

what software engineers should know about

# Security and Privacy

17-313 Fall 2024

Foundations of Software Engineering

<https://cmu-17313q.github.io>

Eduardo Feo Flushing

Sources:

- Some slides adapted from CMU 17-437/637 Web Application Development
- "What software engineers should know about privacy". MSE Seminar. Hana Habib. CMU
- "Ethics, Fairness, Responsibility, and Privacy in Data Science". CMSC 25900. U. Chicago

# Learning goals

- Explain why software is vulnerable to attacks
- Use the right secure software terminology
- Discuss a wide range of security attacks that can target software systems and tools and techniques to identify, prevent, and mitigate them
- Explain why privacy is not dead
- Describe common privacy principles and techniques
- Differentiate privacy threats

# Smoking Section

- Last **two** full rows



# Security

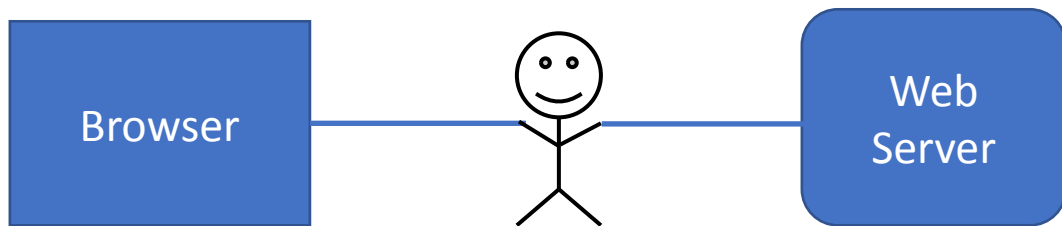


# Attacking the Network

- Examples
  - Person-in-the-middle attack
  - Sniffing
  - Spoofing
- We must assume the network is not secure
- We must guard against a compromised network

# Person-in-the-Middle

- Someone that can intercept network traffic
- Can read the messages (coming and going)
- Can change the messages before sending them on (to the correct or incorrect destination)



# Sniffing, Eavesdropping, etc

- You can listen to the traffic going by on the net
- This is typically traffic on your subnet
  - Still it can be most interesting
  - If you can plug in to the backbone...

# Spoofing

- Pretending to be someone you're not
- IP spoofing
  - Pretending to a "client" you're not (with a specific IP address)
- E-mail Spoofing
- DNS spoofing
  - Pretending to be a server that you're not
  - Fool a DNS server to give out incorrect IP addresses for DNS Names
- Note: also be careful of typos or similar characters attacks:
  - <http://mytimes.com>
  - <http://paypal.com>



# Attacks can be expensive

FT Financial Times

## UK regulator hits Equifax with £11mn fine over cyber breach

The Financial Conduct Authority has fined credit reporting agency Equifax just over £11mn for failing to protect the data of nearly 14mn UK...



The Guardian

## Uber fined \$148m for failing to notify drivers they had been hacked

Failure to report 2016 data breach 'one of the most egregious cases we've ever seen', says Illinois attorney general.



CBS News

## PlayStation Network breach has cost Sony \$171 million

(CBS/AP) TOKYO - Sony has spent 14 billion yen, which translates to roughly \$171 million, to cover the costs of the massive security breach...



May 24, 2011

# The big three

The “big three” concepts in network security

- Authentication
- Authorization
- Confidentiality

# Terms Defined

- Authentication

- Knowing with whom you are communicating
  - User knowing the server and/or server knowing the user
    - Which is more important??

- Authorization

- User having privilege to perform an operation on server

- Confidentiality

- Communicating without others knowing what's been said
- Intermediaries cannot change what was said
- Typically includes protection from replay attack

(Typically does *\*not\** provide secrecy of communication. Others can know communication occurred)

# Poll

- Which of the “big three” protect you from:
  - Sniffing? **Confidentiality**
  - Spoofing? **Authentication**
  - Person-in-the-middle Attack? **Authentication + Confidentiality**

# Concepts every SWE need to know

- One-way Hashing
- Secret Key Encryption
- Public Key Encryption
- Certificates

# Hashing

## (aka Message Digests, One-Way Hashing)

- A hash function is a one-way encoding of data
  - Same input, same output
  - Different output, different input
- Easy (relatively) to compute the hash function
- Hard to compute the hash function's inverse
- We only store hashed passwords on disk
  - To prevent passwords from being compromised if our servers are broken into

pbkdf2\_sha256\$180000\$pP3DfkAYXSS1\$L9JMQtFygrKbT246E/ZEaFScCTaXlp2v2ANN14ryXLY=


Algorithm

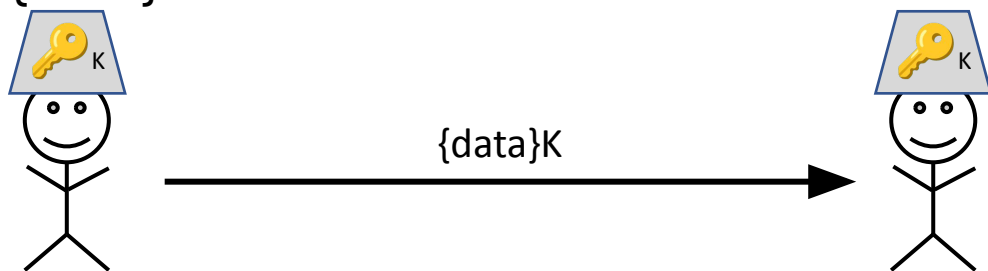
Iterations

Salt

Hashed Password

# Secret Key Cryptography (aka Symmetric, Private Key Crypto)

- Like in the old movies and spy books
- One key (K)
  - Shared Secret  K
  - Used to encrypt and decrypt
  - Notation: {data}K

