

# Sicurezza dei sistemi e delle reti[1]

Mattia Monga

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

a.a. 2014/15

[1] © 2011–15 M. Monga. Creative Commons Attribuzione — Condividi allo stesso modo 4.0 Internazionale. http://creativecommons.org/licenses/by-sa/4.0/deed.it. Derivato con permesso da © 2010 M. Cremonini.





# Lezione XI: Rilevamento delle intrusioni

# Intrusion detection system



**IDS**

Un sistema di monitoraggio (generalmente del tutto passivo) che genera allarmi

Tre fasi:

1. Raccolta dati
2. Analisi dei dati
3. Generazione degli allarmi

# Perché usare un IDS?


In generale i sistemi di monitoraggio sono utili perché:

- le tecnologie di prevenzione degli eventi indesiderati o pericolosi possono fallire
- è utile avere un un meccanismo di segnalazione che permetta di attivare procedure di correzione o di emergenza
- l'uso di strumenti che permettano di monitorare lo stato corrente di un sistema, sia esso un componente che una rete, per accumulare conoscenza statistica sulle modalità d'uso

# HIDS e NIDS

In base al punto in cui avviene la raccolta dati

- **HIDS** (Host-based Intrusion Detection System) sistemi che analizzano informazioni relative all'attività locale di un singolo host (log di sistema, accesso a file critici, ...)
- **NIDS** (Network Intrusion Detection System) sistemi che utilizzano le informazioni raccolte da analizzatori di traffico di rete.

# Rilevazione di eventi critici


L'analisi dei dati raccolti per rilevare una situazione d'allarme:

**Misuse detection** Si caratterizza l'abuso: si rilevano le situazioni che ricadono nella descrizione di un attacco (sono detti anche signature based)

**Anomaly detection** Si caratterizza l'uso normale: si rilevano le situazioni che si scostano dal "normale" funzionamento in modo da poter rilevare anche attacchi ancora sconosciuti

# Misuse detection



Elencare le situazioni illecite

**Signature Detection**

L'amministratore definisce pattern (signature) predefiniti di usi non conformi e il sistema analizza gli eventi monitorati(di rete, di sistema, nei log) rispetto all'elenco di pattern

È la tecnica piú affermata e diffusa

# Problemi



I misuse detection rilevano solo attacchi che corrispondono a schemi noti

- Regole rigide non rilevano attacchi non noti e varianti (a volte l'IDS è cosí rigido che basta cambiare un bit per evadere la rilevazione)
- Piú le regole sono flessibili e piú aumenta la complessità di gestione/configurazione
- l'elenco di firme deve essere adattato alle specificità della rete monitorata

# Anomaly detection

Anomalie rispetto

- ad eventi singoli: esempio azioni "anomale" di un utente rispetto un profilo d'uso predefinito
  es: `http://example.com/##&<>623??%`
- a dati aggregati: tipicamente, deviazioni rispetto a parametri statistici
  es: il traffico con sorgente 123.45.67.88 è di 5GB al minuto

# Anomaly detection

+ Non dipende dalla conoscenza puntuale di tutte le modalità di intrusione
– Molto complesso da realizzare (come fare il modello dell'uso "normale"?) e oneroso da gestire
• Non forniscono informazioni su quale vulnerabilità l'attaccante intende colpire

# Modelli di uso normale



Il sistema monitorato inizialmente durante gli usi normali.

- L'ipotesi di assenza di compromissione e normalità non è facile da verificare: in fase di test si rischia l'anomalia
- Impiego di tecniche come data mining, analisi bayesiana, ecc. . .
- È molto complesso tenere il passo con l'evoluzione del sistema
- il monitoraggio introduce inefficienze maggiori rispetto ai misuse

# Usi consolidati dell'anomaly detection

Ci sono casi in cui l'anomaly detection funziona bene e il loro uso è ormai standard

- hashing dei file (HIDS): l'integrità dei file del sistema controllata con hash da una distribuzione originale (es. Tripwire)
- Protocol Anomaly Detection (NIDS): analizzato il traffico di rete rispetto alle specifiche del protocollo applicativo

