Metodi e tecniche per la rilevazione e il contrasto di botnet in reti universitarie

Andrea Balboni

andreabalboni@gmail.com

7° Borsisti Day

20/01/2016
Roma – Consortium GARR

Borse di Studio
Orio Carlini
Consortium GARR

Obiettivi del progetto

Studio di tecniche e metodi innovativi per l'individuazione e il contrasto di botnet e di server di Comando e Controllo

- Elementi originali:
  - Acquisizione e integrazione dati da **molteplici sorgenti eterogenee**
  - Focalizzazione sulla correlazione e analisi dei **log DNS**

Borse di Studio Orio Carlini Consortium GARR

Fasi del progetto

1. **Acquisizione**
2. **Analisi**
3. **Risultati**

Acquisizione: *requisiti della ricerca*

Fonti eterogenee

- log di analizzatori di traffico di rete
- log di un server DNS GARR
- blacklist DNS

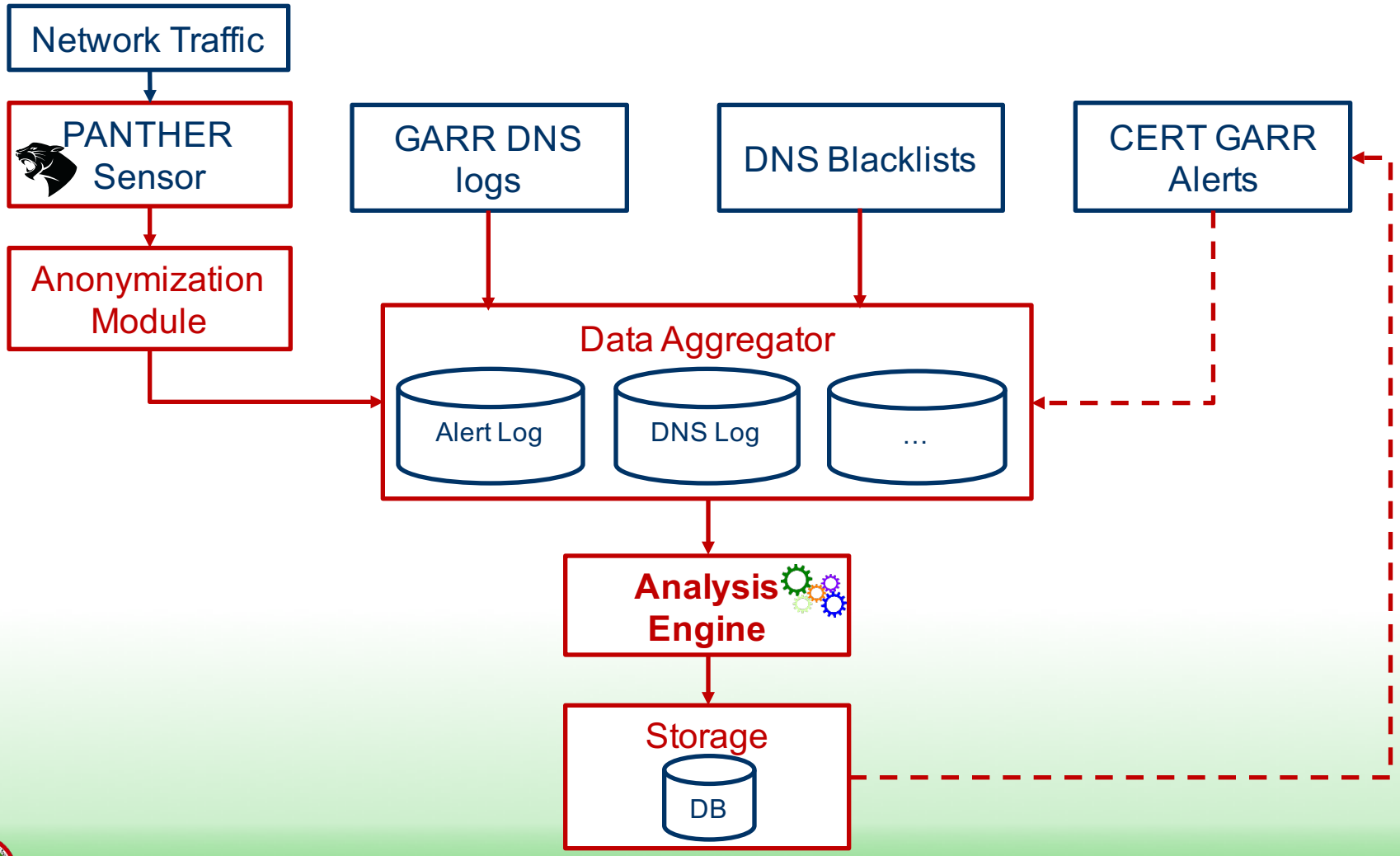
→ necessario individuare un **formato univoco**

