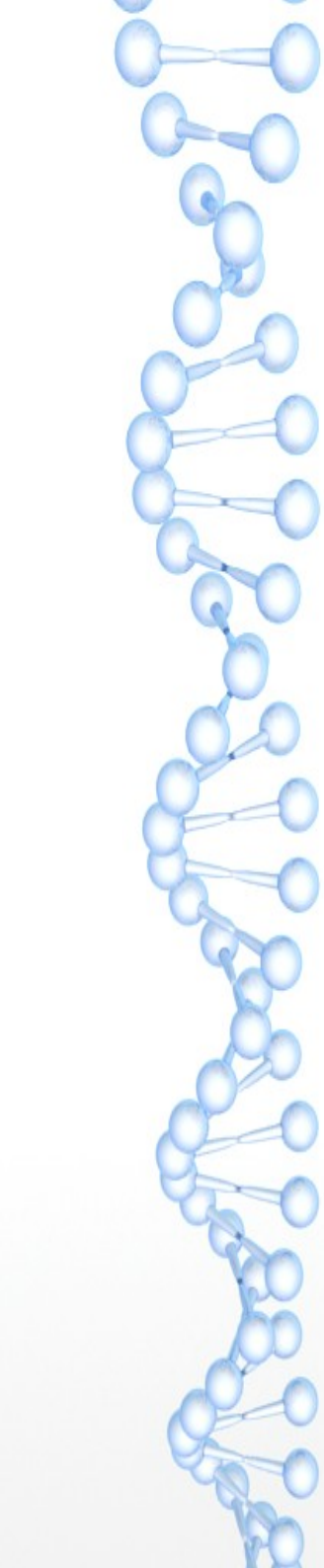


La sicurezza informatica
e
il data mining

Costantina Caruso
Costantina.caruso@uniba.it
27 novembre 2014
aula 2 Palazzo delle Aule
Campus Universitario
Bari

Outline della presentazione

Sicurezza informatica

Il lavoro svolto sinora

Da fare

Sicurezza informatica

Cos'è?

- Disciplina che studia e sviluppa metodologie e tecnologie per la protezione di informazioni e servizi ICT da uso non autorizzato e/o improprio, da distruzione e manomissione, da mancata disponibilità

Che cosa vuole garantire?

- Riservatezza (confidentiality), integrità (integrity), disponibilità (availability)

Come?

- Attuando **misure di controllo/prevenzione** sulle potenziali **vulnerabilità** del sistema per bloccare **minacce** esterne fisiche e logiche

Vulnerabilità

Una vulnerabilità è un punto debole del sistema che potrebbe consentire ad un aggressore di compromettere la sicurezza dell'intero sistema.

Che cosa è candidato ad possedere vulnerabilità?

- I sistemi operativi
  - windows, unix
- Le infrastrutture
  - dialup, voip, wireless, hardware
- Le applicazioni e i dati
  - programmi, web, database, mobile
- L'organizzazione e le persone

Cioè tutto :((((

Minacce

- Attacchi diretti ed indiretti (exploit)
- Denial of service (Dos e DDos)
- Intercettazioni
- Social engineering, errori umani, insider threat
- Esempi di strumenti
  - RAT (Backdoor, Rootkit)
  - Keylogger
  - Malware (virus, worm, trojan, ...)

Misure di controllo/prevenzione

- Security by design
- Protezione fisica
- Crittografia
- Dispositivi di protezione logica (firewall, IDS, IPS, etc.)
- Controllo accessi/autenticazione/autorizzazione
- Policy, informazione/formazione del personale

La sicurezza informatica e il data mining

Concetto 1.

La **sicurezza informatica** è la disciplina che studia e sviluppa metodologie e tecnologie per la protezione di informazioni e servizi ICT da uso non autorizzato e/o improprio, da distruzione e manomissione, da mancata disponibilità

Concetto 2.

Il **data mining** è l'insieme di metodologie e tecniche che hanno per oggetto l'estrazione di nuova conoscenza a partire da grandi quantità di dati (statistica, intelligenza artificiale, basi di dati).

Anomaly detection (data mining) e IDS (sicurezza informatica)

Concetto 3.

In data mining, il **rilevamento delle anomalie** (anomaly detection) è l'identificazione di eventi o osservazioni che non sono conformi a un modello previsto in un insieme di dati.

