



Pattern Recognition
and Applications Lab

# Ingannati dagli algoritmi
# Algoritmi tratti in inganno

*Giorgio Giacinto*
*giacinto@unica.it*


3 Ottobre 2018

cini
Cyber Security National Lab

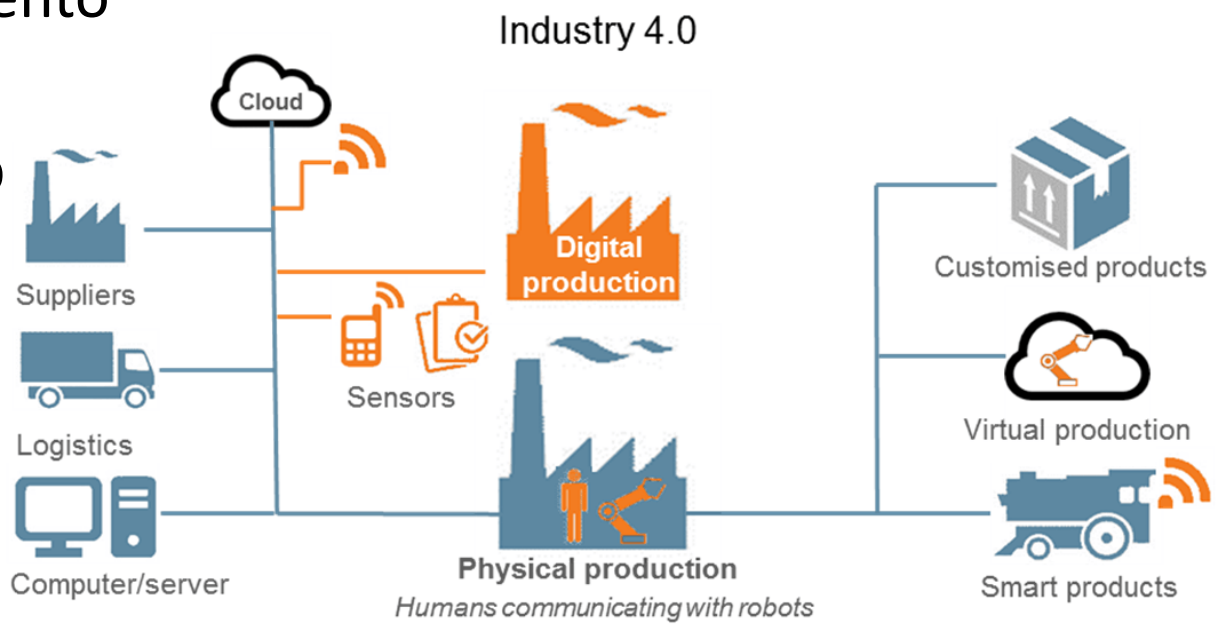
University
of Cagliari, Italy

Department of Electrical
and Electronic
Engineering

# Lo scenario

# Hyperconnected world

- **Il computer, protagonista in tutte le funzioni di una organizzazione**
  - Amministrazione
  - Rapporti con i clienti
  - Produzione
  - Approvvigionamento
  - Cloud
  - Presenza sul web

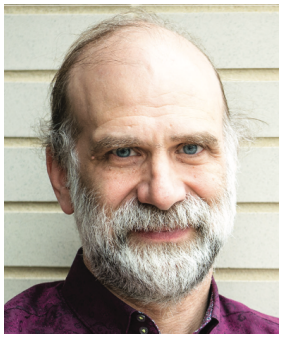





http://blog.euromonitor.com

# La complessità è nemica della sicurezza

- Internet è un sistema estremamente **complesso**
  - **persone**
  - **dispositivi e applicazioni**
  - **interazioni**
  - maggiore probabilità di **errore** in fase di **progetto** e **sviluppo**



