Threat Modeling

AI Engineering - Recitation 9

Threat Modeling

- Threat modeling works to identify, communicate, and understand threats and mitigations within the context of protecting something of value.
  - Can be applied to a wide range of things, including software, applications, systems, networks, etc.
  - Done preferably early, so that findings can inform the design.

- Why?
  - Build a secure design
  - Identify threats, and evaluate their risk
  - Define and build required controls
  - Document threats and mitigation strategies

Source: https://owasp.org/www-community/Application_Threat_Modeling

Threat Modeling

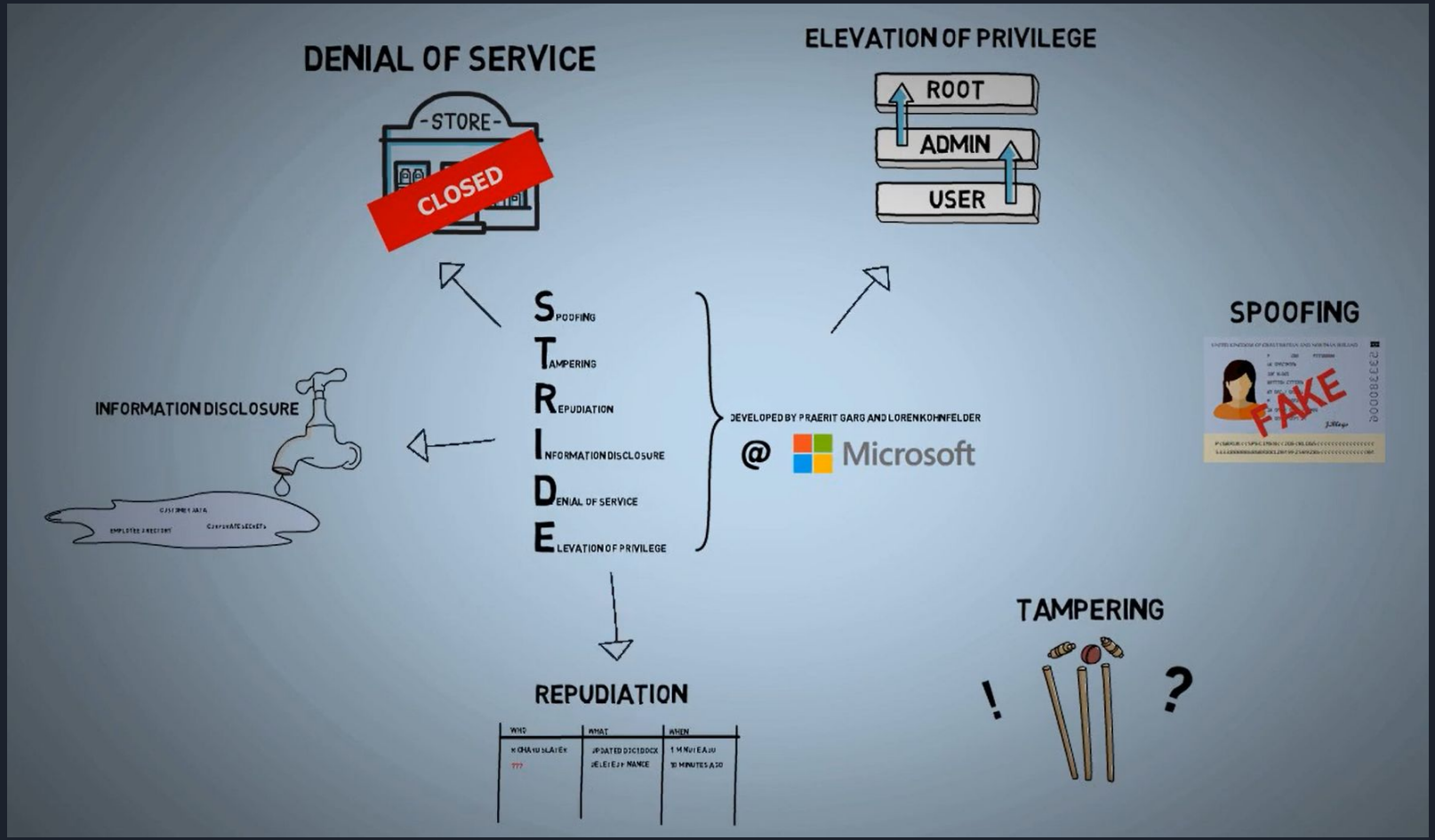
- Create a profile of an attacker
  - What are they trying to achieve?
  - What do they already know?
  - What can they do?
  - How much effort can they spend?
  - Why do they want to do this?

