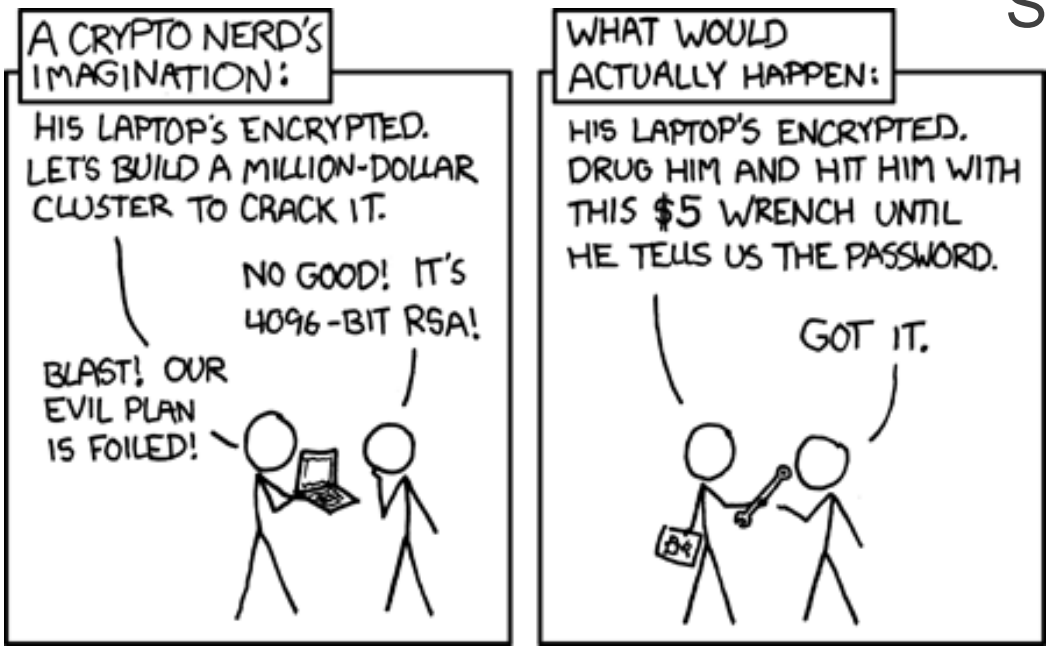
Lecture 3: Assurance

CS 138

Spring 2026



Bases for Trust

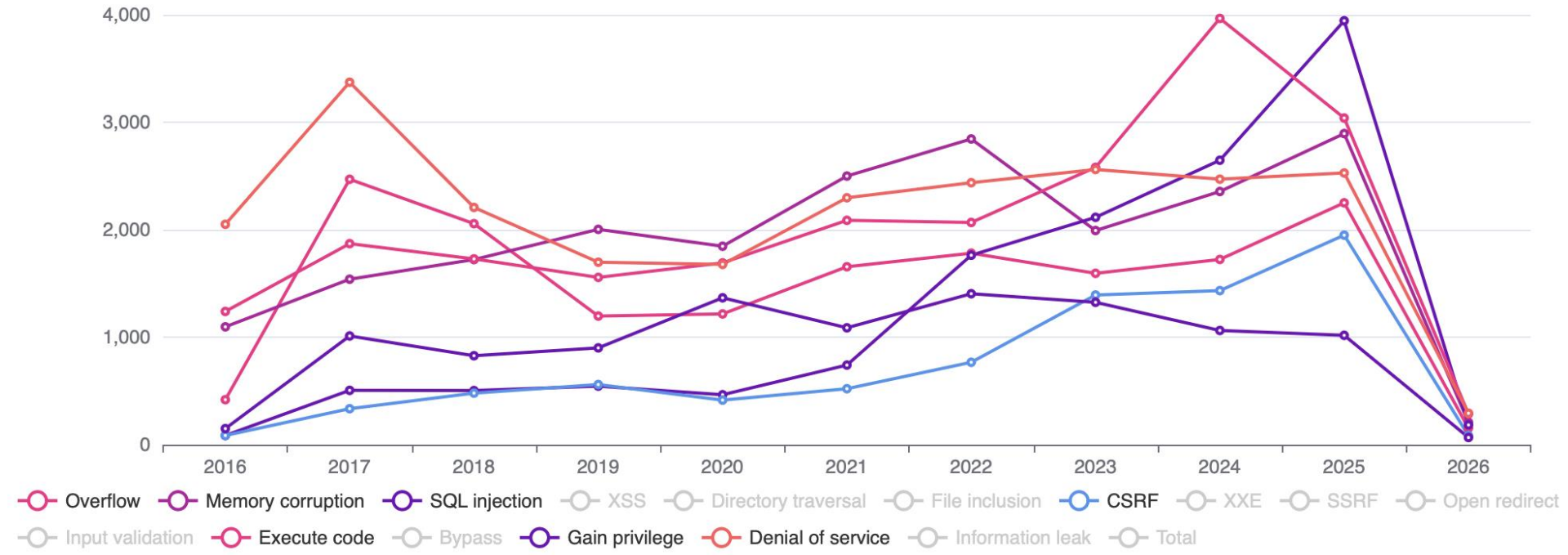
Axiomatic Trust

Analytic Trust

Synthetic Trust

- **Axiomatic Trust:** Trust derived from beliefs that we accept on faith. We might trust some hardware or software, for example, because it is built or sold by a given company. We are putting our faith in the company's reputation.
-
-

Vulnerabilities by Year



Bases for Trust

Axiomatic Trust

Analytic Trust

Synthetic Trust

- **Axiomatic Trust:** Trust derived from beliefs that we accept on faith. We might trust some hardware or software, for example, because it is built or sold by a given company. We are putting our faith in the company's reputation.
- **Analytic Trust:** Trust derived from testing and/or reasoning to justify conclusions about what a component or system will and/or will not do. Trust in an artifact is justified by trust in some method of analysis.
-

Testing

- Goal is to expose existence of faults, so that they can be fixed
  - **Unit testing:** isolated components
  - **Integration testing:** combined components
  - **System testing:** functionality, performance, acceptance

- When do you stop testing?
  - **Bad answer:** when you run out of time
  - **Bad answer:** what all tests pass
  - **Better answer:** when methodology is complete (code coverage, paths, boundary cases, etc.)

Penetration testing

- Experts attempt to attack
  - Internal vs. external
  - Overt vs. covert
- Typical vulnerabilities exploited:
  - Passwords (cracking)
  - Buffer overflows
  - Bad input validation
  - Race conditions / TOCTOU
  - Filesystem misconfiguration
  - Kernel flaws

Fuzz testing

- Generate random inputs and feed them to programs:
  - Crash? hang? terminate normally?
- Of ~90 utilities in '89, crashed about 25-33% in various Unixes
