Môn học
An Toàn & Bảo mật HTTT

Giảng viên: Hà Lê Hoài Trung

Email: trunghlh@uit.edu.vn

Mục tiêu

- Nội dung chính của môn học là các kỹ thuật mã hoá DES, AES, mã hoá công khai, kiểm soát quyền truy cập, truy vết…Các kỹ thuật tấn công và phòng chống SQLi, và các kỹ thuật tấn công và bảo vệ tính riêng tư trong cơ sở dữ liệu thống kê.
- Đưa ra những đánh giá về rủi ro trong các hệ thống phân tích dữ liệu – chống lại các cuộc tấn công phá hủy.
- Xem xét đánh giá mức độ an toàn hệ thống trong quá trình thiết kế CSDL.

Nội dung bài học

1. Giới thiệu tổng quan – B01
2. Mã Hoá – B01 & B02
3. Xác thực người dùng – B03
4. Kiểm soát truy cập – B04

Kiểm tra giữa kỳ – tại lớp – thi đề mở

5. Kiểm toán bảo mật – B05
6. Thiết kế bảo mật CSDL – B05
7. Tính riêng tư trong CSDL – B06

Tài liệu tham khảo

1. Stallings W., Brown L., Bauer M.D. and Howard M. Computer security: principles and practice, 4th Edition, Pearson, 2018.
2. Matt Bishop. Computer security: art and science, 2nd Edition, Addison-Wesley Professional, 2018.
3. Mark Stamp. Information security: principles and practice, 2nd Edition, JohnWiley & Sons, 2011.
4. Joseph, A.D., Nelson, B., Rubinstein, B.I. and Tygar, J.D., 2018. Adversarial machine learning. Cambridge University Press.
5. Vorobeychik, Y. and Kantarcioglu, M., 2022. *Adversarial machine learning*. Springer Nature.

Đánh giá

- Quá trình: 30%
  - Assignment: 25%
  - Bài tập về nhà - trên lớp: 5%
  - Cộng tối đa 2đ cho seminar (slide + report) – tài liệu 4,5,6. Thông qua form (tùy chọn – không bắt buộc, số lượng có hạn)
- Giữa kỳ: 20%:
  - Sử dụng 1 tờ giấy A4
  - Lý thuyết & Bài tập
  - Tự luận
- Cuối kỳ: 50%
  - Được đem 2 tờ giấy A4
  - Lý thuyết & Bài tập
  - Tự luận & trắc nghiệm

Dạng bài tập Assignment

- Machine Learning Security
  - Malware
  - Instruction Detection
  - Zero – day
  - Authentication Biometrics
  - Phising Domain
  - SQL Injection
- Web
- Security for machine learning
- Privacy for machine learning

Web – 3 nhóm

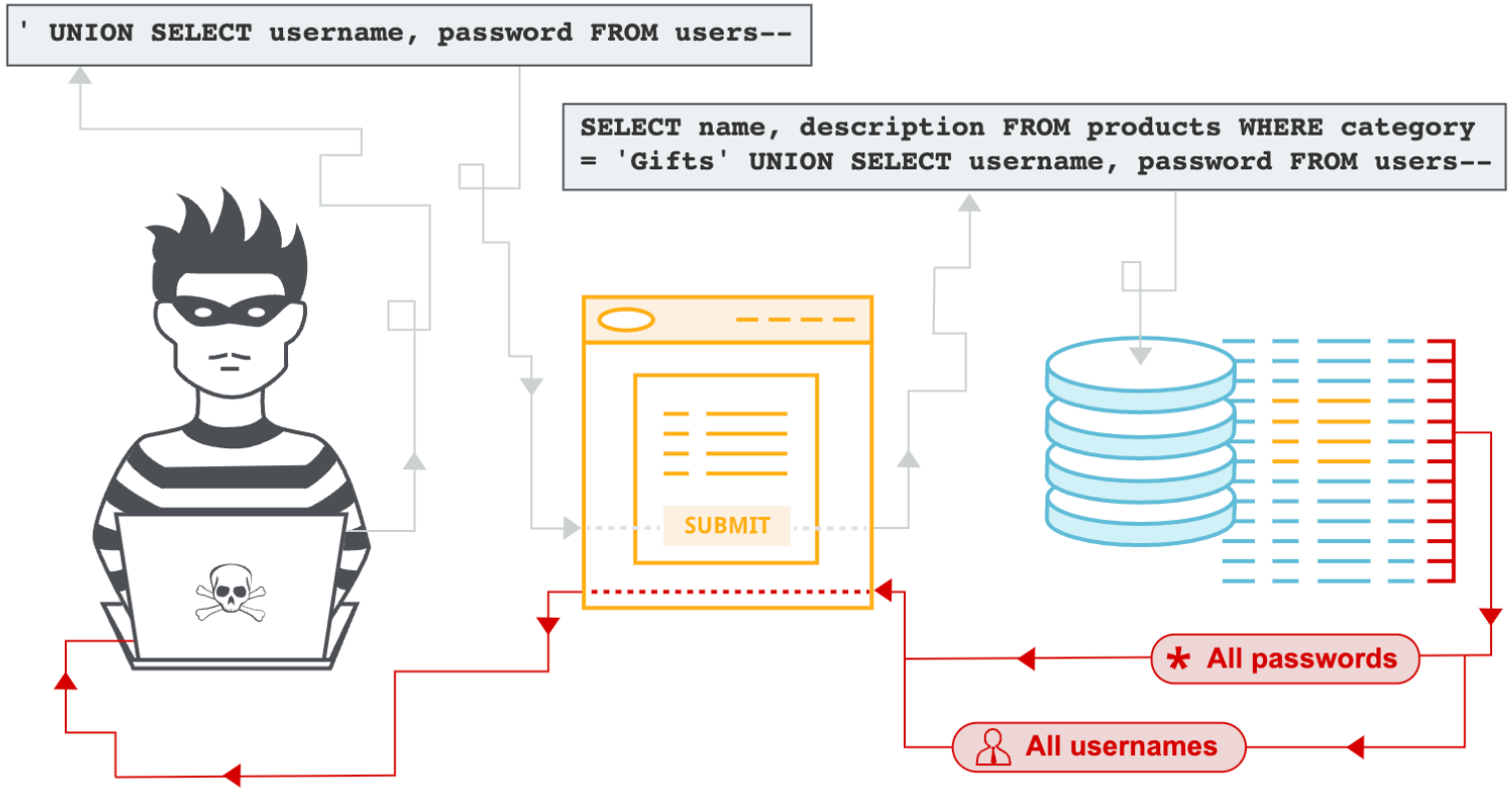
- Các công việc cần thực hiện
- Dựng trang web: tối thiểu có chức năng sau
  - Login – Signup
  - Xem danh sách món hàng
  - Xem chi tiết món hàng
  - Thực hiện thêm món hàng – chỉnh sửa món hàng
- Thực hiện 2 kịch bản tấn công trên web server:
  - SQLi, prototype population, cross-site scripting, cross-site request forgery, XML external entity injection, clickjacking, Cross-origin resource sharing (CORS), Server-side request forgery (SSRF), HTTP request smuggling, Server-side template injection, Access control vulnerabilities and privilege escalation, Authentication vulnerabilities, OAuth 2.0 authentication vulnerabilities, Testing for WebSockets security vulnerabilities, DOM-based vulnerabilities, Web cache poisoning, HTTP Host header attacks, File upload vulnerabilities, JWT attacks.

