Sicurezza dei sistemi e delle reti[1]

Mattia Monga

Dip. di Informatica
Università degli Studi di Milano, Italia
mattia.monga@unimi.it

a.a. 2015/16

Lezione IX: Falsi allarmi

Falsi allarmi

In generale in tutti gli IDS (misuse e anomaly) occorre bilanciare

falsi negativi attacchi non rilevati

falsi positivi attacchi rilevati corrispondenti a situazioni normali

Falsi allarmi

- Quanto piú la rilevazione è **specifica** (es. firme molto dettagliate) tanto piú aumenta il carico computazionale e la rilevazione diventa sensibile a variazioni dell'evento analizzato.
- Quanto piú la rilevazione si fa **lasca** (es. firme generiche) tanto piú il carico computazionale cala, la rilevazione dell'evento analizzato risulta poco influenzata da varianti ma tanto piú vengono rilevati eventi simili ma non pericolosi.

Relazione fra FP e FN

FP e FN risultano correlati inversamente: agendo per diminuire l'una, tipicamente l'altra aumenta.
Il problema si ripropone in moltissime discipline (information retrieval, farmacologia, ...): ogni volta che si ha una decisione binaria (test)

	positivo	negativo
attacco	TP	FN
non attacco	FP	TN

$$TP + TN + FP + FN = totale$$

Relazione fra FP e FN

FP Type I error, falso allarme

FN Type II error, miss

sensibilità del test, recall, hit rate, TPR $\frac{TP}{TP+FN}$

specificità del test $\frac{TN}{TN+FP}$

accuratezza del test $\frac{TP+TN}{totale}$

precisione del test $\frac{TP}{TP+FP}$

FPR $\frac{FP}{TN+FP} = 1 - \text{specificità}$

Quali regole di IDS hanno lavorato meglio?

A (FPR:0.28,TPR:0.63)	allarme	¬allarme
attacco	TP=63	FN=37
¬attacco	FP=28	TN=72

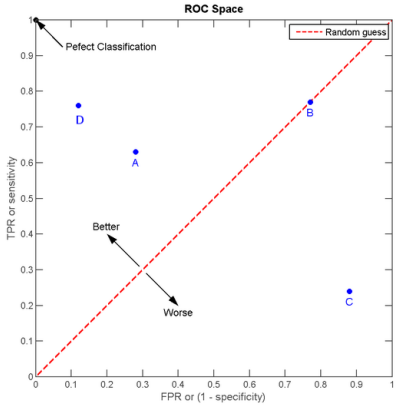
B piú sensibile(0.77,0.77)	allarme	¬allarme
attacco	TP=77	FN=23
¬attacco	FP=77	TN=23

C (0.88,0.24)	allarme	¬allarme
attacco	TP=24	FN=76
¬attacco	FP=88	TN=12

D piú specifico, accurato, preciso(0.12,0.76)	allarme	¬allarme
attacco	TP=76	FN=24
¬attacco	FP=12	TN=88

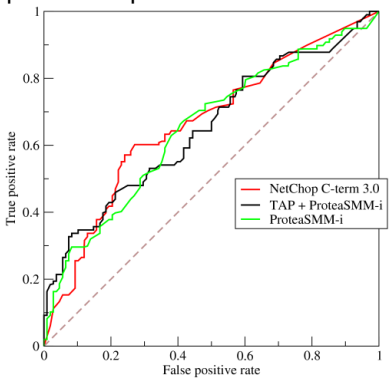
ROC

receiver operating characteristic (ROC): sensibilità vs. tasso dei falsi positivi



E un insieme di regole?

E come valutare l'efficacia di un insieme di regole per tutti i parametri possibili? A volte si usa l'Area under curve.



Riassumendo

Il problema degli IDS sono i falsi allarmi e le mancate segnalazioni

- la qualità di un IDS sta nella relazione fra FP e FN
- per valutarla ci si serve di ROC e AUC

I falsi allarmi

Gli strumenti di analisi come ROC permettono di valutare l'efficacia *ex-post*, valutando il peso di FP e FN in un determinato contesto sperimentale.
Un IDS genera migliaia di allarmi al giorno: qual è la probabilità che un allarme sia davvero relativo a un attacco?
L'intuito inganna perché dipende in maniera complessa dalla **probabilità a priori di un attacco**.

Esempio dalla letteratura medica

La probabilità che una donna sviluppi un cancro al seno è 0.8%. Se una donna **ha** il cancro al seno, la probabilità che il suo mammogramma sia positivo è 90%; se **non ha** il cancro al seno, c'è comunque una probabilità del 7% che il mammogramma sia positivo. Se il mammogramma di una donna è positivo, qual è la probabilità che abbia effettivamente il cancro al seno?

Esempio da "Quando i numeri ingannano", di G. Gigerenzer: studi su medici mostrano che la risposta piú frequente al problema cosí posto è 90%.

