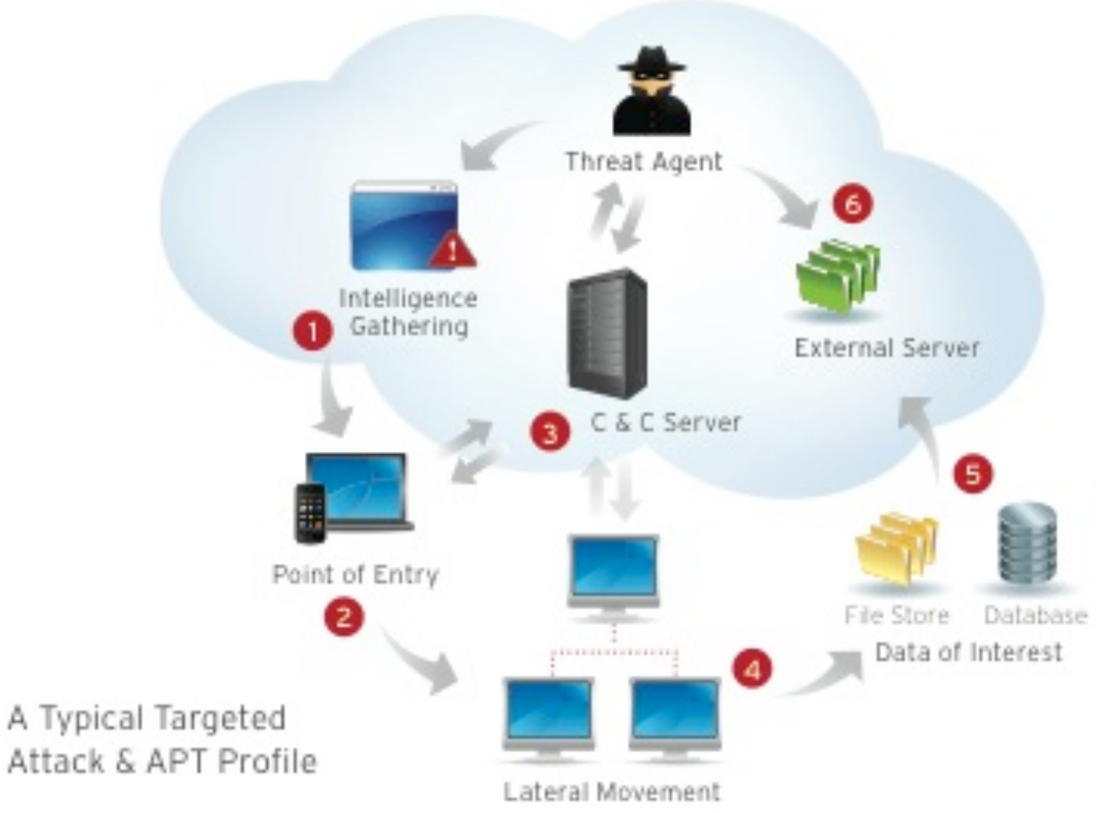
APT Discovery Beyond Time and Space

Gabriela F. Ciocarlie, PhD
Phil Porras
Vinod Yegneswaran, PhD
Shalini Ghosh, PhD

Computer Science Laboratory
SRI International

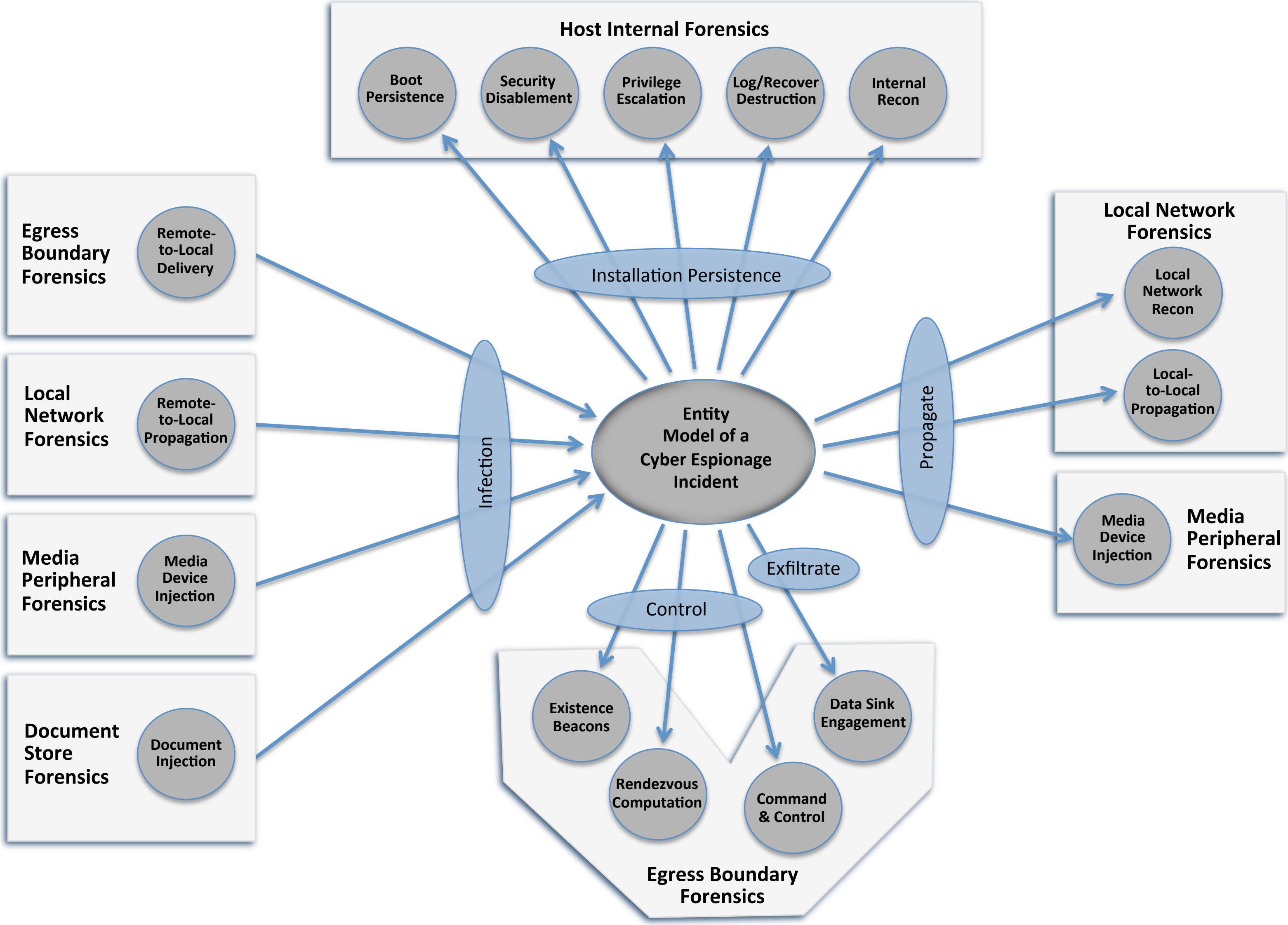
Advanced Persistent Threats

- Continue to increase in sophistication
  - Stealth
  - Persistence
- Initial Infection
  - Watering Hole attacks
  - Spearphishing
  - Social attacks
- Objective
  - Data Exfiltration
  - IP / Identity Theft



http://www.trendmicro.com/cloud-content/us/pdfs/business/white-papers/wp_custom-defense-against-targeted-attacks.pdf

APT Behavioral Lifecycle Model



Examples of APT Families Behaviors

	Aurora with Hydraq	RSA Breach	Stuxnet	Shady RAT	1.php APT (Mandiant)	Lurid	Gauss	HeartLand Hannaford	Flame	Zegost CN-RAT
Egress Delivery	yes			yes	yes		unknown	yes	unknown	yes
Beacons		yes		yes	yes					
Rendezvous Probing	yes					yes				
Command & Control	yes	yes	yes	yes	yes		yes	yes	yes	yes
Data Sink Engagement	yes	yes		yes	yes		yes	yes		
Media Injection			yes				yes		yes	
Document Injection		yes	yes		yes	yes				
LocalNet Recon	yes	yes	yes	yes		yes		yes	yes	
Local Propagation		yes	yes	yes				yes	yes	
Boot Persistence	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Security Disablement			coexist	stopped			halts for some	undetected	undetected	yes
Privilege Escalation		yes	yes	yes	yes		yes	yes	yes	yes
Log Delete	yes		yes				yes		yes	
Internal Recon	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Automated Analysis Beyond Time and Space