# Riassumendo

- HIDS e NIDS
- Misuse e anomaly detection

# 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 | allarme | ¬allarme |
|---|---|---|
| attacco | TP=63 | FN=37 |
| ¬attacco | FP=28 | TN=72 |

| B | allarme | ¬allarme |
|---|---|---|
| attacco | TP=77 | FN=23 |
| ¬attacco | FP=77 | TN=23 |

| C | allarme | ¬allarme |
|---|---|---|
| attacco | TP=24 | FN=76 |
| ¬attacco | FP=88 | TN=12 |

| D | allarme | ¬allarme |
|---|---|---|
| attacco | TP=76 | FN=24 |
| ¬attacco | FP=12 | TN=88 |

# Quali regole di IDS hanno lavorato meglio?

| A | allarme | ¬allarme |
|---|---|---|
| attacco | TP=63 | FN=37 |
| ¬attacco | FP=28 | TN=72 |

| B piú sensibile | allarme | ¬allarme |
|---|---|---|
| attacco | TP=77 | FN=23 |
| ¬attacco | FP=77 | TN=23 |

| C | allarme | ¬allarme |
|---|---|---|
| attacco | TP=24 | FN=76 |
| ¬attacco | FP=88 | TN=12 |

| D piú specifico, accurato, preciso | allarme | ¬allarme |
|---|---|---|
| attacco | TP=76 | FN=24 |
| ¬attacco | FP=12 | TN=88 |

# 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 |

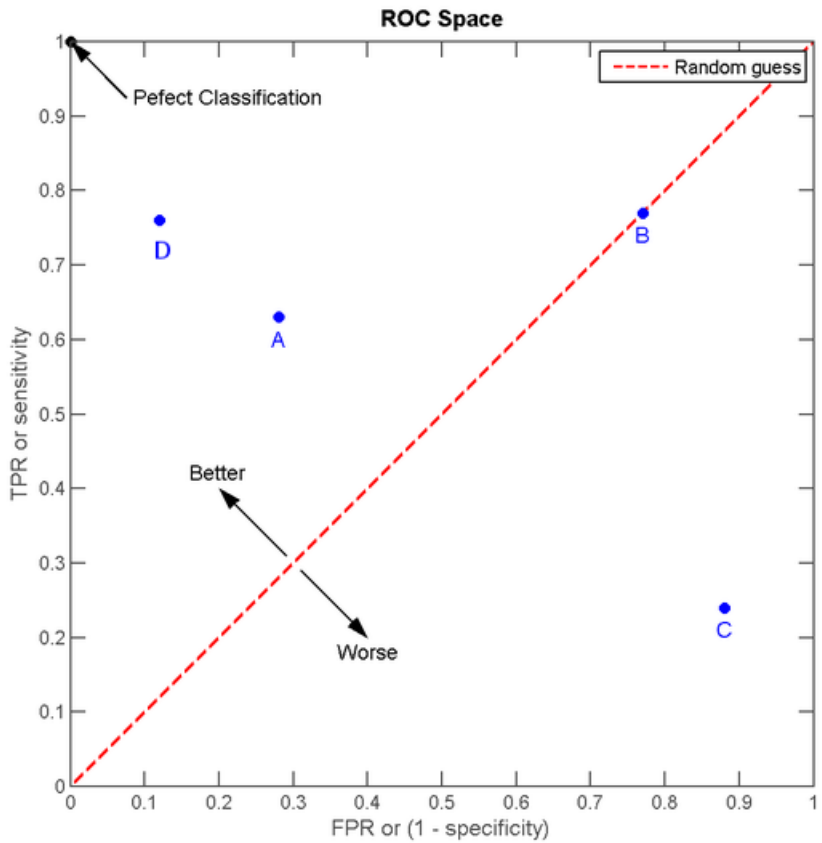
| B (0.77,0.77) | allarme | ¬allarme |
|---|---|---|
| attacco | TP=77 | FN=23 |
| ¬attacco | FP=77 | TN=23 |

