CODATA – RDA

Data Schools

Information Security

Raphael Cobe raphaelmcobe@gmail.com

Why Security?

Data Security Concepts

Security Objectives

Guidelines and Principles

Introduction to Encryption

Hash Functions

Certificates

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**

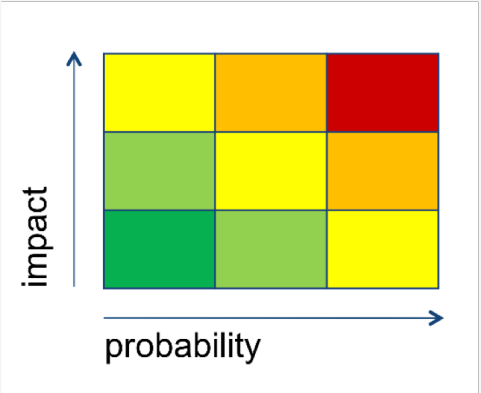
- homeland security
- airport and aircraft security
- information and computer security

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.

Typical Threats

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 course :)

Processes

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

Processes

Processes

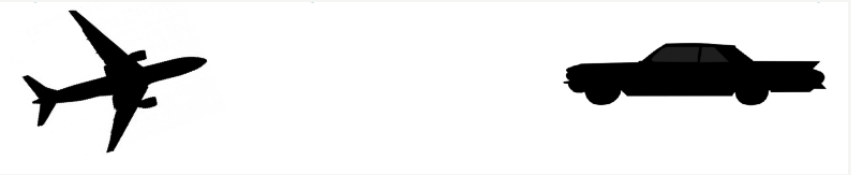
Security solutions often degrade with time - they need to be verified periodically!

People

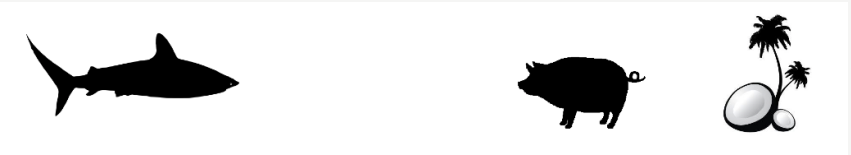
- Have flawed risk perception
- Are bad at dealing with exceptions and rare cases
- Put too much trust in their computers
- Easily fall for social engineering
- Sometimes turn malicious
- Prefer convenience and bypass security measures
- Often make mistakes...

Risk Perception

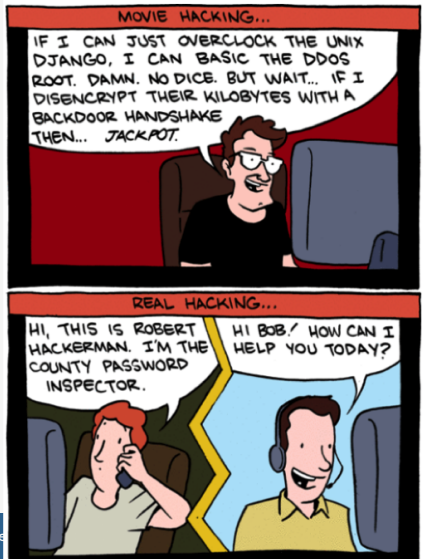
Is flying more dangerous than traveling by car?



