Security Data Science

Introduction to Cyber Security

CIFF

Centro Internacional
de Formación Financiera

Instructor

- José Carlos Baquero is a Software Engineering Executive with over eighteen years experience in the ITC sector. He leads Big Data & Analytics initiatives for GMV in areas such as Fraud Prevention, Earth Observation, Social Media, Anomaly Detection, etc.

- Due to his years of experience of designing high performance solutions, Jose Carlos has acquired a high level of expertise in hardware platforms, security and open source solutions

- jbaquerot@gmail.com
- jcbaquero@gmv.com
- @jbaquerot

CIFF
Centro Internacional de Formación Financiera

Why Cyber Security?

What about Data Science?

CIFF

Centro Internacional
de Formación Financiera

Real Time Cyber-Attacks

Data Science can help to Cyber Security to uncover **unknown risks**

CIFF
Centro Internacional
de Formación Financiera

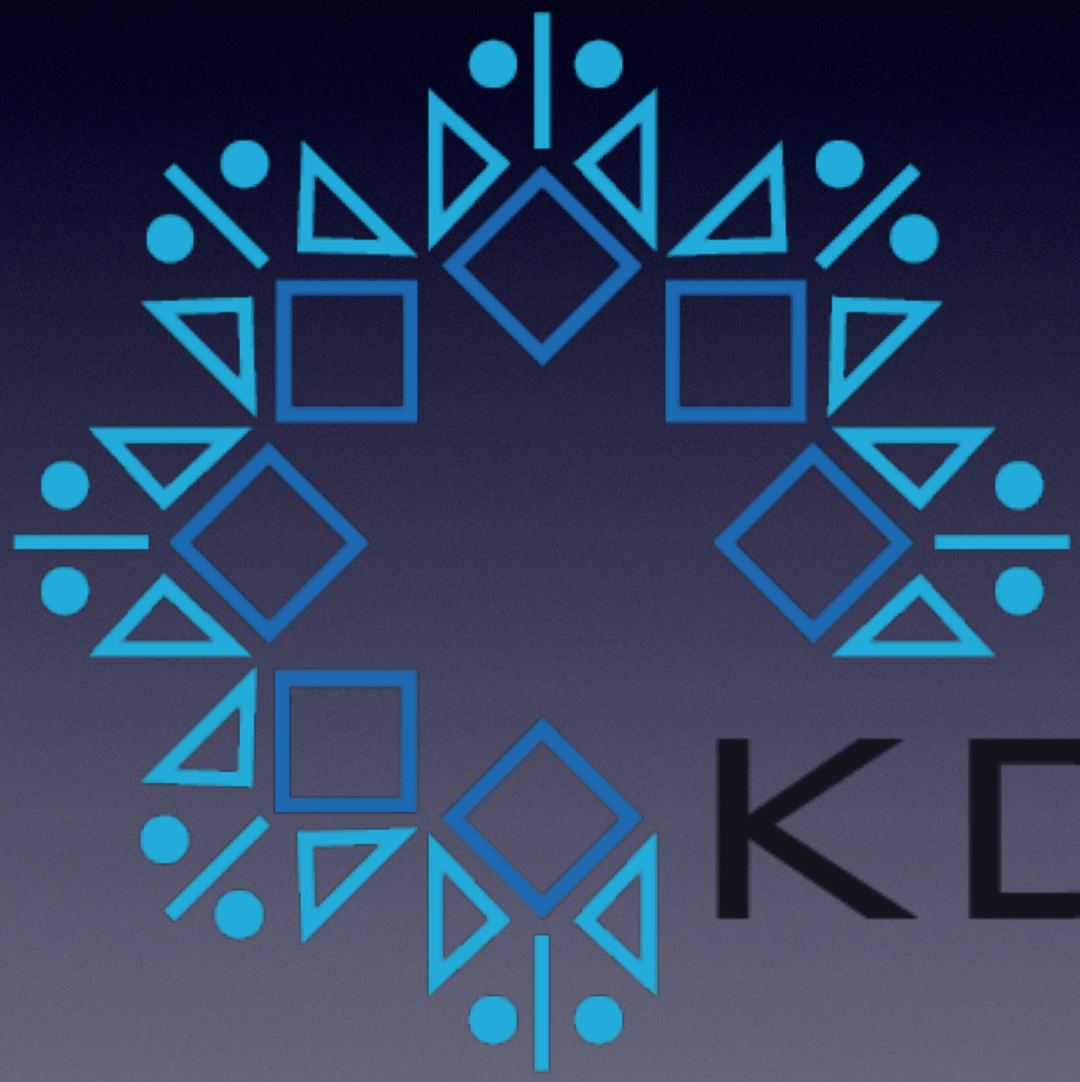
How to learn Cyber Security?

In a Data Science Course?

CIFF

Centro Internacional
de Formación Financiera

KDD 99 Competition



KDD

CIFF

Centro Internacional
de Formación Financiera

What?

- Introduction to Cyber Security
- Cyber Security Topics
- Security Data Science
- KDD99 Data Set

CIFF
Centro Internacional
de Formación Financiera

Introduction to Cyber Security

CIFF
Centro Internacional
de Formación Financiera

Cyber Security Definition (Wikipedia)

- Cyber Security is the process of applying **security measures** to ensure **confidentiality, integrity, and availability** of data.
- Cyber Security attempts to assure the **protection of assets**, which includes data, desktops, servers, buildings, and most importantly, humans.
- The goal of Cyber Security is to **protect data** both in transit and at rest.
- **Countermeasures** can be put in place in order to increase the security of data. Some of these measures include, but are not limited to, access control, awareness training, audit and accountability, risk assessment, penetration testing, vulnerability management, and security assessment and authorization

http://en.wikipedia.org/wiki/Computer_security

Cyber Security Topics

1. The threat landscape
2. Vulnerabilities
3. Cyber attacks, stages and patterns
4. Reducing the impact of an attack

CIFF
Centro Internacional
de Formación Financiera

The Thread Landscape

CIFF
Centro Internacional
de Formación Financiera

1. The Threat Landscape

- The risk to information and computer assets comes from a **broad spectrum of threats** with a **broad range of capabilities**.

- The **impact** (and therefore the harm) on your business will depend on:
  - the **opportunities** you present to an attacker (in terms of the vulnerabilities within your systems),
  - the **capabilities** of the attackers to exploit them, and ultimately
  - their **motivation** for attacking you

The attackers need an **opportunity** to deliver a **successful attack**.