Web

- Thực hiện 2 kịch bản tấn công trên web server
  - Giới thiệu kỹ thuật tấn công – kỹ thuật phòng chống
  - Thực hiện bài lab
  - Áp dụng vào trang web miêu tả
- Tài liệu tham khảo
  - https://portswigger.net/web-security/all-materials
  - Web Application Security: Exploitation and Countermeasures for Modern Web Applications

Web

- SQLi




```
https://insecure-website.com/products?category=Gifts
```

```
https://insecure-website.com/products?category=Gifts'--
```

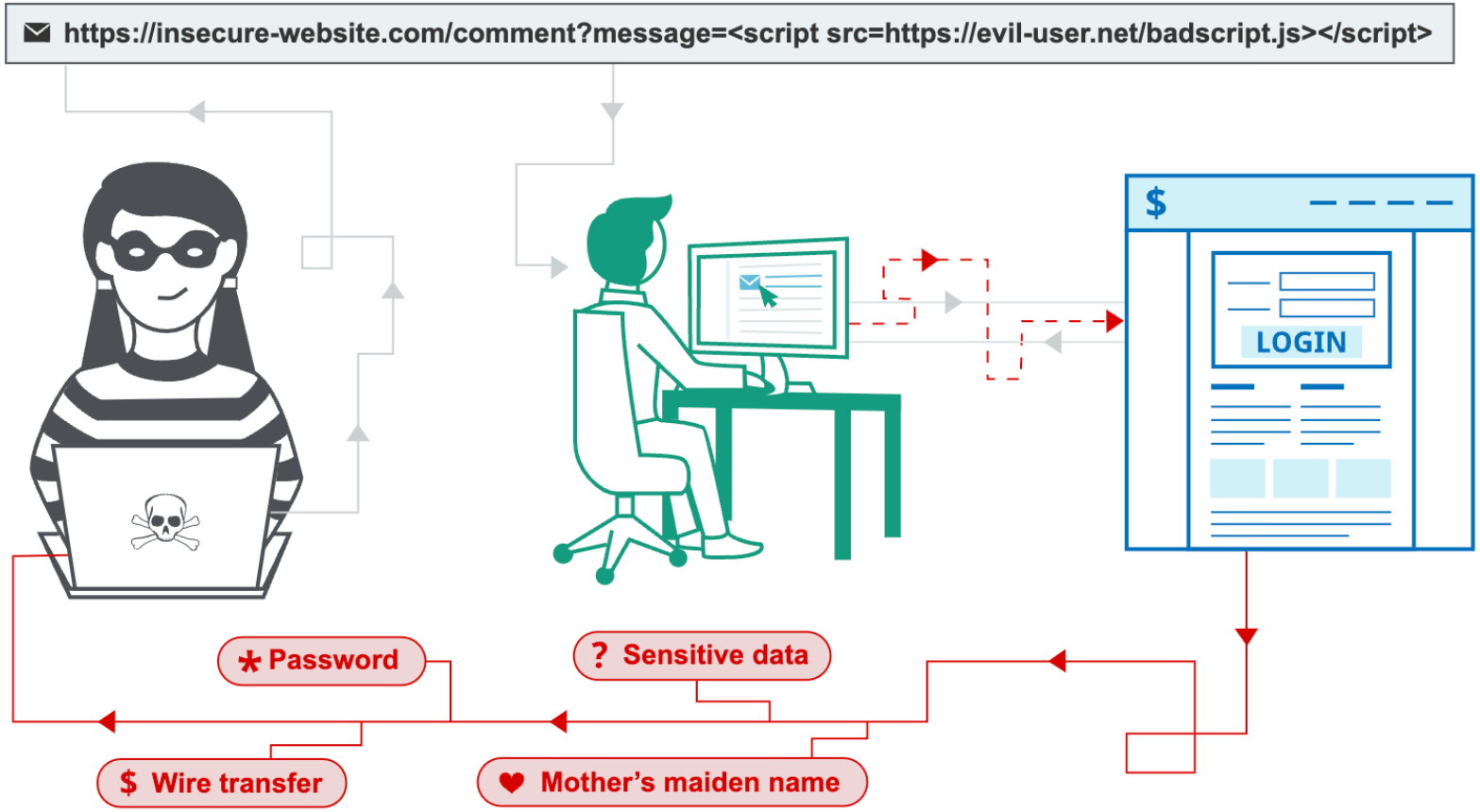
```
SELECT * FROM products WHERE category = 'Gifts'--' AND released = 1
```

Web

```
https://insecure-website.com/status?message=All+is+well.
<p>Status: All is well.</p>
```

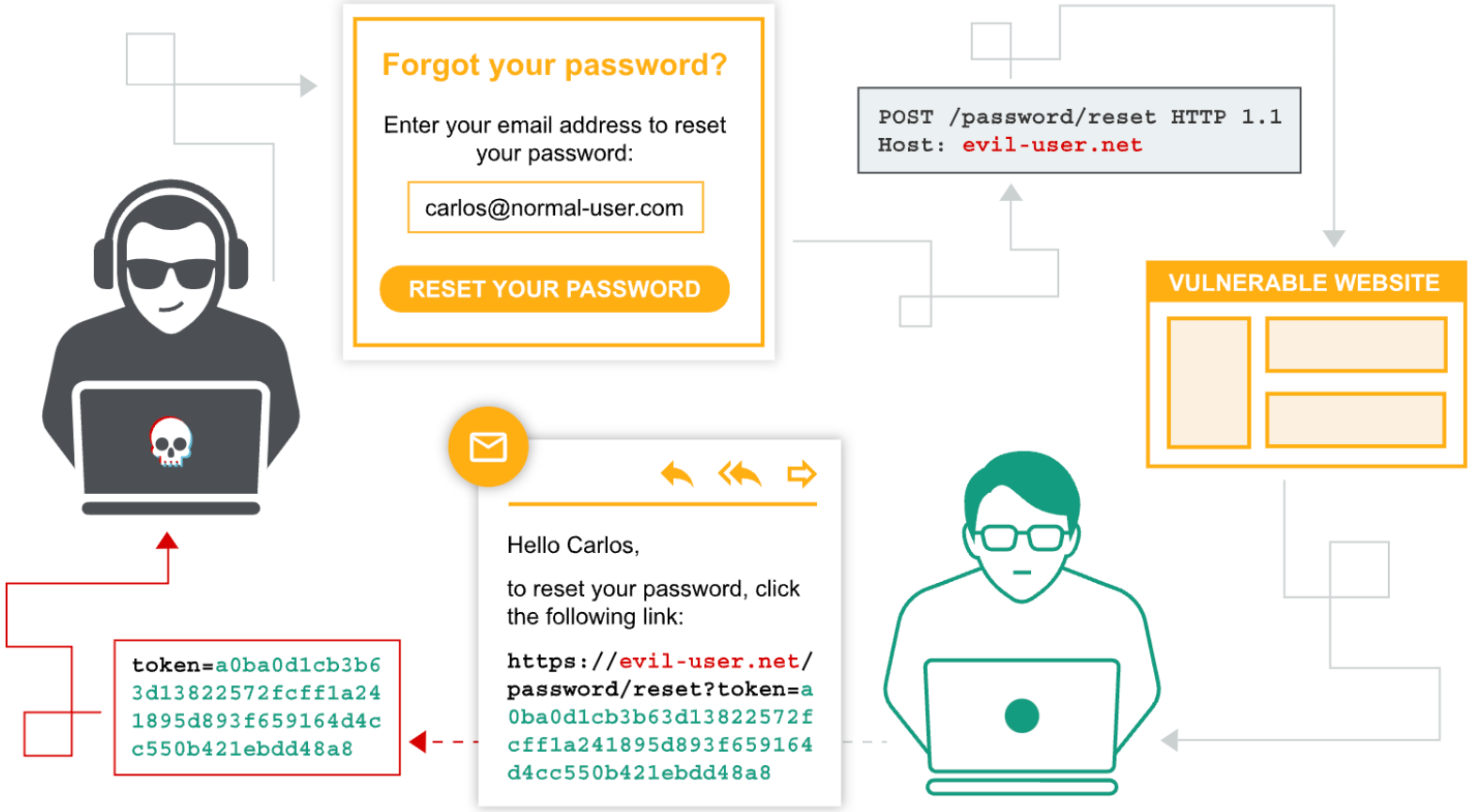
```
https://insecure-website.com/status?message=<script>/*+Bad+stuff+here.
<p>Status: <script>/* Bad stuff here... */</script></p>
```