  - Results have been repeated for Windows, Mac OSX
  - Results keep getting **worse** in GUIs but better on command line
- Since then, "fuzzing" has become a standard practice for security testing

- How to generate random inputs:
  - Use grammar to generate inputs
  - Or randomly mutate good inputs in small ways
    - especially for testing of network protocols

Type Checking

```
public class Main {
    public static void main(String[] args) {
        String s = 5;
        System.out.println(s);
    }
}
```

```
$ javac Main.java
Main.java:3: error: incompatible types: int cannot be converted to String
String s = 5;
           ^
1 error
```

SpotBugs

- Looks for *patterns* in code that are likely faults and that are likely to cause failures
- Categorizes and prioritizes bugs for presentation to developer

Formal Verification

- prove program is correct with respect to some formal specification
- Examples: seL4, CompCert
- Problems: correctness of specification, scale

Bases for Trust

Axiomatic Trust

Analytic Trust

Synthetic Trust

- **Axiomatic Trust:** Trust derived from beliefs that we accept on faith. We might trust some hardware or software, for example, because it is built or sold by a given company. We are putting our faith in the company's reputation.
- **Analytic Trust:** Trust derived from testing and/or reasoning to justify conclusions about what a component or system will and/or will not do. Trust in an artifact is justified by trust in some method of analysis.
- **Synthetic Trust:** Trust derived from modification of the system. Trust in the whole derives from how components are combined. Examples: *OS* isolation, reference monitors, firewalls

Engineering Countermeasures

Attacks
are perpetrated by
threats
that inflict
harm
by exploiting
vulnerabilities
which are controlled by
countermeasures.

