mese europeo della sicurezza informatica

convegno

6 cyber sicuro security 2018

Cyber Security e tecnologie intelligenti nella PA: dal machine learning all'intelligenza artificiale

Roma, 28 Novembre 2018 h. 9.00-13.30

Nuovi Paradigmi basati sull'intelligenza artificiale nella protezione degli asset informatici

Fabio Sammartino

Cybersecurity Expert, Kaspersky Lab Italia

organizzato e promosso da

MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

SAPIENZA UNIVERSITÀ DI ROMA DIPARTIMENTO DI INFORMATICA

PRS Planning Ricerche e Studi

Agenda

- Le minacce informatiche di nuova generazione: come siamo arrivati allo stato attuale
- Come le nuove tecnologie e le nuove minacce impattano utenti e infrastrutture
- Perché il Machine Learning è indispensabile e quali sono le difficoltà di applicazione nella Cybersecurity

Evoluzione delle minacce informatiche

1994

1

NEW VIRUS EVERY HOUR

2006

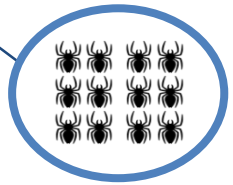
1

NEW VIRUS EVERY MINUTE

2011

1

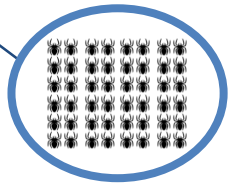
NEW VIRUS EVERY SECOND



2018

360,000

NEW SAMPLES EVERY DAY



Numeri delle minacce 2017

www.kaspersky.com

KASPERSKY lab

Kaspersky Lab:
The Big Numbers of 2017

Online threats

Information for : **2017** | 2016

A billion malicious online attacks: 1 billion ↑ | 758 million

15,714,700 unique malicious objects (scripts, exploits, executable files, etc.) detected by Kaspersky Lab's web antivirus in 2017

22% of computers where our web antivirus was triggered hit by advertising programs and their components

88% of attacks originated in 10 countries

Ransomware

More than **96,000** modifications of crypto-ransomware detected

38 new families discovered

939,722 unique KSN users attacked by encryptors, including

>240,000 corporate users

Banking malware

1,126,701 devices saw attempted attacks to launch malware capable of stealing money via online banking

