CODATA – RDA

Data Schools

Information Security

Raphael Cobe raphaelmcobe@gmail.com

Why Security?

Data Security Concepts

Security Objectives

Guidelines and Principles

Introduction to Encryption

Thanks

- These slide are a (small) variation of the presentation by Hannah Short

About me

- Member of the Advanced Institute for Artificial Intelligence
- Member of the Sao Paulo Research and Analysis Center
- Experience with High Performance Computing and Artificial Intelligence
- Very amateur Climber, Cyclist, Runner, Swimmer, etc; (-:

Course Objectives

- Understand why Security is important for you as a Data Scientist
- Familiarise yourself with the basic principles of Information Security

Note:
If the slide title is in red, the slide is considered an advanced topic

Data Schools

Section 1

Why Security?

Why Security?

- You are constantly exposed to reputational, financial and even physical risks online
- The aim is to **minimise your exposure to risk** through
  - Secure online activity
  - Secure software design

Safety vs Security

Safety is about protecting from **accidental risks**

- road safety
- air travel safety

Security is about mitigating risks of dangers caused by **intentional, malicious actions**

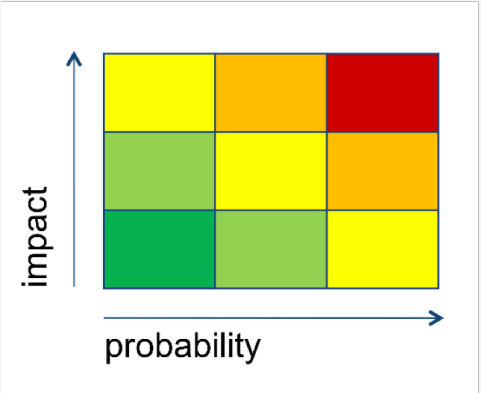
- homeland security
- airport and aircraft security
- information and computer security

Why is security difficult?

Security is as strong as the weakest link. There is no 100% security!

What is risk?

- Probability * impact
- Risks should be: Assessed, Prioritised, Mitigated, Avoided and finally Accepted



Typical Threats

But we're Scientists, surely we're not a target...!

Typical Threats

BBC Sign in | News | Sport | Weather | Capital | TV | Radio

NEWS — Watch ONE-MINUTE WORLD NEWS

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science & Environment
Technology
Entertainment
Also in the news

Video and Audio

Page last updated at 11:24 GMT, Monday, 15 September 2008 12:24 UK

E-mail this to a friend Printable version

'Big bang' experiment is hacked

Part of the computer system of the Large Hadron Collider (LHC) was hacked into as the world's most powerful physics experiment got under way.

The CMS detector was not affected by the computer hackers

A group calling itself the "Greek Security Team" hacked into a computer connected to the system last Wednesday.

A spokesman for Cern, the lab that houses the LHC, said the hackers put up a message on the facility's website.

No harm was done but the incident has highlighted the need for security in the LHC's network, the spokesman said.

http://news.bbc.co.uk/2/hi/technology/7616622.stm

Typical Threats

https://www.reuters.com/business/healthcare-pharmaceuticals/health-rounds-dna-researchers-are-warned-beware-hackers-2025-04-18

Why Security - Summary

- Security = mitigating risk of malicious actions
- Science is an interesting target for bad guys/girls

Section 2

Data Security Concepts

Data Security Concepts

At the heart of Security we have three key components:

- Technology
- Processes
- People

Technology

We will come back to some of this in part 2 of our lecture :-)

Processes

"Security is a process, not a product" - Bruce Schneier

Processes

Processes

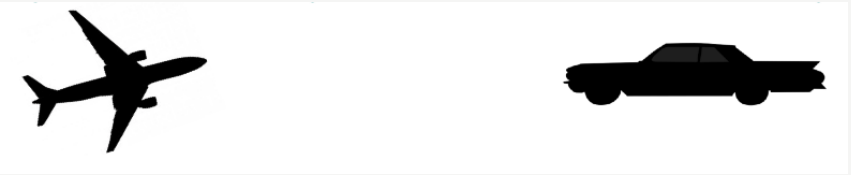
Security solutions often degrade with time - they need to be verified periodically!

People

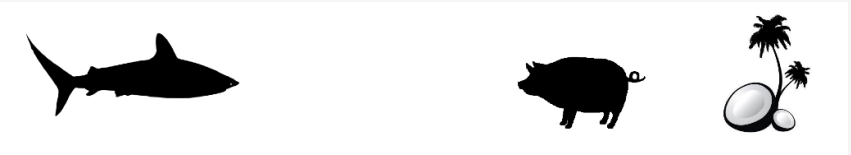
- Have flawed risk perception
- Are bad at dealing with exceptions and rare cases
- Put too much trust in their computers
- Easily fall for social engineering
- Sometimes turn malicious
- Prefer convenience and bypass security measures
- Often make mistakes...