Threats

A principal that has potential to cause harm to assets

- Adversary or attacker: a human threat, motivated and capable
- Sometimes humans aren't malicious: accidents happen
- Sometimes non-humans cause harm: floods, earthquakes, power outage, hardware failure

Threat Models

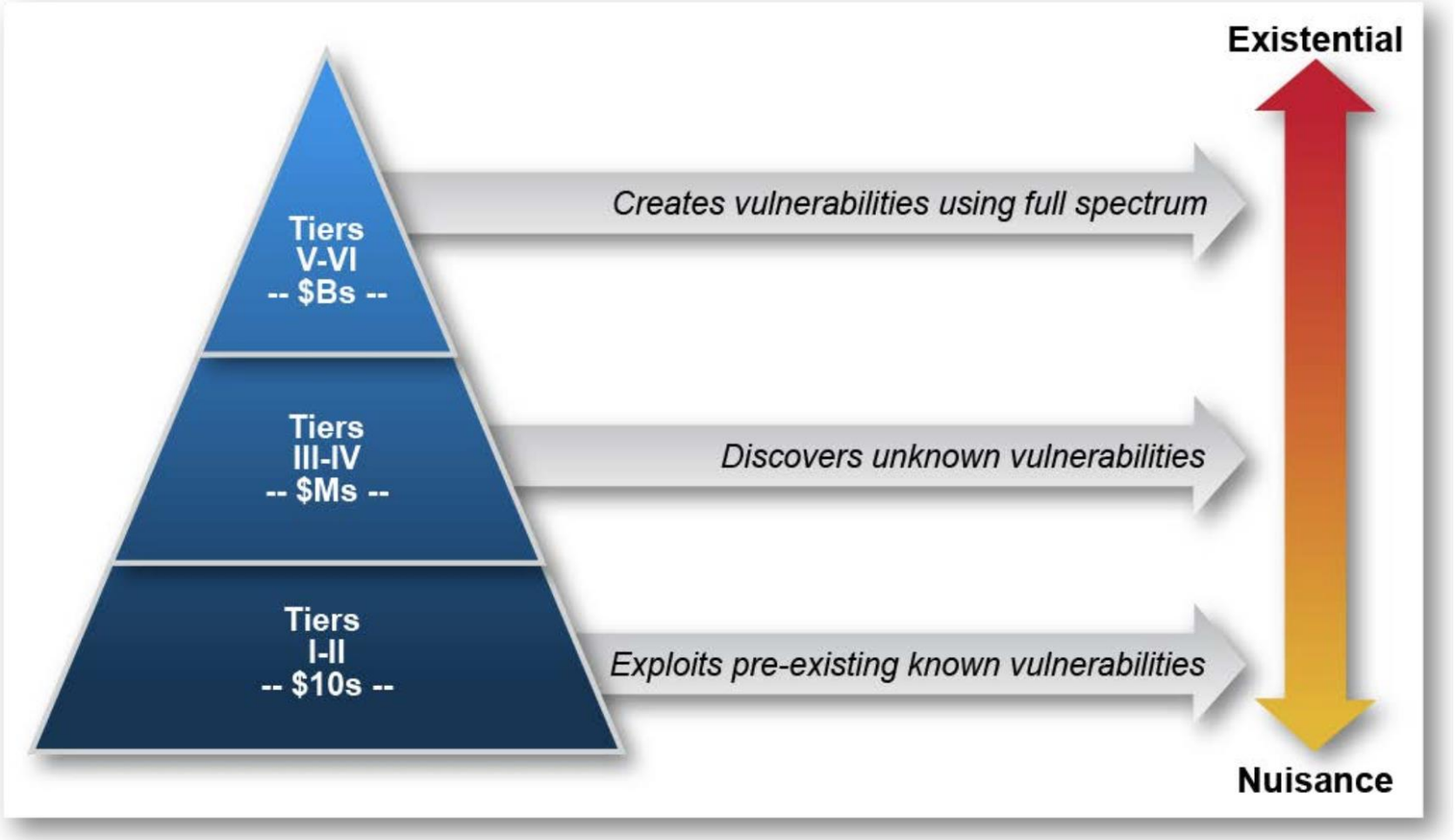
- Identify threats of concern to system
  - Especially malicious, human threats
  - What kinds of attackers will system resist?
  - What are their **motivations**, **resources**, and **capabilities**?
- Best if analysis is specific to system and its functionality