You have **no control** over their **capabilities** and **motivations**, but you can make it harder for attackers by reducing your **vulnerabilities**.

CIFF
Centro Internacional de Formación Financiera

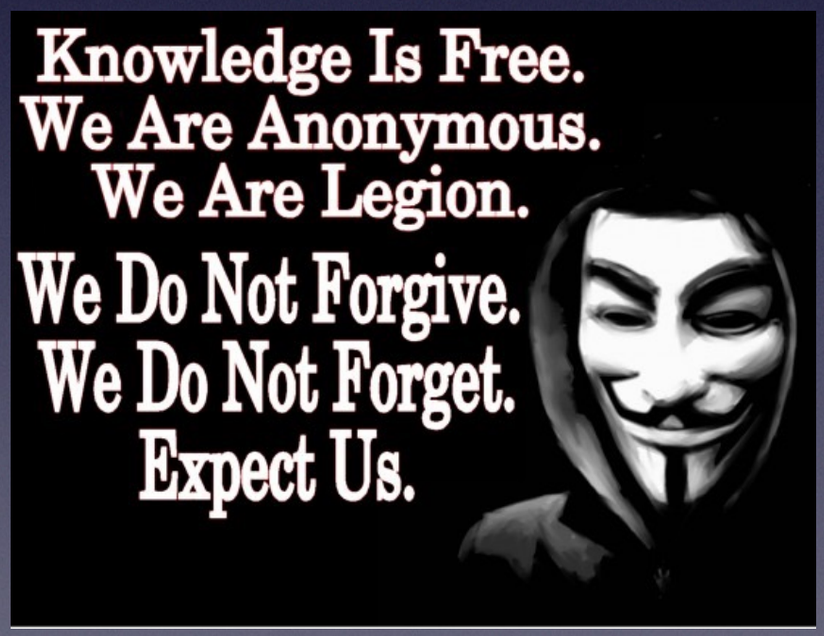
Risk Definition

Risk is the potential for a **threat** (a person or thing that is likely to cause damage) to exploit a **vulnerability** (a flaw, feature or user error) that may result in some form of **negative impact**

CIFF
Centro Internacional de Formación Financiera

Who might be attacking you?

- Cyber criminals
- Industrial competitors and foreign intelligence services
- Hackers
- Hacktivists
- Employees, or those who have legitimate access



Commodity vs bespoke capabilities

- **Commodity capability** involves tools and techniques openly available on the Internet (off-the-shelf) that are relatively simple to use.

- **Bespoke capability** involves tools and techniques that are developed and used for specific purposes, and thus require more specialist knowledge.

- Bespoke capabilities usually become commodity capabilities once their use has been discovered

Un-targeted vs targeted attacks

- In **un-targeted attacks**: attackers indiscriminately target as many devices, services or users as possible:
  - phishing
  - water holing
  - ransomware
  - scanning
- In a **Targeted attack**, your organisation is singled out because the attacker has a specific interest in your business, or has been paid to target you.
  - spear-phishing
  - deploying a botnet
  - subverting the supply chain

The Insider Threat

- **Insiders** (anyone who has legitimate access to your systems as an employee or a contractor) should also be considered as part of a holistic security regime.

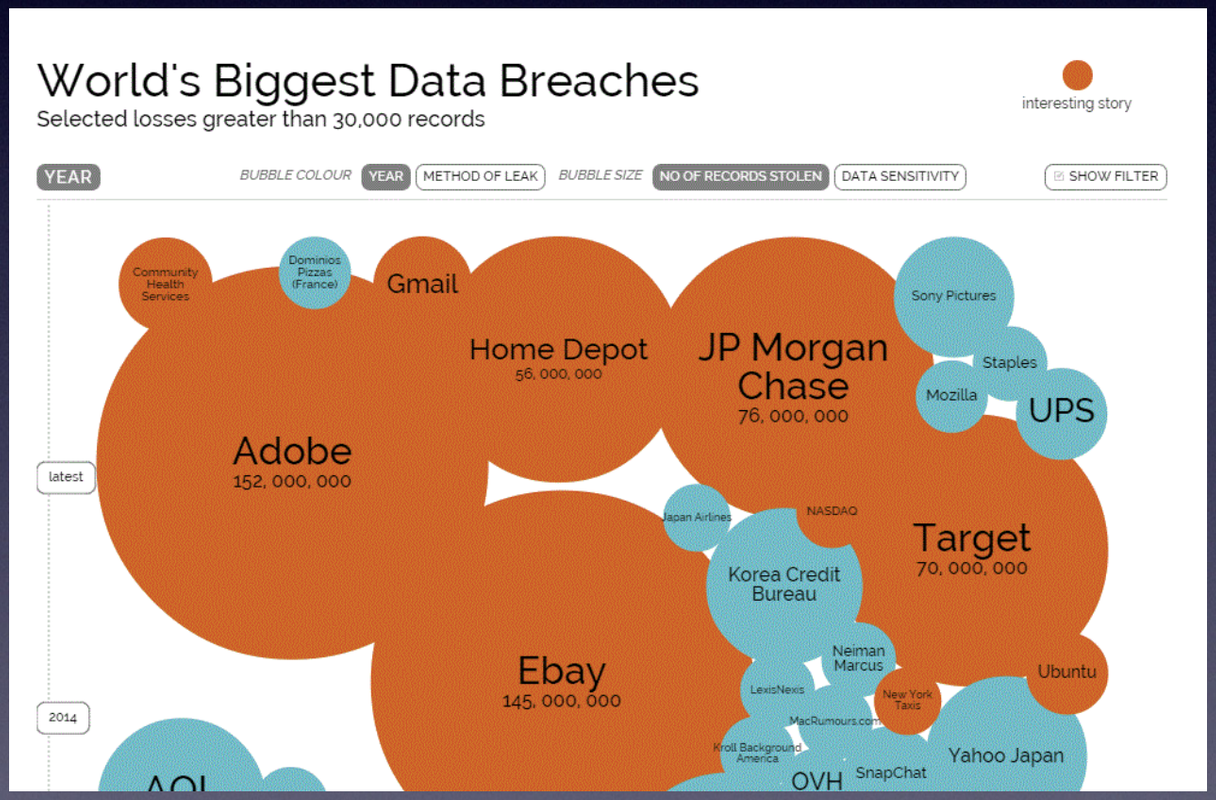
- They may be **motivated** by personal gain or redress against grievances.