Risk Perception

Is flying more dangerous than traveling by car?



Are you more likely to be killed by a shark, a pig or a coconut?

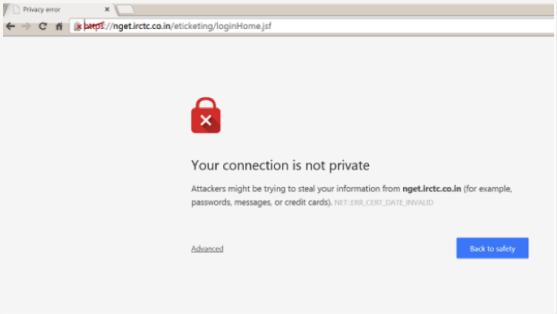


Social Engineering

https://www.smbc-comics.com

Taking security decisions

Users typically make poor security choices despite systems trying to protect them!



And sometimes it's just plain difficult

Which links point to eBay?

- secure-ebay.com
- www.ebay.com\cgi-bin\login?ds=1%204324@%31%32%34.%31%33%36%2e%31%30%2e%32%30%33/p?uh3f223d
- www.ebay.com/ws/eBayISAPI.dll?SignIn
- scgi.ebay.com/ws/eBayISAPI.dll?RegisterEnterInfo&siteid=0&co_partnerid=2&usage=0&ru=http%3A%2F%2Fwww.ebay.com&rafId=0&encRafId=default

...

Data Security Concepts - Summary

- Processes must be ongoing, security degrades with time
- People often provide the easiest way for an attacker to compromise the system
- Security is only as strong as the weakest link - don't lock the front door but leave the back door open!

Data Schools

Section 3

Security Objectives

Security Objectives

Computer Security aims to meet these objectives:

- Confidentiality
- Integrity
- Availability

We will start with a quick look at Identity, as this is essential for meeting security objectives!

Identity

Online Identity is really no different from your real life Identity! Your Identity is the answer to the question: **“who are you?"**

- It could be a username for a website
- It could be a government ID
- It could be a digital certificate

Authentication and Authorisation

Authentication = How can I prove my Identity? Authorisation = What am I able to do?

Multifactor Authentication

Factor	Description	Example
1	Something you know	Password, pin
2	Something you have	Phone, Yubikey
3	Something you are	Fingerprint, iris scan

Which is most secure?

Security Objectives

- **Confidentiality**
- Integrity
- Availability

Can the correct people access the data at the correct time?

Security Tip: Pay attention to where your data is stored and how it is shared!

Confidentiality

- Your online identity is as valuable as your passport
- Your authorisation may be misused if it falls into the wrong hands

Security Tip: Store your secrets safely, not in the public domain, e.g. github

The Register®
Biting the hand that feeds IT

Security

Dev put AWS keys on Github. Then BAD THINGS happened

Fertile fields for Bitcoin yields - with a nasty financial sting

By Darren Pauli 6 Jan 2015 at 13:02 25 SHARE ▼

Bots are crawling all over GitHub seeking secret keys, a developer served with a $2,375 Bitcoin mining bill found.

DevFactor founder Andrew Hoffman said he used Figaro to secure Rails apps which published his Amazon S3 keys to his GitHub account.

He noticed the blunder and pulled the keys within five minutes, but that was enough for a bot to pounce and spin up instances for Bitcoin mining.

"When I woke up the next morning, I had four emails and a missed phone call from Amazon AWS - something about 140 servers running on my AWS account," Hoffman said.

"I only had S3 keys on my GitHub and they were gone within five minutes!"

"As it turns out, through the S3 API you can actually spin up EC2 instances, and my key had been spotted by a bot that continually searches GitHub for API keys."

Most read


Fork it! Google fined €4.34bn over Android, has 90 days to behave


Official: The shape of the smartphone is changing forever


Trump wants to work with Russia on infosec. Security experts: lol no


Boss helped sysadmin take down horrible client with swift kick to the nether regions


British Airways' latest Total Inability To Support Upwardness of Planes* caused by Amadeus system outage

How bad can it be?

- 5 minutes exposure
- $2,375
- Plus it could have been avoided, Amazon has a service (IAM) to manage keys securely...

`https://www.theregister.co.uk/2015/01/06/dev_blunder_shows_github_crawling_with_keyslurping_bots/`

Security Objectives

- Confidentiality
- **Integrity**
- Availability

Can we be sure that the data is reliable and hasn't been altered?

Security Tip: Reduce the risk of impersonation, enable multi-factor authentication wherever possible!