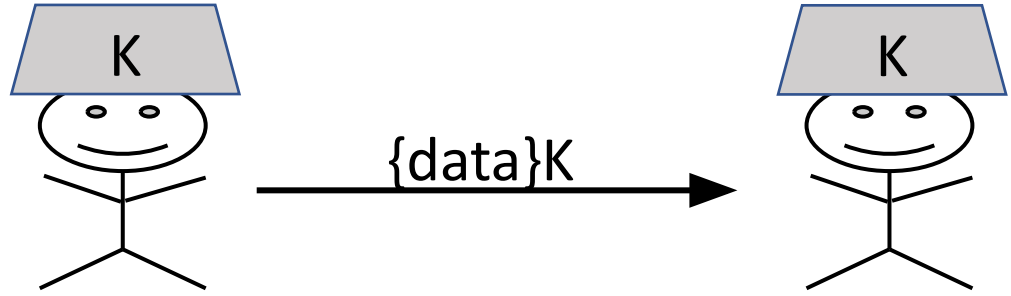


# Public Key Cryptography (aka Asymmetric Key Crypto)

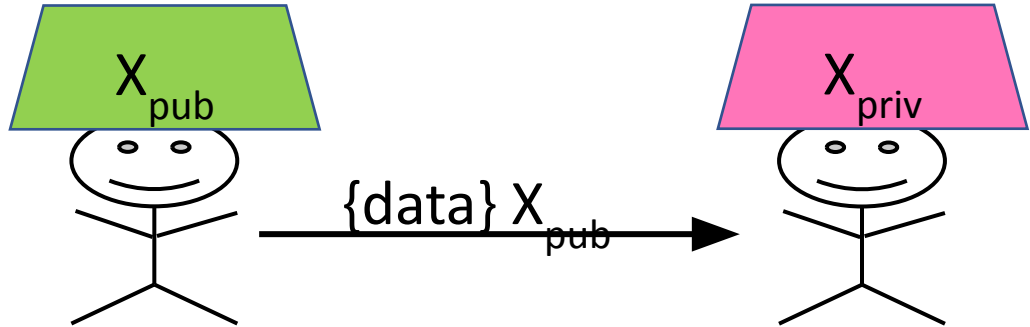
- Key Pair (key 1 & key 2)
  - Either key can be used to encrypt (key 1 or key 2)
  - You can only decrypt using the “other key” (key 2 or key 1)
  - One key is given out (**the public key**)
  - The other key is kept secret (**the private key**)
  - Notation: For entity X, we have keys  $X_{pub}$  &  $X_{priv}$
- A public key can be given out freely to
  - Encrypt data sent to the holder (X) of the private key
    - Notation:  $\{data\}_{X_{pub}}$



Secret Key  
Crypto

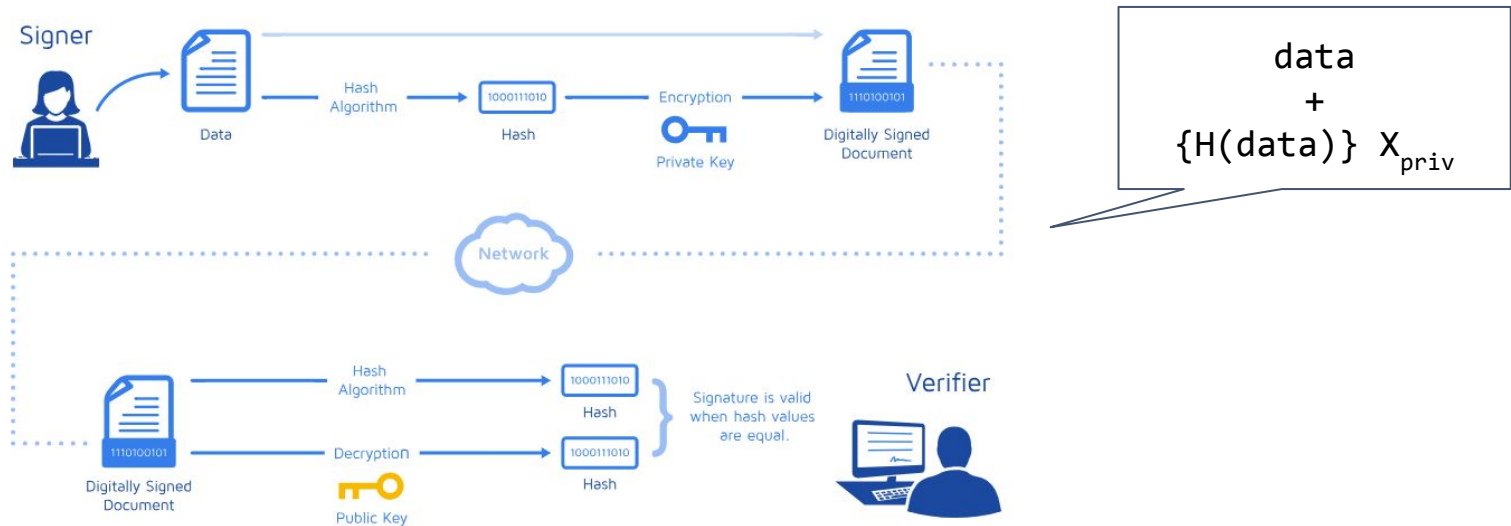


Public Key  
Crypto



# Public key cryptography: Authentication

- Digital signature on a message gives a recipient confidence that the message came from a sender known to the recipient.



