      - If port open, target sends SYN-ACK to zombie and zombie sends RST (to target) with incremented sequence ID
    - Attacker probes zombie with another unsolicited SYN-ACK
      - *When the attacker receives the RST from the zombie, he knows that the target port is open **if** seq ID from zombie was doubly incremented*

Honeypots

- Another component of intrusion detection technology
- Can detect intrusions, including port scans
- Involves setting up a “bait” computer that appears tempting to attackers
  - Containing software with known vulnerabilities
  - Dummy data that appears valuable (attractive content)
  - Set up so that legitimate users would never connect to the honeypot

→ All attempts to connect are intrusions
  - Easier to identify intrusion and intruder

→ Can also distract intruders from truly valuable resources



source: Fig. 8.8 (annotated)