- Cross-site scripting



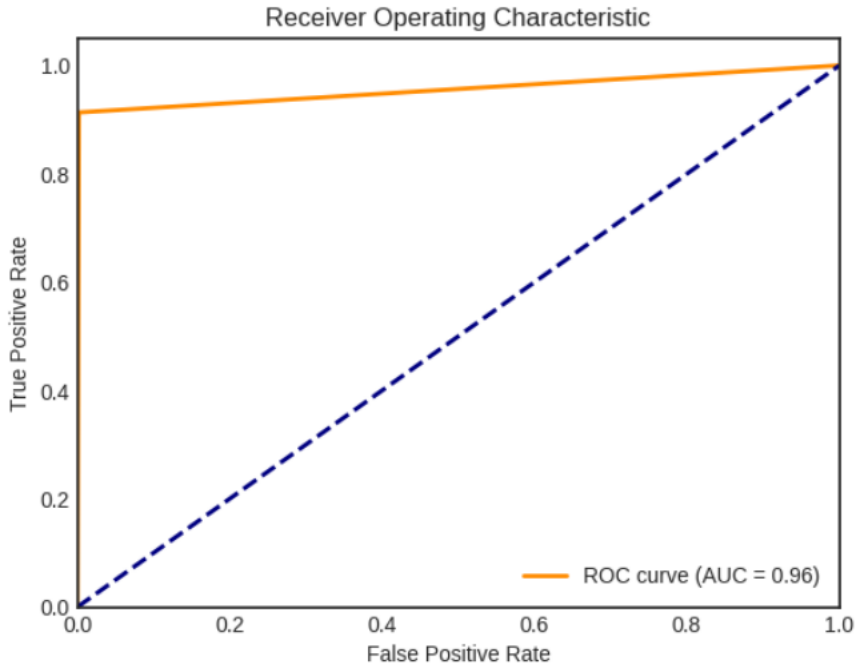
Web

- Authentication



Machine Learning Security

- Các công việc thực hiện:
  - Chọn tập dataset: Instruction, Virus
  - Giới thiệu giải thuật data mining
  - Thực hiện các giải thuật data mining thực hi
  - Trực quan kết quả thực hiện được



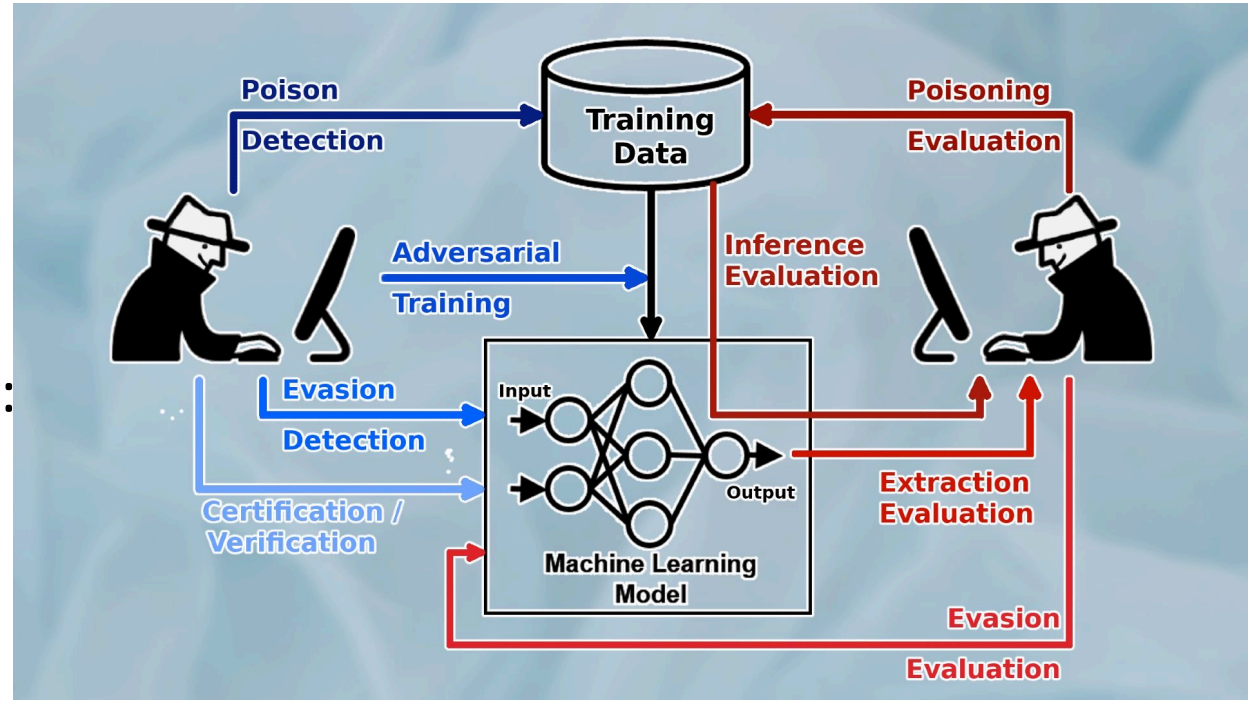
Scenarios	Scenarios 1	Scenarios 2	Scenarios 3	Scenarios 4	Scenarios 5
Learning rate	0.1	0.01	0.01	0.01	0.01
Activation Function	sigmoid	Softmax	sigmoid	sigmoid	sigmoid
Loss function	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy
Epoch	10	20	10	10	20
Batch Size	32	32	32	32	32
Dataset Split	6-2-2	6-2-2	6-2-2	7-2-1	7-2-1
Dense Layer	256	64	64	64	64
Training time	85.13804	105.9209	**50.9064**	85.6601	110.8019
Inference time	1.6612	1.4660	1.5476	0.6701	**0.6264**
Training Accuracy	0.9668	0.9659	0.9658	**0.9674**	0.9674
Testing Accuracy	0.9571	0.9334	0.9614	**0.9627**	0.9611
Validation Accuracy	0.9267	0.9338	0.9320	**0.9367**	0.9367
F1 Score	0.9390	0.9014	0.9452	**0.9462**	0.9437
Precision	0.8949	0.8394	0.9090	**0.9110**	0.9056
Recall	**0.9877**	0.9733	0.9845	0.9843	0.9851

Machine Learning Security

- Tài liệu tham khảo:
  - Hands-On Machine Learning for Cybersecurity: Safeguard your system by making your machines intelligent using the Python ecosystem
  - Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies
  - Machine Learning for Cybersecurity Cookbook: Over 80 recipes on how to implement machine learning algorithms for building security systems using Python
  - Kapoor, A., Gulli, A., Pal, S. and Chollet, F., 2022. Deep Learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd.