- Non threats?
  - Trusted hardware
  - Trusted environment (e.g., physically secured machine room reachable only by trustworthy system operators)

Threats (DSB)

Tier	Description
I	Practitioners who rely on others to develop the malicious code, delivery mechanisms, and execution strategy (use known exploits).
II	Practitioners with a greater depth of experience, with the ability to develop their own tools (from publically known vulnerabilities).
III	Practitioners who focus on the discovery and use of unknown malicious code, are adept at installing user and kernel mode root kits[10], frequently use data mining tools, target corporate executives and key users (government and industry) for the purpose of stealing personal and corporate data with the expressed purpose of selling the information to other criminal elements.
IV	Criminal or state actors who are organized, highly technical, proficient, well funded professionals working in teams to discover new vulnerabilities and develop exploits.
V	State actors who create vulnerabilities through an active program to "influence" commercial products and services during design, development or manufacturing, or with the ability to impact products while in the supply chain to enable exploitation of networks and systems of interest.
VI	States with the ability to successfully execute full spectrum (cyber capabilities in combination with all of their military and intelligence capabilities) operations to achieve a specific outcome in political, military, economic, etc. domains and apply at scale.

Threats (DSB)


Existential
Tiers
V-VI
-- $Bs --
Creates vulnerabilities using full spectrum
Tiers
III-IV
-- $Ms --
Discovers unknown vulnerabilities
Tiers
I-II
-- $10s --
Exploits pre-existing known vulnerabilities
Nuisance

Exercise 1: DSB Threat Model

- For each of the following systems, how would you classify the threats that are likely to perpetrate attacks against that system using the DSB threat model?
  1. A game app
  2. A social network
  3. A classified government database

Threats (Motives)

- Harm
- Gain

Threats (Motives)

- Inquisitive people, unintentional blunders
- Hackers driven by technical challenges
- Disgruntled employees or customers seeking revenge
- Criminals interested in personal financial gain, stealing services, or industrial espionage
- Organized crime with the intent of hiding something or financial gain
- Organized terrorist groups attempting to influence policy by isolated attacks
- Foreign espionage agents seeking to exploit information for economic, political, or military purposes
- Tactical countermeasures intended to disrupt specific weapons or command structures
- Multifaceted tactical information warfare applied in a broad orchestrated manner to disrupt major military missions
- Large organized groups or nation states intent on overthrowing a government

Exercise 2: Motives Threat Model

- For each of the following systems, how would you classify the threats that are likely to perpetrate attacks against that system according to their motives?
  1. A game app
  2. A social network
  3. A classified government database

Threats (Goals)

Type	Description	Counter
Spoofing	Accessing and using another user's credentials, such as username and password.	Authentication
Tampering	Changing or modifying persistent data, such as records in a database, or alterating of data in transit over an network, such as the Internet.	Integrity
Repudiation	Performing prohibited operations in a system that lacks the ability to trace the operations.	Audit
Information disclosure	Reading a file that one was not granted access to, or reading data in transit.	Confidentiality
Denial of service	Denying access to valid users, such as making a web server temporarily unavailable or unusable.	Availability
Elevation of privilege	Gaining privileged access to resources in order to gain unauthorized access to information or to compromise a system.	Authorization

Exercise 3: Goals Threat Model

- Consider your favorite social network. What is one concrete example of a goal an adversary attacking that network might have for each of the following categories:
  - Spoofing
  - Tampering
  - Repudiation
  - Information disclosure
  - Denial of Service
  - Elevation of Privilege

Threats (Capabilities)