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

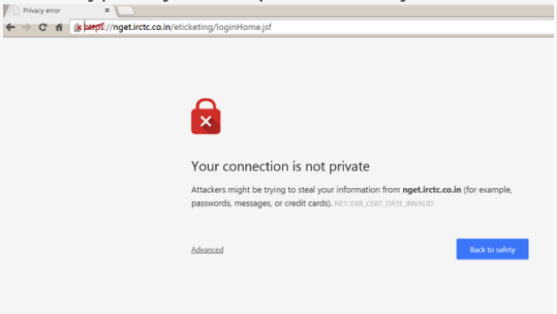


Social Engineering



Taking security decisions

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



And sometimes it's just plain difficult

Data Schools

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

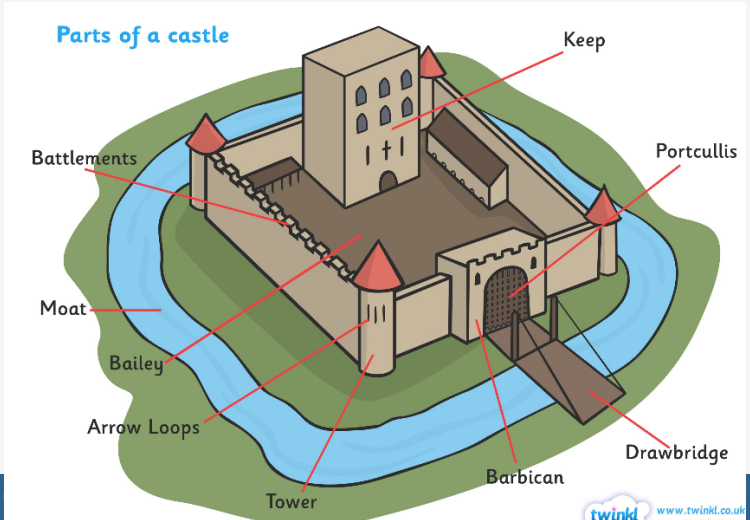
- What problem is it trying to solve?
- Does it help?
- Does it introduce new problems?
- What are the costs?

Security Design Principles

- Defense in depth
- Deny by default
- Least privilege principle
- Complex = insecure
- Security, not obscurity

Defense in depth

How can you avoid a single point of failure? Where should you keep your assets?



Deny by default

Use whitelisting rather than blacklisting

```
def isAllowed(user):
  allowed = true
  try:
    if (!inFile(user, "admins.xml")): allowed = false
  except IOError: allowed = false
  except: pass
  return allowed
```

No!

What if XMLError is thrown instead?

```
def isAllowed(user):
  allowed = false
  try:
    if (inFile(user, "admins.xml")): allowed = true
  except: pass
  return allowed
```

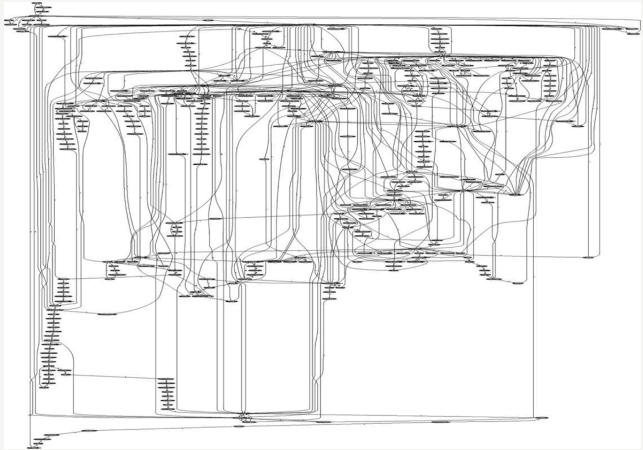
Yes

Least privilege principle

"Need to know" basis: require, grant and use only the privileges that are really needed

Complex = insecure

Maintenance of complex code leads to vulnerabilities



System calls in Apache

Security by obscurity

What is it? Hiding design or implementation details to gain security:

- e.g. hiding a DB server under a name different from “db”, etc.
- e.g. keeping the encryption algorithm secret, instead of the key

Security by obscurity

The idea doesn't work

- It's difficult to keep secrets (e.g. source code gets stolen, Google indexes hidden pages...)
- If security of a system depends on a secret that's revealed, the whole system is compromised
- Secret algorithms, protocols etc. will not get reviewed, flaws won't be spotted and fixed, less security

Systems should be secure by design, not by obfuscation!