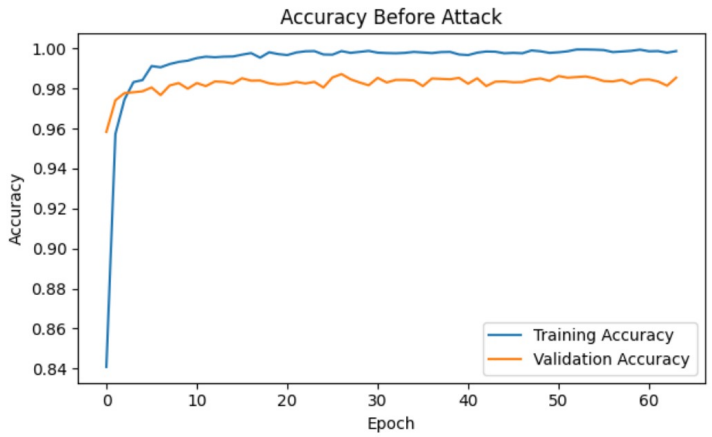
Security Machine Learning

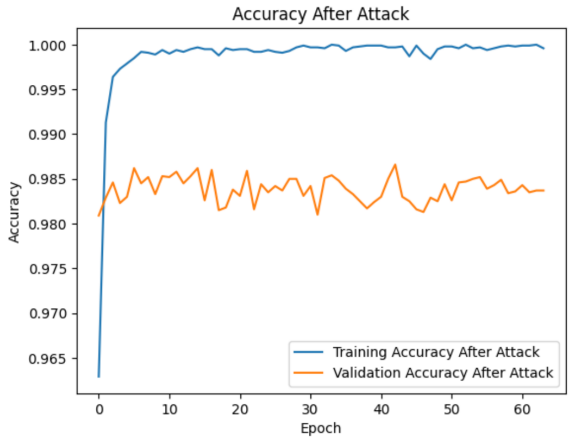
- Các công việc thực hiện
  - Chọn tập dataset
  - Thực hiện training model
  - Chọn một trong các kỹ thuật tấn công:
    - Evasion Attacks
      - ✓ Fast Gradient Sign Method (FGSM)
      - ✓ Projected Gradient Descent (PGD)
    - Poisoning Attacks
    - Model Inversion Attacks
    - Model Extraction Attacks
  - Thực hiện tấn công phá hủy mô hình
  - Chọn một trong các kỹ thuật phòng chống:
    - Adversarial Training, Gradient Masking, Defensive Distillation, Gradient Masking
  - Trực quan kết quả



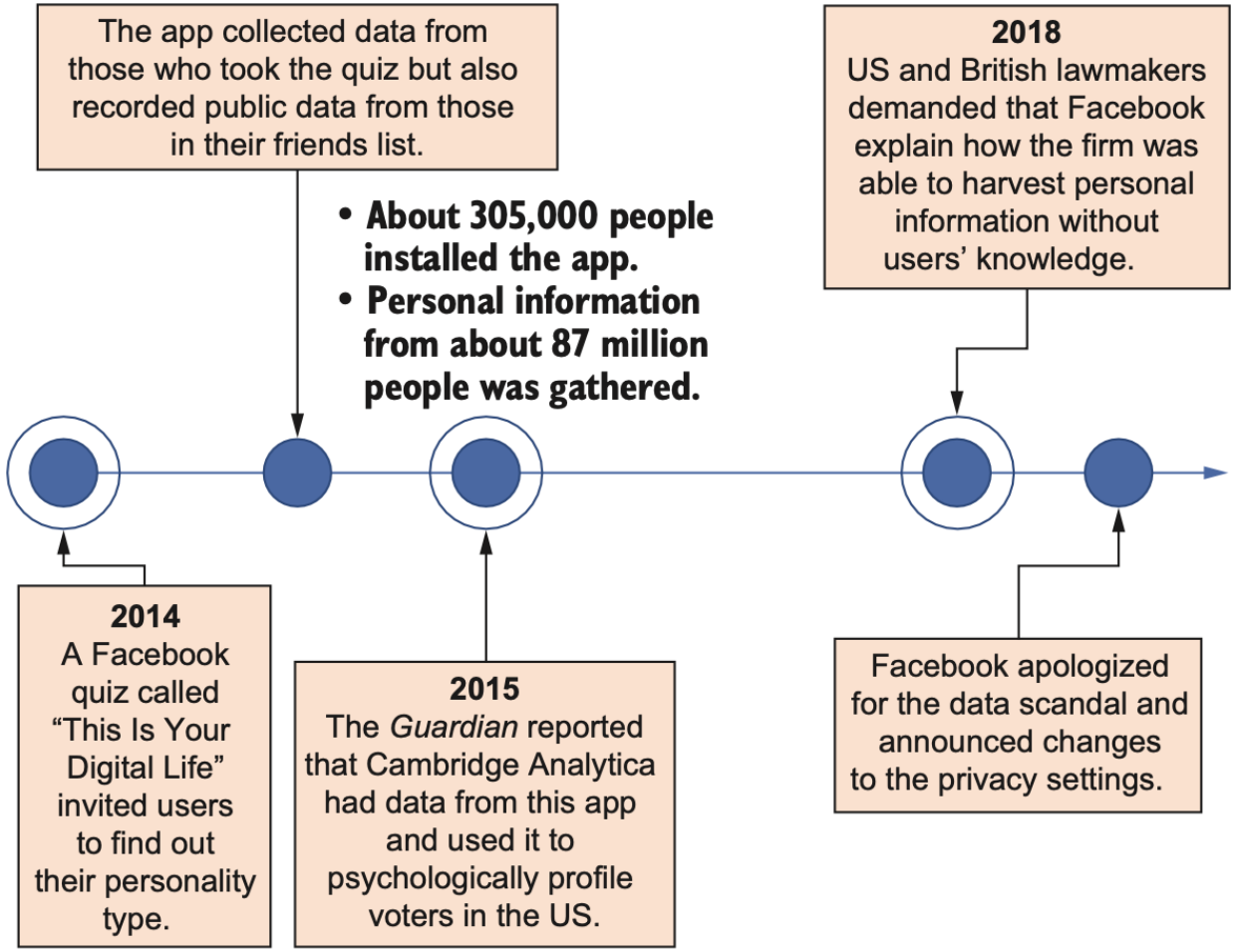
Security Machine Learning

- Tài liệu tham khảo
  - Joseph, A.D., Nelson, B., Rubinstein, B.I. and Tygar, J.D., 2018. Adversarial machine learning. Cambridge University Press.
  - Vorobeychik, Y. and Kantarcioglu, M., 2022. Adversarial machine learning. Springer Nature.
  - Kapoor, A., Gulli, A., Pal, S. and Chollet, F., 2022. Deep Learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd.



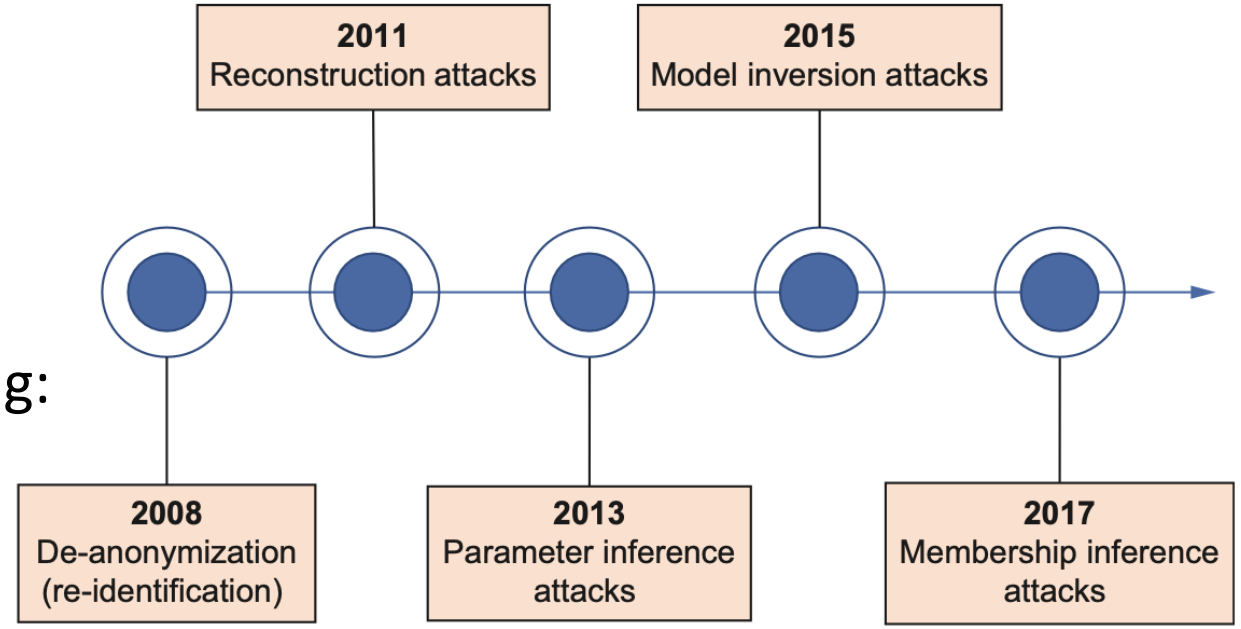


Privacy machine learning



Privacy machine learning

- Các công việc thực hiện
  - Chọn tập dataset
  - Thực hiện training model
  - Chọn một trong các kỹ thuật tấn công:
    - MIA: data, model parameter
    - Model Inversion Attack
    - Reconstruction Attack
    - De-anonymization or re-identification attacks
  - Thực hiện tấn công phá hủy mô hình
  - Chọn một trong các kỹ thuật phòng chống
  - Trực quan kết quả