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

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

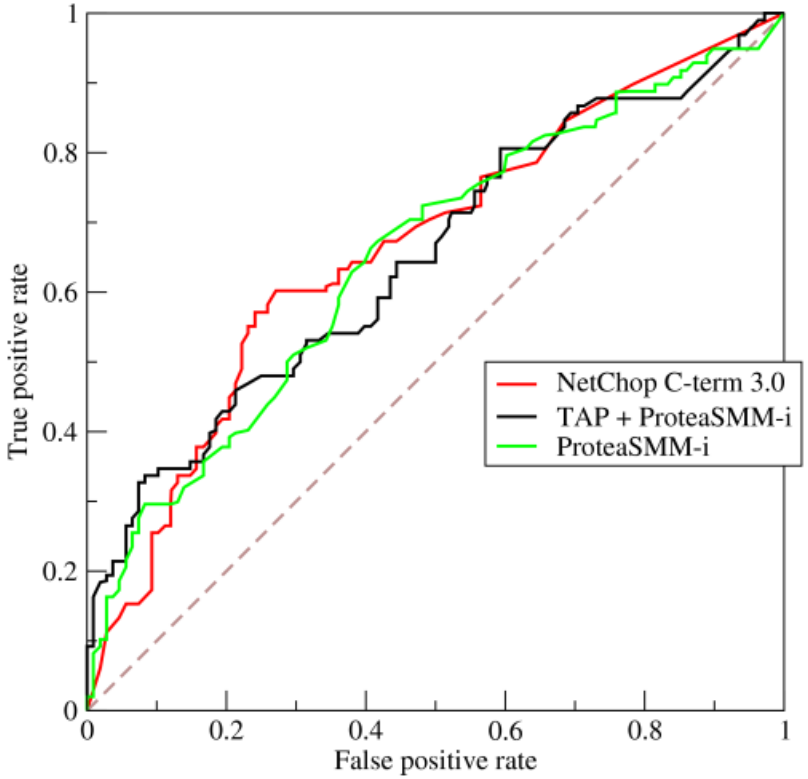
# ROC

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 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

$$\Pr(\text{attacco}|\text{allarme}) = \frac{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco})}{\Pr(\text{allarme})}$$

# Teorema di Bayes

## Teorema di Bayes

$$\Pr(\text{attacco}|\text{allarme}) = \begin{cases} \dfrac{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco})}{\Pr(\text{allarme})} \\ \dfrac{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco})}{\Pr(\text{allarme}|\text{attacco}) \cdot \Pr(\text{attacco}) + \Pr(\text{allarme}|\neg\text{attacco}) \cdot \Pr(\neg\text{attacco})} \end{cases}$$

# Teorema di Bayes



## Teorema di Bayes

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

# Teorema di Bayes

## Teorema di Bayes

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

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.