- Physical access

Cold-boot attack

- https://youtu.be/E6gzVVjW4yY

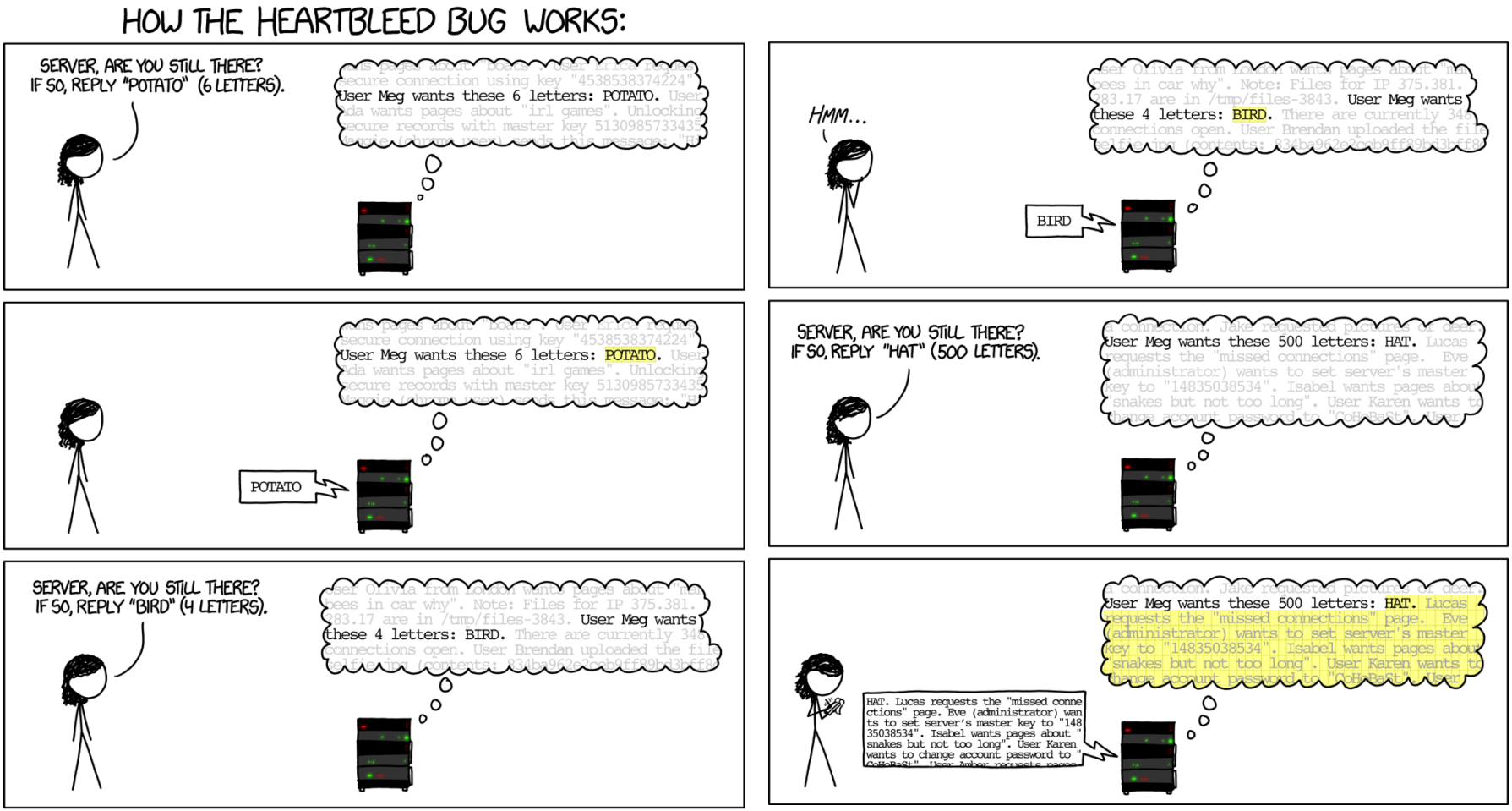
Threats (Capabilities)

- Physical access
- Software access
  - disk access
  - memory access
  - privilege levels