Guidelines and Principles - Summary

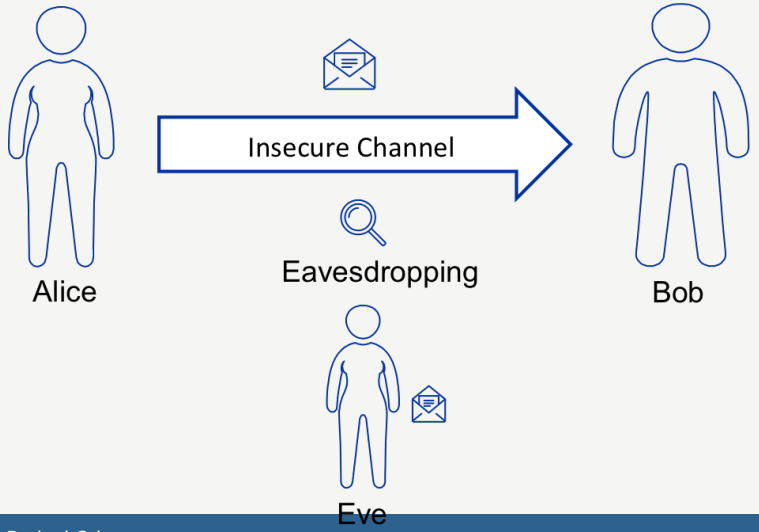
- Security is a balance of risk, usability and cost
- The Security Design Principles discussed will help you prioritise security
- Ensure Security Design Principles are included from the very beginning of a software project

Data Schools

Section 5

Introduction to Encryption

Why Encryption?



Why Encryption?

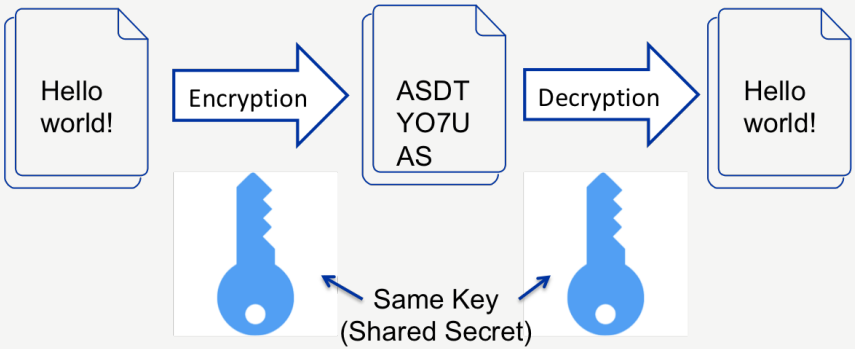
What are the goals?

- **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**: To prevent adversaries from denying an action