Concetto 4.

Nella sicurezza informatica, **l'Intrusion Detection System o IDS** è un dispositivo software e hardware utilizzato per identificare accessi non autorizzati (anomali) ai computer o alle reti locali.

L'IDS meglio spiegato

Il controllo della sicurezza delle reti è per lo più affidato agli Intrusion Detection System (IDS). Questi sistemi analizzano le informazioni riguardanti le attività che si svolgono sui computer e che si propagano nella rete, prestando attenzione a comportamenti sospetti.

L'obiettivo di un Intrusion Detection System è quello di analizzare uno o più sorgenti (come workstation, server e segmenti di rete) ed identificare tentativi di attacchi.

Quando viene individuato un attacco, l'IDS produce un allarme che descrive il tipo d'attacco e le entità coinvolte (es. host, processi, utenti).

Questi allarmi vengono consultati da un utente umano che li valuta e intraprende contromisure adeguate.

Caratteristiche principali di un IDS

Anomaly Detection vs. Misuse Detection

Nel primo caso le anomalie sono rilevate confrontando il comportamento di un sistema con un profilo “normale”. Nel secondo caso le anomalie sono rilevate direttamente in base a delle “firme” (signature) di attacchi noti.

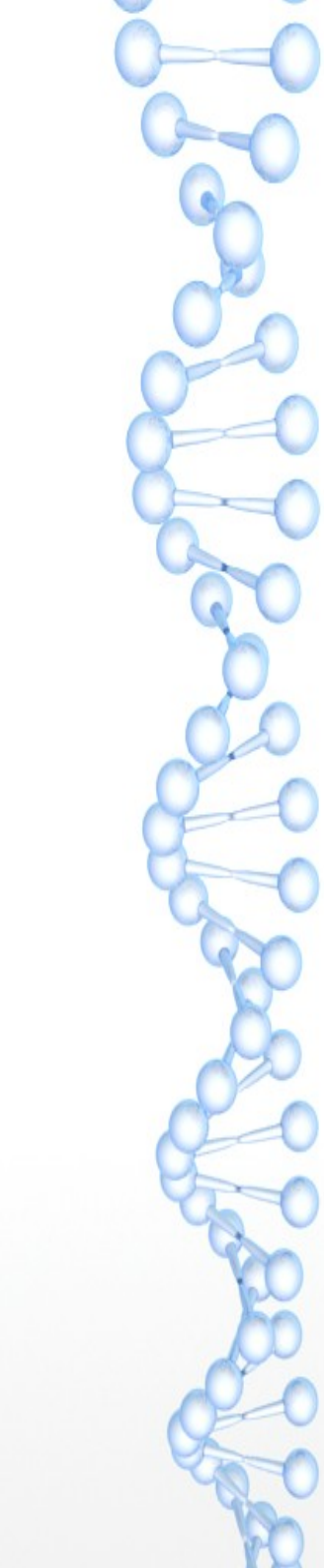
Host-Based vs. Network-Based

I sistemi Host-Based (HIDS) controllano una singola macchina o una singola applicazione, spesso sono costruiti “ad-hoc”. I sistemi Network-Based (NIDS) controllano un’intera rete; di solito sono situati sui gateway.

On-line vs. Off-line

Gli IDS On-line analizzano i dati in tempo reale, durante l’esecuzione dei processi. Quelli Off-line analizzano i dati a posteriori.

Es. Snort (NIDS)



La sicurezza informatica e il data mining: *IDS anomaly based*

1. Aggregazione e correlazione di allarmi di un IDS (NIDS/HIDS)
2. Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie (HIDS)
3. Analisi di file di log di un web server per l'individuazione di attacchi (HIDS)
4. Tecniche di content filtering e applicazione allo spamming (HIDS)
5. Applicazione del local outlier factor per l'intrusion detection (NIDS/HIDS)

Aggregazione e correlazione di allarmi di un IDS
descrizione del problema

1. Quando viene individuato un attacco, l'IDS produce un allarme che descrive il tipo d'attacco e le entità coinvolte.

2**.** Gli allarmi generati da un IDS vengono consultati da un utente umano che li valuta e intraprende contromisure adeguate.