CL0P Data Breach

Heartbleed

Heartbleed


HOW THE HEARTBLEED BUG WORKS:
SERVER, ARE YOU STILL THERE? IF SO, REPLY "POTATO" (6 LETTERS).
User Meg wants these 6 letters: POTATO.
HMM...
BIRD
User Meg wants these 4 letters: BIRD.
POTATO
SERVER, ARE YOU STILL THERE? IF SO, REPLY "HAT" (500 LETTERS).
User Meg wants these 500 letters: HAT.
SERVER, ARE YOU STILL THERE? IF SO, REPLY "BIRD" (4 LETTERS).
HAT. Lucas requests the "missed connections" page. Eve (administrator) wants to set server's master key to "14835038534". Isabel wants pages about "snakes but not too long". User Karen wants to change account password to

Speculative Execution

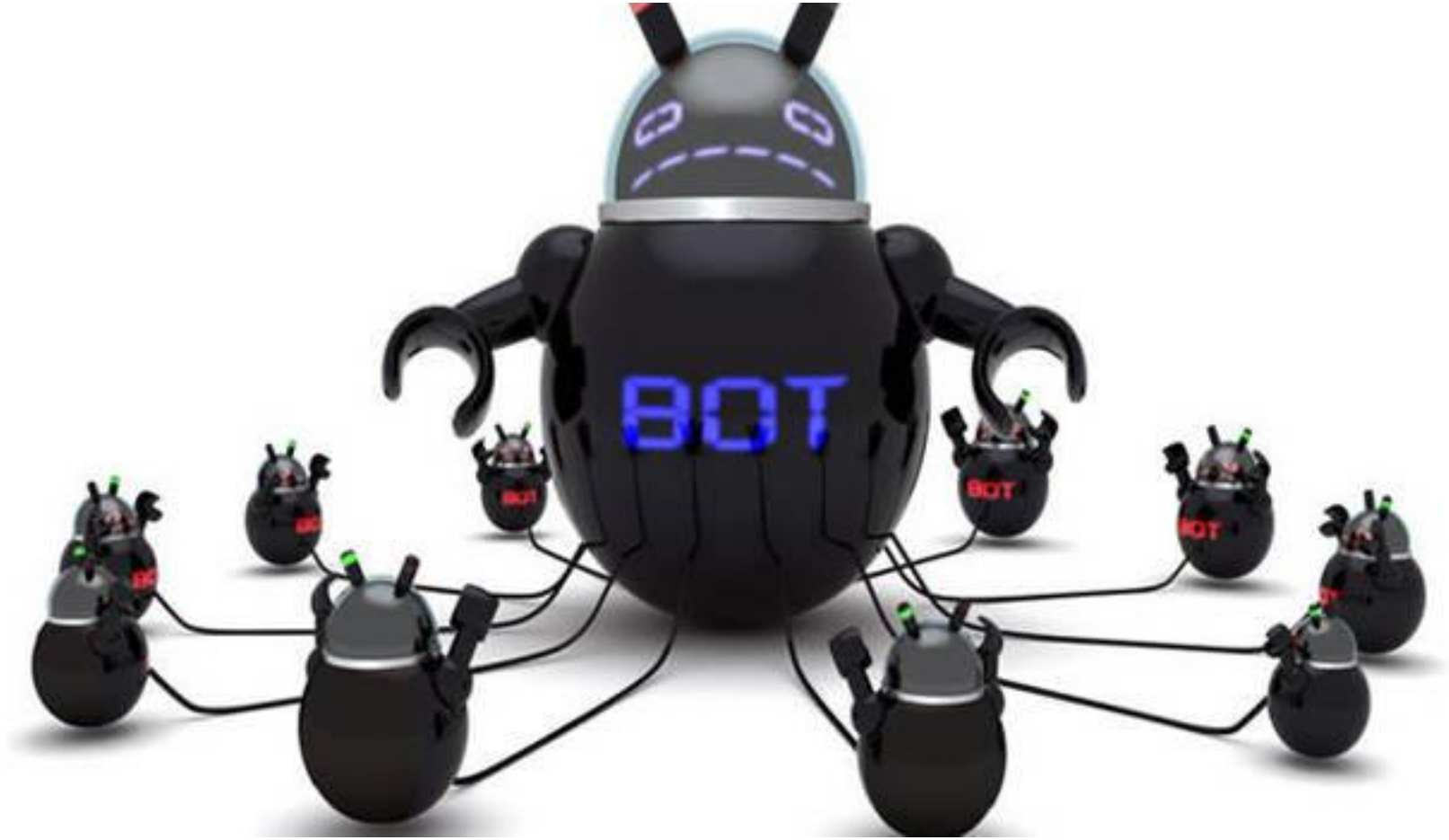
```
int i1, i2;
boolean b1,b2;
boolean[] a1,a2;
```

```
if (i1 < a1.length()) {
    boolean bval= a1[i1];
    if(bval){i2= 1;} else{i2= 0;}
    if(i2 < a2.length()){
        b2 = a2[i2];
    }
}
```