# Activity: Try public key cryptography

- Encrypt a message using my public key and post it on slack **#lecture-security**, tagging the bot @pgp\_checker
- To encrypt the message, you can use the online PGP tool 8gwifi.org
- The message that you encrypt should start with your andrew id
  - The bot will reply “Hello <andrew id>!” if you did it correctly

Notes:

My public key looks like this:

```
-----BEGIN PGP PUBLIC KEY BLOCK-----  
Version: OpenPGP.js v.1.20130420  
Comment: http://openpgpjs.org  
...  
-----END PGP PUBLIC KEY BLOCK-----
```

Make sure you copy and paste the full PGP block:



```
-----BEGIN PGP MESSAGE-----  
Version: OpenPGP.js v.1.20130420  
Comment: http://openpgpjs.org
```

```
wUwDPgIyJu9LonkBAf9SpEoknt7ryM9kobfXB/8fduSAHx2C6b5Fdes1+jx  
wn4iRkganSC6c7DNktZ+hSRp8JLRI6u483DkpXU0Fky001EBw17vF9i+/cLj  
U+1z+QpvUJp/FBKIFGmKQ+mMSvDSWU+0wd+DcKORHNJPZIxjUIzTTGRKqXu  
0JQH7VLPYTFEQIgTueqBxDJUD+uQ0Gex5E=  
=duirj  
-----END PGP MESSAGE-----
```

See the example on slack

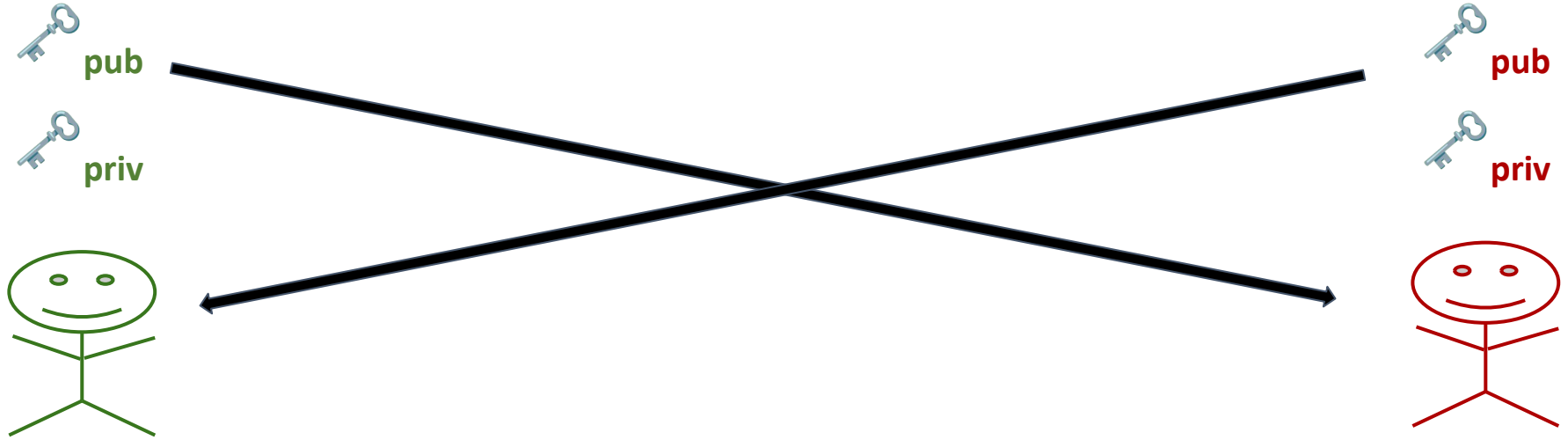
## Some considerations

- Who can read the messages?
- Do we have confidentiality?
- How can I respond to your message if needed?
- How can we confirm the sender's identity?

## Activity: Try public key cryptography (Part 2)

- Go to slack and find an encrypted message I just posted on Slack
  - I used my private key to encrypt it
- Got to [webencrypt.org/openpgpjs/](https://webencrypt.org/openpgpjs/)
- Decrypt the message using my public key