3. **Valutare un attacco e concepire un'appropriata contromisura è risultato un compito difficoltoso perché gli IDS generano migliaia di allarmi al giorno, il 99% dei quali sono falsi positivi (== falsi allarmi)**

4. Questo sovraccarico di falsi positivi rende molto difficile l'individuazione dei messaggi più importanti ovvero allarmi generati da veri attacchi.

5. L'investigazione manuale degli allarmi è molto costosa in termini di tempo e risorse umane e presenta un'alta tendenza agli errori.

Aggregazione e correlazione di allarmi di un IDS
una possibile approccio risolutivo

Causa scatenante di un allarme: motivo per il quale l'allarme viene generato.

Fatti osservabili.

- Numero relativamente piccolo di cause scatenanti predominanti. Queste tendono ad essere persistenti cioè si ripetono finché non vengono rimosse.
- Sono le principali responsabili della generazione dei falsi positivi.
- Gli allarmi che condividono la stessa causa scatenante sono simili.

Principio: raggruppare gli allarmi che hanno simili proprietà strutturali in cluster, assumendo che questi allarmi condividano la stessa causa scatenante

Aggregazione e correlazione di allarmi di un IDS
l'algoritmo proposto

Dato un log d'allarmi, è necessario trovare gruppi (cluster) d'allarmi di grandi dimensioni che possano essere adeguatamente modellati da allarmi generalizzati

L'algoritmo inizia prendendo in input il log *L* e iterativamente generalizza gli allarmi in *L*.

La generalizzazione è eseguita scegliendo un attributo *Ai* e sostituendo il valore che assume in tutti gli allarmi con il suo genitore nella gerarchia *Gi*.

Il processo continua finché si trova uno ed un solo allarme a cui possono essere generalizzati un numero prefissato degli allarmi originali.

Questo allarme costituisce l'output dell'algoritmo.

Aggregazione e correlazione di allarmi di un IDS
esempio pratico e risultati

I dati grezzi

1) ip web server & DestPort=80
2) ip smtp server & DestPort=25
3) 192.168.30.3-192.168.30.12

Generalizzazione

Tutte le occorrenze di 1) nel log dell'IDS verranno sostituite da WEBSERVERLAN1.
Tutte le occorrenze di 2) nel log dell'IDS verranno sostituite da SMTPSERVERLAN1.
Tutte le occorrenze di 3) nel log dell'IDS verranno sostituite da LAN1.

Dopo la generalizzazione, si raggruppano gli allarmi simili al fine di isolare e riconoscere la causa scatenante e ridurre il numero totale di elementi da analizzare e da visualizzare al security officer.