Applications most targeted by exploits

MS Office: 17.6% ↑ | 13%

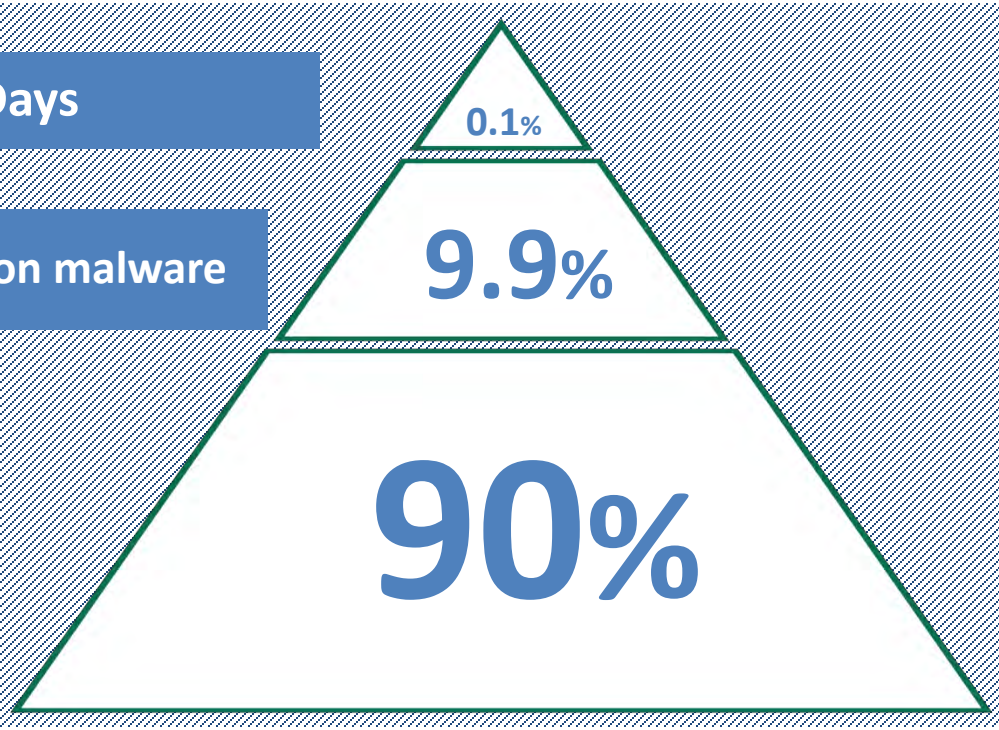
Adobe Flash: 4.5% ↓ | 8%

Fabio Sammartino
Kaspersky Lab

Tassonomia delle minacce

APTs ; 0-Days

Attacchi Non malware



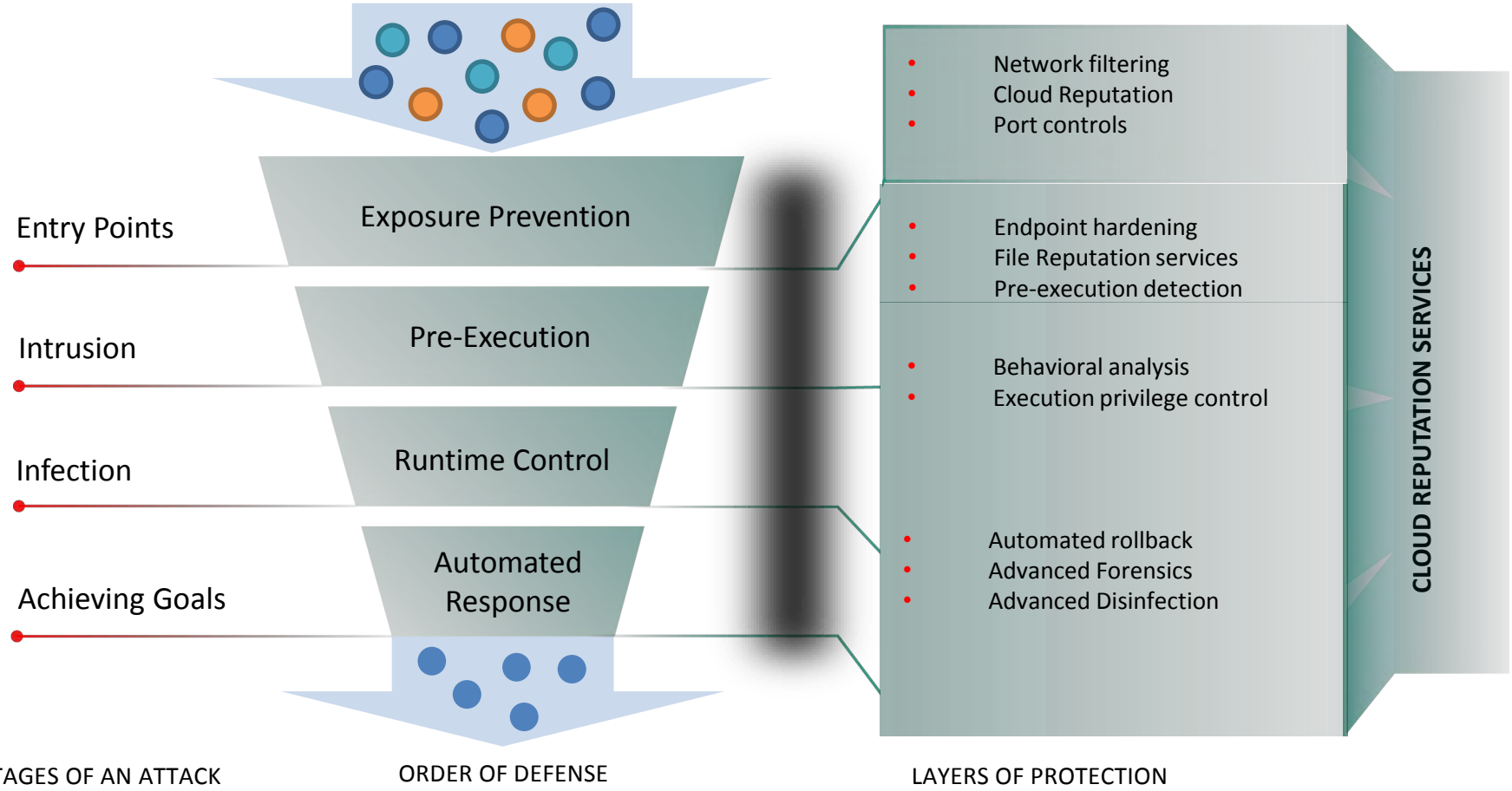
Cyber-weapons

Attacchi Mirati

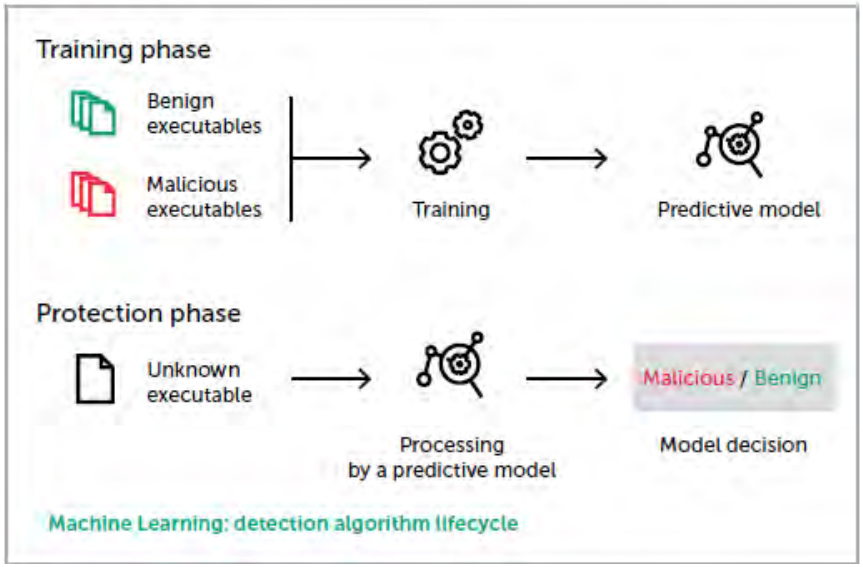
Cybercrime Tradizionale

- Maggiore complessità delle minacce informatiche dovuta a:
  - Business model efficienti aperti a criminali non avanzati tecnicamente
  - Malware avanzato reperibile in vendita o sottratto a grandi organizzazioni
  - Utilizzo di tecniche proprie degli attacchi mirati per attacchi su larga scala
  - Semplicità di anonimizzazione (Tor – Bitcoin – Darkweb)
  - Utilizzo di minacce non sofisticate abbinate a tecniche di Social Engineering

Catena di attacco e tecnologie di protezione Endpoint



- Unsupervised Learning
  - Ha lo scopo di individuare la struttura dei dati o la legge di generazione del dato
  - Apprendimento basato su data set privi delle risposte corrette
- Supervised Learning
  - Ha lo scopo di individuare il modello che produrrà la risposta corretta per i nuovi oggetti
  - Viene usato quando sono disponibili sia i dati che le risposte giuste
  - Si compone di due stadi:
    - Apprendimento e adattamento di un modello ai dati disponibili