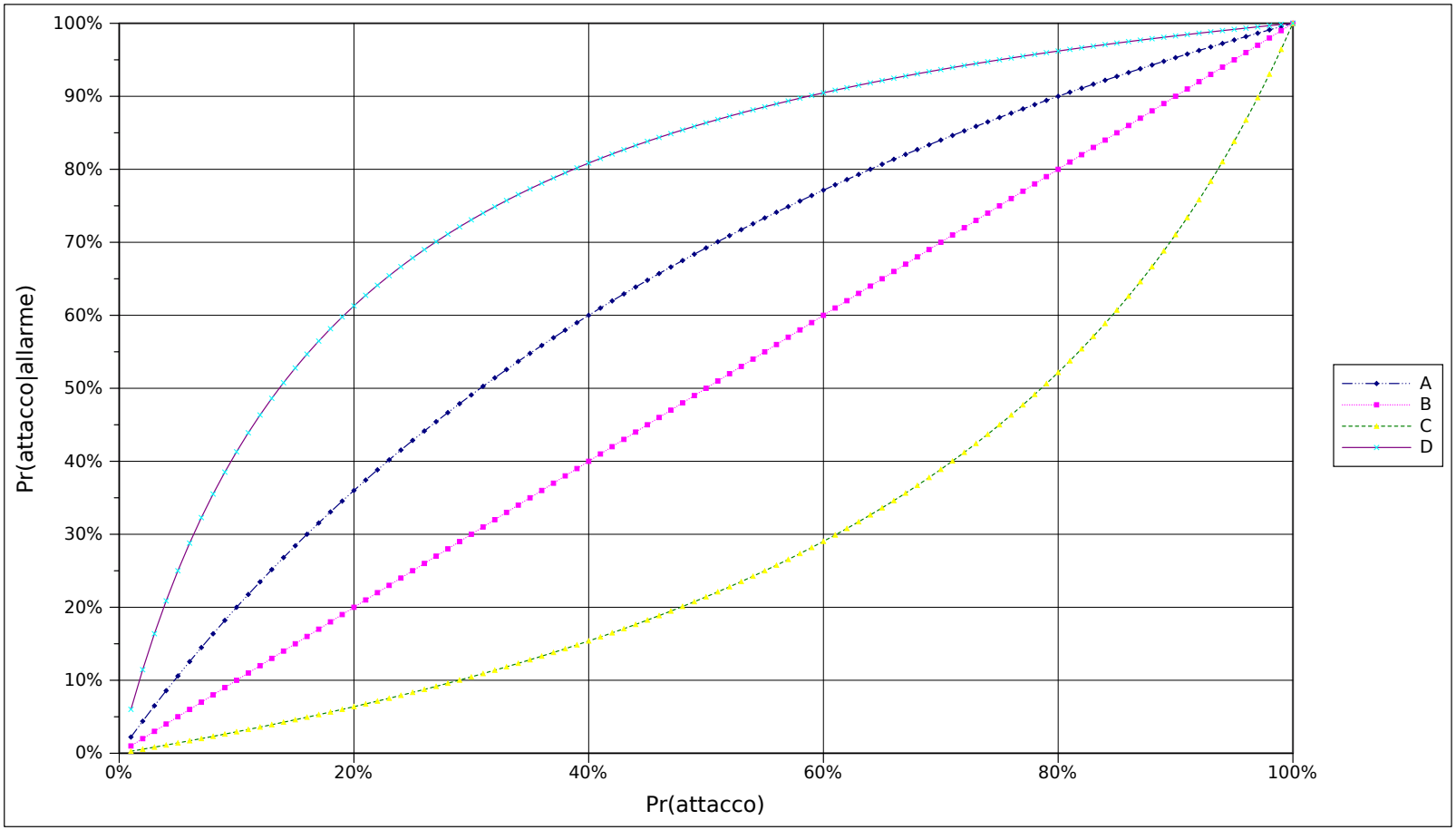
# Esempio IDS

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
  $\Pr(\text{attacco}) = 1\%$
  $\rightsquigarrow \Pr(\text{attacco}|\text{allarme}) = 6\%$

# Esempi


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Pr(attacco|allarme)
0%
20%
40%
60%
80%
100%
Pr(attacco)
A
B
C
D