Risultato: la riduzione ottenuta è spesso dell'ordine di 10^3

Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie

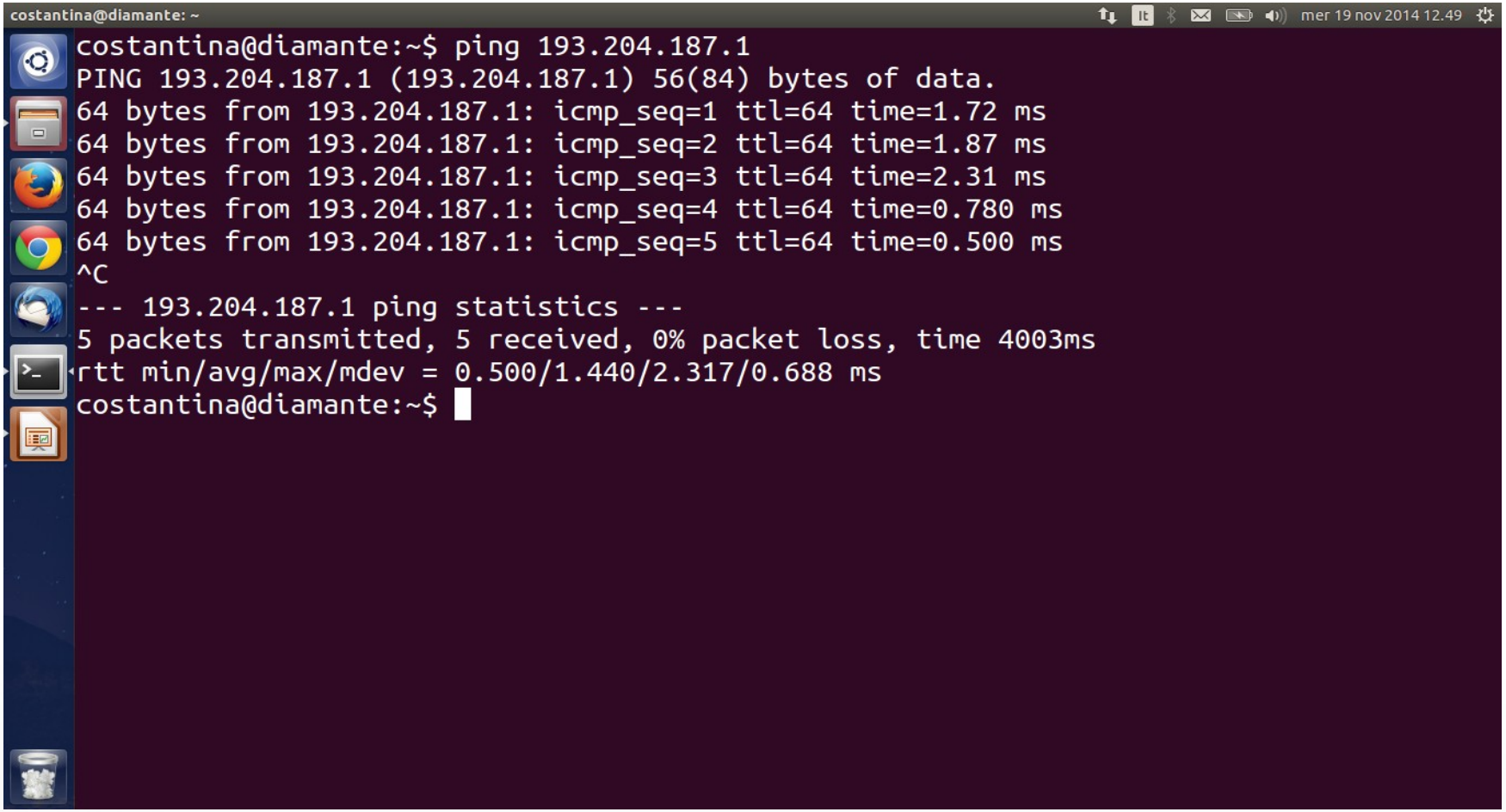
il problema affrontato

1. Ogni comando eseguito da un server chiama sempre le stesse routine di sistema (syscall)

2. Se un server viene hackerato, il comando viene “forzato” ad eseguire una lista di syscall differente

Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie

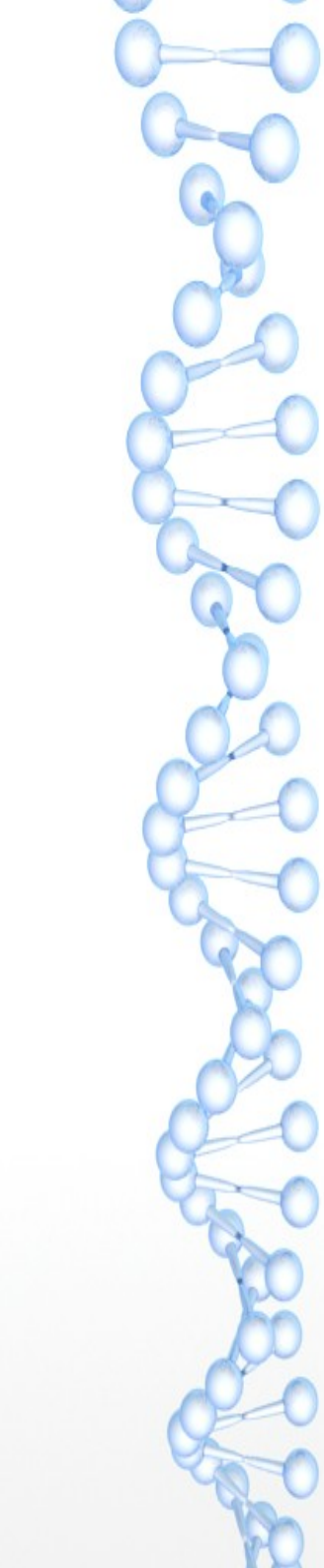
esempio



Esecuzione del ping con cattura delle chiamate a sistema

“ping 193.204.187.1”

output



Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie
un possibile approccio risolutivo