- Insiders can also **accidentally compromise** a system or the information it holds

CIFF
Centro Internacional
de Formación Financiera

Every organisation is a potential victim

- All organisations have **something of value** that is worth something to others.
- If you openly demonstrate **weaknesses** in your approach to cyber security by failing to do the basics, you will experience some form of cyber attack.
- You should **be assessing** whether you are likely to be the victim of a targeted or un-targeted attack



CIFF
Centro Internacional de Formación Financiera

Understanding Vulnerabilities

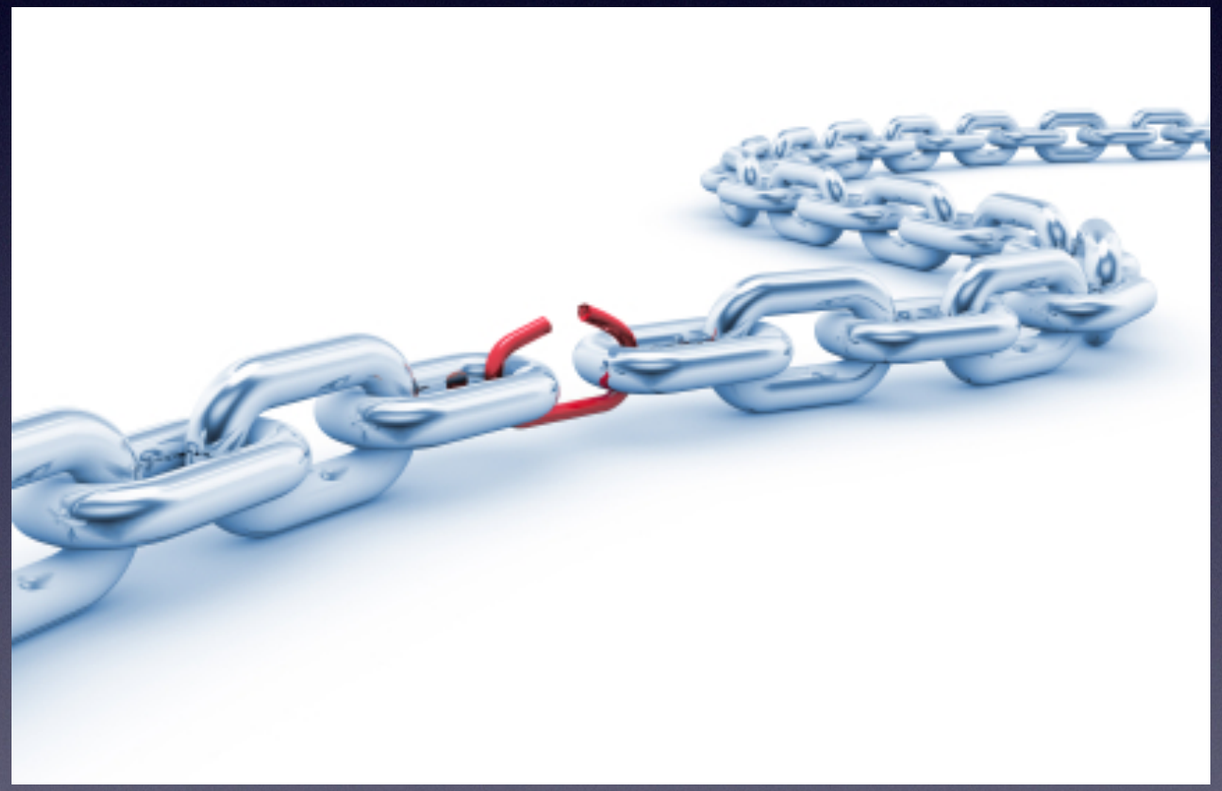
CIFF
Centro Internacional
de Formación Financiera

2. Understanding Vulnerabilities

- Vulnerabilities provide the **opportunities** for attackers to gain **access to your systems**.

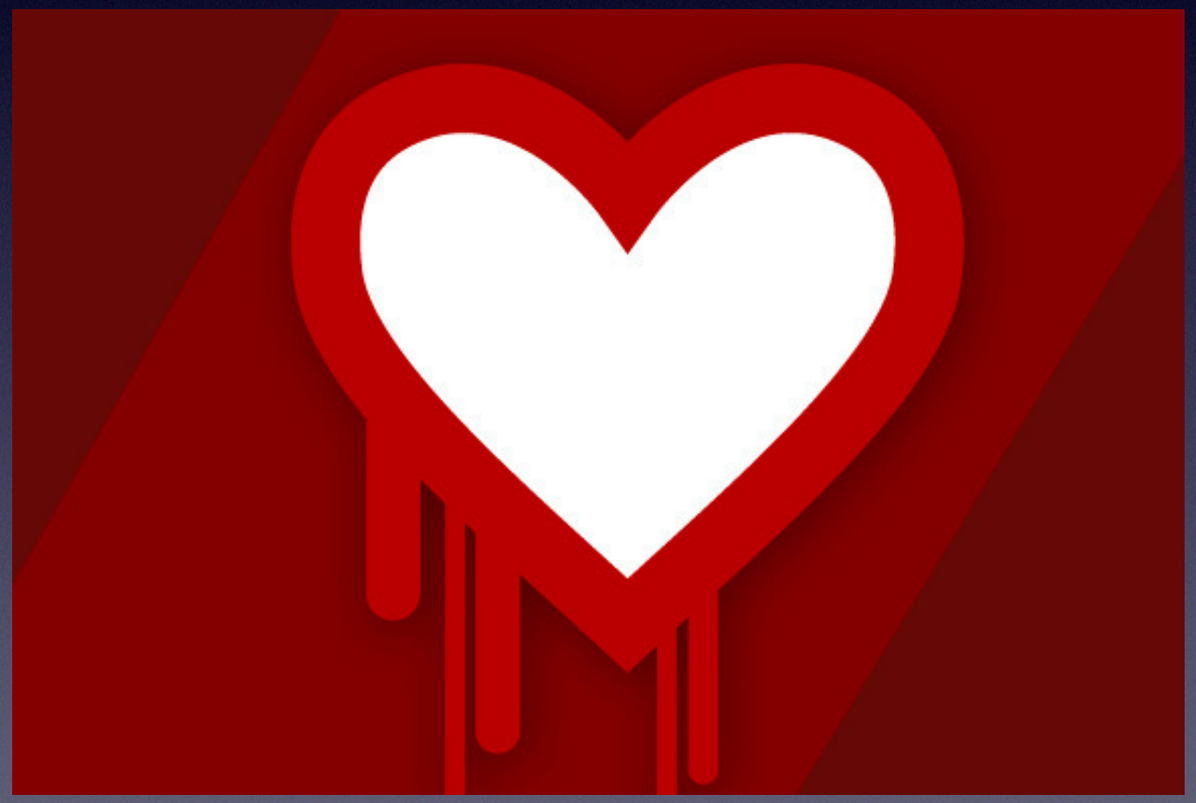
- They can occur through **flaws**, **features** or **user error,** and attackers will look to exploit any of them, often combining one or more, to achieve their end goal.

