la rilevazione automatica dei problemi: IDS e affini

applicabilità di ciò che segue

- log/event audting
- Network Intrusion Detection Systems
- Host Intrusion Detection Systems
- Intrusion Prevention Systems
- antivirus

l'automatismo necessario

rilevare problemi richiede…

- scandagliare una grande quantità di dati
  - pacchetti di rete
  - "eventi" o righe di log
    - anche da migliaia di fonti distinte
  - system call
    - normalmente non loggate perché troppe e troppo veloci
- trovare le parti "problematiche"

le parti problematiche?

- la “definizione” di parte problematica è…
  - soggettiva
  - può variare nel tempo
    - per le esigenze dell’organizzazione
    - per nuovi attacchi
    - per cambiamenti delle apparecchiature o del business
  - non essere chiara all’organizzazione o a chi si occupa di sicurezza

approcci

- rule-based
  - regole formalizzate in qualche linguaggio
    - es. espressioni regolari
  - matching automatico

- anomaly-based
  - apprendimento di un comportamento “normale”
    - tecniche di AI o statistiche
  - rilevazione di comportamenti “anomali”

il grande problema dei “falsi”

- **falsi positivi**: il sistema **rileva** anomalie che non esistono
  - noiosi
  - costosi
  - misurati in % su totale degli allarmi forniti
- **falsi negativi**: il sistema **NON rileva** anomalie che dovrebbe rilevare
  - **pericolosi!**
  - misurati in % sul totale delle anomalie... **molto difficili da misurare!**
- tutti i sistemi possono sbagliare dando falsi positivi o negativi
  - idealmente vorremmo 0% di falsi positivi e negativi

rule-based

- tutti gli strumenti basati su regole vanno “tarati” per le esigenze del caso
  - aggiungendo e togliendo regole
- la taratura...
  - richiede risorse umane
    - la taratura viene fatta da un “amministratore”
  - complessa e quindi può essere errata
  - si basa sull'esperienza e quindi richiede tempi lunghi
    - il sistema non sarà efficace sin da subito
- **molto più facile ridurre i falsi positivi che i falsi negativi!**

rule-based

- l'aggiornamento del set di regole può essere automatica
- alcuni sistemi sono associati ad un servizio in cloud e non sono configurabili esplicitamente
  - semplicità
  - riduzione costi
  - difficile stabilire il comportamento e l'efficacia

anomaly-based

- generalmente hanno due modalità operative
  - **learning mode**: per apprendere il comportamento normale
  - **detection mode**: per rilevare anomalie
- rappresentazione interna non accessibile
  - potrebbe non essere “esplicita”, es. bayesian network
- impossibile effettuare tarature

anomaly-based

- learning mode: input deve essere rappresentativo
  - non deve contenere attacchi!
  - deve contenere tutti i casi “normali”
- quando il sistema cambia comportamento deve essere rilanciato il learning
- *continuous learning*
  - il sistema si adatta lentamente ai cambiamenti
  - rilevato solo l’inizio dell’anomalia, poi il sistema assorbe l’anomalia nel comportamento normale appreso
  - non rileva attacchi molto “lenti”

anomaly-based

- difficile da adottare in pratica in ambienti IT standard
  - difficile da prevedere e correggere
- adottabile in ambienti molto statici
  - es. sistemi di controllo industriale, sistemi militari

approccio ibrido

- una combinazione dei due precedenti
- la combinazione può essere fatta in vari modi
  - anomaly-based con possibilità di inibire certi allarmi mediante regole (per evitare falsi positivi)
  - unione di allarmi derivante dall'approccio rule-based e anomaly-based
    - per rilevare anomalie non presenti nelle regole
    - per avere un livello maggiore di "certezza" quando si attiva una regola

parametri di qualità/prestazioni

- tempestività nella rilevazione dell'intrusione
- accuratezza nella rilevazione dell'intrusione
  - nessuno o pochi falsi positivi
  - nessuno o pochi falsi negativi (più pericolosi!)
- semplicità di configurazione
- eventuale risposta immediata all'attacco
- throughput massimo
  - ... senza che l’elevato carico introduca maggiori falsi negativi

la comparazione è un problema

dati due sistemi di rilevazione automatica dei problemi dire **quale dei due è "migliore"**

- spesso non si conoscono…
  - i metodi di funzionamento interno
  - le regole o gli algoritmi sottostanti
- non esiste metodologia semplice per fare una comparazione black-box
  - c'è un problema di copertura dello "spazio delle anomalie"
  - le anomalie importanti domani possono non essere note oggi