Dati sensibili → necessario adottare tecniche di anonimizzazione **privacy preserving**

Volumi elevati di dati in costante crescita → necessario progettare un **sistema scalabile** per strorage e indexing

Borse di Studio Orio Carlini Consortium GARR

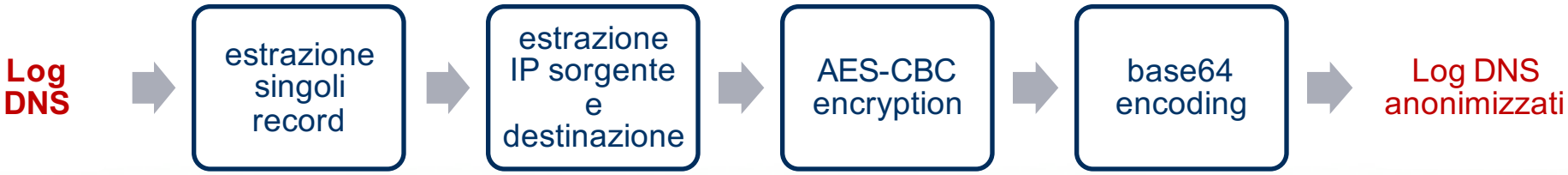
Acquisizione dati: *Architettura*



Anonimizzazione dei dati sensibili: *soluzione adottata*

Utilizzo di algoritmi di cifratura AES

- chiave conosciuta solo dall'amministratore di rete
- mancato utilizzo di initialization vector garantisce che a ogni valore in chiaro corrisponda sempre lo stesso valore cifrato