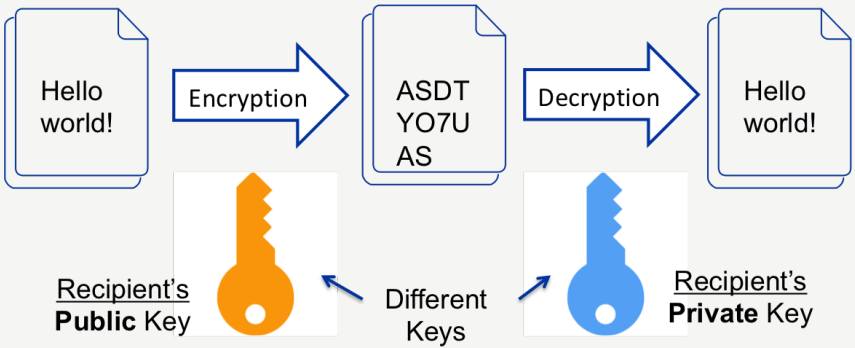
Encryption in practice

- 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**

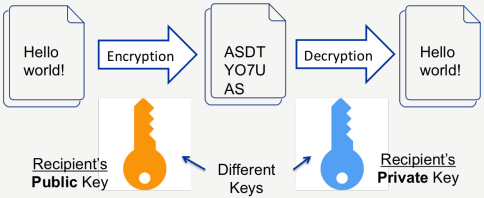
Symmetric Encryption



Asymmetric Encryption



Asymmetric Encryption



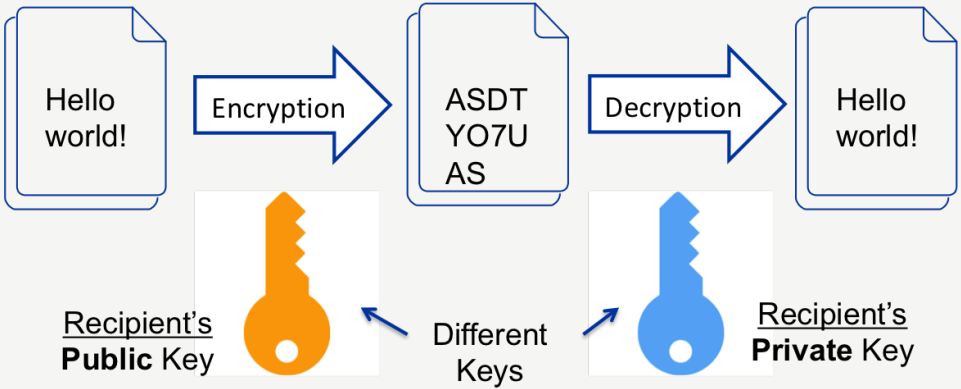
2 interchangeable (linked) keys

- Public + Private
- Mathematically difficult to compute one from the other
- 1 for encryption and the other decryption

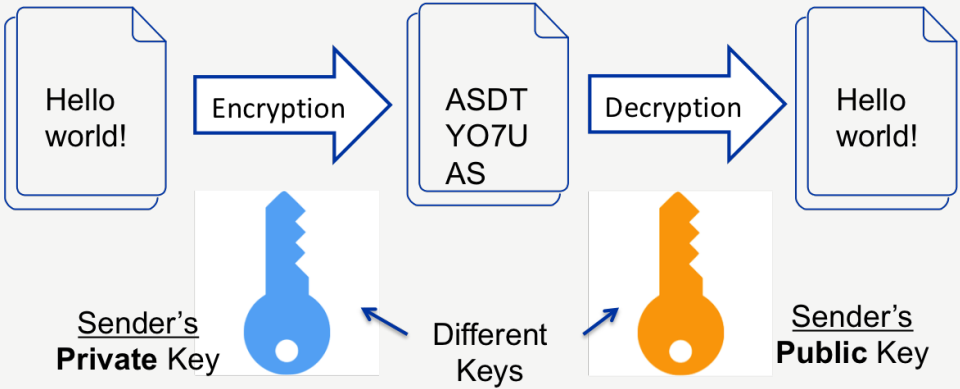
Asymmetric Encryption

- Relies on the fact that it is **easy** to multiply primes but hard to factorise their product
- Consider the number 221... what are its factors?
- 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. - Authentication



Encryption - Summary

- Encryption can be used for confidential, authenticated communication
- Symmetric and Asymmetric Ciphers have evolved over time

Data Schools

Section 6

Hash Functions

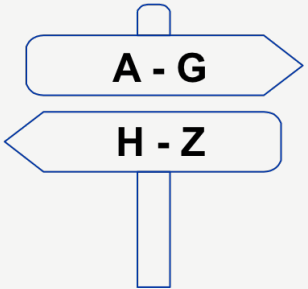
Hash Functions

A hash function is any function that can be used to map data of arbitrary size to data of a fixed size.

Hash Functions

How can we make something of a fixed length?

- We want the input (however long) to turn into one of a smaller range of possible outputs
- Consider a registration process where attendees pick up their badges based on the first letter of surname...



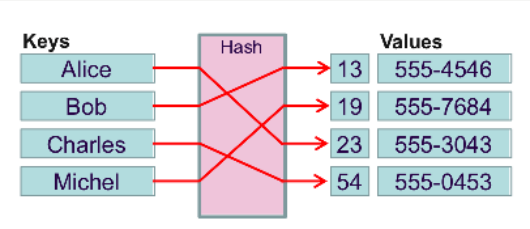
Hash Functions

E.g. *"Today is gonna be the day that they're gonna throw it back to you. By now you should've somehow realized what you gotta do. I don't believe that anybody feels the way I do, about you now"* becomes **0283bf5eb0c60213a99f011a89300179** using the MD5 hashing algorithm
`https://passwordsgenerator.net/md5-hash-generator/`