# 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

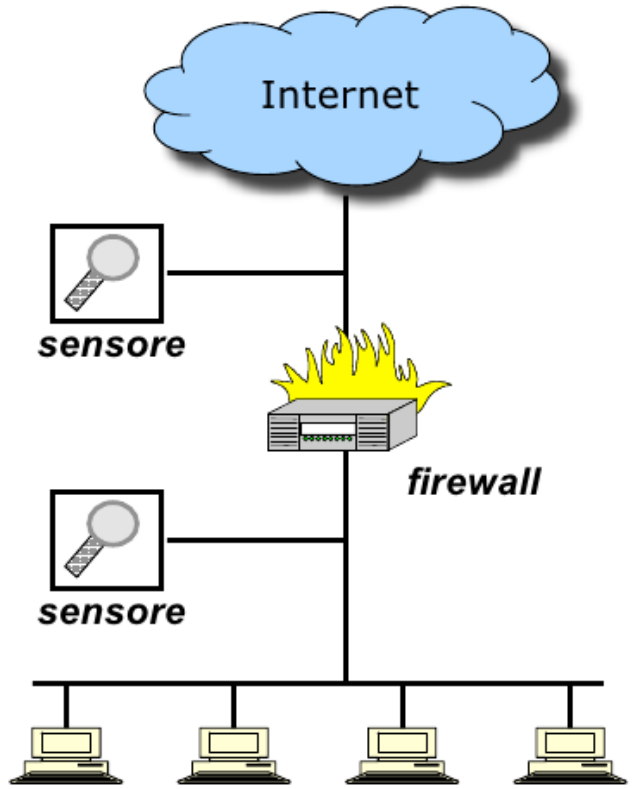
# Sensori e firewall

## Esterno

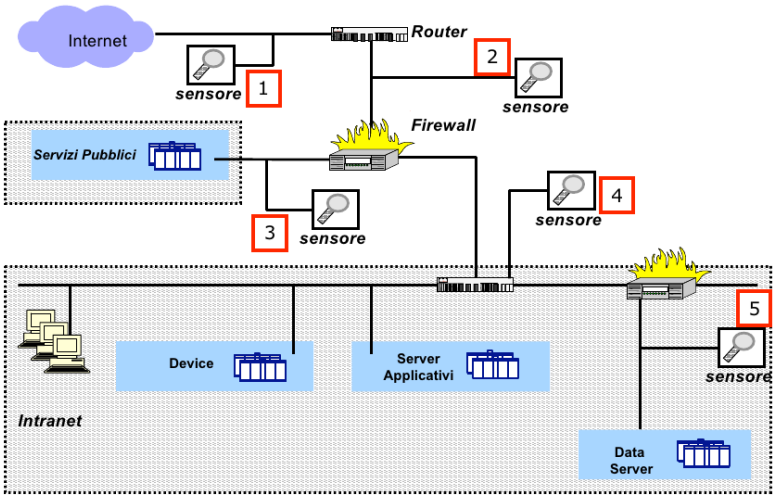
- Rileva l'intero traffico diretto alla rete
- Piú dati
- Piú allarmi

## Interno

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

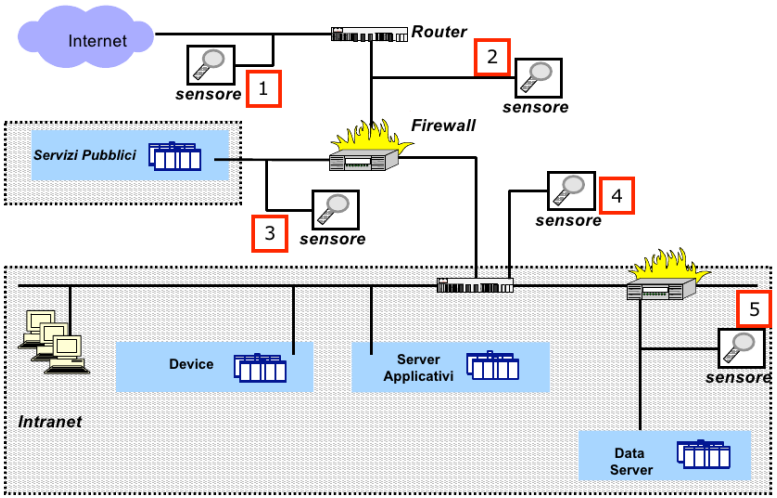


# Posizionamento nella rete aziendale



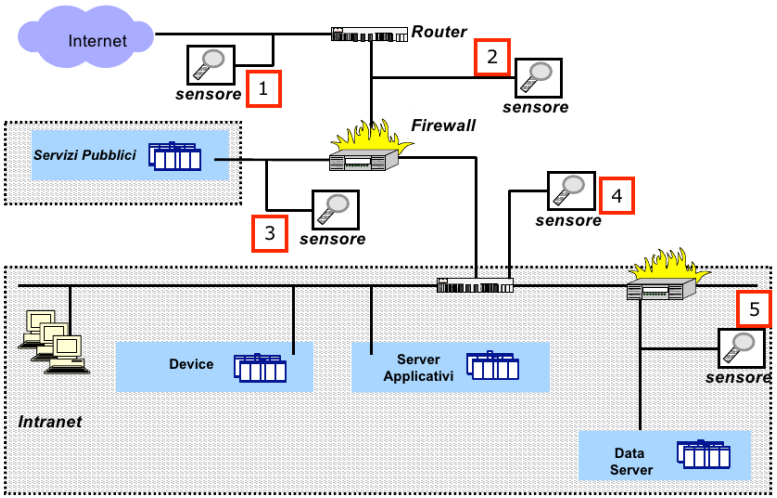

1. Esterno al border router
   - Tutto il traffico diretto alla rete aziendale.
   - Informazione completa e non filtrata.
   - Tanti dati e allarmi