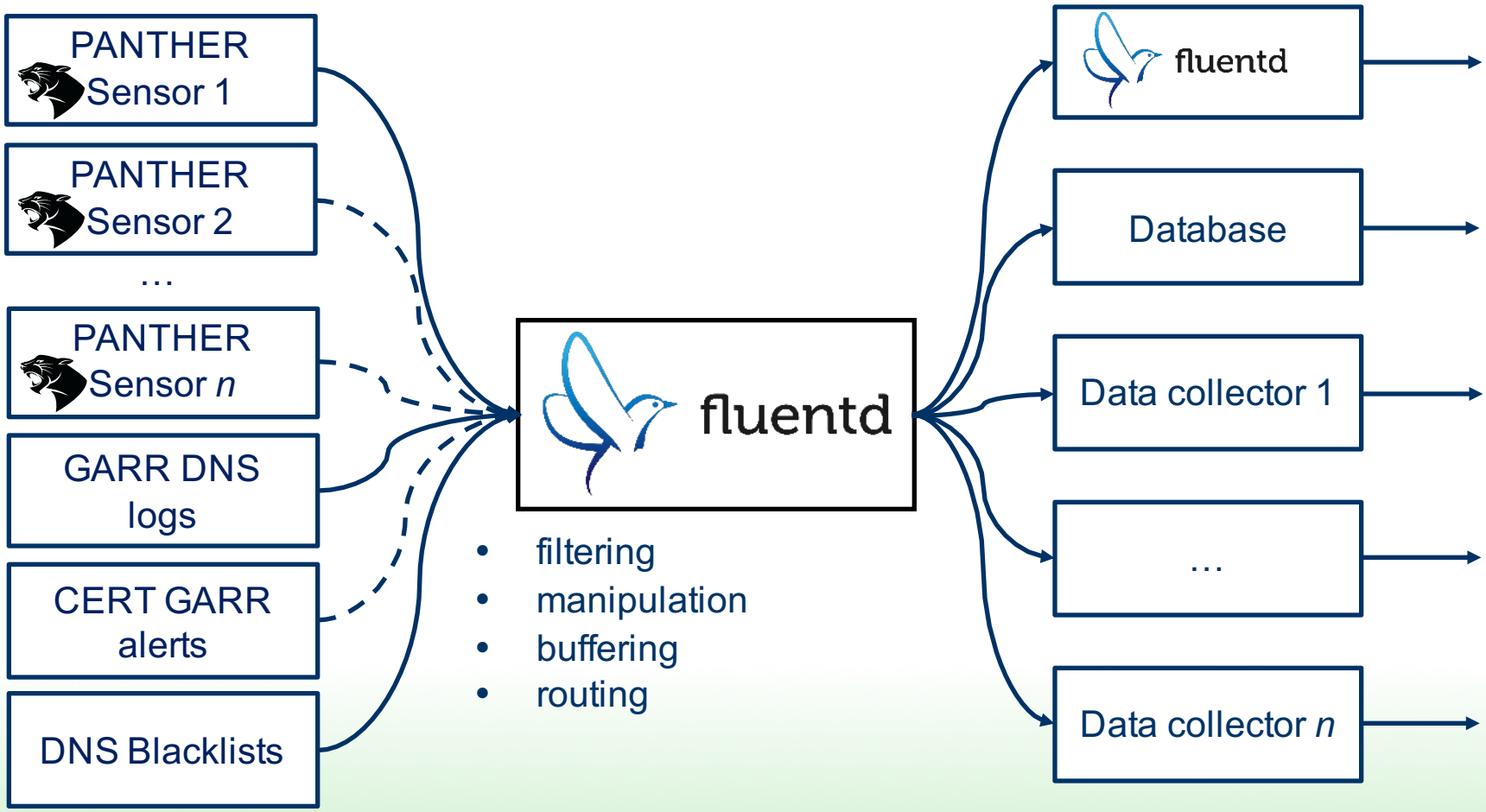
Aggregazione dati da sorgenti eterogenee: *sottosistemi*

Data Aggregator → 2 sottosistemi

1. **Data collection e routing**
   - parsing
   - manipolazioni
   - conversioni di formato → formato univoco JSON
   - inoltro a diverse destinazioni

2. **Storage e indicizzazione**
   - memorizzazione
   - indicizzazione multipla
   - ricerche e filtri

Data collection: schema



Storage e indicizzazione: elasticsearch

- Storage distribuito
- Funzionalità avanzate di indicizzazione
  - Distribuzione indici su più nodi di elaborazione e storage
- Motore di ricerca
- Filtri avanzati
- Aggregazione

Fasi del progetto

1. Acquisizione
2. Analisi
3. Risultati

Analisi focalizzata sui log DNS

- Le moderne botnet coinvolgono il DNS in diversi ambiti:
  - **Comunicazioni** tra bot e server di comando e controllo (C&C)
  - Attività di perlustrazione del bot-master per verificare l'eventuale blacklisting DNS dei **nomi di dominio associati ai bot o ai server di controllo**
- Possibilità di rilevare botnet **indipendentemente dal protocollo** che utilizzano (HTTP, IRC, …)

Individuazione DGA

Domain Generation Algorithms (DGA) ⇒ per diminuire le probabilità di essere individuati, i bot contattano i server C&C e/o i peer attraverso nomi di dominio auto-generati. Es.,