L'idea è:

costruire il profilo normale delle sycall dei demoni potenzialmente soggetti ad attacchi e utilizzare tale profilo per il rilevamento delle anomalie ovverosia di eventuali attacchi *live*

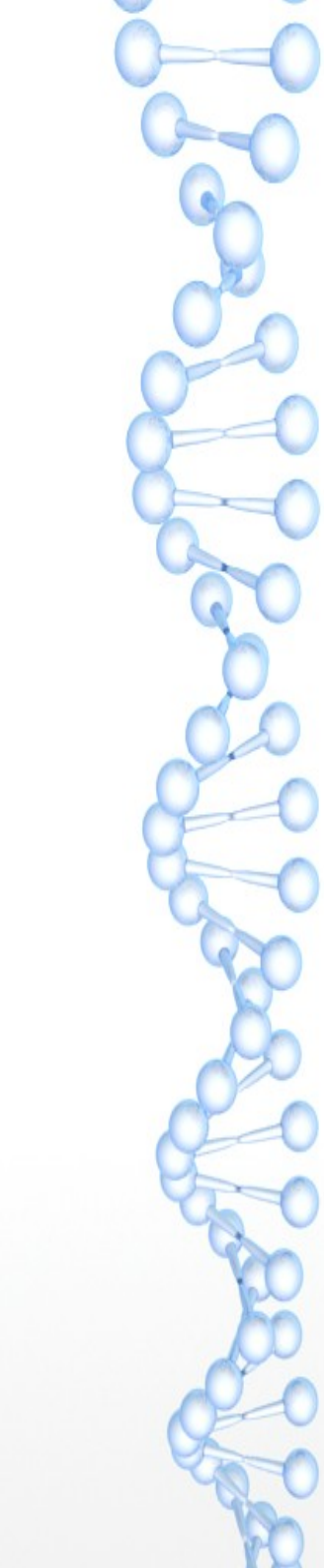
Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie
la sperimentazione

1. I demoni di cui si è costruito il profilo:

DNS – Bind

SMTP – Postfix

2. Il metodo utilizzato per definire il modello normale consiste nel far scorrere una finestra di dimensione prefissata lungo tutto il tracciato prodotto dall'esecuzione di un processo, memorizzando in un database ogni nuova sequenza incontrata.



Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie

la sperimentazione

Per rilevare le anomalie viene seguito lo stesso metodo, applicandolo ai nuovi tracciati da testare.

I tracciati sono analizzati confrontando le sequenze incontrate con quelle presenti nel database.

Normal DB

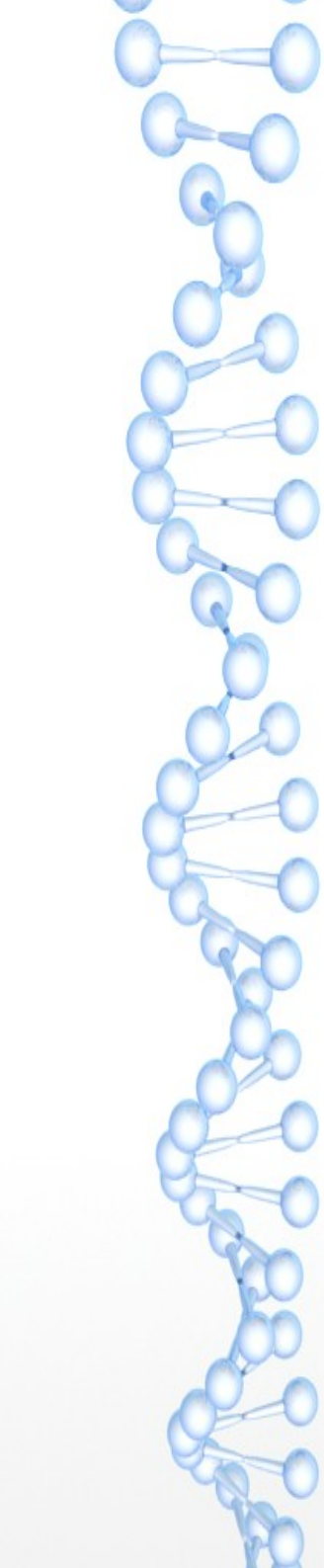
SC1	SC2	SC3	SC4
open	read	read	mmap
read	read	mmap	fork
read	mmap	fork	select
read	write	write	close

Utilizzo di sequenze di chiamate di sistema nella individuazione delle anomalie
i risultati

L'utilizzo di brevi sequenze di system call si è rivelato un buon metodo per la definizione di modelli di comportamento normale, comportandosi meglio con processi non troppo complessi quali il demone del DNS.