Bruce Schneier

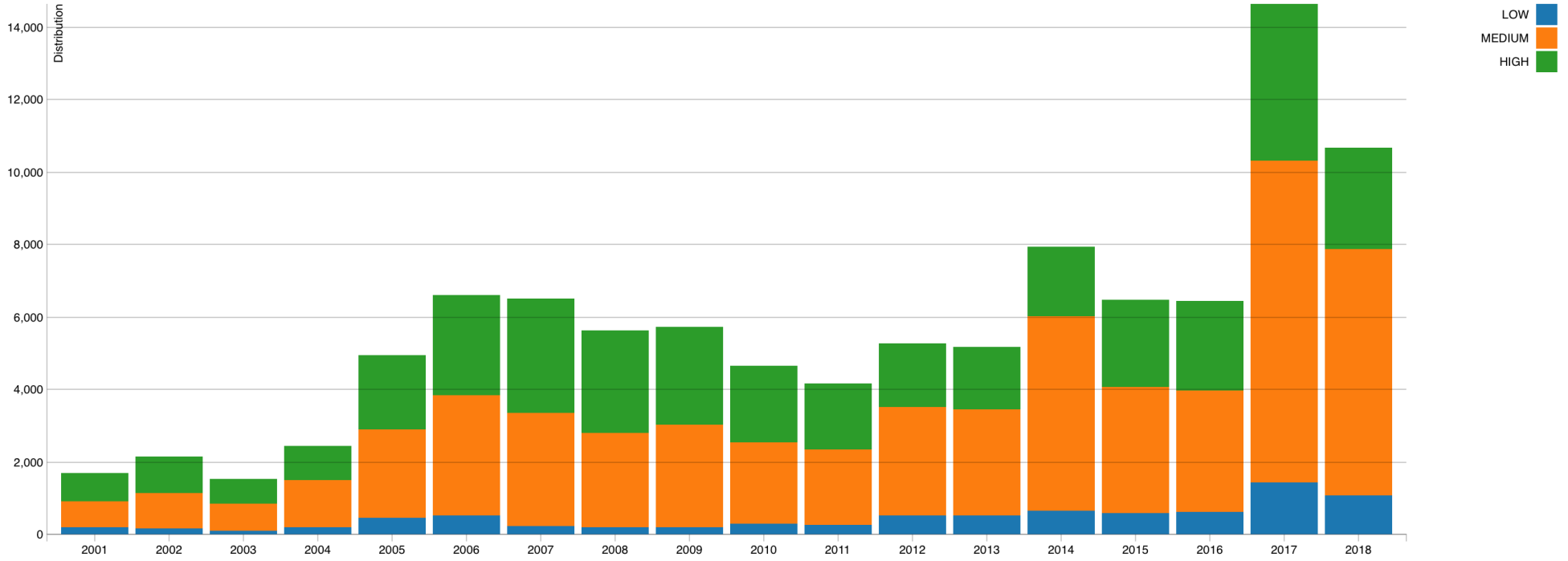
- Un sistema complesso ha una **superficie di attacco** ampia
- **Difenderlo** significa rendere **sicura l'intera superficie** di attacco**.**
- Ma per l'**avversario** è sufficiente trovare **una sola vulnerabilità.**



http://nymag.com/selectall/2017/01/the-internet-of-things-dangerous-future-bruce-schneier.html

# Vulnerabilità nei sistemi software

CVSS (Common Vulnerabilities Scoring System)



# Facebook security issue 50m+ accounts

**Facebook Security Update**

Originally published on September 28, 2018 at 9:41AM PT

*Our investigation is still in its early stages. But **it's clear that attackers exploited a vulnerability in Facebook's code** that impacted "View As" a feature that lets people see what their own profile looks like to someone else. **This allowed** them **to steal Facebook access tokens** which they could then use to take over people's accounts.*



https://newsroom.fb.com/news/2018/09/security-update/

# Il ”cerchio della fiducia”

Robert De Niro e Ben Stiller in “Ti presento i miei”

# Fiducia nel mondo virtuale

- Nel mondo virtuale si può concedere **fiducia**
  - fra persone
  - fra dispositivi
  - fra applicativi software

  ***limitatamente*** ad alcuni ambiti

- **Pericolo**: Si può *facilmente* concedere ***ampia* fiducia** in base a pochi *indizi*