# Authentication



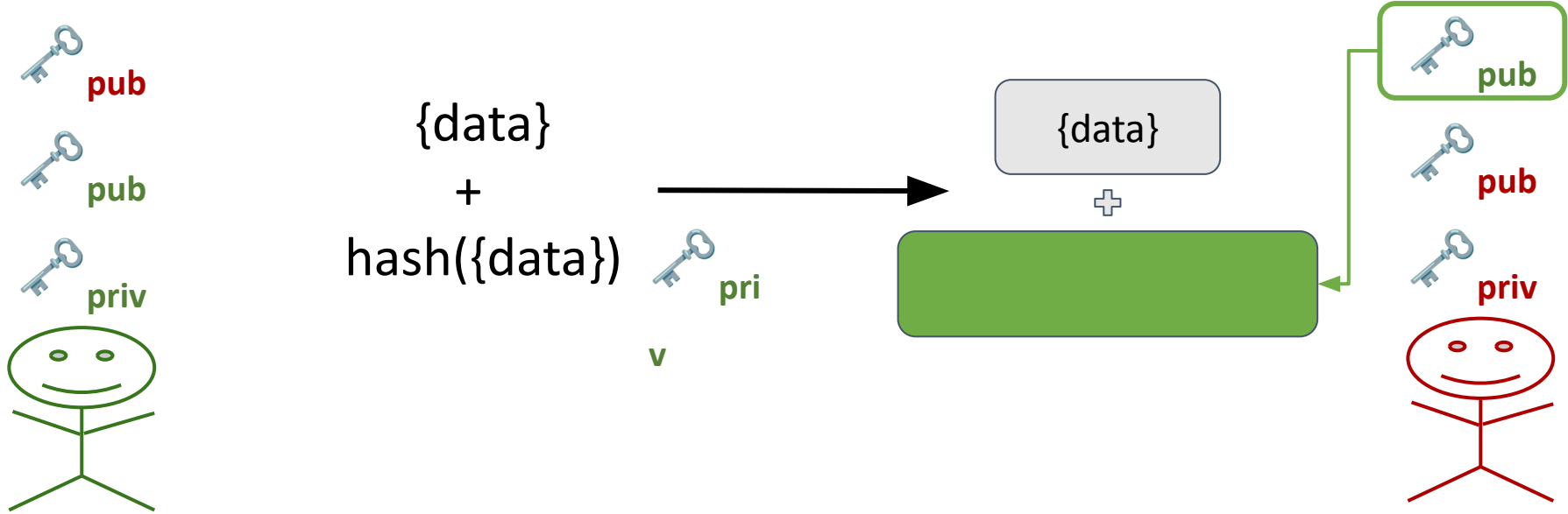
# Authentication



{data}  
+  
hash({data})

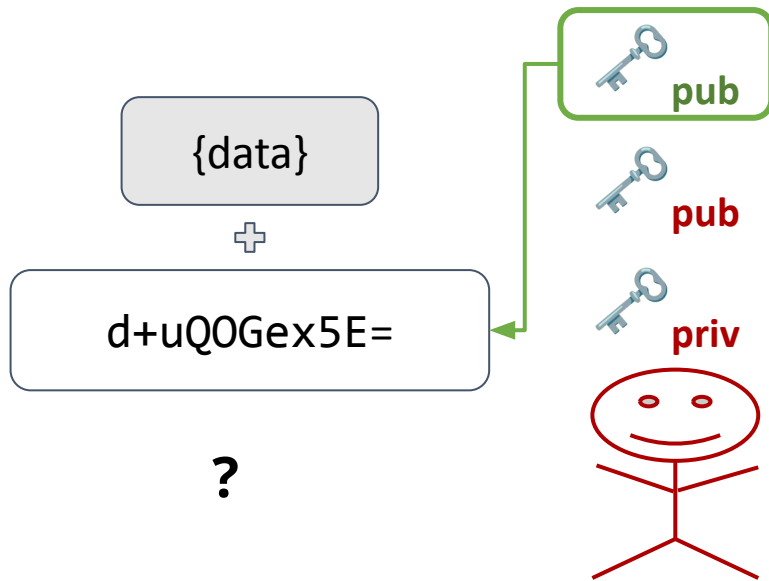
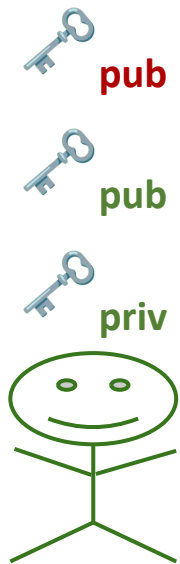