# Posizionamento nella rete aziendale



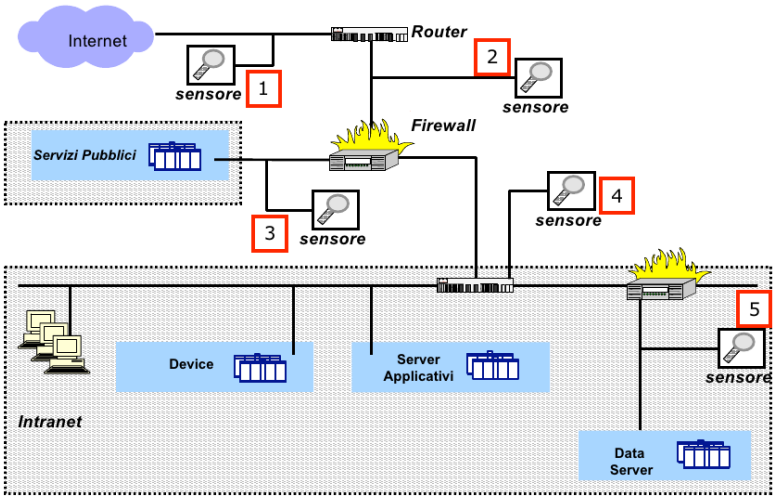

2. Tra border router e firewall
   - Tutto il traffico meno quello filtrato
   - Tanti dati, molti falsi allarmi.

# Posizionamento nella rete aziendale



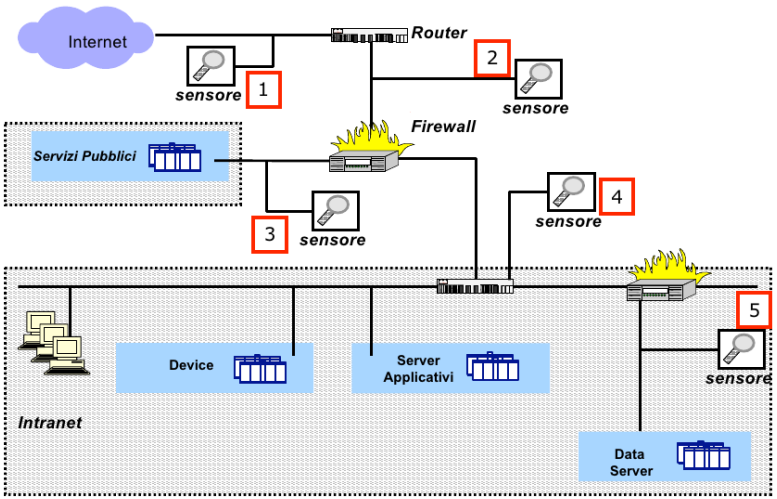

3. Sulla rete dei servizi pubblici, dietro il firewall
   - Tutto il traffico autorizzato dal firewall e diretto ai servizi pubblici.
   - Possibilità filtraggio mirato.
   - Eventuale traffico illecito dai server pubblici.

# Posizionamento nella rete aziendale




4. Sulla Intranet
   - Sia il traffico da reti piú esposte (es. DMZ) che interno alla Intranet.
   - Rileva eventuali usi non leciti interni.
   - Difficile dare firme, molti falsi allarmi.

# Posizionamento nella rete aziendale




5. Su di un segmento critico della rete aziendale

- Monitora le connessioni dirette ad alcune risorse particolarmente critiche della rete aziendale per le quali si richiede un livello di sicurezza piú elevato (es. i server contenenti dati aziendali sensibili).
- Servizi specifici, quindi possibilità di configurazione mirata delle firme.

# 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