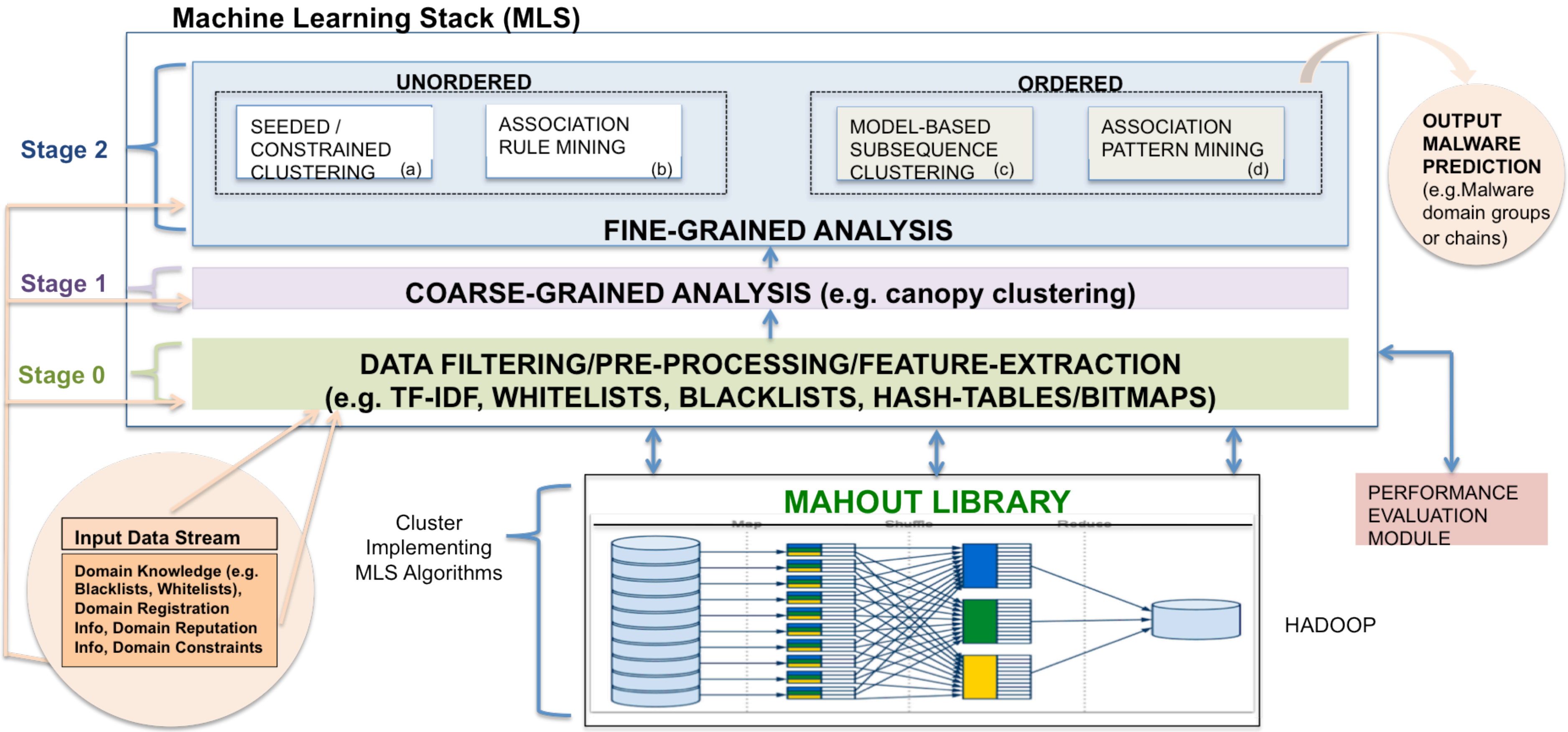
Desiderata for Defeating APTs

- Multi-faceted (multi-sensor) data collection
- Continuous and pervasive monitoring
- Large-scale, long-term and multi-dimensional data analysis
- Automated mitigation