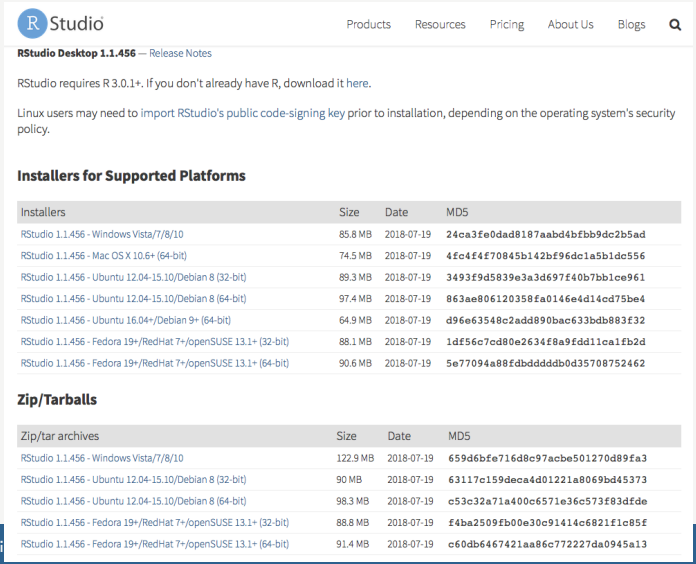
Hash Functions

What is a *good* hash function? For $h = hash(m)$

- Difficult to find any message m with a given hash value h
- Difficult to find 2 messages $m1$, $m2$ such that: $hash(m1) = hash(m2)$



Hash Functions - a Use Case



RStudio Products Resources Pricing About Us Blogs

RStudio Desktop 1.1.456 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it here.

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	85.8 MB	2018-07-19	24ca3fe0dad8187aabd4bfbb9dc2b5ad
RStudio 1.1.456 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-07-19	4fc4f4f70845b142bf96dc1a5b1dc556
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-07-19	3493f9d5839e3a3d697f40b7bb1ce961
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-07-19	863ae806120358fa0146e4d14cd75be4
RStudio 1.1.456 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.9 MB	2018-07-19	d96e63548c2add890bac633bdb883f32
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-07-19	1df56c7cd80e2634f8a9fdd11ca1fb2d
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-07-19	5e77094a88fdbdddddb0d35708752462

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	122.9 MB	2018-07-19	659d6bfe716d8c97acbe501270d89fa3
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-07-19	63117c159deca4d01221a8069bd45373
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-07-19	c53c32a71a400c6571e36c573f83dfde
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-07-19	f4ba2509fb00e30c91414c6821f1c85f
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-07-19	c60db6467421aa86c772227da0945a13

Hash Functions - another Use Case

- Instead of storing passwords, secure services store their hashes!
- What would happen if the database is compromised?

Hash Functions - another Use Case

- Instead of storing passwords, secure services store their hashes!
- What would happen if the database is compromised?
  - **Dictionary attacks still valid**
  - This can be avoided by "salting" passwords, the system adds digits before hashing