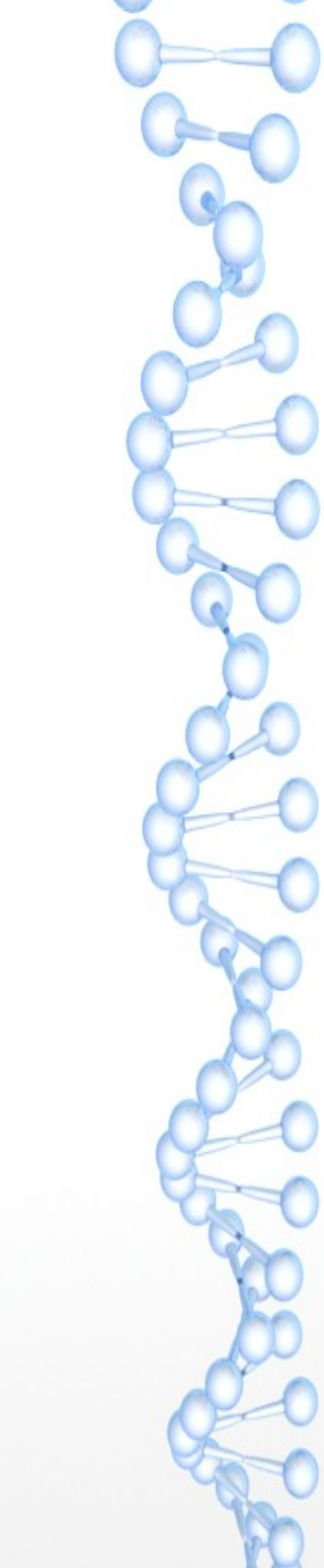
Per BIND l'insieme di sequenze prodotte è limitato, fornendo un ottimo modello. Difatti nella fase di test le anomalie vengono rilevate correttamente.

Postfix ha generato molte più sequenze, nonostante siano state "catturate" solo alcune delle modalità normali. Questo poiché l'applicazione è molto complessa e i processi implicati sono numerosi. Tracciare tutte le modalità normali avrebbe potuto influire negativamente sull'efficienza dell'algoritmo.



Analisi di file di log di un web server
per l'individuazione di attacchi
descrizione del problema

Quando un web server viene attaccato attraverso le proprie pagine web (e non tramite servizio), è possibile rilevare gli attacchi usando i log prodotti dal server stesso?



Analisi di file di log di un web server per l'individuazione di attacchi
un possibile approccio risolutivo

- i log sono essenzialmente sequenze di eventi semplici, di numero finito
- la loro aggregazione costituisce un episodio tipico
- i singoli eventi sono determinati da un lato dal server analizzato e dall’altro dal contesto
- dal parsing dei log si ricavano gli “episodi frequenti” che rappresentano il profilo normale del sistema in base al quale vengono analizzati i log di periodi diversi per il riconoscimento delle anomalie.

Analisi di file di log di un web server per l'individuazione di attacchi

- il log
- gli eventi ricorrenti
- il profilo normale
- le anomalie individuate

Individuazione delle email di spam

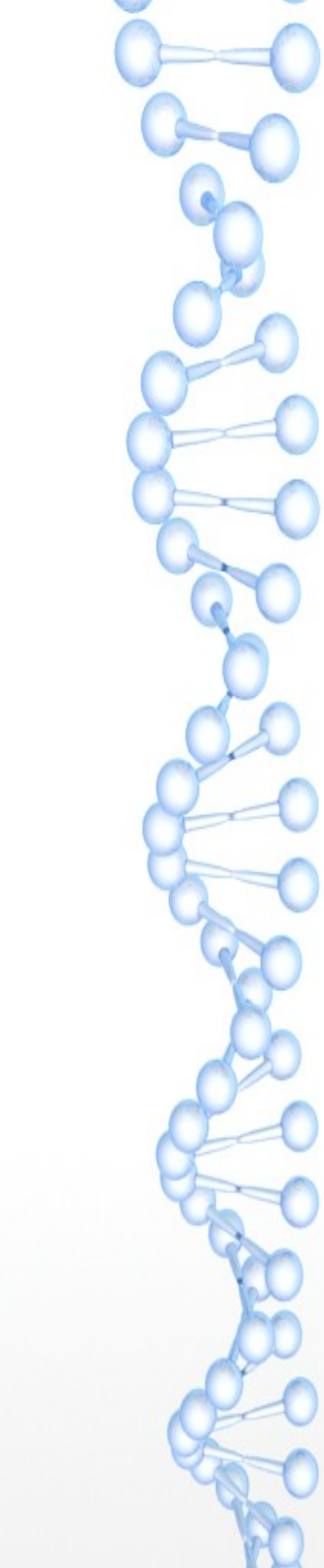
Tecniche di content filtering e applicazione allo spamming
un possibile approccio risolutivo