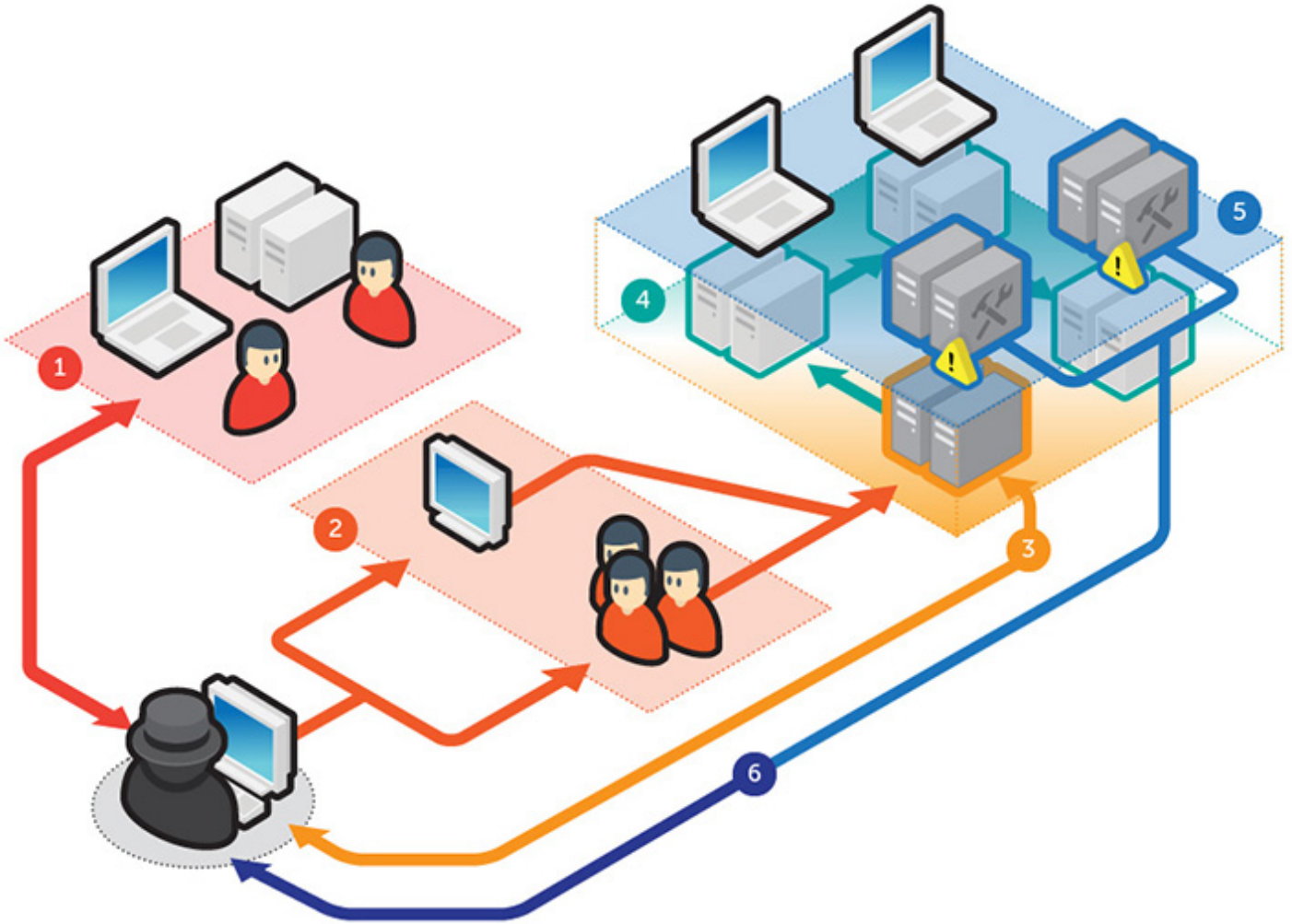
- **Attenzione**: La **fiducia *non*** può essere **transitiva**
  - fra coppie di entità, relazioni di fiducia in ambiti diversi
  - difficile, se non *impossibile*, da controllare

# Entrare nel cerchio della fiducia

# Come ispirare fiducia?

"Carini e Coccolosi"