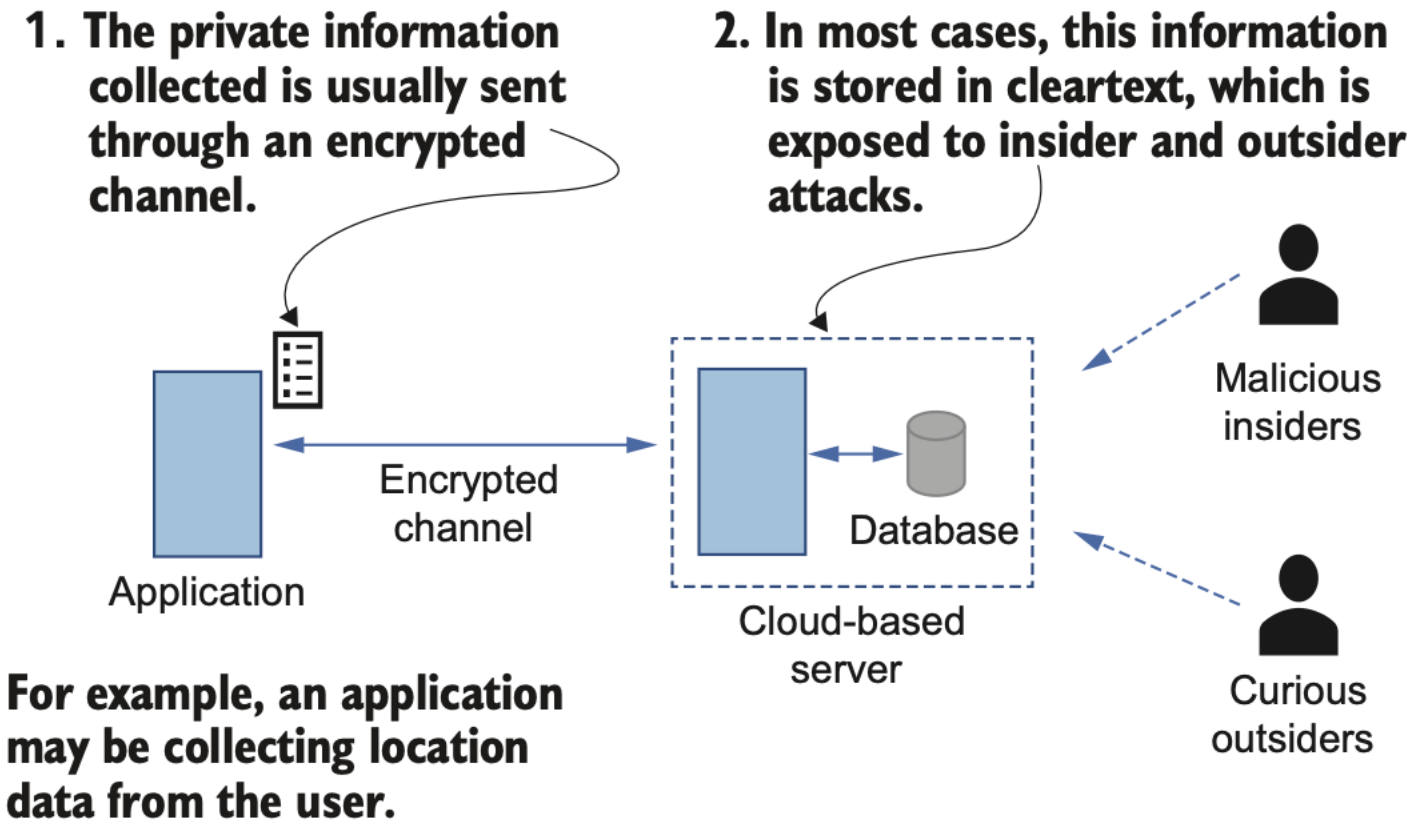


Privacy machine learning

- Tài liệu tham khảo:

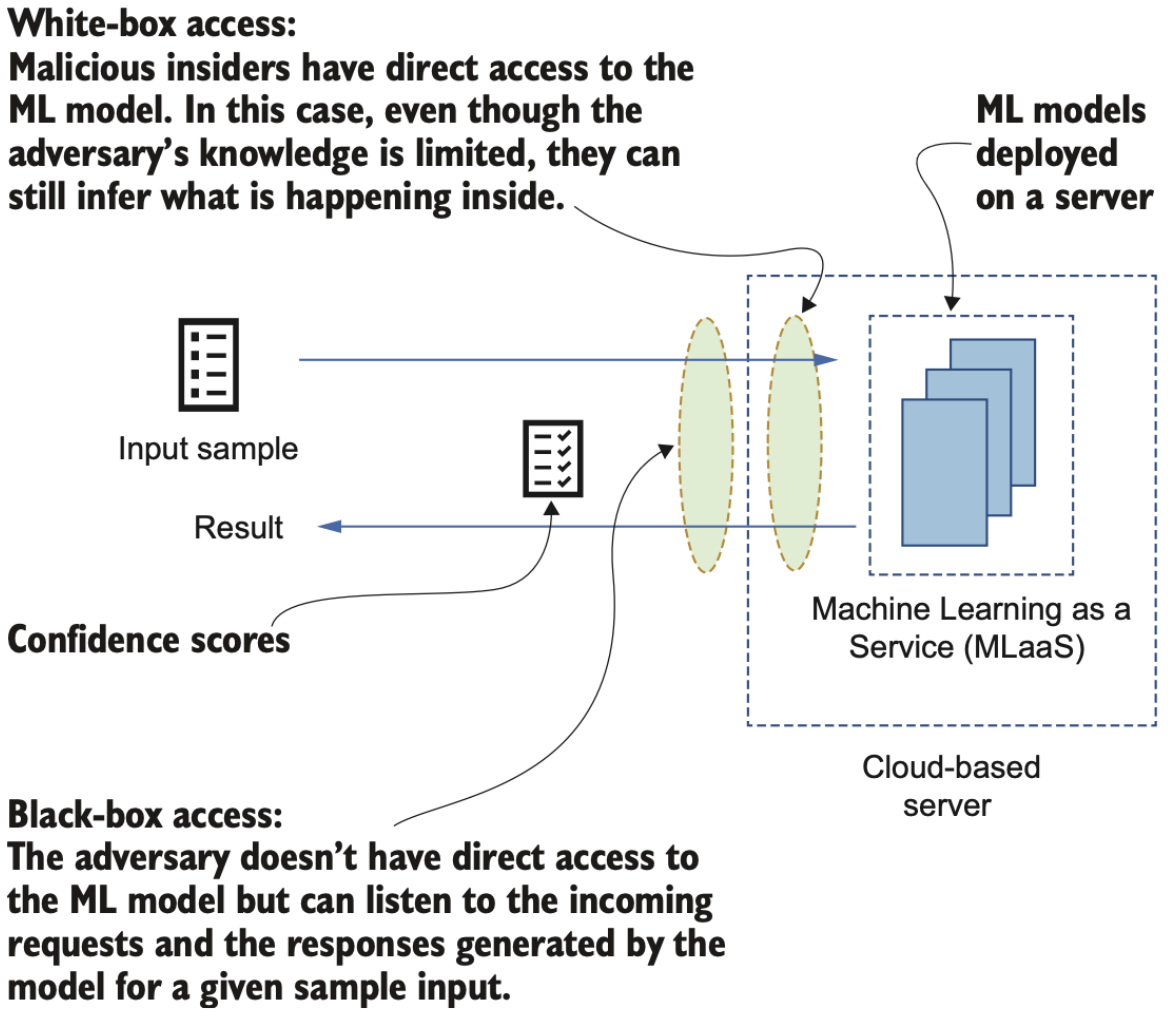
  - Near, J.P. and Abuah, C., 2021. Programming differential privacy. URL: https://uvm.
  - Kapoor, A., Gulli, A., Pal, S. and Chollet, F., 2022. Deep Learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd.
  - Chang, J.M., Zhuang, D., Samaraweera, G. and Samaraweera, G.D., 2023. Privacy-Preserving Machine Learning. Simon and Schuster.