- `razzrwsbzum.org`
- `jbfygzlcjak.info`

Obiettivo della ricerca

Classificazione automatica dei nomi di dominio ⇒ {*sospetti, non sospetti*}

Analisi DGA

- Classificatore realizzato mediante euristiche sulle label dei nomi di dominio
  - lunghezza
  - rapporto vocali/consonanti inconsueto
  - numerosità vocali/consonanti inconsueta
  - assenza di parole di senso compiuto
- Classificatore implementato come plugin custom per fluentd
  - dati analizzati e annotati prima di essere memorizzati e indicizzati

Integrazione con blacklist DNS

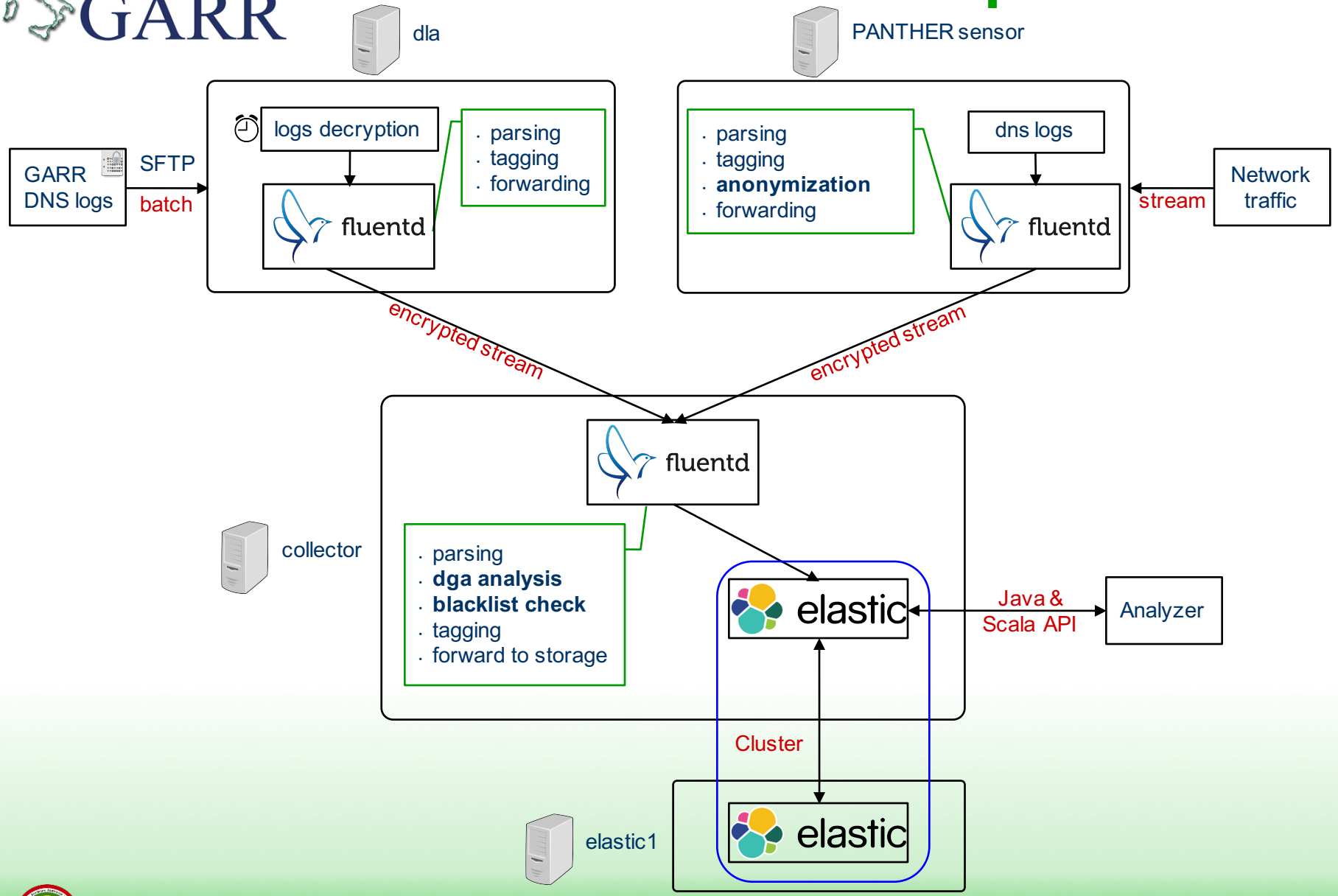
- Ogni record DNS viene analizzato per verificare se il nome di dominio richiesto è presente in blacklist
- Blacklist configurabili e aggiornate quotidianamente
  - Malware Domains: http://www.malwaredomains.com/
  - AAMS: http://www.servizi.garr.it/aams/aams_block_domain.txt