- A **vulnerability** is a weakness in an IT system that can be exploited by an attacker to deliver a **successful attack**



Flaws

- A **flaw** is **unintended functionality.**
- This may either be a result of poor design or through mistakes made during implementation.
- Flaws may go **undetected** for a significant period of time (**“zero-day”** vulnerabilities).
- The **majority of common attacks** we see today exploit these types of vulnerabilities.

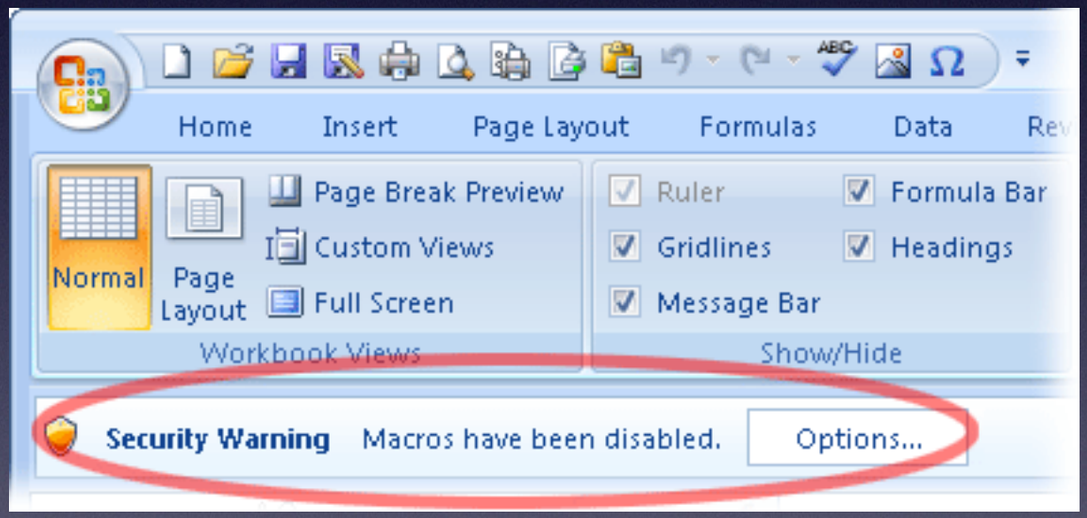


Features

- A **feature** is **intended functionality** which can be misused by an attacker to breach a system.

- Features may improve the user's experience, help diagnose problems or improve management, but they can also be exploited by an attacker.

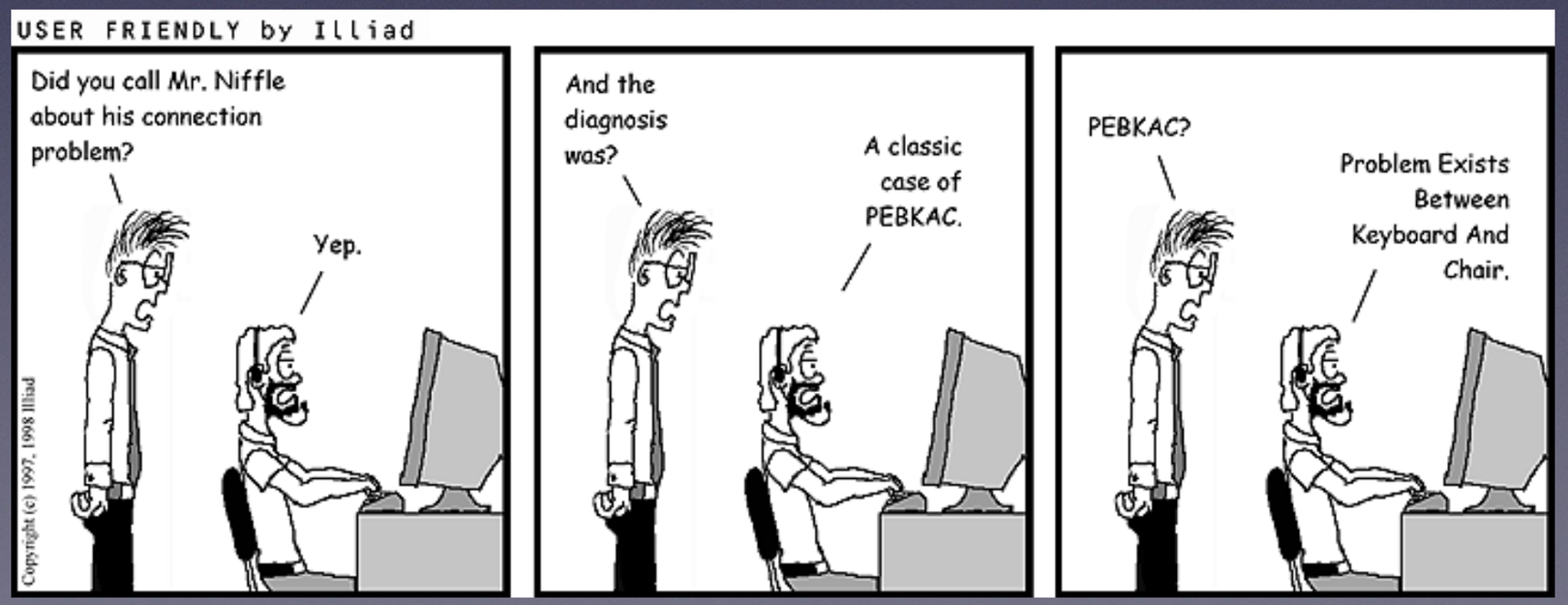
- Ex. MS Office Macros



CIFF
Centro Internacional de Formación Financiera

User Error

- Users can be a significant **source of vulnerabilities**.
- They make mistakes, such as choosing a common or easily guessed password, or leave their laptop or mobile phone unattended.
- A computer or system that has been carefully designed and implemented can minimise the vulnerabilities of exposure to the Internet. Unfortunately, such **efforts can be easily undone by an inexperienced system administrator**.