    - Applicazione del modello ai nuovi sample e ottenimento di predizioni
- Deep Learning
  - Approccio che facilita l'estrazione di funzionalità ad un alto livello di astrazione da dati di basso livello (Ex. spech recon, face recon.)
  - Utilizzato per identificare malware dai dati di basso livello (Ex. Gerarchie di funzioni, detection multi stadio…)



- Necessità di un Dataset molto ampio
  - La necessità è di addestrare il nostro modello attraverso un dataset rappresentativo delle condizioni reali in cui dovrà operare.
- Trained Model deve essere interpretabile
  - Solitamente i modelli di ML sono delle Black Box non interpretabili dagli umani, questo non va bene in cybersecurity perché non permette l'analisi di una serie di casi (ex. Falsi positivi)
- False Positive Rate pari a zero
  - Necessità non comune nel ML, che richiede di impostare requisiti molto alti nella fase di training, sia per il modello che per le metriche.
  - Il modello deve permettere la correzione dei FP «on the fly» in caso di rilevamenti errati basati su dati sconosciuti, senza dover addestrare di nuovo il modelloGli algoritmi devono potersi adattare alle reazioni dei Cybercriminali – distribuzione dei dati variabile
  - Gli avversari scrivono nuovi malware con nuove tecniche
  - Migliaia di Software House producono nuovi tipi di eseguibili benigni