Fasi del progetto

1. Acquisizione
2. Analisi
3. Risultati

Borse di Studio Orio Carlini
Consortium GARR

Prototipo realizzato

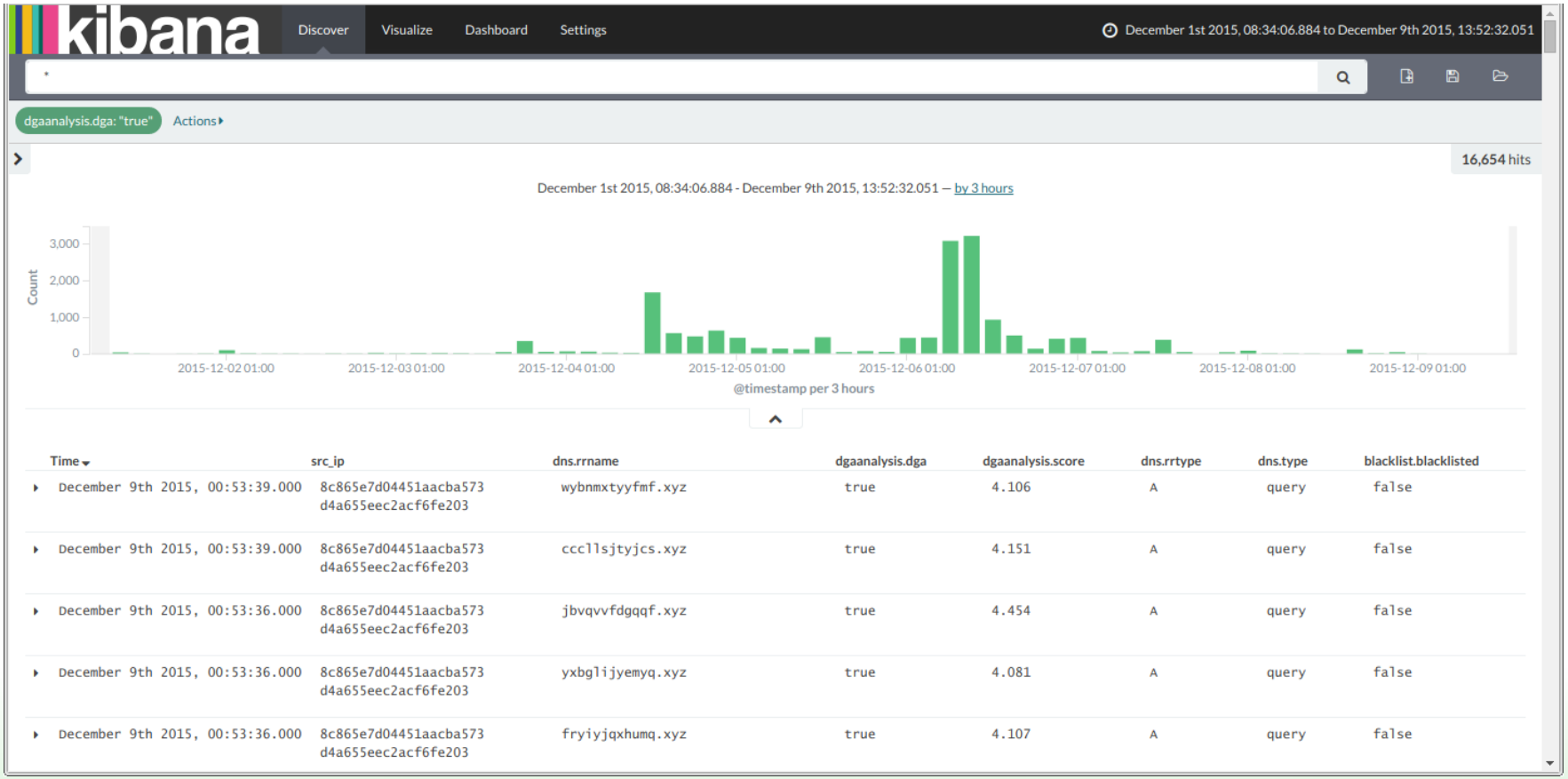


Risultati sperimentali

- **6.6 miliardi** di query DNS analizzate e indicizzate
  - circa 1.3 TB di dati
- Segnalati numerosi DGA in 2 mesi di traffico:
  - GARR: 31.480 query per **21.911 DGA**
  - UNIMORE: 12.510 query per **10.500 DGA**
- Falsi positivi: 2.9% con pre-filtering
- Tempi di estrazione dati:
  - 30 secondi - 3 minuti