CIFF
Centro Internacional de Formación Financiera

Common Cyber Attacks

Stages and Patterns

CIFF

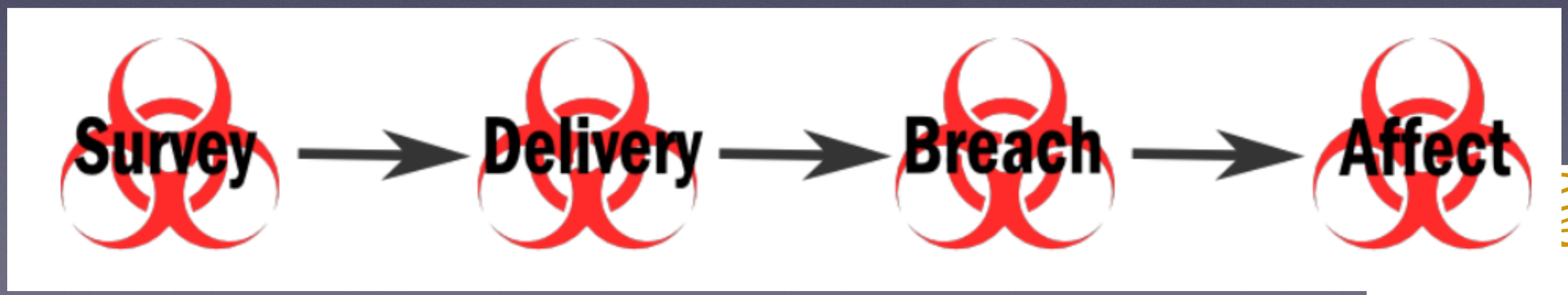
Centro Internacional
de Formación Financiera

3. Common Cyber Attacks - Stages and Patterns

- Regardless of whether an attack is targeted or un-targeted, or the attacker is using commodity or bespoke tools, **cyber attacks have a number of stages in common.**

- The attacker is effectively **probing your defences** for weaknesses that, if exploitable, will take them closer to their ultimate goal.

- **Understanding** these stages will help you to **better defend yourself.**



Stages of an attack

- **Survey** - investigating and analysing available information about the target in order to identify potential vulnerabilities

- **Delivery** - getting to the point in a system where a vulnerability can be exploited

- **Breach** - exploiting the vulnerability/ vulnerabilities to gain some form of unauthorised access

- **Affect** - carrying out activities within a system that achieve the attacker's goal

CIFF
Centro Internacional de Formación Financiera

Reducing Your Exposure to Cyber Attack

CIFF
Centro Internacional
de Formación Financiera

4. Reducing Your Exposure to Cyber Attack

- **Preventing**, **detecting** or **disrupting** the attack at the earliest opportunity limits the business impact and the potential for reputational damage.

- Once the attacker has consolidated their presence they will be **more difficult to find and remove**

CIFF
Centro Internacional de Formación Financiera

Breaking the attack pattern

- Attackers will frequently use **commodity tools and techniques**, which are cheaper and easier for them to use.

- So putting in place **security controls and processes that can mitigate** these will go some way to making your business a hard target.

- Equally, adopting a **defence-in-depth approach** to mitigate risks through the full range of potential attacks will give your **business more resilience** to cope with attacks that use more bespoke tools and techniques.

Reducing your exposure using essential security controls

- Boundary firewalls and Internet gateways
- Malware protection
- Patch management
- Whitelisting and execution control
- Secure configuration
- Password policy
- User access control
- **Security monitoring**
- User training
- Security incident management

Security Data Science

CIFF

Centro Internacional
de Formación Financiera

Security Data Science

- Security Data Science is the application of advanced analytics to activity and access data to **uncover unknown risks**.

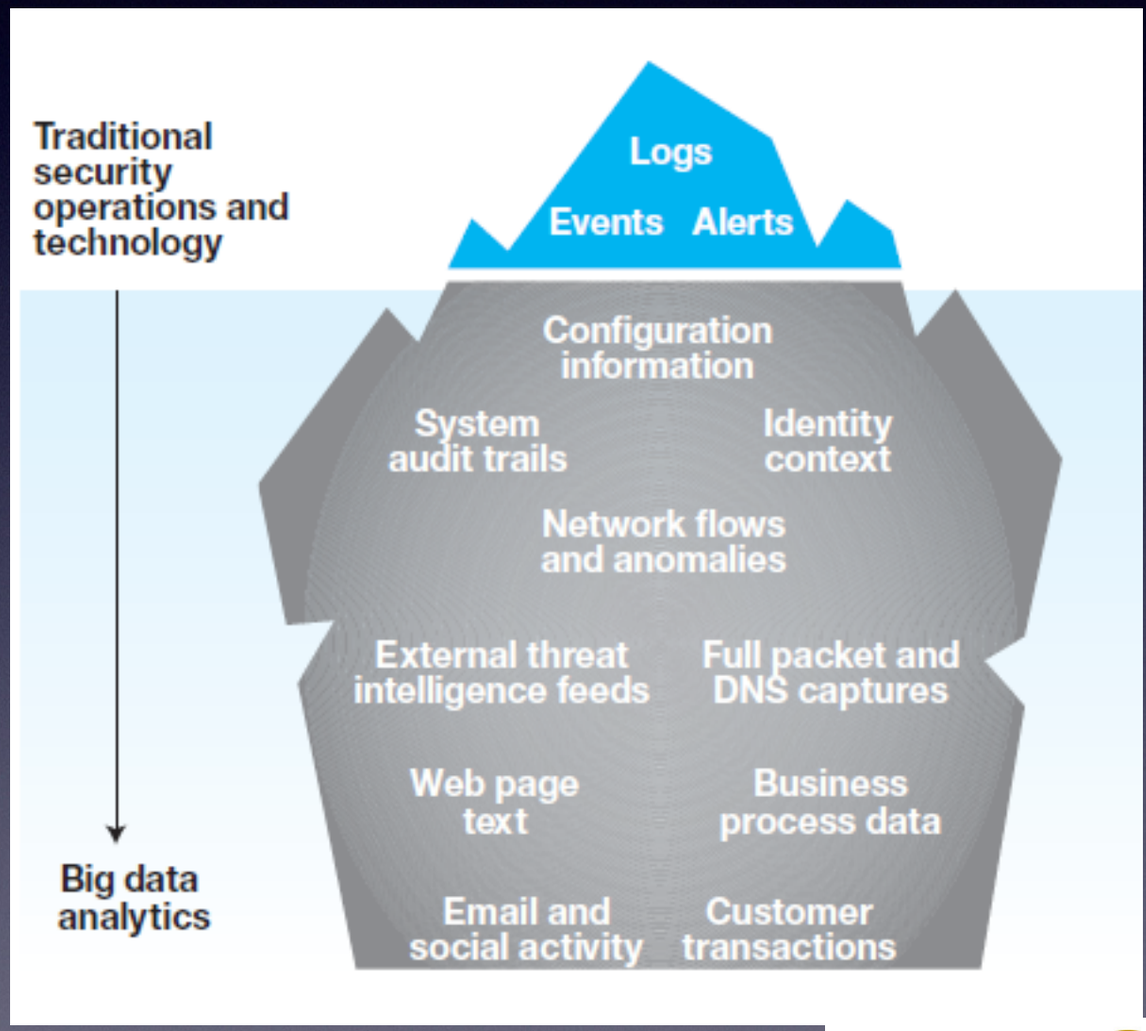
- Generally Data Science is the practice of deriving valuable insights from data. In Security the valuable insight leads to **reduced risk.**