Kaspersky Lab Machine Learning Multi Livello

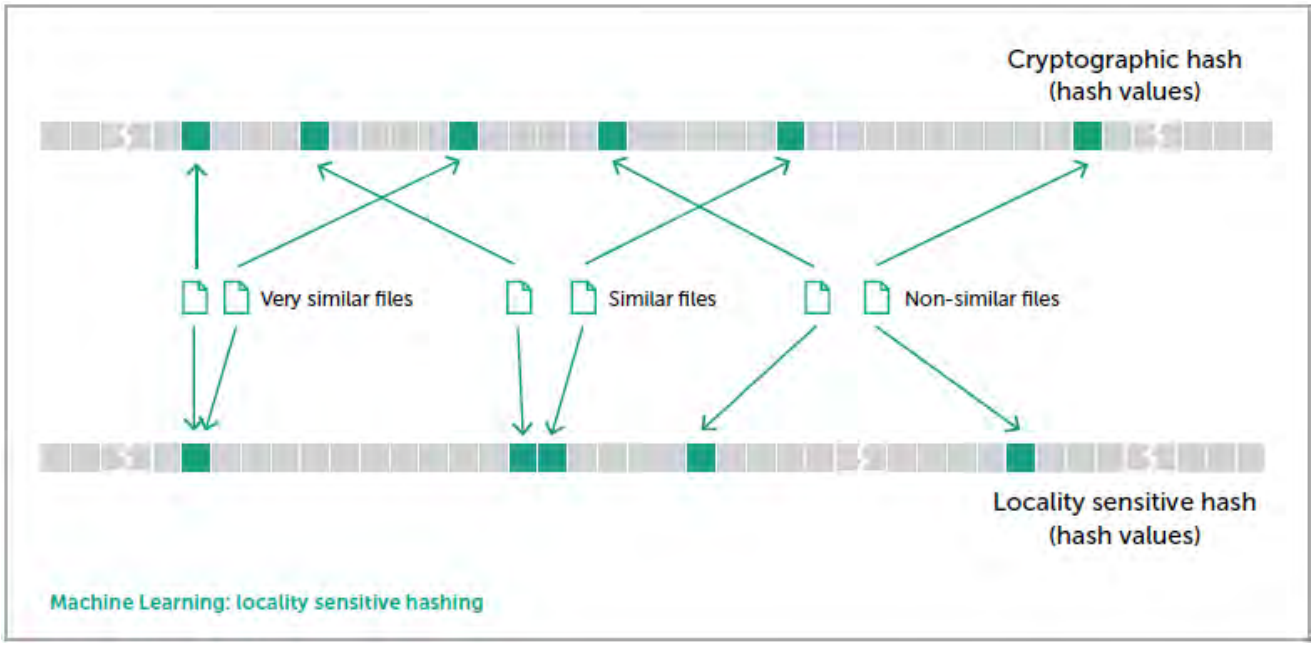
	IN CLOUD	ON CLIENT
PRE EXECUTION	**Astraea** Big Data Expert system for metadata aggregation **ML Cloud requests w/o a model on client** ML	**Decision tree ensemble** ML Built by gradient boosting technique **Behavioral** ML Expert by emulation logs **Locality-Sensitive Hashing** ML Built by orthogonal projection learning **Cure and detection routines** Expert
POST EXECUTION		**Behavioral** Expert by execution logs **Behavioral Model** ML Deep learning utilizing execution logs

ML — Content is generated automatically by machine-learning techniques

Expert — Content is generated by experts

Big Data — Suspicious files' metadata from millions of endpoints collected and processed

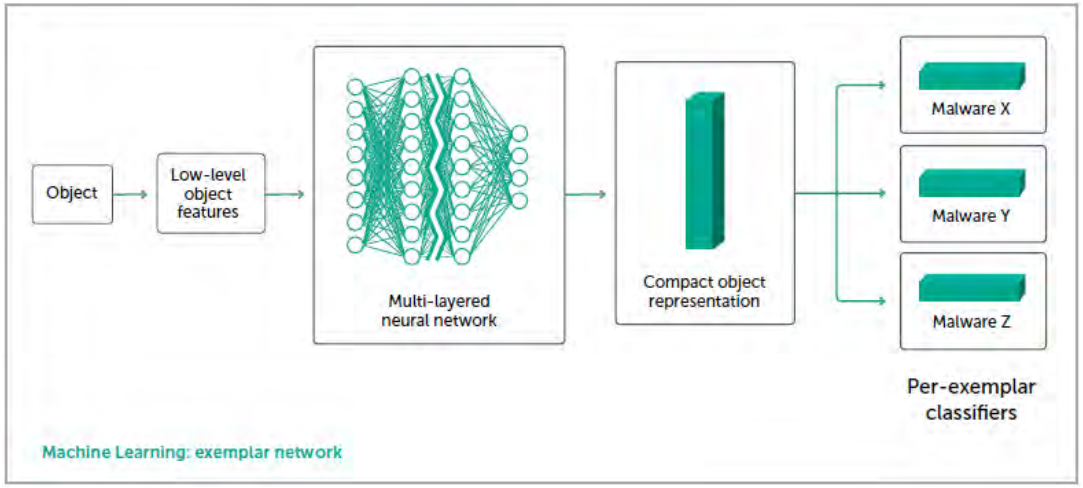
- Pre-Execution:
  - Similarity Hashing (LSH) basato sulla struttura del file, molto utile per identificare varianti di varie tipologie di malware.
  - Viene utilizzato in combinazione con altri algoritmi in uno schema a due stadi, allo scopo di ridurre il carico computazionale sull'endpoint e limitare i falsi positivi.