# Authentication

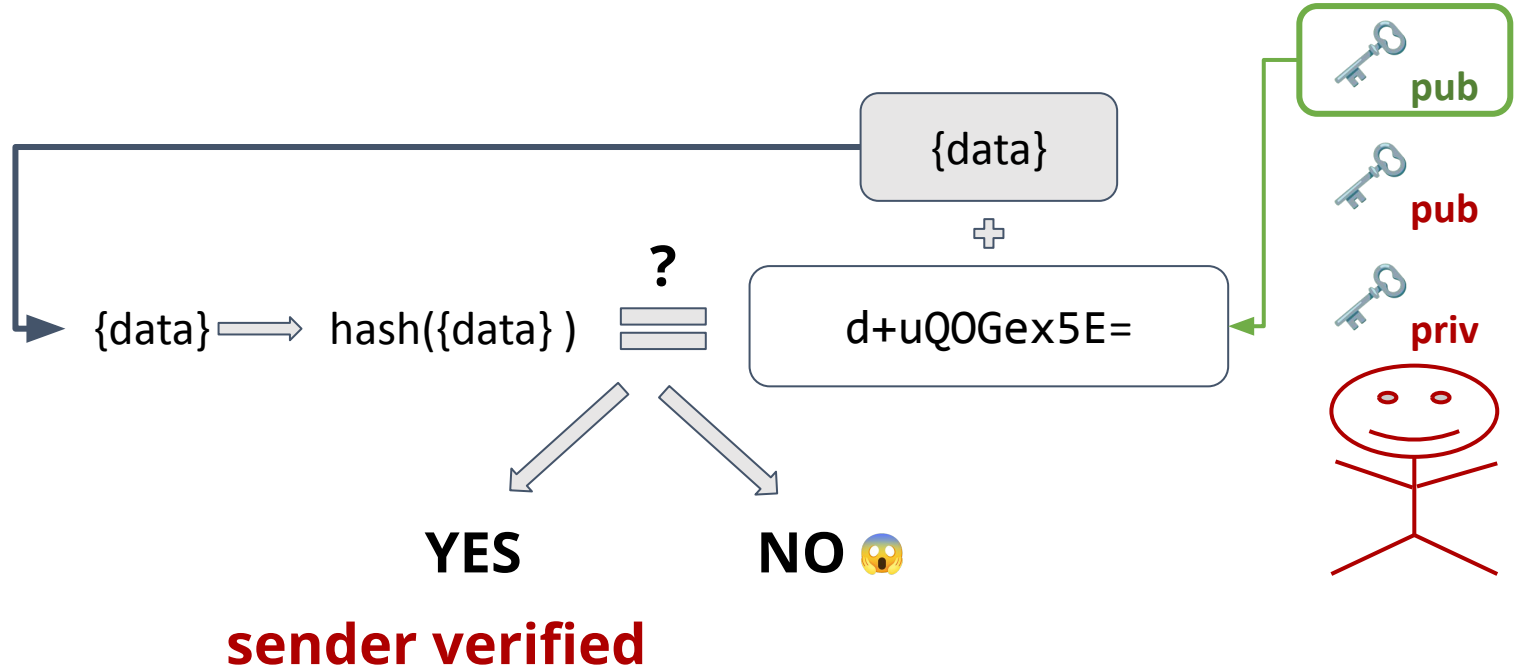




# Authentication



# Authentication



# Comparison

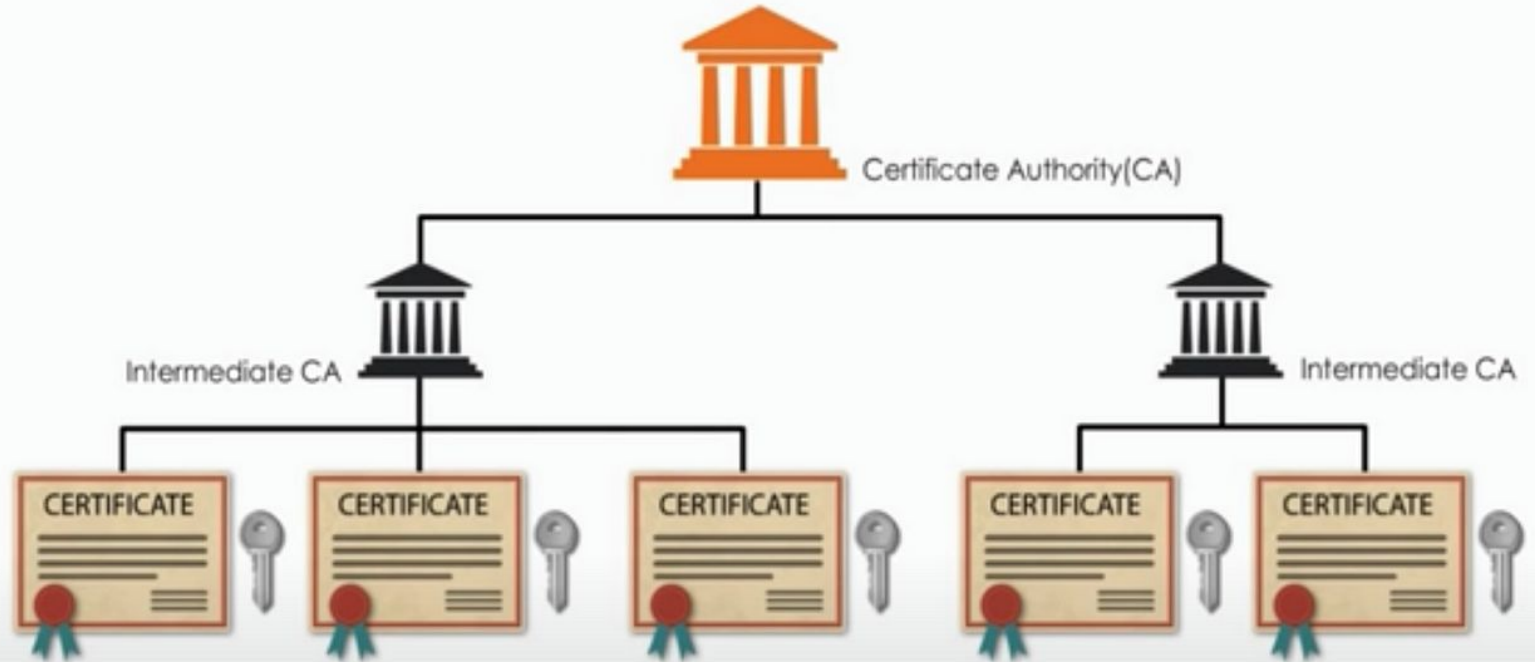
- Public & secret key crypto are (both) very secure
  - Unless the keys are compromised
- Public key crypto is computationally expensive
  - Secret key crypto is relatively fast
- It's hard to distribute the secret key between communicating parties
  - This is why we like public key cryptography
  - We just use public key to distribute secret keys which are then used

**How to distribute the public keys?**

# Certificate Authority

- A Certificate Authority (CA) confirms an entity's public key
  - Usually this will be a server's public key
- Companies get paid to do this
  - They “check out” the requestor
    - Now-a-days domain registrars provide this service
  - They issue a “certificate” with the information
  - Certificates are signed with the **CA's private key**
- **CA's public key** is “well-known”
  - It's in an additional certificate
  - Pre-installed or added to your configuration

# Certificate Chain of Trust



# Certificate Chain of Trust

What happens if the private key is compromised?

What happens if the private key is compromised?