- **Context**
  - Application of scalable, multi-perspective DNS traffic analysis for malware domain group detection

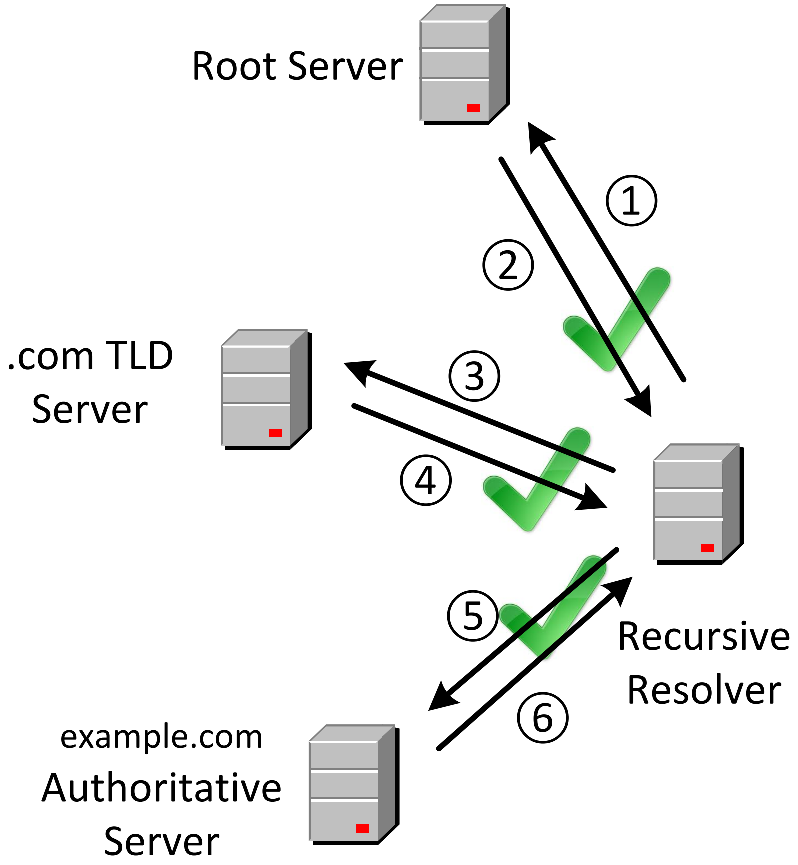
- **Motivating Examples**
  - Detecting traffic-redirection chains for watering hole attacks
  - Identification of drop zone domains

ML Stack for Large Scale DNS Traffic Analysis



Machine Learning Stack (MLS)
Stage 2
Stage 1
Stage 0
UNORDERED
SEEDED / CONSTRAINED CLUSTERING (a)
ASSOCIATION RULE MINING (b)
ORDERED
MODEL-BASED SUBSEQUENCE CLUSTERING (c)
ASSOCIATION PATTERN MINING (d)
FINE-GRAINED ANALYSIS
COARSE-GRAINED ANALYSIS (e.g. canopy clustering)
DATA FILTERING/PRE-PROCESSING/FEATURE-EXTRACTION (e.g. TF-IDF, WHITELISTS, BLACKLISTS, HASH-TABLES/BITMAPS)
OUTPUT MALWARE PREDICTION (e.g.Malware domain groups or chains)
Input Data Stream
Domain Knowledge (e.g. Blacklists, Whitelists), Domain Registration Info, Domain Reputation Info, Domain Constraints
Cluster Implementing MLS Algorithms
MAHOUT LIBRARY
Map
Shuffle
Reduce
HADOOP
PERFORMANCE EVALUATION MODULE