Threats (Capabilities)

- Physical access
- Software access
  - disk access
  - memory access
  - privilege levels
- Network access

Dyn DDoS



Recent DDoS Attacks

Record-breaking DDoS attack against Microsoft Azure mitigated

The attack was linked to the Aisuru botnet, which targets compromised home routers and cameras.

Published Nov. 19, 2025

Record 29.7 Tbps DDoS Attack Linked to AISURU Botnet with up to 4 Million Infected Hosts

Ravie Lakshmanan Dec 04, 2025 DDoS Attacks / Network Security

Czechia Under Coordinated DDoS Assault: Weekly DDoS Threat Intelligence Analysis

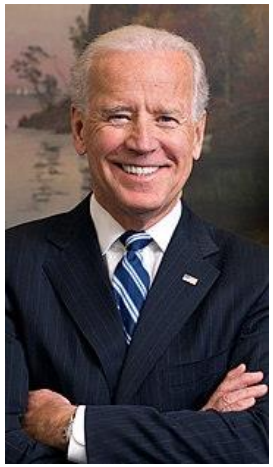
Analysis Period: January 19–25, 2026

Threats (Capabilities)

- Physical access
- Software access
  - disk access
  - memory access
  - privilege levels
- Network access
- User access

Phishing

twitter

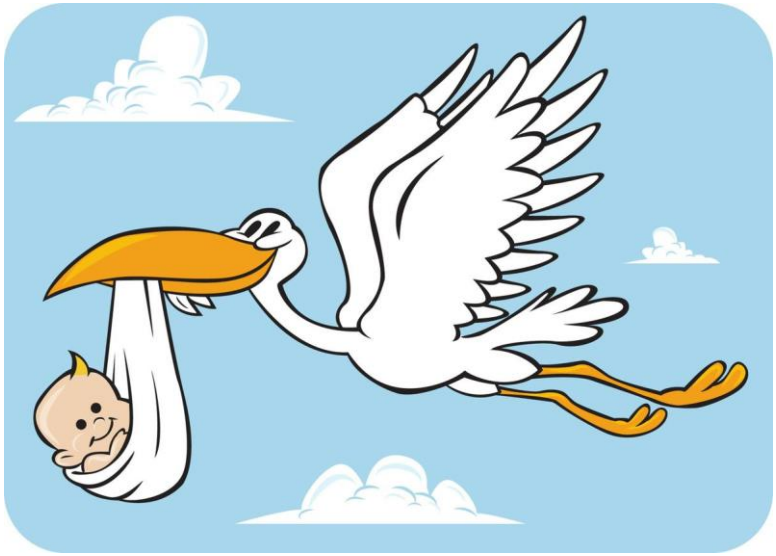


Formal Models of Capabilities

- PPT
- Dolev-Yao

Example

Threat model: The adversary desires to prevent baby deliveries. The adversary has access to radio equipment that transmits and receives on the same frequencies that providence uses for communication with a stork. The adversary also controls weapons systems that can destroy a stork in flight.



Exercise 4: Capability Threat Model

- For each of the following systems, how would you classify the threats that are likely to perpetrate attacks against that system according to their access?
  1. A game app
  2. A social network
  3. A classified government database

Threat Models