# Generating keys for a Root Certificate Authority





# SSL: Combining two ciphers

- The expensive public-key cipher
  - Consists of two keys: one public, one private
  - These are each typically 1024-bit or 2048-bit keys
  - But has great key distribution properties
- The inexpensive symmetric cipher
  - These are typically 128-bit or 256-bit keys
  - Need to distributed the symmetric key
    - SSL uses public-key encryption to distribute the symmetric key

# What does SSL Give You?

- SSL can be used for any TCP/IP communication
- Once you have SSL
  - You have confidentiality
  - You have server authentication
- User authentication can be done using
  - Your own userids and passwords

# User authentication using passwords?



# Other factors that can lead to security breaches

- Injection Flaws, XML External Entities (XXE) Attacks
- Deployment misconfigurations
  - Default accounts with unchanged passwords
  - Unnecessary features enabled
  - Improperly configured permissions on cloud services
- Cross-Site Scripting (XSS)
  - Allows attackers to execute scripts in the victim's browser
- **Using dependencies with known vulnerabilities (CVE)**
- **Insufficient Logging and Monitoring**

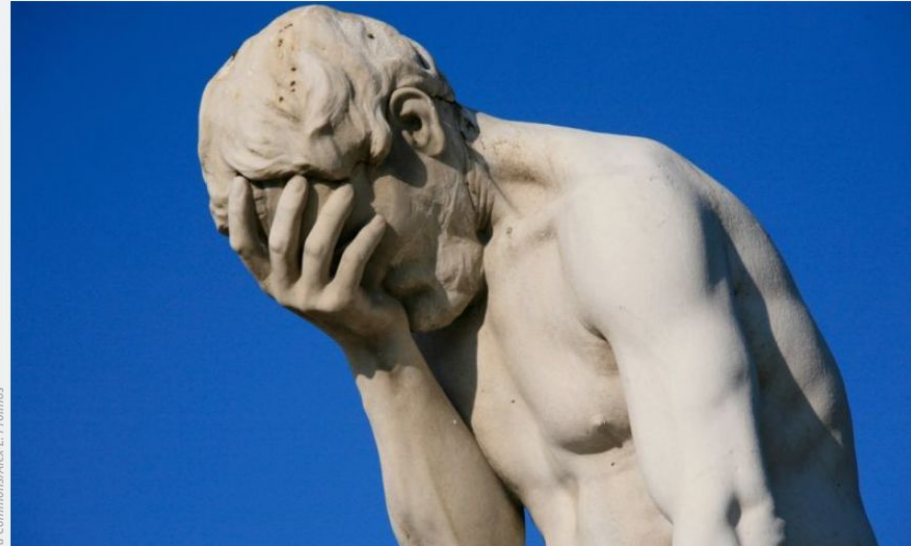
# Common Vulnerabilities and Exposures (CVE)

BIZ & IT —

## Failure to patch two-month-old bug led to massive Equifax breach

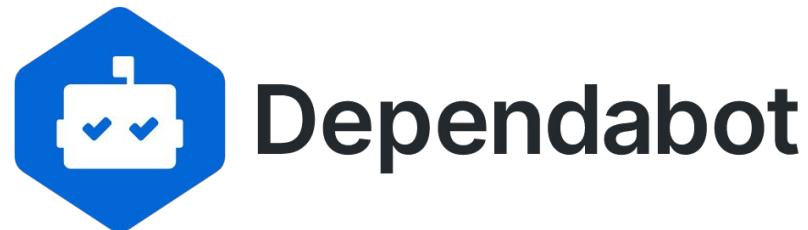
Critical Apache Struts bug was fixed in March. In May, it bit ~143 million US consumers.

DAN GOODIN - 9/13/2017, 11:12 PM



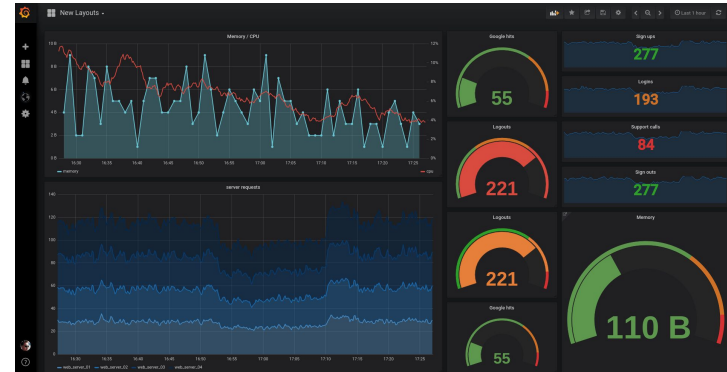
to Commons/Alex E. Proimos

# Common Vulnerabilities and Exposures (CVE)

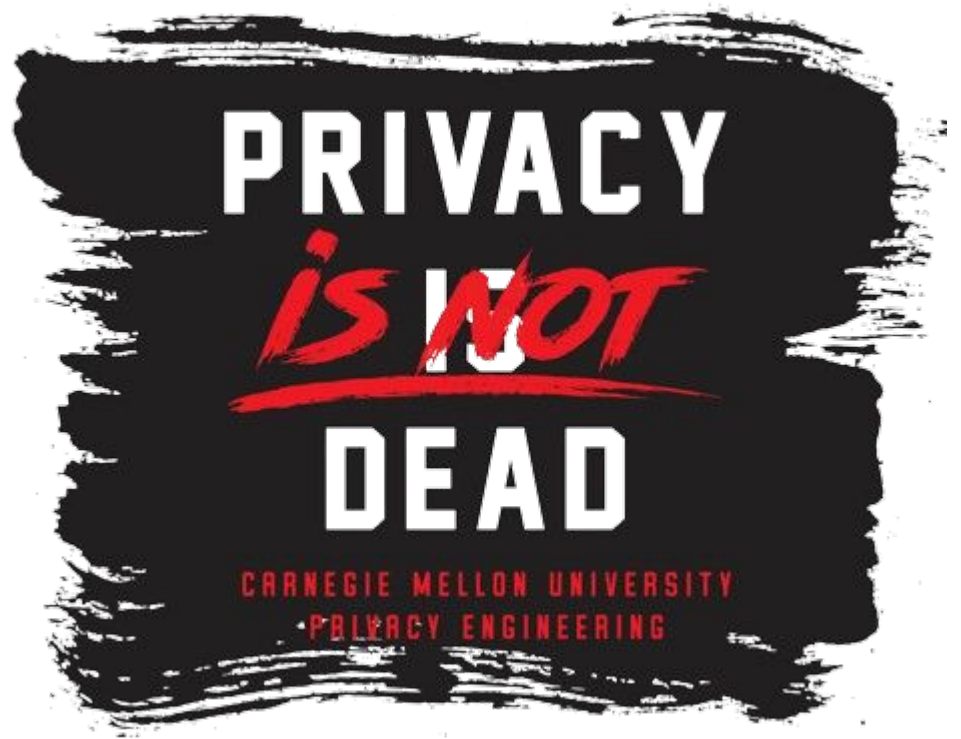


# Logging and Monitoring

- Early Detection of Security Incidents
- Forensic Analysis and Compliance
- Proactive Threat Hunting
- Audit Trails for Accountability