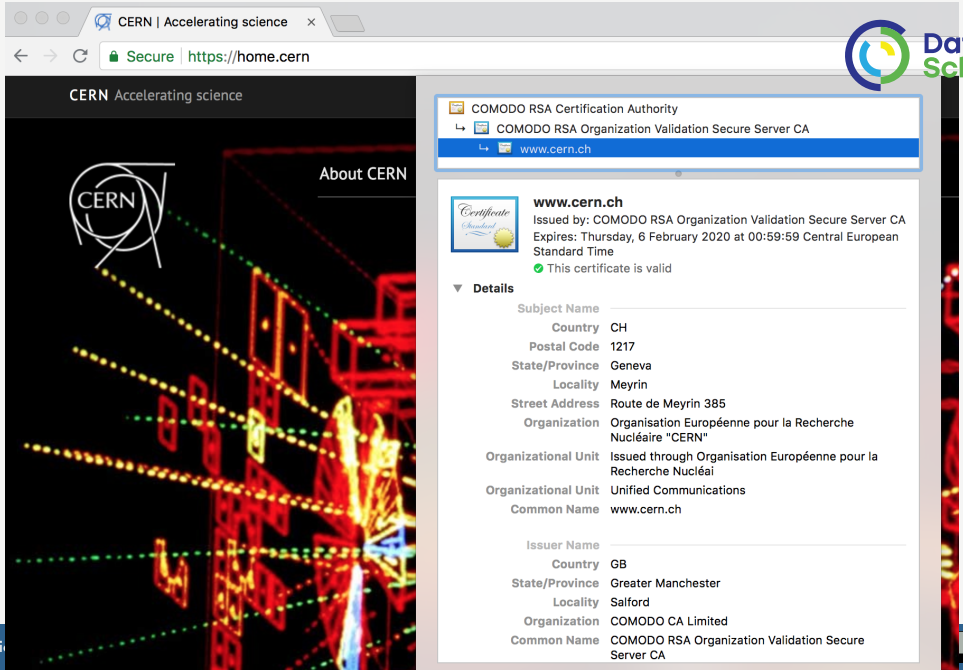
Hash Functions - Summary

- Hash functions transform arbitrary data to a fixed size in a deterministic (repeatable) way
- There are multiple applications, e.g. file integrity & password storage

Data Schools

Section 7

Certificates


CERN | Accelerating science
Secure
https://home.cern
CERN Accelerating science
About CERN
COMODO RSA Certification Authority
COMODO RSA Organization Validation Secure Server CA
www.cern.ch
www.cern.ch
Issued by: COMODO RSA Organization Validation Secure Server CA
Expires: Thursday, 6 February 2020 at 00:59:59 Central European Standard Time
This certificate is valid
Details
Subject Name
Country CH
Postal Code 1217
State/Province Geneva
Locality Meyrin
Street Address Route de Meyrin 385
Organization Organisation Européenne pour la Recherche Nucléaire "CERN"
Organizational Unit Issued through Organisation Européenne pour la Recherche Nucléai
Organizational Unit Unified Communications
Common Name www.cern.ch
Issuer Name
Country GB
State/Province Greater Manchester
Locality Salford
Organization COMODO CA Limited
Common Name COMODO RSA Organization Validation Secure Server CA

Certificates

What is a certificate? A digital document that:

- Contains identity information
- Contains a public key (this is public information)
- Is digitally “signed" by a trusted body

Certificates are accompanied by private keys (kept secret by the owner!)

Certificate Authentication

Owning a Certificate of Alice does not mean that you are Alice

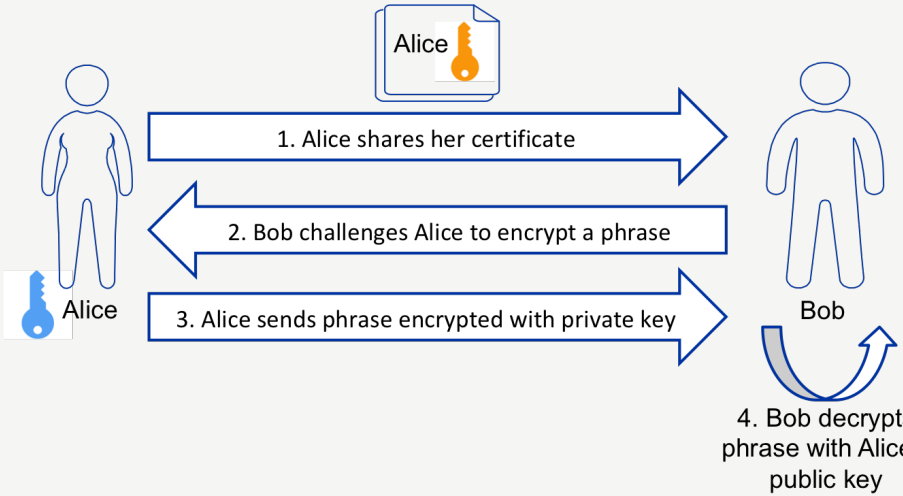
- Holding a Certificate does not imply you are authenticated
- How would you verify that the person who comes to you pretending to be Alice and showing you a certificate of Alice is really Alice ?

Certificate Authentication

Owning a Certificate of Alice does not mean that you are Alice

- Holding a Certificate does not imply you are authenticated
- How would you verify that the person who comes to you pretending to be Alice and showing you a certificate of Alice is really Alice ?
  - **You have to challenge her!**
  - Only the real Alice has the private key that goes in pair with the public key in the certificate.

Certificate Authentication



Certificates - Summary

- Contain a public key and identity information
- Certificates are validated by Certificate Authorities
- Certificates & private keys together allow asymmetric encryption and authentication

Questions?

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