Esempio: soluzione

- Su 1000 donne, 992 sono sane e 8 malate.
- Delle 8 malate, il 90% ($\simeq 7$) risulteranno positive e il 10% negative ($\simeq 1$).
- Delle 992 sane, il 7% ($\simeq 70$) risulteranno positive e il 93% negative ($\simeq 922$).
- Le positive saranno quindi $7 + 70 = 77$, delle quali sono malate 7: la probabilità che un mammogramma positivo sia indice di malattia è quindi $\frac{7}{77} = 9\%$

Teorema di Bayes

Teorema di Bayes

$$\Pr(\text{attacco}|\text{allarme}) = \begin{cases} \frac{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco})}{\Pr(\text{allarme})} \\ \frac{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco})}{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco}) + \Pr(\text{allarme}|\neg\text{attacco}) \cdot \Pr(\neg\text{attacco})} \\ \frac{\frac{\text{TP}}{\text{TP+FN}} \cdot \Pr(\text{attacco})}{\frac{\text{TP}}{\text{TP+FN}} \cdot \Pr(\text{attacco}) + \frac{\text{FP}}{\text{FP+TN}} \cdot \Pr(\neg\text{attacco})} \end{cases}$$

Per calcolare la probabilità di un allarme veritiero occorre sempre stimare **la probabilità a priori di un attacco,** che è spesso (fortunatamente!) piuttosto bassa: **i falsi allarmi** sono inevitabilmente comuni, a meno di avere un IDS straordinariamente preciso o asset particolarmente appetibili.

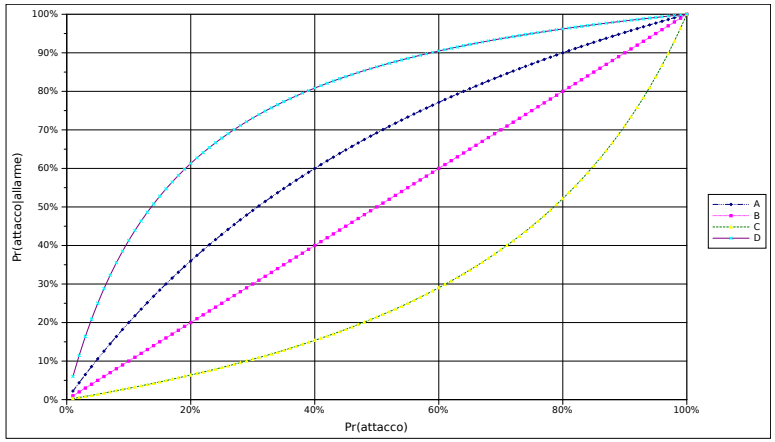
Esempio IDS

Riprendiamo il migliore IDS esaminato con la curva ROC

D (0.12,0.76)	allarme	¬allarme
attacco	TP=76	FN=24
¬attacco	FP=12	TN=88

- Nell'esperimento:
  $\Pr(\text{attacco}) = \frac{76+24}{76+24+12+88} = 50\%$
  $\rightsquigarrow \Pr(\text{attacco}|\text{allarme}) = 86\%$
- Attacchi poco frequenti
  $\Pr(\text{attacco}) = 1\%$
  $\rightsquigarrow \Pr(\mathbf{attacco}|\mathbf{allarme}) = 6\%$

Esempi



Riassumendo

- L'amministratore di un IDS è interessato a stimare la probabilità che un allarme sia davvero relativo a un attacco.
- dipende però dalla probabilità *a priori* di un attacco.

Aspetti architetturali

Un NIDS **complementa** altre soluzioni, con un'architettura a diversi livelli (defense in-depth).
Importanti aspetti architetturali:

- Quanti sensori installare nella rete
  - costi e complessità di gestione
- Dove installarli
  - quantità vs. ridondanza di informazioni
- Come gestire i dati
  - analisi e logging centralizzato vs. distribuito

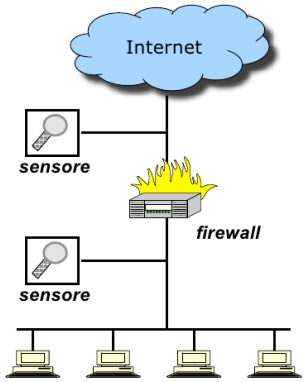
Sensori e firewall

Esterno

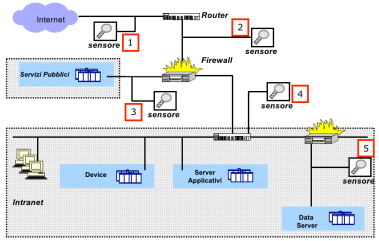
- Rileva l'intero traffico diretto alla rete
- Piú dati
- Piú allarmi

Interno

- Rileva solo il traffico che entra effettivamente
- Verifica l'efficacia del firewall
- Non fornisce info sugli attacchi bloccati dal fw



Posizionamento nella rete aziendale



1. Esterno al border router
   - Tutto il traffico diretto alla rete aziendale.
   - Informazione completa e non filtrata.
   - Tanti dati e allarmi
2. Tra border router e firewall
   - Tutto il traffico meno quello filtrato
   - Tanti dati, molti falsi allarmi.
3. Sulla rete dei servizi pubblici, dietro il firewall
   - Tutto il traffico autorizzato dal firewall e diretto ai servizi pubblici.
   - Possibilità filtraggio mirato.
   - Eventuale traffico illecito dai server pubblici.

Riassumendo

Il posizionamento dell'IDS influisce molto sulla sua efficacia

1. Esterno
2. Tra router e firewall
3. Vicino ai servizi
4. Sulla intranet
5. In punti critici