Security Objectives

- Confidentiality
- Integrity
- **Availability**

Is the data available? Are our systems reliable?

Security Tip: Keep backups!

Security Objectives - Summary

- Key objectives: Confidentiality, Integrity and Availability
- Consider disaster scenarios and plan for them
- Online Identity is critical to meeting security objectives

Data Schools

Section 4

Guidelines and Principles

Security Measures

Is this a good security measure?

Security Measures

- What problem is it trying to solve?
- Does it help?
- Does it introduce new problems?
- What are the costs?

Security Measures
How much security?

It's a balance of risk, usability and cost

Security Design Principles

- Defense in depth
  - How can you avoid a single point of failure? Where should you keep your assets?
- Deny by default
  - Use Allowlist rather than Denylist
- Least privilege principle
  - Least privilege principle: "Need to know" basis: require, grant and use only the privileges that are really needed

Security Design Principles

- Complex = insecure
  - Maintenance of complex code leads to vulnerabilities
- Security, not obscurity
  - Hiding design or implementation details to gain security
  - Systems should be secure by design, not by obfuscation!

Guidelines and Principles - Summary

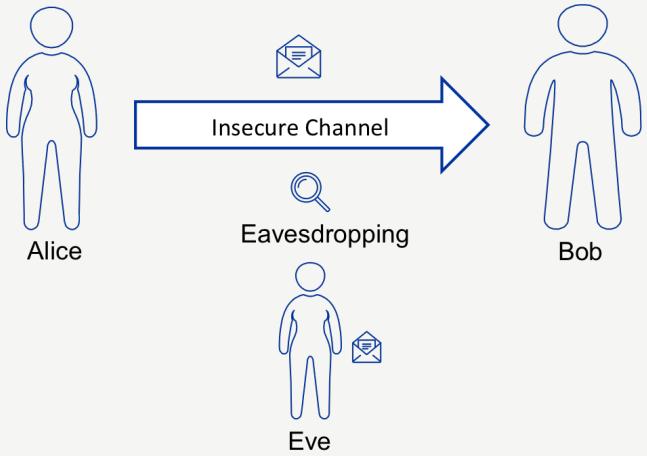
- Security is a balance of risk, usability and cost
- The Security Design Principles discussed will help you prioritise security
- Ensure Security Design Principles are included from the very beginning of a software project

Data Schools

Section 5

Introduction to Encryption

Why Encryption?



Why Encryption?

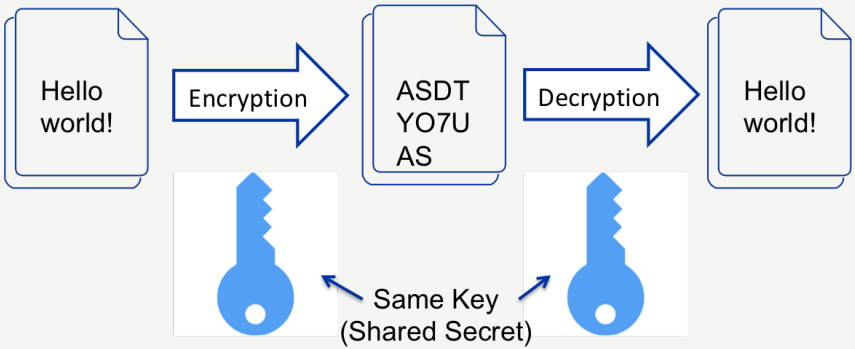
What are the goals?

- **Confidentiality**: To prevent adversaries from viewing/accessing messages
- **Integrity**: To prevent adversaries from silently modifying messages
- **Authentication**: To prevent adversaries from impersonating an identity
- **Non-repudiation**: Ensures that someone cannot deny (i.e., "repudiate") having performed an action, such as:

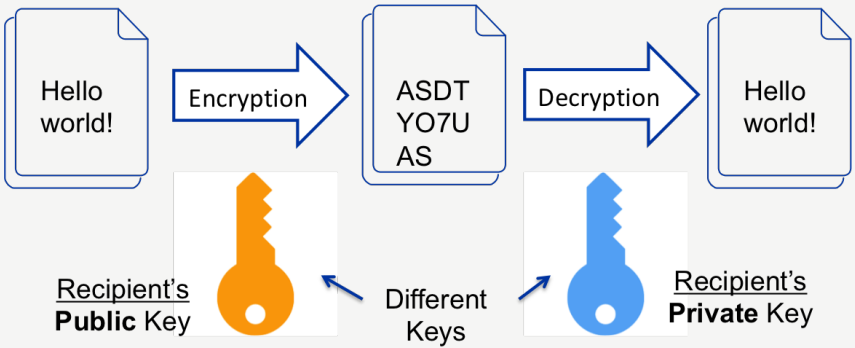
Encryption in practice

- There are several common, robust encryption algorithms available (e.g. AES and RSA)
- A good encryption algorithm relies on keeping the key secret, not the cipher algorithm itself!
- Choose a well known, secure algorithm and keep the key secure (do not trust proprietary algorithms)
- There are two main types; **Symmetric and Asymmetric**

