Sicurezza delle reti[1]

Mattia Monga

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

a.a. 2010/11

Lezione XII: Rilevamento delle intrusioni

I falsi allarmi

Gli strumenti di analisi come la curva ROC ci permettono di valutare l'efficacia di uno strumento di rilevamento delle intrusioni *ex-post*, valutando il peso di falsi positivi e falsi negativi in un determinato contesto sperimentale.
Un IDS genera spesso centinaia di allarmi al giorno: qual è la probabilità che un allarme sia davvero relativo ad un attacco?
La risposta a questa domanda è difficile da dare in maniera intuitiva, perché dipende in maniera complessa dalla probabilità a priori di un attacco.

Esempio dalla letteratura medica

Esempio

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 risulti positivo è del 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?

L'esempio è tratto da "Quando i numeri ingannano", di Gerd Gigerenzer: studi su medici tedeschi e americani 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 ($\simeq 992$).
- 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
  $\Pr(\text{attacco}) = 1\% \rightsquigarrow \Pr(\text{attacco}|\text{allarme}) = 6\%$
- Attacchi molto frequenti
  $\Pr(\text{attacco}) = 80\% \rightsquigarrow \Pr(\text{attacco}|\text{allarme}) = 96\%$

Esempi