- Threat modeling is a formal process to do this

STRIDE

- S - Spoofing - violates authentication
- T - Tampering - violates integrity
- R - Repudiation - violates non-repudiation
- I - Information disclosure - violates confidentiality
- D - Denial of Service - violates availability
- E - Elevation of privilege - violates authorization

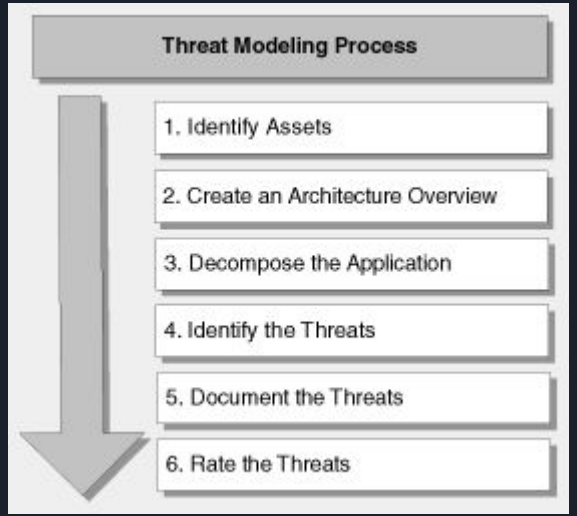
Source: https://docs.microsoft.com/en-us/archive/msdn-magazine/2006/november/uncover-security-design-flaws-using-the-stride-approach



Source: https://www.youtube.com/watch?v=HI_DK1S_TNU

STRIDE Process

- Identify valuable assets
- Construct simple architecture diagrams with every component and connections
  - Show data flow, trust boundaries
- For each component, identify threats
  - Document and rate threats
- For each threat, devise a mitigation strategy



Source: https://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10)

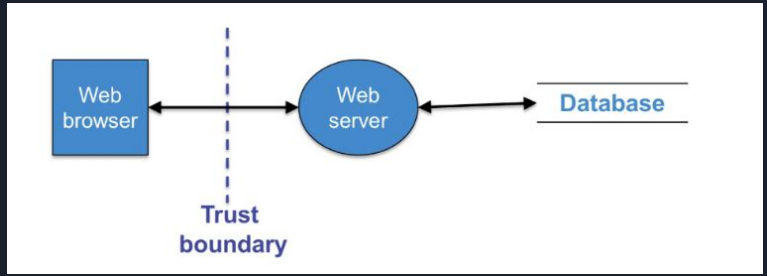
Data Flow Diagram

Item	Purpose	Symbol
Data Flow	Data in motion over network	Arrow
Data store	File, database, etc.	Parallel Lines
Process	Computation or program	Circle
Multi-process	Multiple processes	Two Circles
Trust Boundary	Border between trusted/untrusted entities	Dotted Line
Interactor	System endpoints	Rectangle

Source: https://docs.microsoft.com/en-us/archive/msdn-magazine/2006/november/uncover-security-design-flaws-using-the-stride-approach

Example - Web Application

- Scenario
  - Web server running a website, with a database.
  - Users need to login to view the content
- Assets & Security Objectives
  - User credentials
  - Maintain availability



Example - Web Application

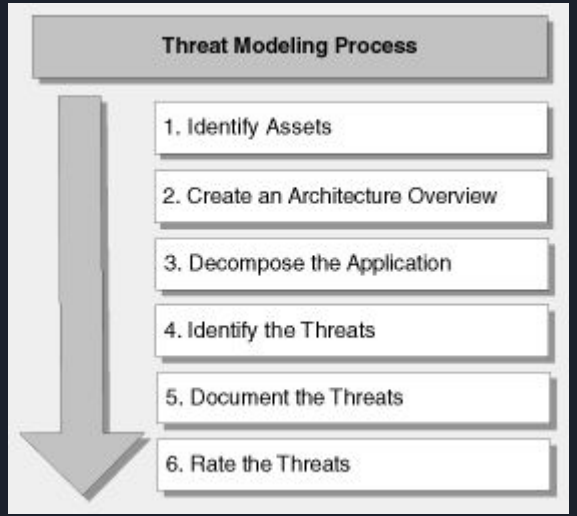
Threat	Stride Categories
Malicious user views or tampers with personal profile data en route from the web server to the client	Tampering, information disclosure
Attacker denies access to web server by flooding it with TCP/IP packets	Denial of service
Failure to validate cookie input	Tampering, information disclosure
Failure to sanitize data read from database	Information disclosure

Scenario

- System
  - Amazon-like online shopping platform
  - ML component recommends products based on user ratings
- Context
  - Several vendors are in close competition for selling products of similar types
- Attacker’s goal
  - Favor certain vendor’s products to be recommended over the others

Scenario - STRIDE Process

- What are the assets?
- What is our security objective?
- What components are there in our system?
- Where should we draw the trust boundary?
- What data goes in and out via the trust boundary?
    - Includes user interactions via interfaces



Additional Reading

- Microsoft's blogpost on Threat Modeling in AI/ML Systems
  - https://docs.microsoft.com/en-us/security/engineering/threat-modeling-aiml
- An Architectural Risk Analysis of ML Systems
  - https://berryvilleiml.com/docs/ara.pdf