- In the last decade **analytics** has become increasingly important to effective **risk mitigation.**

CIFF
Centro Internacional de Formación Financiera

Reducing The Cyber Risk with Data Science

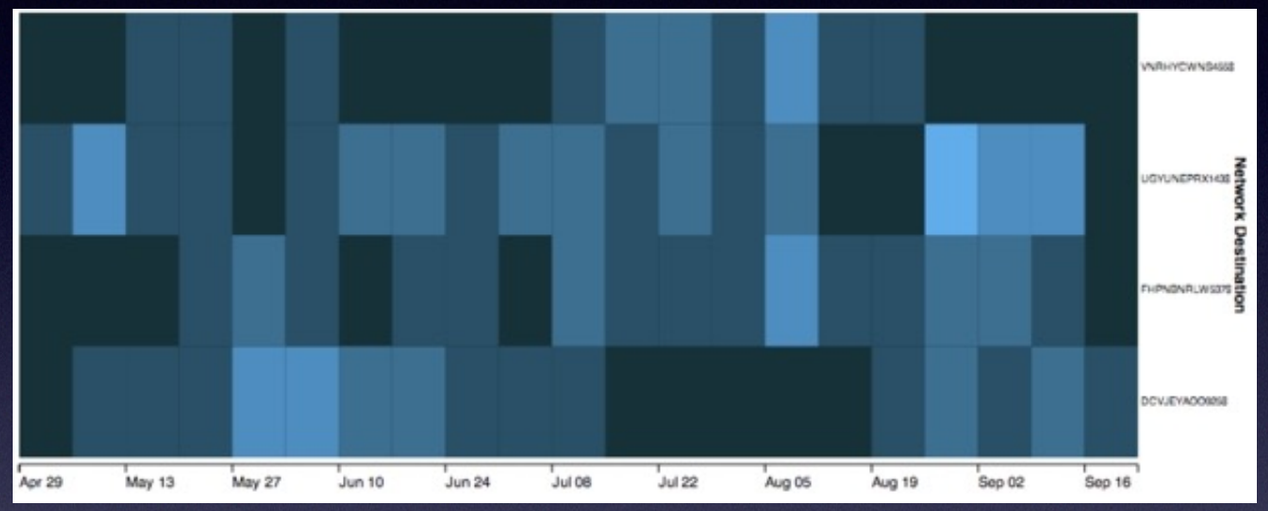
- Use **intrusion monitoring tools** and regularly audit activity logs.
- **Monitor user activity**, particularly access to sensitive information and the use of privileged accounts.
- Ensure that the solution **monitors all networks and host systems** (eg clients and servers).
- Network traffic should be continuously monitored to **identify unusual activity or trends** that could indicate an attack

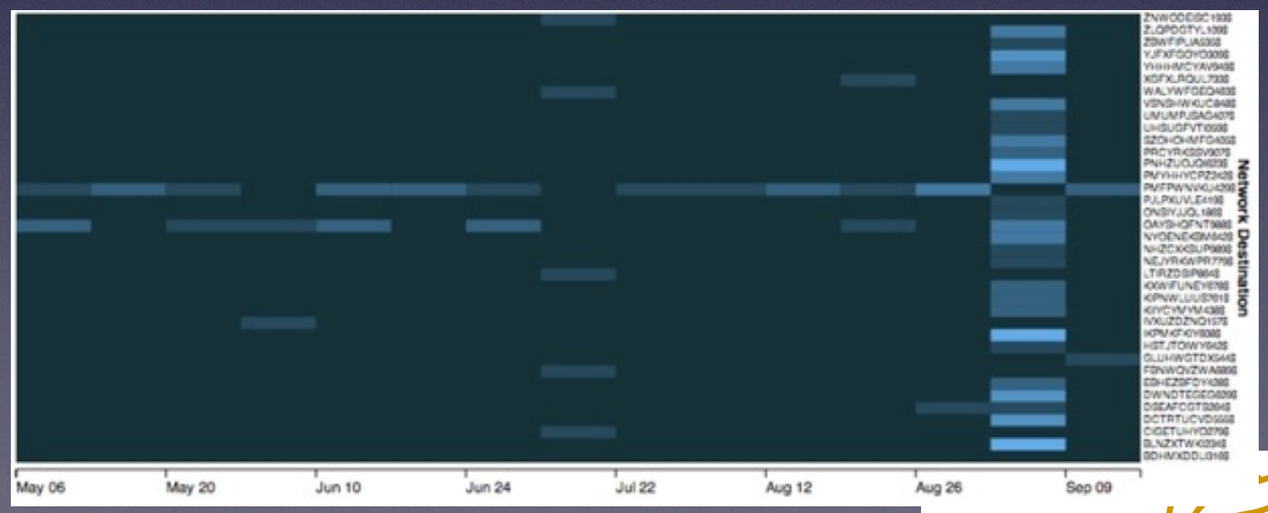


CIFF
Centro Internacional de Formación Financiera

Security Data Science Examples

- Cyber attacks detection
- Botnets identification
- APT detection
- Malware recognition
- Detecting Insider Security Threats
- Digital Forensic