Symmetric Encryption



Asymmetric Encryption



Asymmetric Encryption

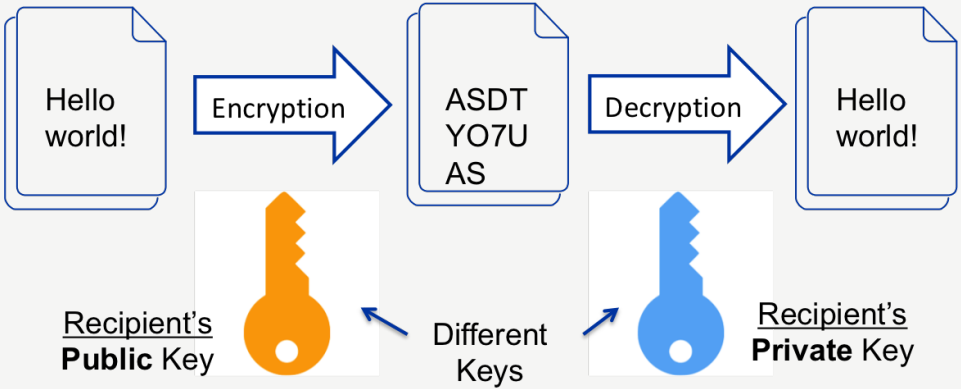
- Relies on the fact that it is **easy** to multiply primes but hard to factorise their product
  - If you take two large primes (e.g., 100+ digits) and multiply them, a computer can do that in milliseconds
  - But if you're only given the product (say, a 200+ digit number), it can take years to figure out what the original primes were
  - What would be the primes for the product 3233?

- Security is dependent on the status of computing technologies (it's secure now, but won't be in 100 years...)

Asymmetric Encryption

- Relies on the fact that it is **easy** to multiply primes but hard to factorise their product
  - If you take two large primes (e.g., 100+ digits) and multiply them, a computer can do that in milliseconds
  - But if you're only given the product (say, a 200+ digit number), it can take years to figure out what the original primes were
  - What would be the primes for the product 3233?
  - Product $= 61 \times 53 = 3233$
- Security is dependent on the status of computing technologies (it's secure now, but won't be in 100 years...)

Asymmetric Enc. - Confidentiality



Asymmetric Enc. - Integrity

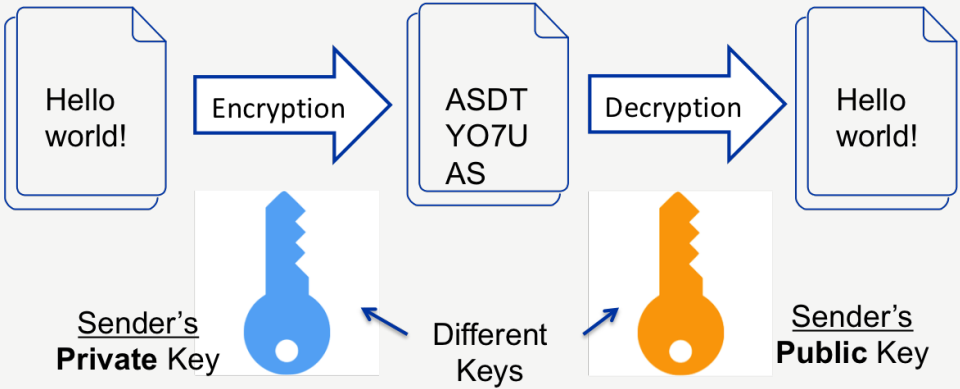
- Sender computes a **hash** of the message
- The hash is then **signed** using their private key
- Receiver uses the sender's **public key** to verify the signature
- Receiver also recomputes the hash of the message and compares

Why this works:

- If the message is altered, the hashes will not match
- The signature will fail verification

This guarantees the message has not been tampered with.

Asymmetric Enc. - Authentication



Asymmetric Enc. - Non-repudiation

- To sign a message, the sender uses their **private key**
- Anyone can verify the signature using the sender's **public key**

Why this matters:

- Only the owner of the private key could have created the signature
- Therefore, the sender **cannot deny** authorship of the message

This cryptographic proof is what enforces non-repudiation.

Encryption - Summary

- Asymetric Encryption can be used for confidential, authenticated communication
- It can also be used to ensure integrity and non-repudiation

Questions?

- Ask now
- Find us during the break
- You are welcome to contact us after the school