# Privacy





## Imagine ...

- You are about to purchase a car insurance policy
- The insurance companies you request quotes from want to know more about you ...

# How comfortable you are disclosing

- How many miles/year you drive
- How fast you drive
- Where you go and when
  - GPS
- How many hours you sleep at night
  - Based on information collected by your smartwatch
- Health history

# Observations

- Not everyone feels the same way about these issues
- Most people have concerns about at least a subset of these scenarios
  - We all care about privacy
- Today all this information is readily available and can be collected by mobile & IoT devices

# Concept of Privacy

- Moral right of individuals to **be left alone**, free from surveillance or interference from other individuals or organizations, including state



*"Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, ... Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threatened to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops."*

**Warren and Brandeis, 1890**

# Information Privacy

- The claim that certain information should not be collected by government or businesses – or possibly under special circumstances (and special rules)
- There are **conflicting considerations**
  - Security and safety
  - Personalization
  - Productivity
- There are **ethical considerations**
  - Under what situations should **government and businesses** be allowed to intrude in the lives of **citizens/consumers**, after all?

# Is Privacy Dead?

## Facebook's Zuckerberg Says The Age of Privacy Is Over

By MARSHALL KIRKPATRICK of  **ReadWriteWeb**  
Published: January 10, 2010

 PRINT

Facebook founder Mark Zuckerberg told a live audience yesterday that if he were to create Facebook again today, user information would by default be public, not private as it was for years until the company changed dramatically in December.

<https://archive.nytimes.com/www.nytimes.com/external/readwriteweb/2010/01/10/10readwriteweb-facebooks-zuckerberg-says-the-age-of-privac-82963.html>

# Is Privacy Dead?

**Forbes**

LEADERSHIP

## Privacy Is Completely And Utterly Dead, And We Killed It

**Jacob Morgan** Contributor ©  
*I write about and explore the future of work!*

Follow

Aug 19, 2014, 12:04am EDT

Privacy...everyone keeps talking about it and apparently everyone is concerned with it, but going forward does it even matter? I recently watched the documentary, “Terms and Conditions may Apply,” which provides a fascinating look at how organizations such as [Facebook](#), [Google](#), [Apple](#), and others have changed the way they look at and approach privacy. After watching the movie it had me wondering, “does privacy even matter anymore?”

<https://www.forbes.com/sites/jacobmorgan/2014/08/19/privacy-is-completely-and-utterly-dead-and-we-killed-it/>

# Is Privacy Dead?

*"You have zero privacy anyway. Get over it."*

Scott McNealy, Former CEO of Sun Microsystems (1999)

<https://www.wired.com/1999/01/sun-on-privacy-get-over-it/>



# Is Privacy Dead?

*“If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place, but if you really need that kind of privacy, the reality is that search engines including Google do retain this information for some time... that information could be made available to the authorities.”*

Eric Schmidt, Former CEO of Google (2009)

[https://www.pcworld.com/article/515472/googles\\_schmidt\\_roasted\\_for\\_privacy\\_comments.html](https://www.pcworld.com/article/515472/googles_schmidt_roasted_for_privacy_comments.html)

# How Privacy is Protected

Laws, self-regulation, technology

- Notice and access
- Control over collection, use, deletion, sharing
- Collection limitation
- Use limitation
- Security and accountability

# US FTC's Fair Information Practice Principles

1. Notice/awareness (core principle)
  - a. Disclose practices
2. Choice/consent (core principle)
  - a. Opt-in, opt-out
3. Access/participation
  - a. Users should be able to review & correct their information
4. Integrity/Security
  - a. Ensure is secure, limited access
5. Enforcement
  - a. Mechanisms for handling violations