CIFF
Centro Internacional de Formación Financiera

KDD 99 Data Set

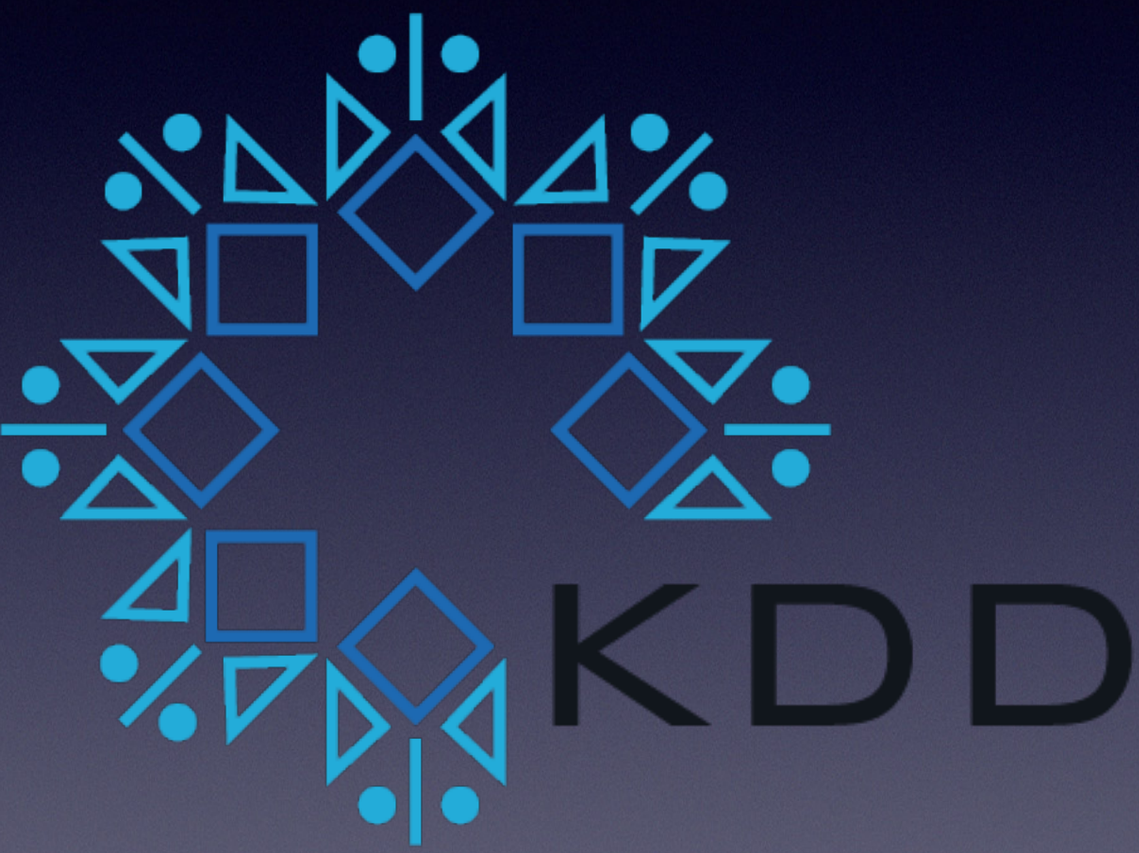
CIFF
Centro Internacional
de Formación Financiera

KDD 99 Competition

- The competition task was to build a **network intrusion detector**, a predictive model capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections

- This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment.

- http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html



The NSL-KDD Data Set

- The NSL-KDD data set advantages over the original KDD 99:
  - It does not include redundant records
  - The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set.
  - The number of records in the train and test sets are reasonable.

STATISTICS OF REDUNDANT RECORDS IN THE KDD TRAIN SET

	Original Records	Distinct Records	Reduction Rate
Attacks	3,925,650	262,178	93.32%
Normal	972,781	812,814	16.44%
Total	4,898,431	1,074,992	78.05%