Borse di Studio
Orio Carlini
Consortium GARR

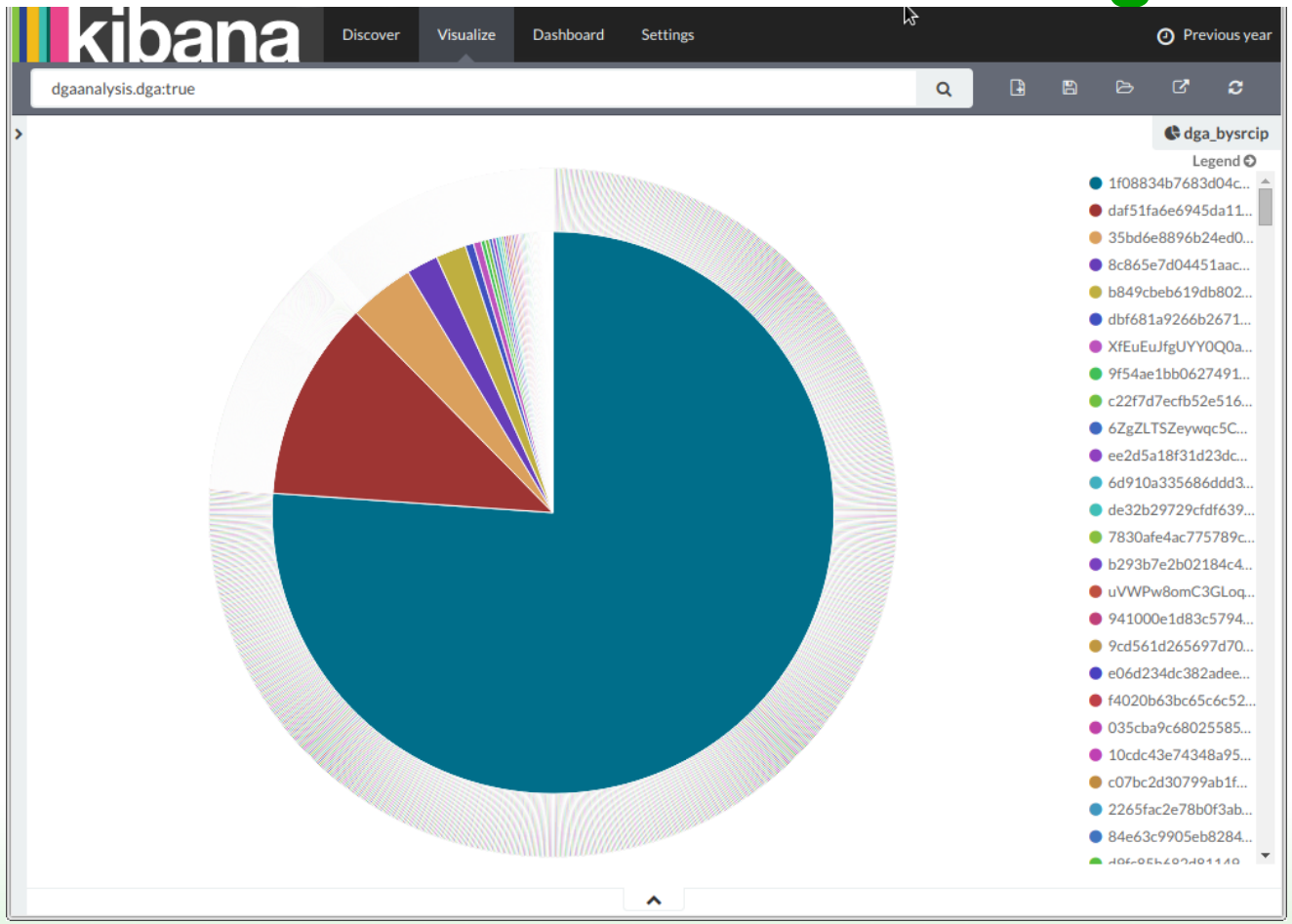
Interfaccia di consultazione



Time	src_ip	dns.rrname	dgaanalysis.dga	dgaanalysis.score	dns.rrtype	dns.type	blacklist.blacklisted
December 9th 2015, 00:53:39.000	8c865e7d04451aacba573 d4a655eec2acf6fe203	wybnmxtyyfmf.xyz	true	4.106	A	query	false
December 9th 2015, 00:53:39.000	8c865e7d04451aacba573 d4a655eec2acf6fe203	cccllsjtyjcs.xyz	true	4.151	A	query	false
December 9th 2015, 00:53:36.000	8c865e7d04451aacba573 d4a655eec2acf6fe203	jbvqvvfdgqqf.xyz	true	4.454	A	query	false
December 9th 2015, 00:53:36.000	8c865e7d04451aacba573 d4a655eec2acf6fe203	yxbglijyemyq.xyz	true	4.081	A	query	false
December 9th 2015, 00:53:36.000	8c865e7d04451aacba573 d4a655eec2acf6fe203	fryiyjqxhumq.xyz	true	4.107	A	query	false