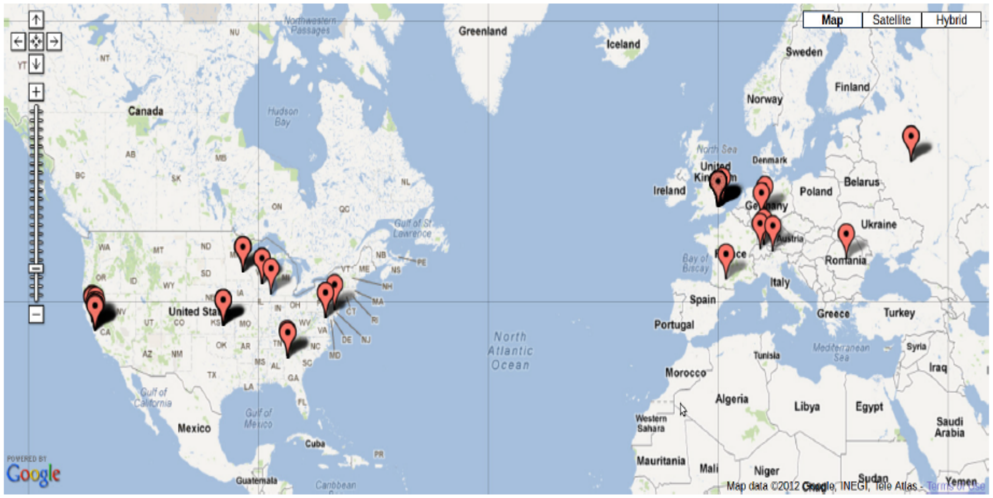
SIE Dataset


Root Server

.com TLD Server

Recursive Resolver

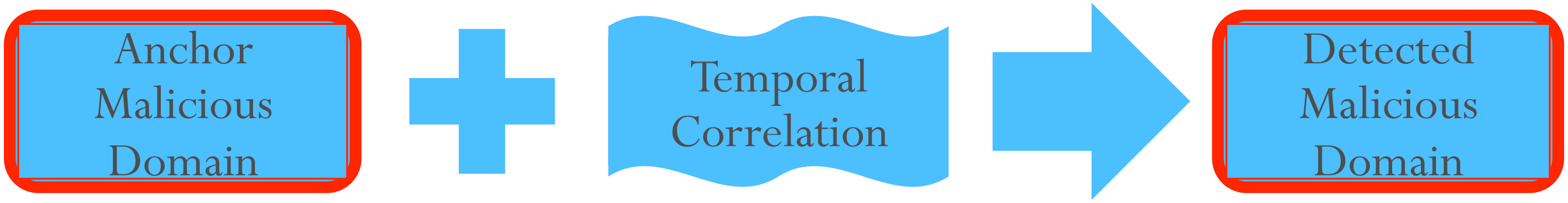
example.com Authoritative Server



- Data size
  - 26 Billion DNS queries and responses
  - 2 TB raw data / day
- 628 of contributing resolvers

Malware Domain Group Detection

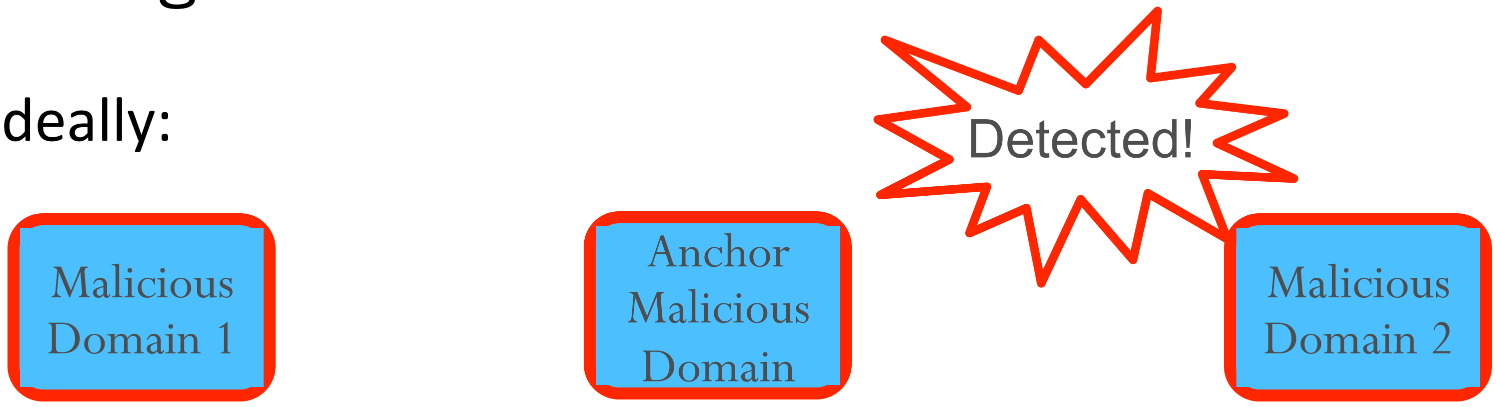
- Key intuition:
  - DNS queries are not isolated instances
- Detection method:



- **Advantages**:
  - Detect malicious domain groups in general (scam, DGA, etc.)
  - Do not need comprehensive labeled training set

Challenge

- Ideally:



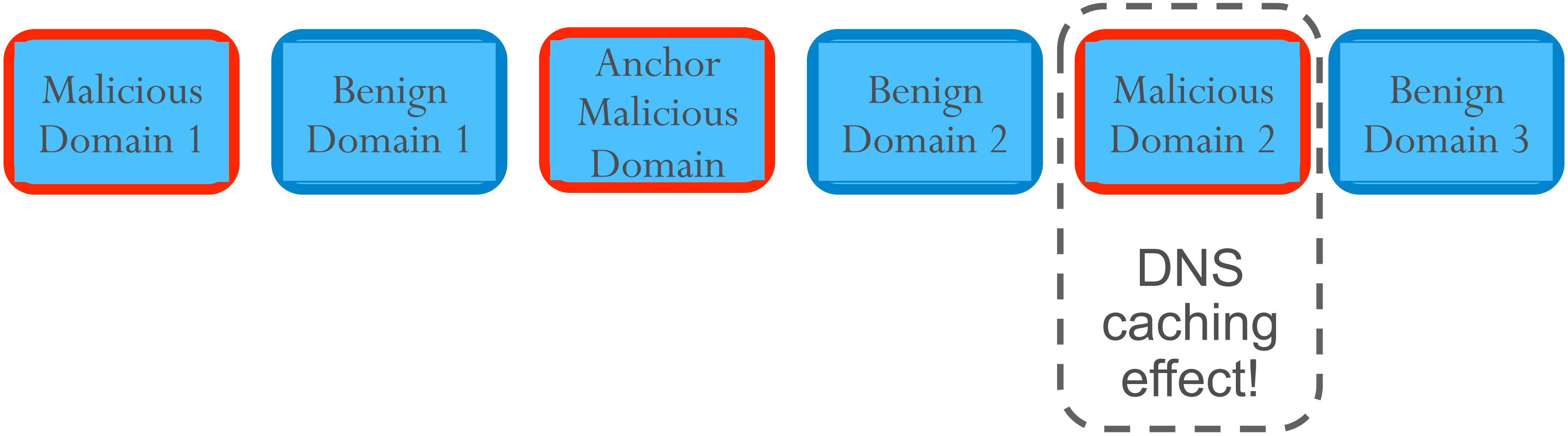
Malicious Domain 1

Anchor Malicious Domain

Detected!

Malicious Domain 2

- In reality:



Malicious Domain 1

Benign Domain 1

Anchor Malicious Domain

Benign Domain 2

Malicious Domain 2

Benign Domain 3

DNS caching effect!