Privacy machine learning



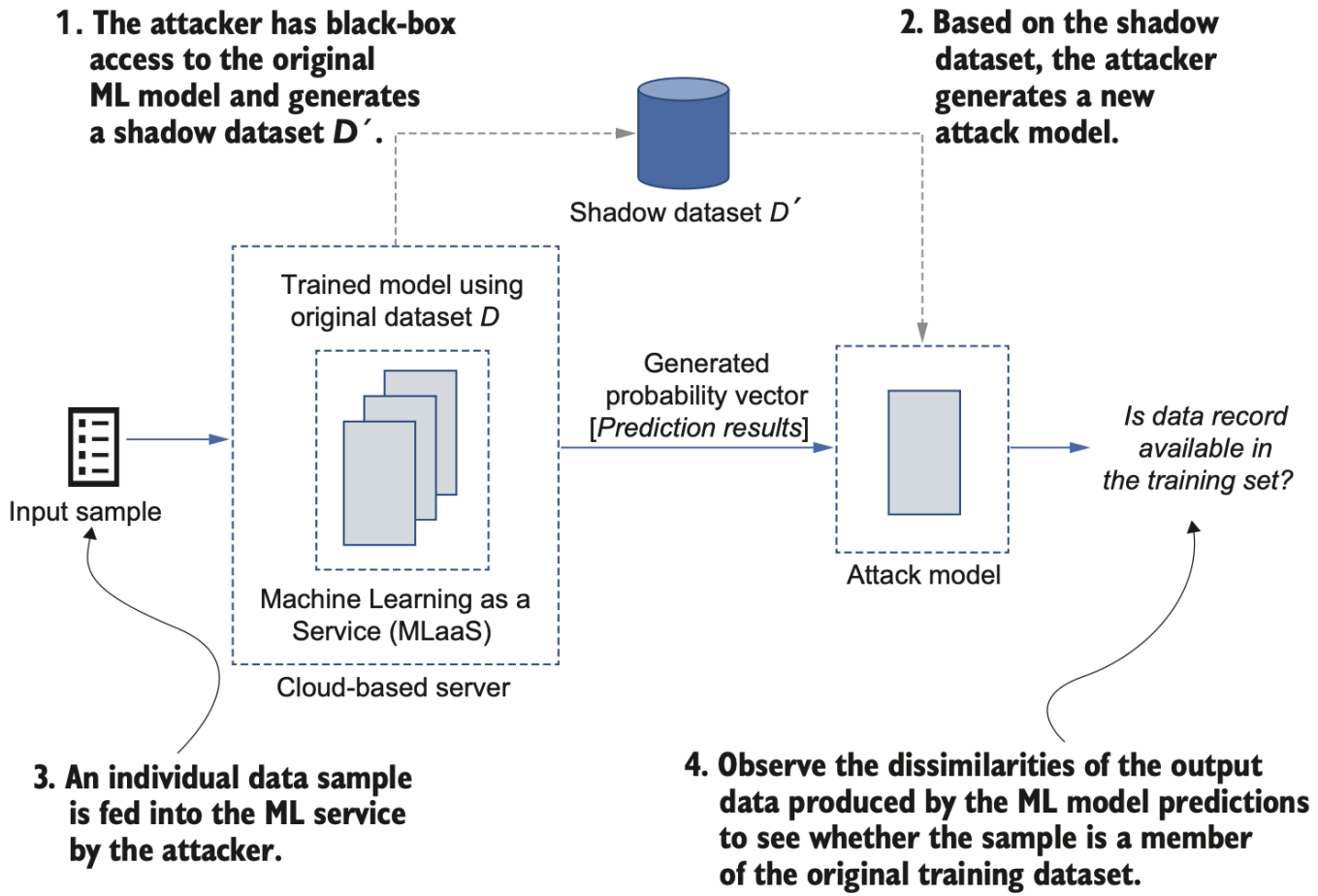
Privacy machine learning

- Model inversion attacks



Privacy machine learning

- Membership inference attacks



Computer Security: Principles and Practice

Chapter 1: Overview

Chapter 1 overview

- Computer Security Concepts
- Threats, Attacks, and Assets
- Security Functional Requirements
- Fundamental Security Design Principles
- Attack Surfaces and Attack Trees
- Computer Security Strategy

Learning objectives

- Describe the key security requirements of confidentiality, integrity and availability
- Discuss the types security threats and attacks that must be dealt with
- Summarize the functional requirements for computer security
- Explain the fundamental security design principles
- Discuss the use of attack surfaces and attack trees
- Understand the principle aspects of a comprehensive security strategy

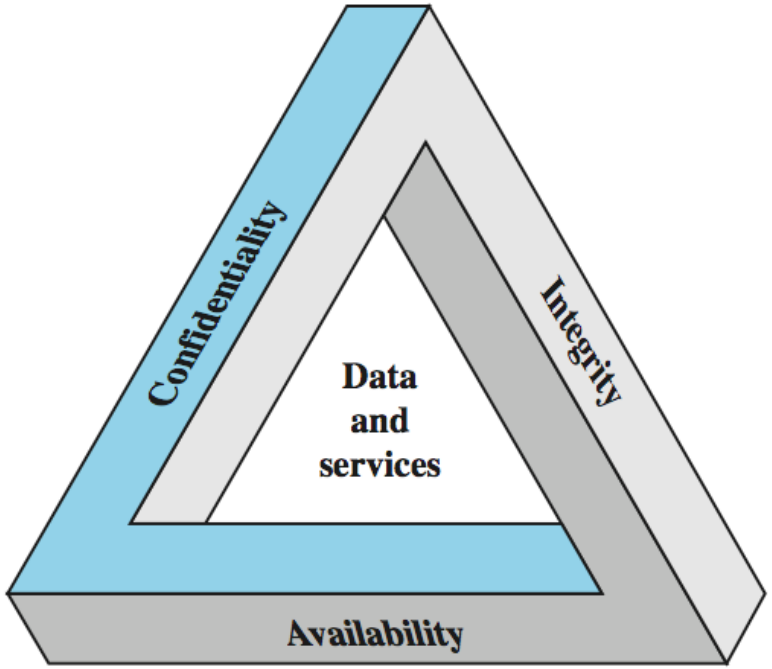
A definition of computer security

- **Computer security:** The protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability and confidentiality of information system resources (includes hardware, software, firmware, information/data, and telecommunications)

NIST 1995

Three key objectives (the CIA triad)

- **Confidentiality**
  - **Data confidentiality**: Assures that confidential information is not disclosed to unauthorized individuals
  - **Privacy**: Assures that individual control or influence what information may be collected and stored
- **Integrity**
  - **Data integrity**: assures that information and programs are changed only in a specified and authorized manner
  - **System integrity**: Assures that a system performs its operations in unimpaired manner
- **Availability**: assure that systems works promptly and service is not denied to authorized users



Other concepts to a complete security picture

- **Authenticity**: the property of being genuine and being able to be verified and trusted; confident in the validity of a transmission, or a message, or its originator
- **Accountability**: generates the requirement for actions of an entity to be traced uniquely to that individual to support nonrepudiation, deference, fault isolation, etc