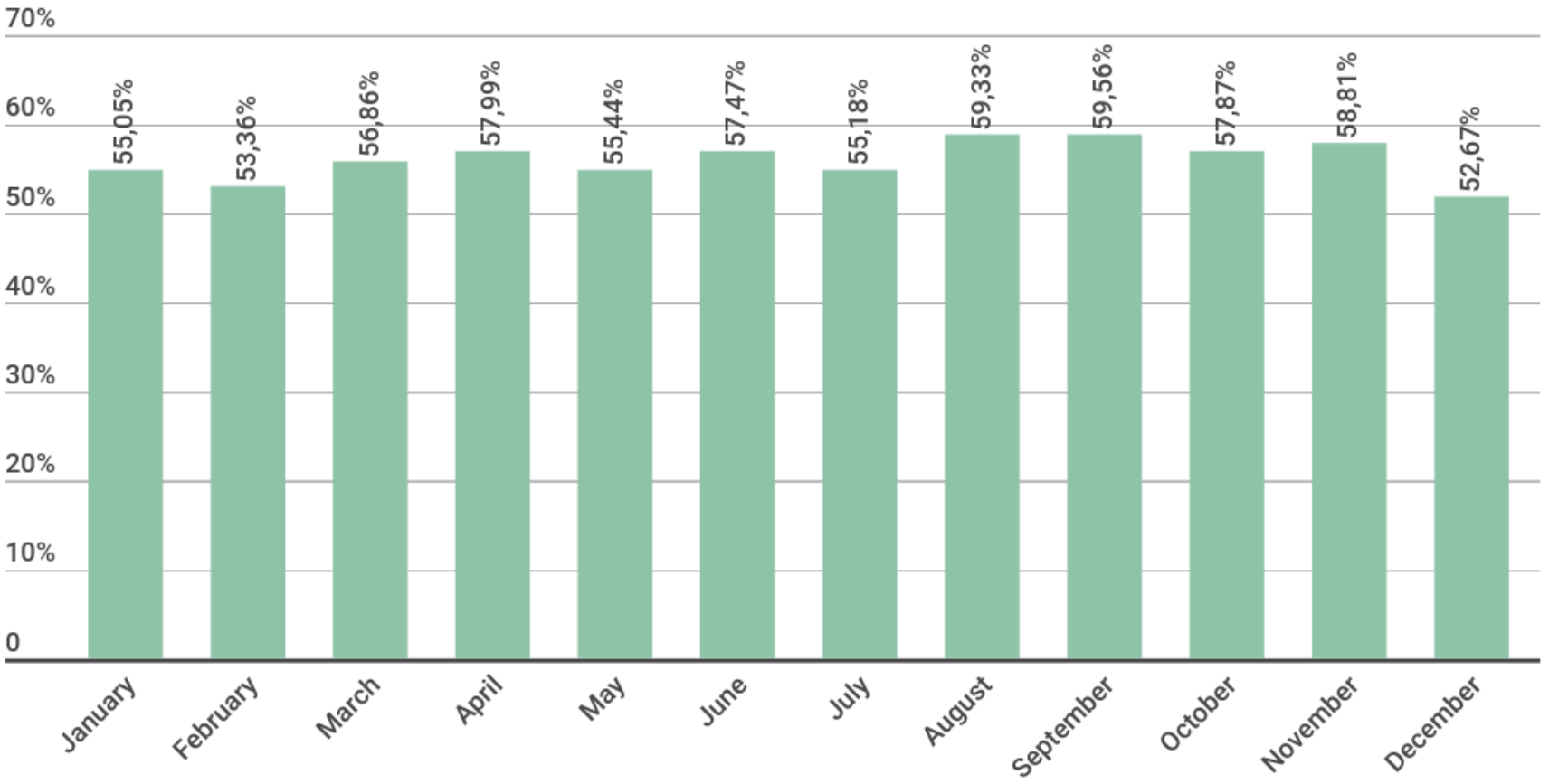
# Kill-chain di un attacco




TrendMicro

# email spam

## 2017 statistics






https://securelist.com/spam-and-phishing-in-2017

# Credibilità dei falsi

- **Ambiguità testo**
- **Manipolazione dei contenuti multimediali**
  - Possono essere state modificate in modo realistico
  - Possono essere vere ma fuori contesto
  - Possono rappresentare una dettaglio marginale
  - ecc.

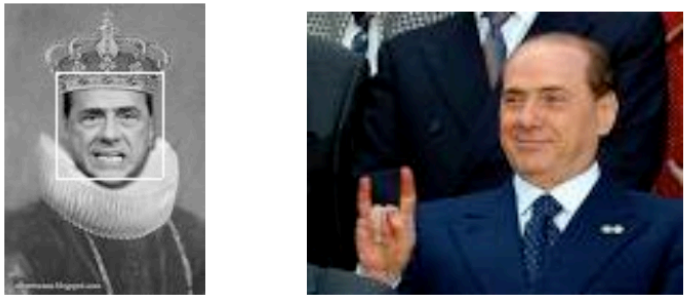
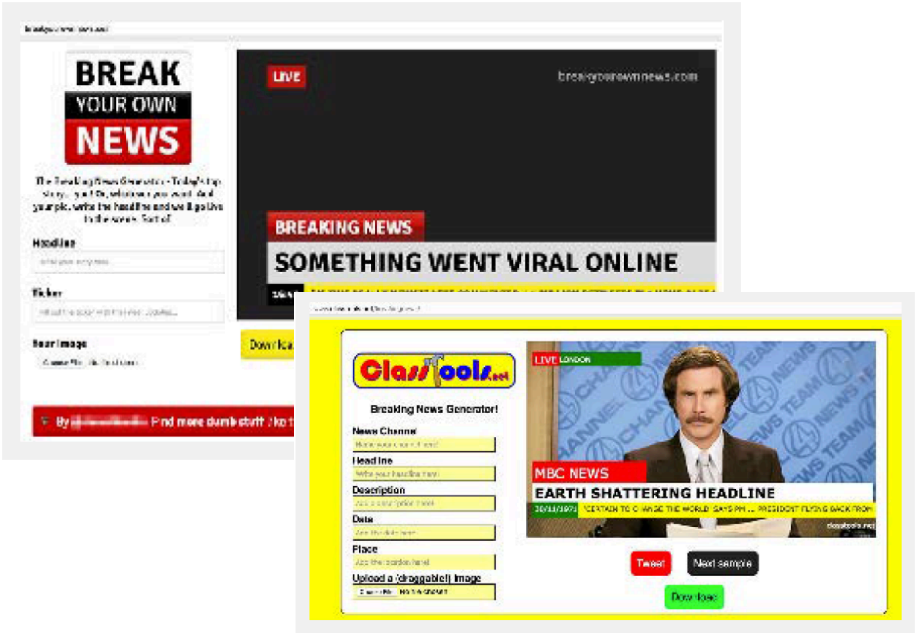