Practical Solution

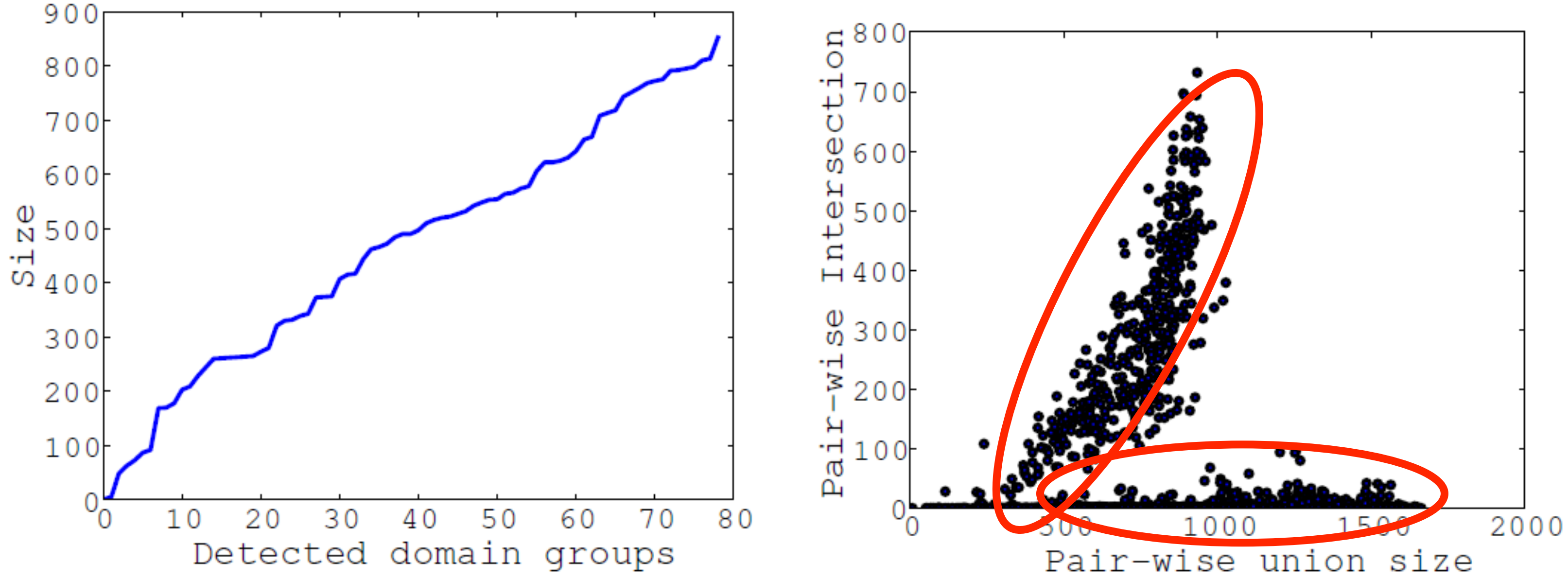
- A 3-step approach to identify the correlated domain group, given an anchor malicious domain
  - Identify the coarse related domain group using a TF-IDF heuristic
  - Cluster the coarse domain group
  - Refine the domain group according to the clustering result

Experimental Evaluation

DNS Data	Size	Anchor Domain #
Dec 16, 2012	1.82B queries	129

- Obtaining anchor domains:
  - Record all domains blacklisted on Dec. 16th from three external blacklists
    - MalwareDomainBlockList, MalwareDomainList, Phishtank
- Validating detected domains:
  - Blacklist matching with 5 external blacklists
    - McAfee SiteAdvisor and MyWot
  - IP address comparison

Domain Group Analysis



(a) Domain group size.

(b) Domain group similarity.

Sample anchor domain pairs deriving highly overlapping groups

surprise-mnvq.tk	surprise-mnvr.tk
vural-electronic.com	vfventura.sites.uol.com
voyeurpornweb.com	vkont.bos.ru

Domain Group Analysis

pill-erectionmeds.ru	pillcheap-med.ru	onlinerxpillhere.ru
medspill-erection.ru	rxpill-medstore.ru	medpillbuy-online.ru

A pharmaceutical domain group, size = 295

uggsbootss.com	niceuggsforsale.com	louisvuittonwhite.net
uggsclassic.org	officialuggsretails.com	nicelouisvuittonbag.com

A counterfeit product domain group, size = 17

lq8p.ru	ol4k.ru	s3po.ru
n5di.ru	p9ha.ru	n4gf.ru

A suspected DGA domain group, size = 71