Levels of security breach impact

- **Low**: the loss will have a limited impact, e.g., a degradation in mission or minor damage or minor financial loss or minor harm
- **Moderate**: the loss has a serious effect, e.g., significance degradation on mission or significant harm to individuals but no loss of life or threatening injuries
- **High**: the loss has severe or catastrophic adverse effect on operations, organizational assets or on individuals (e.g., loss of life)

Examples of security requirements: Confidentiality

- Student grade information is an asset whose confidentiality is considered to be very high
  - The US FERPA Act: grades should only be available to students, their parents, and their employers (when required for the job)
- Student enrollment information: may have moderate confidentiality rating; less damage if enclosed
- Directory information: low confidentiality rating; often available publicly

Examples of security requirements: Integrity

- A hospital patient's allergy information (high integrity data): a doctor should be able to trust that the info is correct and current
  - If a nurse deliberately falsifies the data, the database should be restored to a trusted basis and the falsified information traced back to the person who did it
- An online newsgroup registration data: moderate level of integrity
- An example of low integrity requirement: anonymous online poll (inaccuracy is well understood)

Examples of security requirements: Availability

- A system that provides authentication: high availability requirement
  - If customers cannot access resources, the loss of services could result in financial loss
- A public website for a university: a moderate availably requirement; not critical but causes embarrassment
- An online telephone directory lookup: a low availability requirement because unavailability is mostly annoyance (there are alternative sources)

Challenges of computer security

1. Computer security is not simple
2. One must consider potential (unexpected) attacks
3. Procedures used are often counter-intuitive
4. Must decide where to deploy mechanisms
5. Involve algorithms and secret info (keys)
6. A battle of wits between attacker / admin
7. It is not perceived on benefit until fails
8. Requires constant monitoring
9. Too often an after-thought (not integral)
10. Regarded as impediment to using system

A model for computer security

- Table 1.1 and Figure 1.1 show the relationship
- Systems resources
  - Hardware, software (OS, apps), data (users, system, database), communication facilities and network (LAN, bridges, routers, …)
- Our concern: vulnerability of these resources (corrupted, unavailable, leaky)
- Threats exploit vulnerabilities
- Attack is a threat that is accrued out
  - Active or passive; from inside or from outside
- Countermeasures: actions taken to prevent, detect, recover and minimize risks

Computer security terminology

Adversary (threat agent)
An entity that attacks, or is a threat to, a system.

Attack
An assault on system security that derives from an intelligent threat; that is, an intelligent act that is a deliberate attempt (especially in the sense of a method or technique) to evade security services and violate the security policy of a system.

Countermeasure
An action, device, procedure, or technique that reduces a threat, a vulnerability, or an attack by eliminating or preventing it, by minimizing the harm it can cause, or by discovering and reporting it so that corrective action can be taken.

Risk
An expectation of loss expressed as the probability that a particular threat will exploit a particular vulnerability with a particular harmful result.

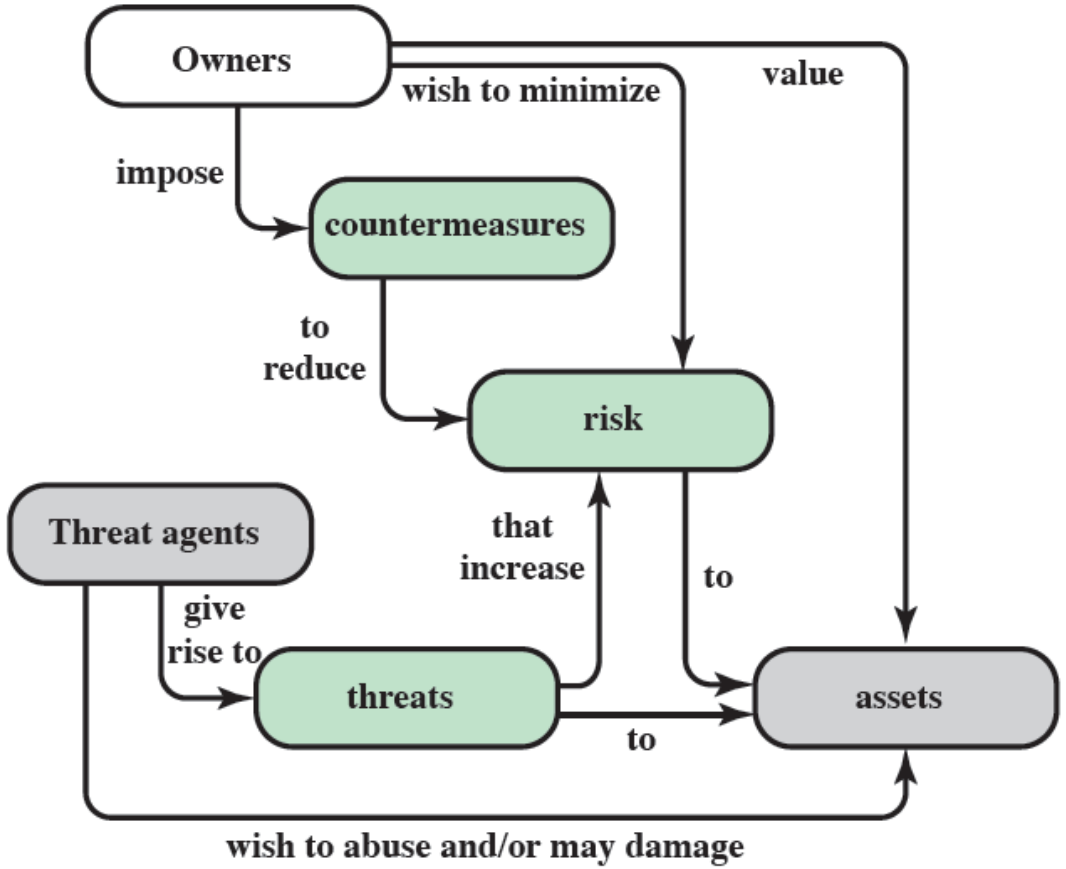
Security Policy
A set of rules and practices that specify or regulate how a system or organization provides security services to protect sensitive and critical system resources.

System Resource (Asset)
Data contained in an information system; or a service provided by a system; or a system capability, such as processing power or communication bandwidth; or an item of system equipment (i.e., a system component--hardware, firmware, software, or documentation); or a facility that houses system operations and equipment.

Threat
A potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm. That is, a threat is a possible danger that might exploit a vulnerability.

Vulnerability
A flaw or weakness in a system's design, implementation, or operation and management that could be exploited to violate the system's security policy.