Un insieme relativamente piccolo di messaggi pre-classificati in modo corretto, affinché se ne possa estrapolare le caratteristiche delle categorie cui appartengono, in termini di frequenze delle parole contenute che rappresentano il profilo normale di un msg

Si utilizzerà l'informazione così acquisita per stimare la possibilità che un nuovo messaggio sia spam o meno.

Tecniche di content filtering e applicazione allo spamming
pretrattamento dei dati

- Un testo non è un dato strutturato e questo rende il compito di classificazione più difficile.
- L'approccio più diffuso nei metodi di content filtering applicati allo spam deriva direttamente dalla disciplina del *text categorization*. Solitamente nei metodi di text categorization, un testo viene rappresentato come un vettore di pesi di termini.
- E, in generale, per rappresentare un documento in una maniera interpretabile da un algoritmo di classificazione si passa attraverso due fasi di preparazione: riduzione della dimensione e indicizzazione.
- Nella riduzione della dimensione si procede ad eliminare le cosiddette “stop words” (connettivi, articoli, etc.) e a riportare le parole ad una forma base tramite le operazioni di stemming e lemmatization.
- L'indicizzazione fornisce una rappresentazione compatta e vettoriale dell'email.



Tecniche di content filtering e applicazione allo spamming

le tecniche utilizzate

Relevance Feedback (Rocchio)

K Nearest Neighbour

Metodo Naive Bayes: *semplice ma molto efficace*

Applicazione del local outlier factor all'intrusion detection
descrizione del problema

Sinora abbiamo visto metodologia di individuazione di anomalie/outlier di tipo”globale” ovverosia utilizzano tutti gli eventi già osservati per decidere se una nuovo evento è anomalo o meno.

L'algoritmo in esame (local outlier factor) calcola invece il grado di anomalia di un nuovo evento confrontandolo solo con i suoi “vicini”.

Un esempio: individuazione di calciatori outlier

ID	Nome e Cognome	RETI	ASSIST	INTERC.	POSIZ.
1	Antonio Cassano	35	10	3	att
2	Francesco Totti	15	5	5	att
3	Simone Inzaghi	14	5	4	att
4	Filippo Inzaghi	16	3	2	att
5	Alessandro Del Piero	14	4	3	att
6	Zlatan Ibrahimovic	17	6	4	att
7	Vincenzo Montella	28	15	2	att
8	Paolo di Canio	0	0	2	att
9	Gianfranco Zola	0	15	3	att
10	Marcelo Zalayeta	12	5	20	att
11	Simone Perrotta	6	4	18	cent
12	Edgar Davids	5	3	16	cent
13	Pavel Nedved	20	4	17	cent
14	Alessio Tacchinardi	4	3	1	cent
15	Juan Sebastian Veron	3	3	14	cent
16	Stephen Appiah	5	5	16	cent
17	Gennaro Gattuso	4	4	18	cent
18	Manuel Rui Costa	4	20	15	cent
19	Esteban Cambiasso	3	3	16	cent
20	Daniele De Rossi	5	6	0	cent
21	Fabio Cannavaro	0	0	30	dif
22	Paolo Maldini	1	1	32	dif
23	Alessandro Costacurta	0	1	33	dif
24	Roberto Carlos	0	2	34	dif
25	Gianluca Zambrotta	1	1	31	dif
26	Nicola Legrottaglie	2	0	33	dif
27	Lilian Thuram	15	1	35	dif
28	Marco Materazzi	2	2	0	dif
29	Paolo Montero	1	1	30	dif
30	Alessandro Nesta	0	0	60	dif