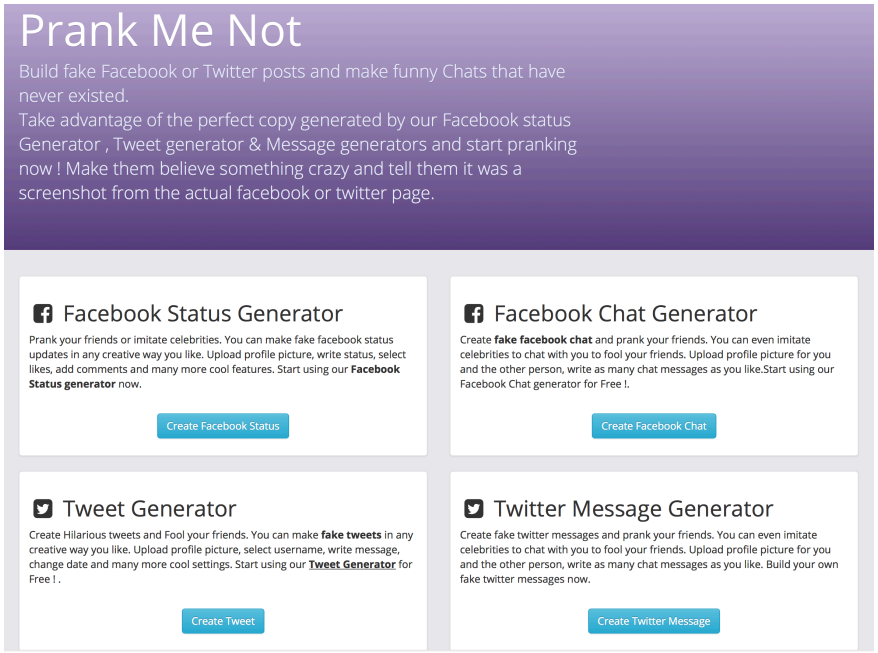
Fig. 2 Examples of (a) photomontage and (b) authentic photograph
G. Boato et al, *Towards Multimedia Opinion Mining*, 2009

- **Nei media sociali, credibilità basata su *numero like, condivisioni, ecc.***

# Software creato per ingannare

- **Software sofisticato per creare contenuti multimediali verosimili**
  - contenuti multimediali attribuibili a notiziari attendibili
  - annunci fasulli sui social network attribuiti a personaggi pubblici




TrendMicro, *The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public*, Giugno 2017