Security concepts and relationships



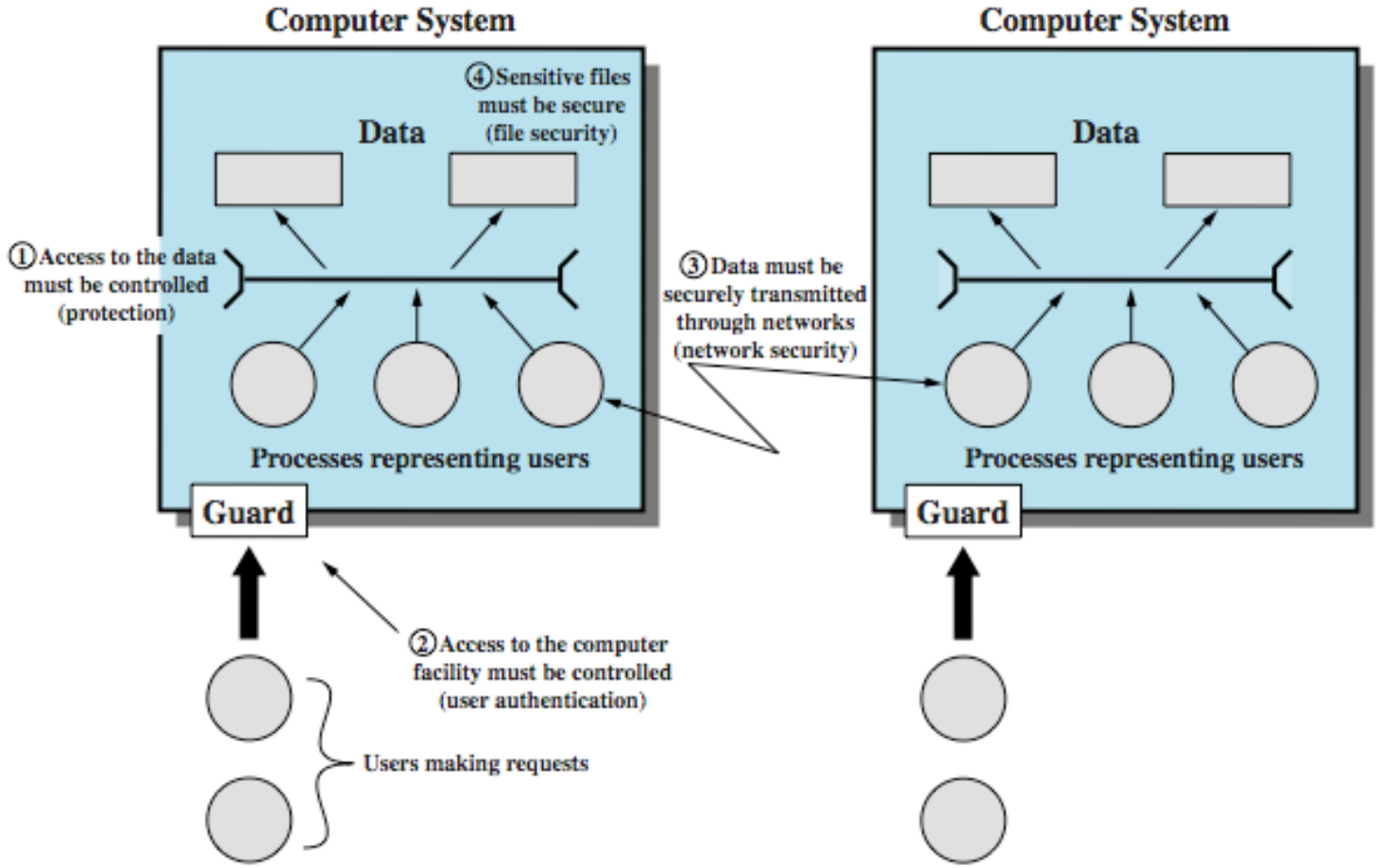
Threat consequences

- Unauthorized disclosure: threat to confidentiality
  - Exposure (release data), interception, inference, intrusion
- Deception: threat to integrity
  - Masquerade, falsification (alter data), repudiation
- Disruption: threat to integrity and availability
  - Incapacitation (destruction), corruption (backdoor logic), obstruction (infer with communication, overload a line)
- Usurpation: threat to integrity
  - Misappropriation (theft of service), misuse (hacker gaining unauthorized access)

Threat consequences (tabular form)

Threat Consequence	Threat Action (Attack)
Unauthorized Disclosure A circumstance or event whereby an entity gains access to data for which the entity is not authorized.	**Exposure:** Sensitive data are directly released to an unauthorized entity. **Interception:** An unauthorized entity directly accesses sensitive data traveling between authorized sources and destinations. **Inference:** A threat action whereby an unauthorized entity indirectly accesses sensitive data (but not necessarily the data contained in the communication) by reasoning from characteristics or byproducts of communications. **Intrusion:** An unauthorized entity gains access to sensitive data by circumventing a system's security protections.
Deception A circumstance or event that may result in an authorized entity receiving false data and believing it to be true.	**Masquerade:** An unauthorized entity gains access to a system or performs a malicious act by posing as an authorized entity. **Falsification:** False data deceive an authorized entity. **Repudiation:** An entity deceives another by falsely denying responsibility for an act.
Disruption A circumstance or event that interrupts or prevents the correct operation of system services and functions.	**Incapacitation:** Prevents or interrupts system operation by disabling a system component. **Corruption:** Undesirably alters system operation by adversely modifying system functions or data. **Obstruction:** A threat action that interrupts delivery of system services by hindering system operation.
Usurpation A circumstance or event that results in control of system services or functions by an unauthorized entity.	**Misappropriation:** An entity assumes unauthorized logical or physical control of a system resource. **Misuse:** Causes a system component to perform a function or service that is detrimental to system security.

The scope of computer security



Examples of threats

	Availability	Confidentiality	Integrity
Hardware	Equipment is stolen or disabled, thus denying service.	An unencrypted CD-ROM or DVD is stolen.	
Software	Programs are deleted, denying access to users.	An unauthorized copy of software is made.	A working program is modified, either to cause it to fail during execution or to cause it to do some unintended task.
Data	Files are deleted, denying access to users.	An unauthorized read of data is performed. An analysis of statistical data reveals underlying data.	Existing files are modified or new files are fabricated.
Communication Lines and Networks	Messages are destroyed or deleted. Communication lines or networks are rendered unavailable.	Messages are read. The traffic pattern of messages is observed.	Messages are modified, delayed, reordered, or duplicated. False messages are fabricated.

Security functional requirements (FIPS 200)

- Technical measures
  - Access control; identification & authentication; system & communication protection; system & information integrity
- Management controls and procedures
  - Awareness & training; audit & accountability; certification, accreditation, & security assessments; contingency planning; maintenance; physical & environmental protection; planning; personnel security; risk assessment; systems & services acquisition
- Overlapping technical and management
  - Configuration management; incident response; media protection

Fundamental security design principles [1/4]

- Despite years of research, it is still difficult to design systems that comprehensively prevent security flaws
- But good practices for good design have been documented (analogous to software engineering)
  - Economy of mechanism, fail-safe defaults, complete mediation, open design, separation of privileges, lease privilege, least common mechanism, psychological accountability, isolation, encapsulation, modularity, layering, least astonishment

Fundamental security design principles [2/4]

- **Economy of mechanism**: the design of security measures should be as simple as possible
  - Simpler to implement and to verify
  - Fewer vulnerabilities
- **Fail-safe default**: access decisions should be based on permissions; i.e., the default is lack of access
- **Complete mediation**: every access should checked against an access control system
- **Open design**: the design should be open rather than secret (e.g., encryption algorithms)

Fundamental security design principles [3/4]

- **Isolation**
  - Public access should be isolated from critical resources (no connection between public and critical information)
  - Users files should be isolated from one another (except when desired)
  - Security mechanism should be isolated (i.e., preventing access to those mechanisms)
- **Encapsulation**: similar to object concepts (hide internal structures)
- **Modularity**: modular structure

Fundamental security design principles [4/4]

- **Layering (defense in depth)**: use of multiple, overlapping protection approaches
- **Least astonishment**: a program or interface should always respond in a way that is least likely to astonish a user

Fundamental security design principles

- **Separation of privilege**: multiple privileges should be needed to do achieve access (or complete a task)
- Least privilege: every user (process) should have the least privilege to perform a task
- **Least common mechanism**: a design should minimize the function shared by different users (providing mutual security; reduce deadlock)
- **Psychological acceptability**: security mechanisms should not interfere unduly with the work of users

Attack surfaces

- Attack surface: the reachable and exploitable vulnerabilities in a system
  - Open ports
  - Services outside a firewall
  - An employee with access to sensitive info
  - …
- Three categories
  - **Network attack surface** (i.e., network vulnerability)
  - **Software attack surface** (i.e., software vulnerabilities)
  - **Human attack surface** (e.g., social engineering)
- Attack analysis: assessing the scale and severity of threats

Computer security strategy

- An overall strategy for providing security
  - **Policy** (specs): what security schemes are supposed to do
    - Assets and their values
    - Potential threats
    - Ease of use vs security
    - Cost of security vs cost of failure/recovery
  - **Implementation/mechanism**: how to enforce
    - Prevention
    - Detection
    - Response
    - Recovery
  - **Correctness/assurance**: does it really work (validation/review)

Summary

- Security concepts
- Terminology
- Functional requirements
- Security design principles
- Security strategy