Applicazione del local outlier factor
alle connessioni di rete

- L'algoritmo di individuazione di outlier locali implementato è stato applicato alle connessioni di rete normalizzate del periodo 29/10/2004 – 05/11/2004, per un totale di circa 230.000 connessioni

- Il Local Outlier Factor sulle connessioni di rete è stato calcolato rispetto a tre set di feature

1) Content Based - attributi basati sul contenuto
2) Time Based - attributi basati sul tempo
3) Connection Based – attributi basati sulle connessioni

Local outlier factor
il formato dei dati originali

RICHIESTE/RISPOSTE ARP

```
01/11/2004;21:09:39;030045;193.204.187.XXX;193.204.187.XXX;ask
01/11/2004;21:09:39;030045;193.204.187.XXX;8:0:20:88:5F:5E;reply
```

PACCHETTI ICMP

```
02/11/2004;01:27:47;770045;0:A0:24:4E:DB:F8;8:0:20:88:6B:5F;0x800;
0x145;193.204.187.XXX;;193.204.187.XXX;;64;0xC0;37144;20;311;3;1;
DESTINATION UNREACHABLE: HOST UNREACHABLE
```

PACCHETTI UDP

```
01/11/2004;21:09:35;790045;8:0:20:88:5F:5E;8:0:20:88:6B:5F;0x800;
0x5E;80.240.114.XXX;1024;193.204.187.XXX;53;44;0x0;18003;20;80;52;;

01/11/2004;21:09:36;660045;0:30:84:87:69:81;FF:FF:FF:FF:FF:FF;0x800;
0x5C;193.204.187.XXX;137;193.204.187.XXX;137;128;0x0;3381;20;78;50;;
```

PACCHETTI TCP

```
01/11/2004;21:09:41;370045;8:0:20:88:5F:5E;0:7:E9:15:2C:BB;0x800;0x3E;62.9
8.174.XXX;3095;193.204.187.XXX;80;117;0x0;1797;20;48
DF;******S*;0xF3D907E9;0x0;0x2238;28;;;4;1460 NOP NOP SackOK

01/11/2004;21:09:37;780045;0:7:E9:15:2C:BB;8:0:20:88:5F:5E;0x800;0x24E;193
.204.187.XXX;80;151.28.148.XXX;1085;64;0x8;28572;20;576
DF;***A****;0x18C8EB47;0x240CE2;0x1920;20;;;;
```

Local outlier factor
Content based features

- Su un dataset di 10.000 connessioni di rete si sono avuti, i seguenti risultati:

Content-Based Features

id	lof
2607	4261,422
8513	3565,405
4770	3525,002
4185	2859,721
4228	2859,721
4279	2859,721
4283	2859,721
4284	2859,721
4414	2859,721
4415	2859,721
4416	2859,721
4463	2859,721
4464	2859,721

La connessione 2607 ha ricevuto lo score più alto; si tratta di una connessione incompleta, proveniente dall'esterno diretta verso la rete dipartimentale, con un elevato numero di ritrasmissioni rispetto al contesto locale.

Local outlier factor
Time based e connection based features

Time-Based Features

id	lof
9172	6,275674
9176	5,711419
9209	3,458373
9179	2,951933
9157	2,569803
400	2,56553
403	2,56553
401	2,531828
5001	2,512792
9153	2,510153

La connessione 9172 ha ricevuto lo score più alto perché, in un intervallo di 5 secondi, si è avuto il massimo numero di servizi richiesti verso lo stesso IP destinazione.

Connection-Based Features

id	lof
9356	9,097276
6756	7,58881
6926	7,33219
3205	3,421169
5221	3,415531
2611	3,38224
7806	3,34911
6209	3,315859
1628	3,298891

La connessione 9356 ha ricevuto lo score più alto, poiché si tratta di una connessione dall'interno del dipartimento verso l'esterno, con un elevato numero di connessioni effettuate dallo stesso IP sorgente nell'intervallo di tempo che va dalla connessione corrente alla centesima precedente.

Da fare

Attività serie:

• sviluppo di strumenti per difesa proattiva, contromisure live e analisi forense

Attività ancora più serie:

• hacking lab con 1liv e 2liv

• partecipare ai contest