Deep Learning contro gli attacchi mirati:

- Nel caso di un sample singolo viene usato l'approccio **Exemplar Network (ExNet)**

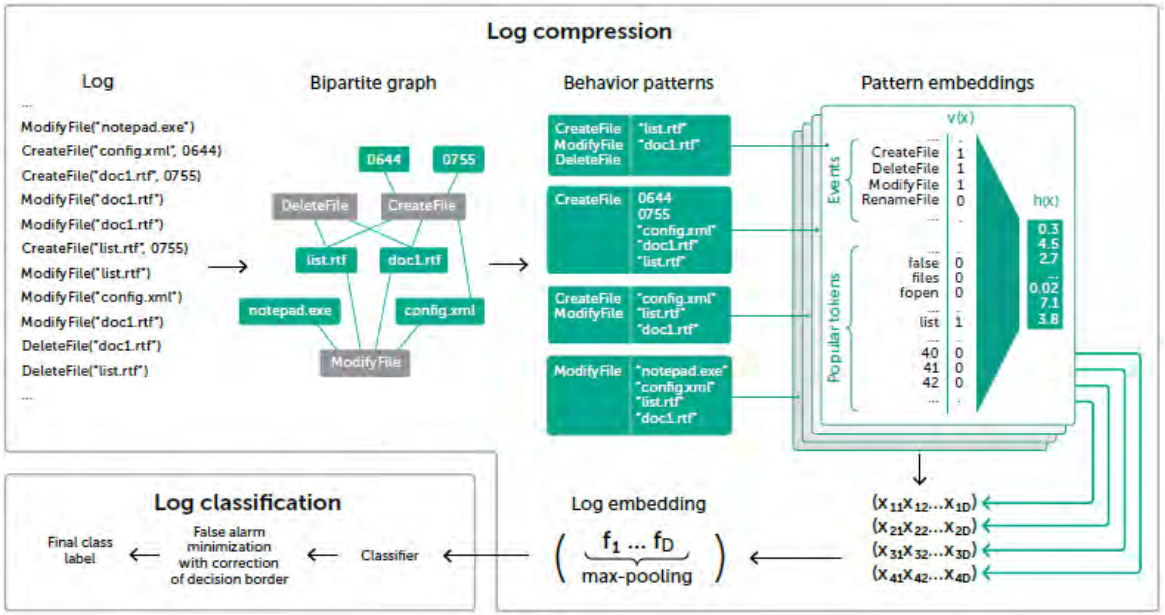
Permette di combinare più passaggi (estrazione di funzionalità, rappresentazione compatta e creazione di un modello per esemplari) **in una singola pipeline con le discriminanti di vari tipi di malware**



Deep Learning in Post-Execution:

- Log forniti dai motori di analisi comportamentale
- **Il modello comprime la sequenza di eventi in un set di vettori binari che vengono forniti ad una rete neurale per distinguere i comportamenti leciti da quelli pericolosi**

Permette di addestrare una rete neurale capace di operare con concetti comportamentali di alto livello. **Può adattarsi a diversi ambienti utente** e incorpora funzionalità di correzioni di falsi allarmi by design.



- Avere i dati giusti
- Conoscere la teoria del ML e come applicarla in cybersecurity
- Conoscere i bisogni degli utenti e essere esperti nella tecnologia
- Avere una giusta quantità di dati
- Applicare un approccio multi livello per il rilevamento delle minacce