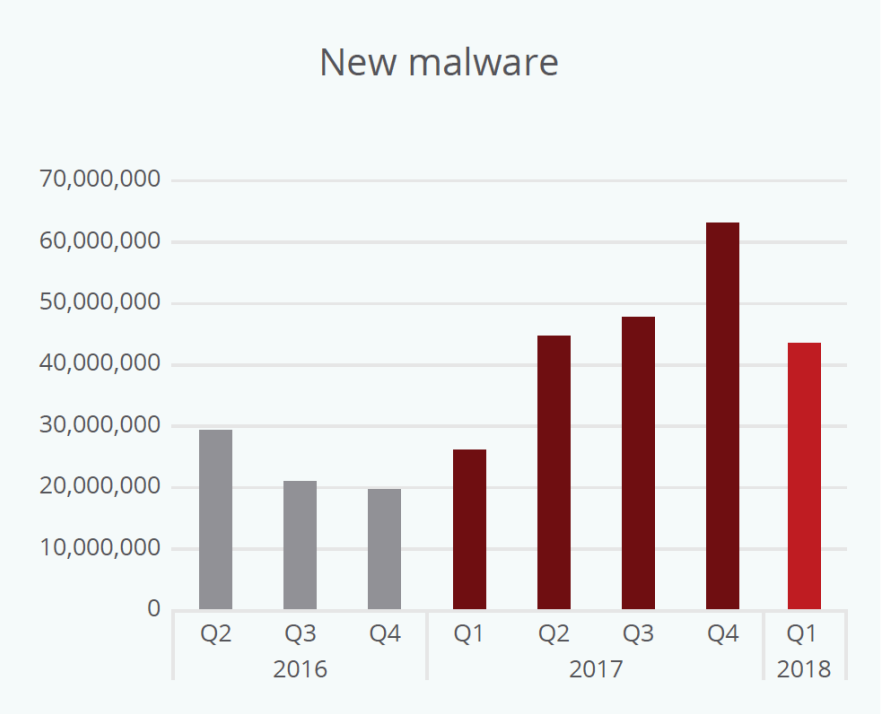
# Threat Landscape 2017

https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017

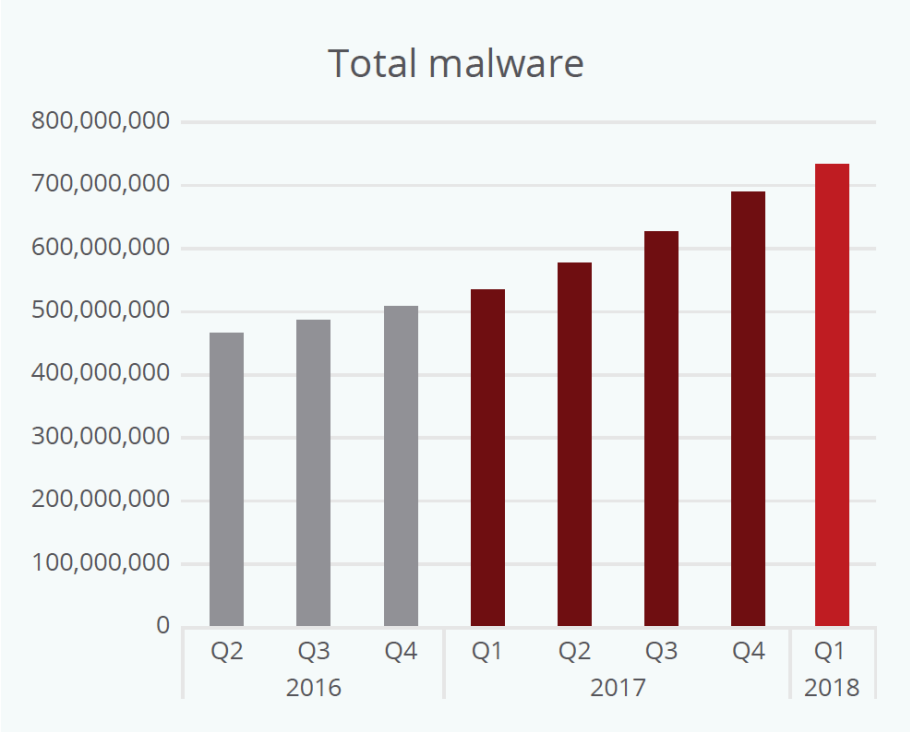
| Top Threats 2016 | Assessed Trends 2016 | Top Threats 2017 | Assessed Trends 2017 | Change in ranking |
|---|---|---|---|---|
| 1. Malware | ↑ | 1. Malware | | → |
| 2. Web based attacks | ↑ | 2. Web based attacks | | → |
| 3. Web application attacks | ↑ | 3. Web application attacks | | → |
| 4. Denial of service | ↑ | 4. Phishing | | ↑ |
| 5. Botnets | ↑ | 5. Spam | | ↑ |
| 6. Phishing | → | 6. Denial of service | | ↓ |
| 7. Spam | ↓ | 7. Ransomware | | ↑ |
| 8. Ransomware | → | 8. Botnets | | ↓ |

Legend: Trends: ↓ Declining, → Stable, ↑ Increasing
Ranking: ↑ Going up, → Same, ↓ Going down

# Malware




Source: McAfee Labs, 2018.

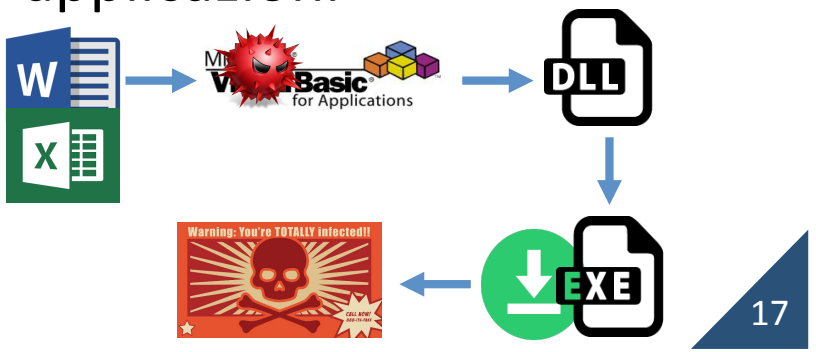



Source: McAfee Labs, 2018.

# Obfuscation and Evasion

- **Tecniche di programmazione di malware**
  - Codice offuscato per ingannare strumenti di *analisi statica*
  - Rilevazione di esecuzione in sandbox per ingannare strumenti di *analisi dinamica*
  - Imitazioni di applicazioni legittime

- **Social Engineering**
  - Spam, Phishing, social networks
  - Malware nascosto in documenti e applicazioni apparentemente legittimi



# Obfuscated and Evasive Malware

**Looking at the Bag is not Enough to Find the Bomb: An Evasion of Structural Methods for Malicious PDF Files Detection**

2013

Davide Maiorca Igino Corona Giorgio Giacinto

SECURITY WEEK
INTERNET AND ENTERPRISE SECURITY NEWS, INSIGHTS & ANALYSIS

Subscribe

Malware & Threats | Cybercrime | Mobile & Wireless | Risk & Compliance | Security

Home › Malware

**Evasive Malware Now a Commodity**

By Siggi Stefnisson on May 04, 2018

**Stealth attacks: An extended insight into the obfuscation effects on Android malware**

2015

*Davide Maiorca*[*], *Davide Ariu, Igino Corona, Marco Aresu, Giorgio Giacinto*

# Machine Learning & Malware Analysis

- AI and ML are important tools in the area of cyber-security. They can be used for context discovery in the area of Threat Intelligence (TI).
- Use of ML and AI in the identification of attack patterns and facilitation of security management/policy implementation supervision.