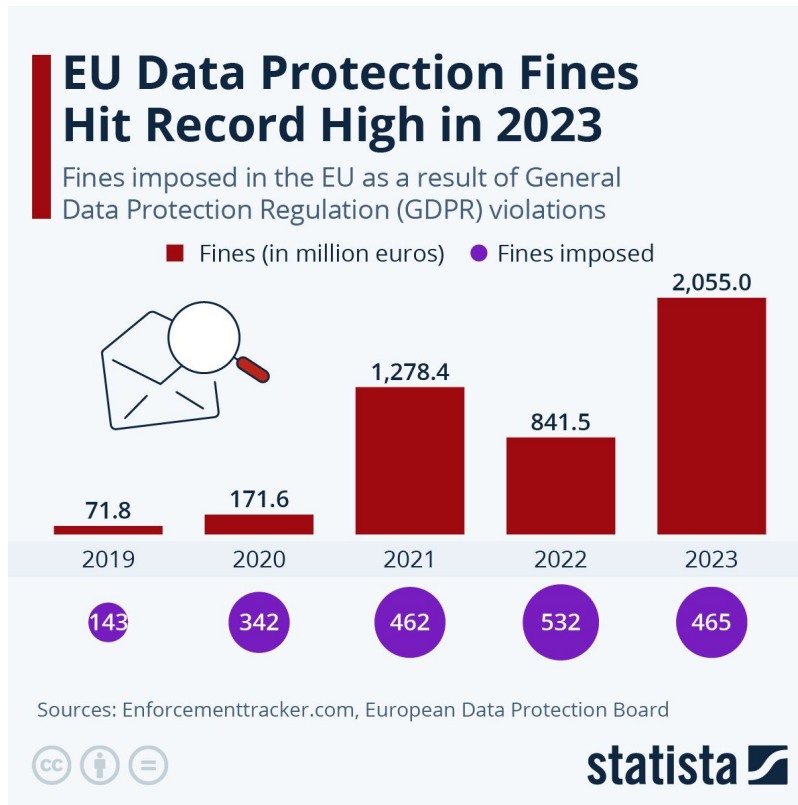
# OECD Fair Information Principles

- Collection limitation
- Data quality
- Purpose specification
- Use limitation
- Security safeguards
- Openness
- Individual participation
- Accountability

# Why SWEs should care about privacy

- Ethical questions (recall **Ethics** lecture)
- Laws restricting data collection by government and agencies
  - Freedom of Information Act (1966), Privacy Act (1974), Electronics and Communications Act (1986), ...
- Laws restricting data collection in different economic sectors
  - COPPA, HIPAA, FERPA, etc
- State Laws (e.g., CCPA) and local laws
- EU - General Data Protection Regulation

# Why SWEs should care about privacy



# Goals of Privacy Engineering

- Ensuring legal compliance
  - GDPR, CCPA, etc.
- Aligning with consumer expectations
  - Transparency about data practices
  - Accurate statements about privacy policies
- Building trust and goodwill
  - Commitment to protecting user data
- Competing on privacy protection
  - Privacy as core value
- Promoting privacy as a societal value
  - Safeguarding privacy rights and advocating for ethical data practices

# Mechanisms in Privacy Engineering

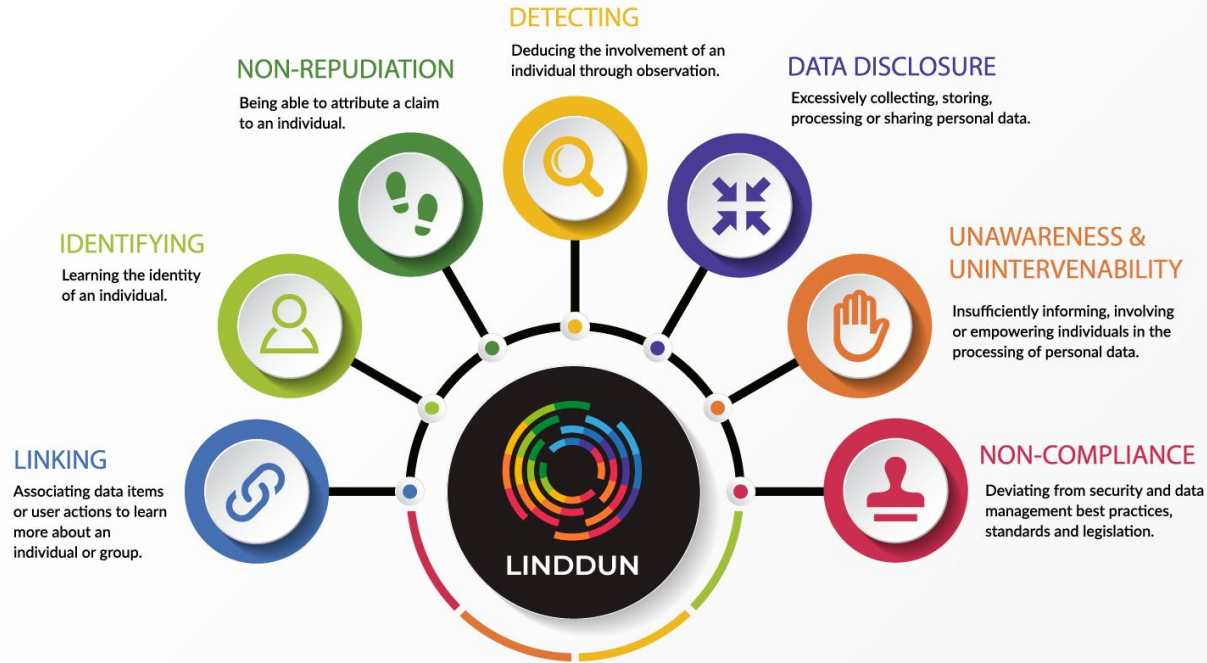
- Selective data collection
  - Purpose-driven, minimize amount of personal information
- Data minimization
  - De-identification, pseudonymization, anonymization
  - Remove sensitive data
- Data retention policies
- Cryptographic tools
  - Confidentiality
- Access controls and secure data storage
- Socio-technical processes and audits
- Privacy-by-design (PbD)
- **Threat modeling**



# Threat Modeling

- Applicable to both security and privacy
- A wide variety of possible security and privacy threats.
  - How can we organize our analysis?
- Basic idea: systematic methodology to identify possible threats and methods to mitigate these threats
- Approach: use a taxonomy of possible threats

# LINDDUN Taxonomy of Privacy Threats



## Group activity: Privacy Threat Identification

- Consider a university admissions system that manages the application process for prospective students, including collecting application materials, evaluating candidates, and making admissions decisions.
- In groups of 2-3, identify **two** privacy threats and propose ways to mitigate them

## Faculty Course Evaluation (FCE)



## TA Evaluation