Distribuzione richieste DGA per IP sorgente

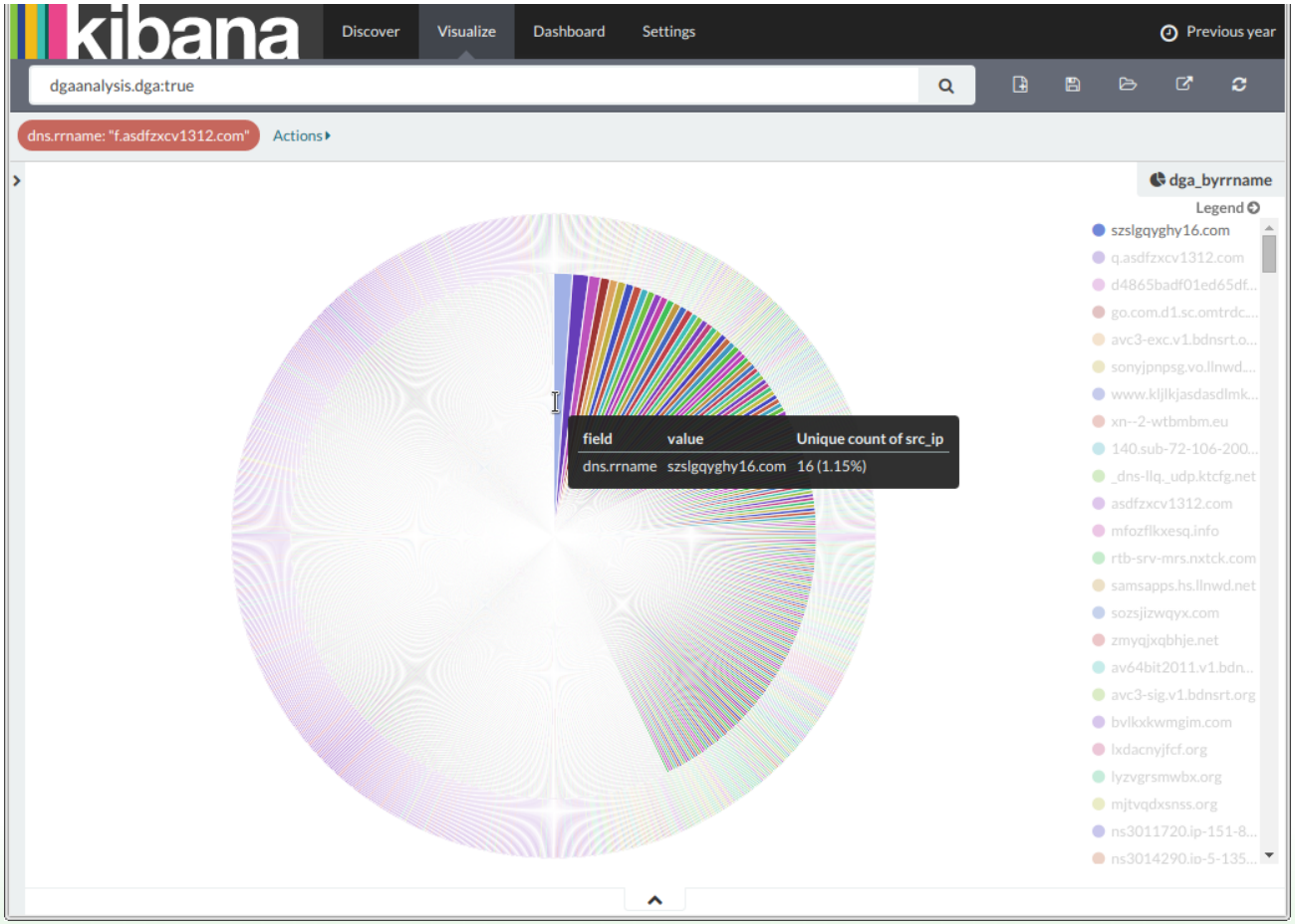
Name server configurati per ricorsione effettuano la maggior parte di richieste per DGA



Distribuzione richieste per ogni DGA

Nomi DGA tipicamente richiesti una sola volta e da un singolo indirizzo sorgente

Richieste multiple possono indicare botnet attiva



Borse di Studio Orio Carlini Consortium GARR

Obiettivi del secondo anno

1. Individuazione di algoritmi e tecniche per **migliorare il riconoscimento di nomi DGA**
   - Riduzione falsi positivi
   - Integrazione dei nomi di **dominio internazionalizzati IDN**
2. **Anomaly detection** basata sui pattern di richieste di host potenzialmente infetti con quelli di host non infetti e sui tempi di inter-arrivo delle richieste a DNS per nomi di dominio sospetti
3. Miglioramento della **scalabilità** della piattaforma per avere risultati delle analisi in tempo reale

Si garantisce l'aspetto sperimentale della ricerca mediante prototipi. Qualche problema nell'accesso a set di dati reali.

Borse di Studio Orio Carlini Consortium GARR

References

- S. Yadav, A. K. K. Reddy, A. L. N. Reddy, and S. Ranjan, “Detecting algorithmically generated malicious domain names,” presented at the the 10th annual conference, New York, New York, USA, 2010, pp. 48–61.
- A. Reddy, “Detecting Networks Employing Algorithmically Generated Domain Names,” 2010.
- Bilge, Leyla, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." In NDSS. 2011.
- S. Yadav, A. K. K. Reddy, A. L. N. Reddy, and S. Ranjan, “Detecting algorithmically generated domain-flux attacks with DNS traffic analysis,” IEEE/ACM Transactions on Networking (TON, vol. 20, no. 5, Oct. 2012.
- Z. Wei-wei and G. Qian, “Detecting Machine Generated Domain Names Based on Morpheme Features,” 2013.
- P. Barthakur, M. Dahal, and M. K. Ghose, “An Efficient Machine Learning Based Classification Scheme for Detecting Distributed Command & Control Traffic of P2P Botnets,” International Journal of Modern …, 2013.