*Looking into the crystal ball - A report on emerging technologies and security challenges* (January 2018)

- **L'efficacia del Machine Learning dipende dal *numero* e dalla *rappresentatività* dei campioni di malware.**

Yes, Machine Learning Can Be More Secure!
A Case Study on Android Malware Detection

Ambra Demontis, *Student Member, IEEE,* Marco Melis, *Student Member, IEEE,* Battista Biggio, *Senior Member, IEEE,* Davide Maiorca, *Member, IEEE,* Daniel Arp, Konrad Rieck, Igino Corona, Giorgio Giacinto, *Senior Member, IEEE,* and Fabio Roli, *Fellow, IEEE*

2017

# Difendersi dagli inganni

- **Quando si scopre un nuovo attacco**
  - Analisi manuale di un piccolo insieme di dati
  - L'esperienza suggerisce le caratteristiche da estrarre dai dati per poter rilevare ulteriori attacchi *simili*
  - Progettazione di sistemi automatici per *generalizzare*
- **Gli algoritmi di AI non devono creare nuove vulnerabilità**
  - Controllo della generalizzazione per evitare di creare opportunità per *ingannare* l'algoritmo di AI
  - *Adversarial Learning*

# Attacchi di Evasione basati sulla Discesa del Gradiente

**Obiettivo:** evasione a massima confidenza

**Conoscenza:** perfetta

**Strategia di attacco:**

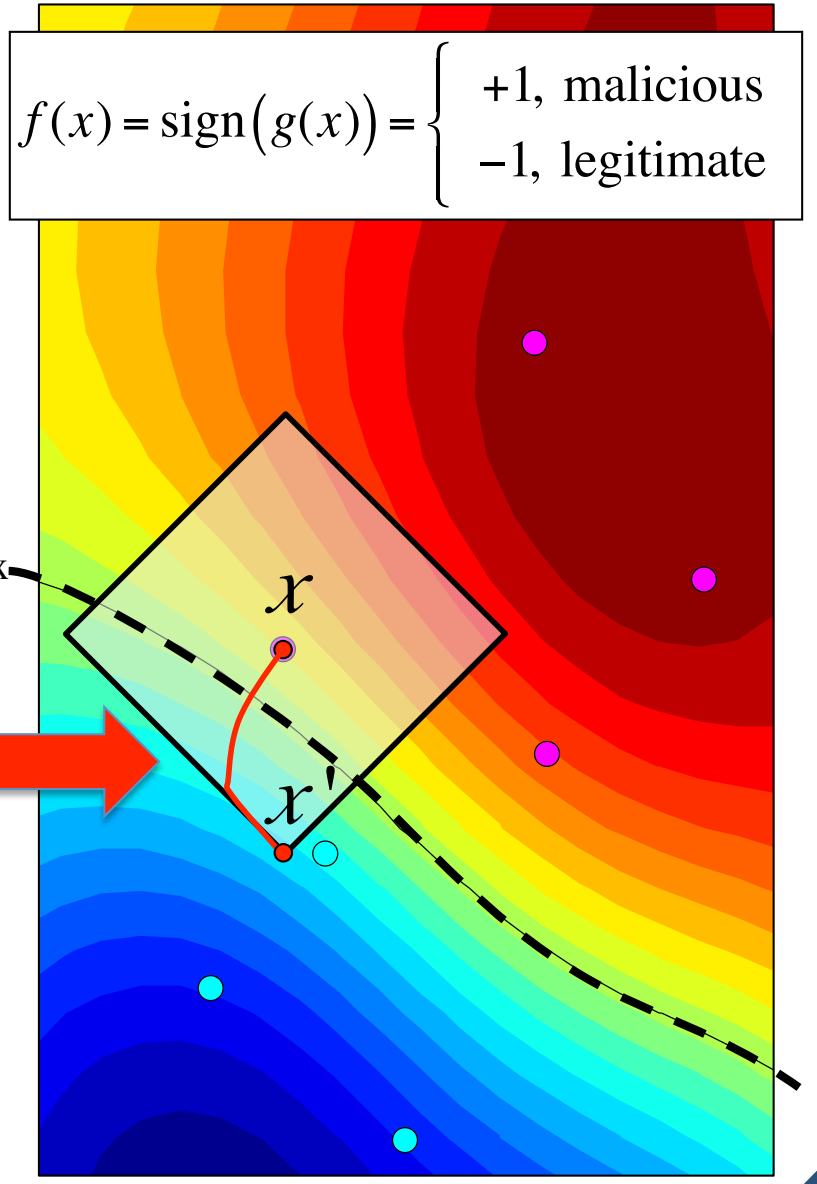
$$\min_{x'} g(x')$$

$$\text{s.t. } d(x, x') \leq d_{\max}$$

**Ottimizzazione basata sulla discesa del gradiente**

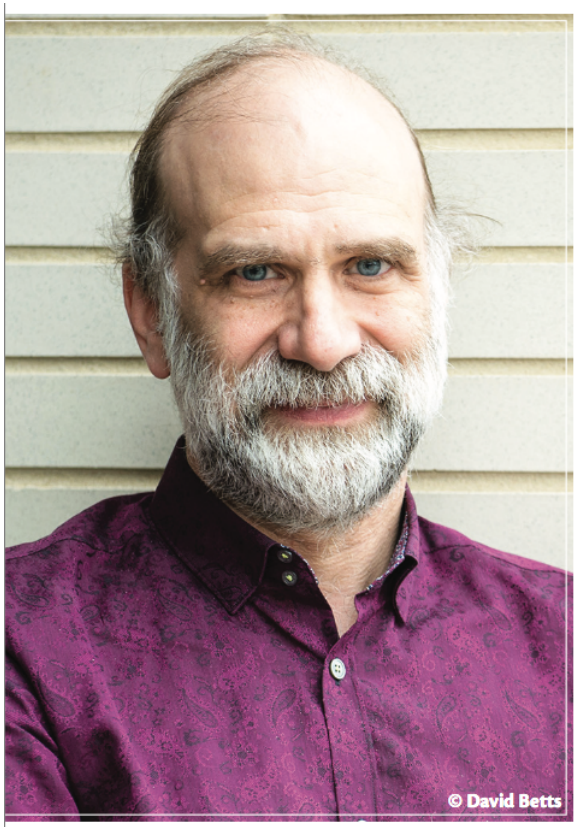
Il gradiente di g(x) può essere calcolato analiticamente

- SVMs, Neural Networks, ecc.



# Gli attori coinvolti

# “Stop Trying to Fix the User”




**Bruce Schneier**
Harvard University

## Stop Trying to Fix the User

IEEE Security & Privacy Sept/Oct 2016

Every few years, a researcher replicates a security study by littering USB sticks around an organization’s grounds and waiting to see how many people pick them up and plug them in, causing the autorun function to install innocuous malware on their computers. These studies are great for making security professionals feel superior. The researchers get to demonstrate their security expertise and use the results as “teachable moments” for others. “If only everyone was more security aware and had more security training,” they say, “the Internet would be a much safer place.”

Enough of that. The problem isn’t the users: it’s that we’ve designed our computer systems’ security so badly that we demand the user do all of these counterintuitive things. Why can’t as a way to bypass the system completely—effectively falling back on the security of their email account.

And finally: phishing links. Users are free to click around the Web until they encounter a link to a phishing website. Then everyone wants to know how to train the user not to click on suspicious links. But you can’t train users not to click on links when you’ve spent the past two decades teaching them that links are there to be clicked.