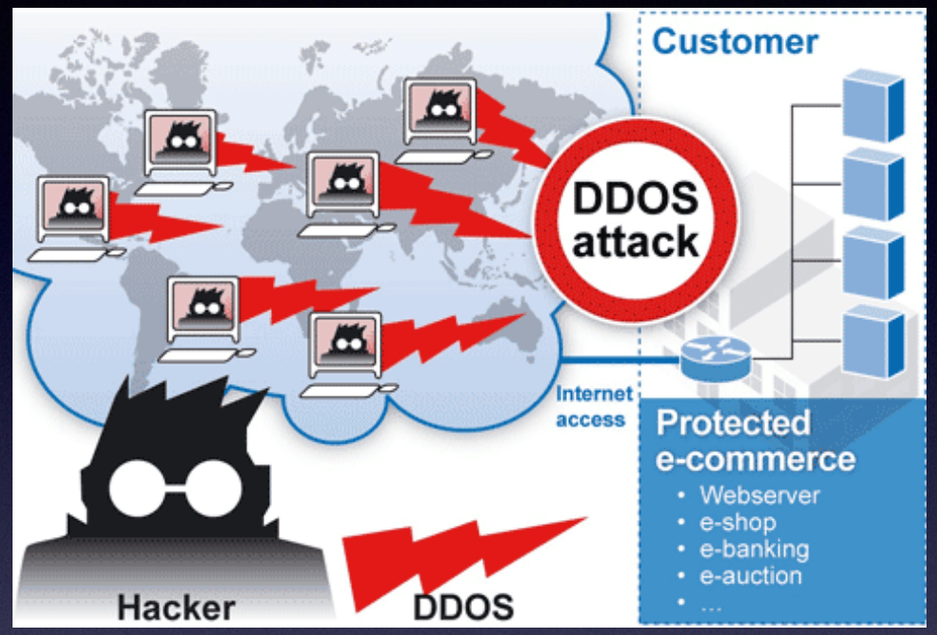
TABLE II

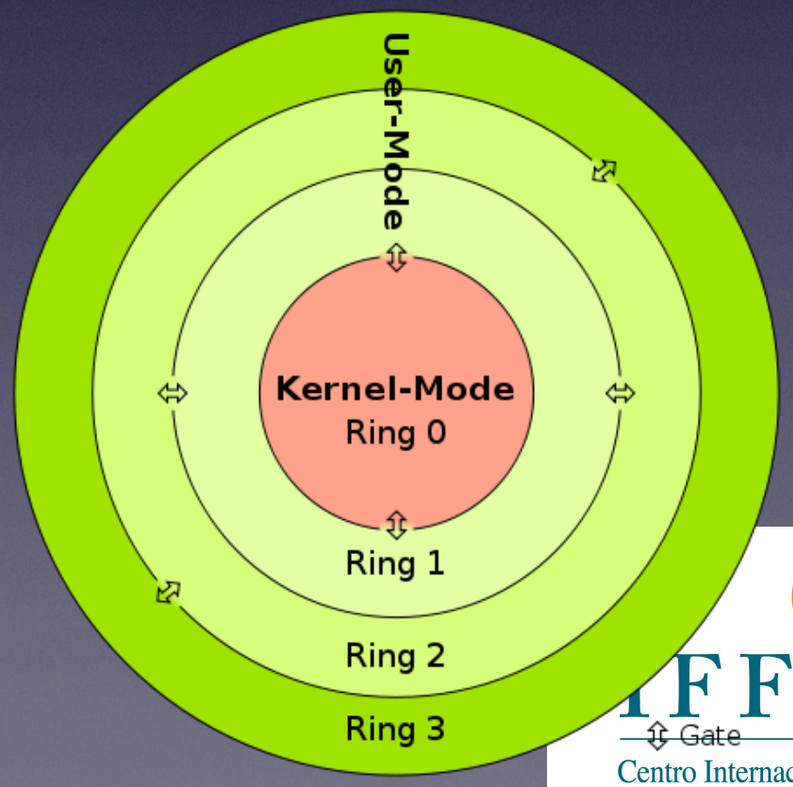
STATISTICS OF REDUNDANT RECORDS IN THE KDD TEST SET

	Original Records	Distinct Records	Reduction Rate
Attacks	250,436	29,378	88.26%
Normal	60,591	47,911	20.92%
Total	311,027	77,289	75.15%

Attack Categories

- Denial of Service Attack (DoS)
- User to Root Attack (U2R)
- Remote to Local Attack (R2L)
- Probing Attack





Features Groups

- Basic features from a TCP/IP connection.
- Content features
- Time traffic from same host (last 2 seconds)
- Time traffic from same service (last 2 seconds)
- Machine traffic from the same host (previous 100 connections)
- Machine traffic from the same service (previos 100 connections)

CIFF
Centro Internacional de Formación Financiera

Basic features from a TCP/IP connection.

Col	Feature name	description	type
1	duration	length (number of seconds) of the connection	continuous
2	protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
3	service	network service on the destination, e.g., http, telnet, etc.	discrete
4	flag	normal or error status of the connection. The possible status are this: SF, S0, S1, S2, S3, OTH, REJ, RSTO, RSTOS0, SH, RSTRH, SHR	discrete
5	src_bytes	number of data bytes from source to destination	continuous
6	dst_bytes	number of data bytes from destination to source	continuous
7	land	1 if connection is from/to the same host/port; 0 otherwise	discrete
8	wrong_fragment	sum of bad checksum packets in a connection	continuous
9	urgent	number of urgent packets. Urgent packets are packets with the urgent bit activated	continuous

Data Mining Process

CIFF
Centro Internacional de Formación Financiera

IPython

NumPy

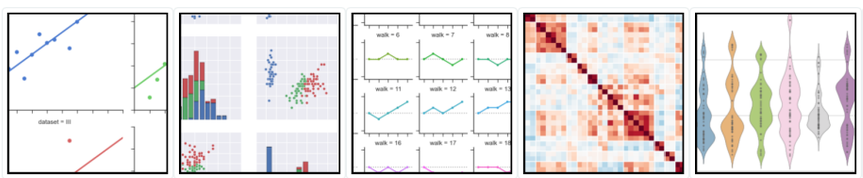
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

Next Week: Data

Exploration

Seaborn: statistical data visualization

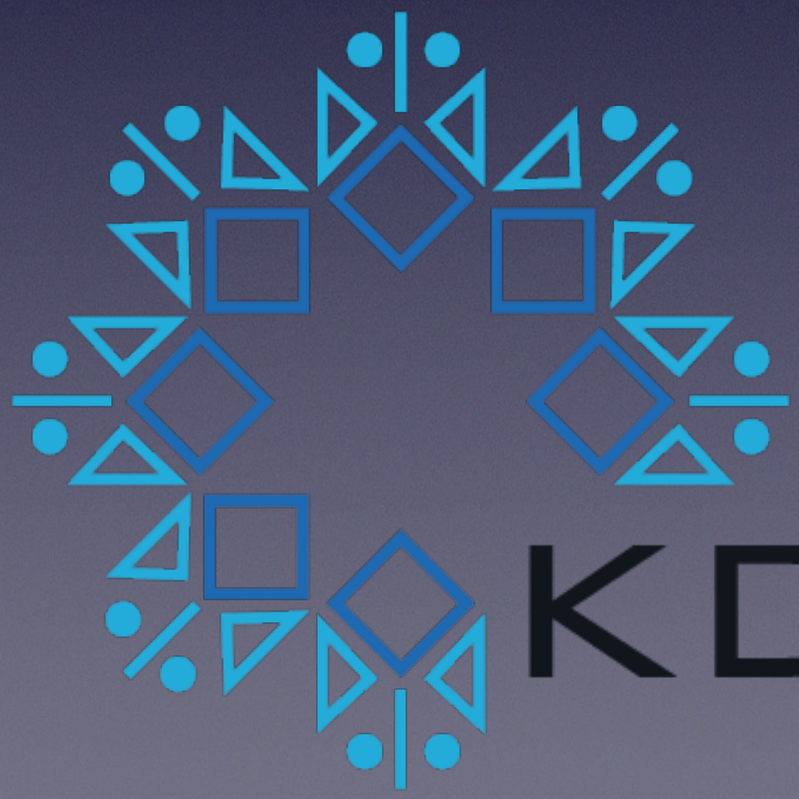


matplotlib

CIFF

Centro Internacional
de Formación Financiera

Download the data set for:
http://nsl.cs.unb.ca/NSL-KDD/



CIFF
Centro Internacional
de Formación Financiera