We must stop trying to fix the user to achieve security. We’ll never get there, and research toward those goals just obscures the real problems. Usable security doesn’t mean “getting people to do what we want.” It means creating security that works, given (or despite) what people do. It means security solutions that

# Usabilità


*A chi non à mai capitato di spingere una porta invece di tirarla o di rinunciare a lavarsi le mani perché non riesce ad azionare il rubinetto?*

*In questi casi la sensazione di incapacità personale è molto forte.*

*Ma **la colpa** non è dell'utente, bensì **di chi ha progettato** questi oggetti d'uso comune **senza considerare le normali attività mentali la cui conoscenza è essenziale per la progettazione di un ambiente ben organizzato e rispondente alle esigenze della mente**.*

# 152 Simple Steps to Stay Safe Online:

## Security Advice for Non-Tech-Savvy Users

Robert W. Reeder, Iulia Ion, and Sunny Consolvo | Google

**IEEE Security and Privacy -** September/October 2017

***Users often don't follow expert advice*** *for staying secure online, but the reasons for users' noncompliance are only partly understood.*

***More than 200 security experts were asked for the top three pieces of advice*** *they would give non-tech-savvy users.*

***The results suggest that****, although individual experts give thoughtful, reasonable answers,* ***the expert community as a whole lacks consensus****.*

# Social Vulnerability Assessment

- Gli attacchi informatici di **data breach** hanno sempre più spesso la loro **porta d'accesso** attraverso l'inganno di chiunque interagisca con un computer
  - si trovano all'interno del **perimetro di fiducia** dell'azienda
  - possono dare fiducia a un **contenuto credibile** ma in realtà **fraudolento**
- Come difendersi?
  - in fase di progettazione dei sistemi
  - semplicità d'uso dei sistemi
  - esercitazioni periodiche sulla prevenzione da attacchi informatici

Dogana

ADVANCED SOCIAL ENGINEERING AND VULNERABILITY ASSESSMENT FRAMEWORK

# Conclusioni

# Inganno

- L'abbondanza di sistemi **connessi** e la *leggerezza* con cui spesso li si configura, rende **vulnerabile** l'intero sistema
- Quando i sistemi prevedono regole stringenti per accesso e controlli sul comportamento degli utenti…

  …violare il sistema richiede entrare in **cerchi della fiducia** sempre più **ristretti**, sia per dimensione, sia per le regole per essere accettati
- Attaccare si può tradurre con ***ingannare***
  - persone
  - software

**L'Intelligenza Artificiale è sia parte del problema, sia della soluzione**

# IA e Cybersecurity

- **Lo spazio virtuale consente a strumenti di IA di costruire contenuti ingannevoli**
  - più facile confondere contenuti artificiali con contenuti reali

- **IA consente di realizzare strumenti di difesa efficaci**
  - ma possono essere a loro volta ingannati

- **Necessità di nuovi paradigmi**
  - trasparenza degli algoritmi di IA
  - adversarial learning


EUROPEAN
CYBER
SECURITY
MONTH