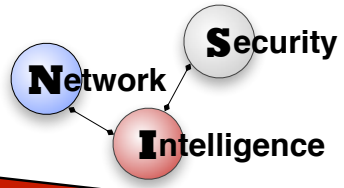
FluxBuster

Early Detection of Malicious Flux Networks via Large-Scale Passive DNS Traffic Analysis

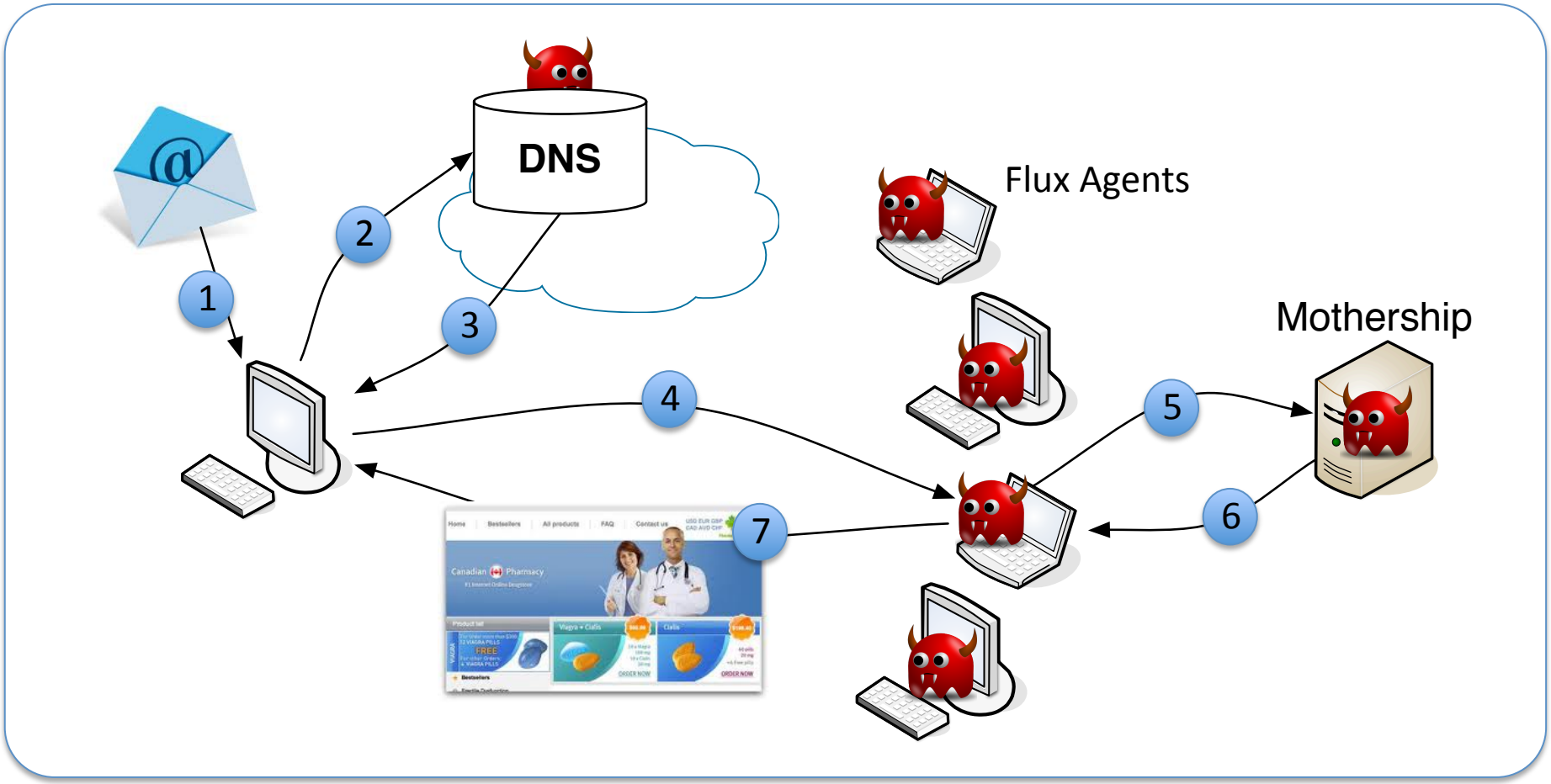
Roberto Perdisci




University of Georgia
Dept. of Computer Science



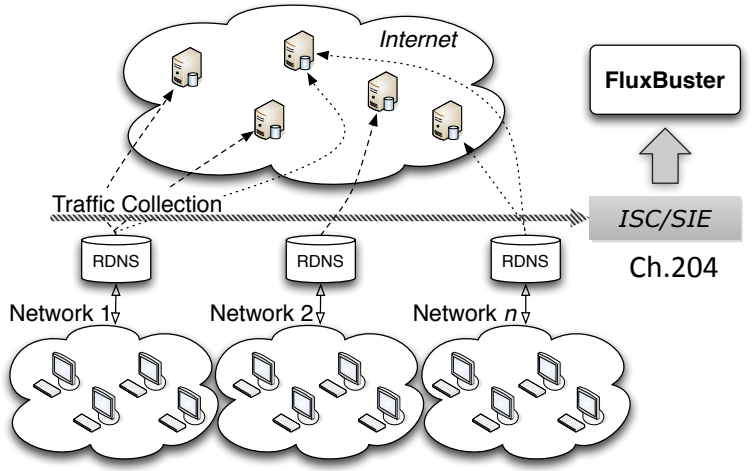
Flux Networks



Research Goals

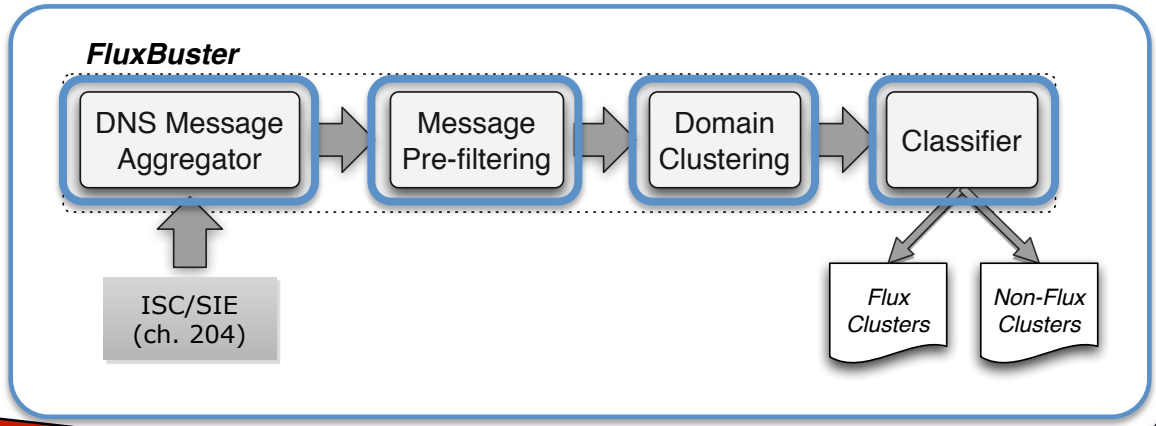
- Previous works on flux detection based mainly on active probing
  - Limited to known bad or suspicious domains
  - Domains treated independently
  - Possible data pollution by attackers

- Passive Detection
  - Monitor "behavior" of *all domains* over time
  - Let other people query for you in a *distributed* way!
  - Only focus on *live* domains
  - Discover *zero-day* flux domains!



FluxBuster System Overview

- Given a domain *d*, aggregate all info about *d* collected during a time T (e.g., 24h)
- Use *conservative heuristics* to filter out domains that are *highly unlikely flux*
- Group domains that are related to each other
  - significant intersection between sets of resolved IPs
  - *Candidate Flux Networks*
- Statistical classifier automatically labels candidate flux networks
  - Each candidate flux networks is described by a number of features
  - *flux* or *non-flux*



Message Pre-Filtering

- Conservative Filtering
  - Objective: reduce burden on following modules
  - Consider only domains for which ***all*** of the following constraints hold
    - *avg*(TTL) <= 3600
    - # of RIPs >= 3 OR *avg*(TTL) <= 30
    - *div*(RIPs) >= 1/3

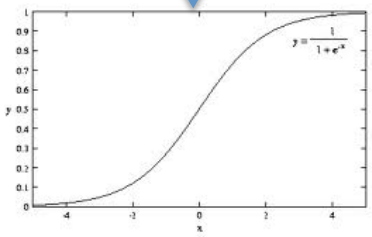
$$div(\text{RIP}) = \frac{\text{\# /16 prefixes in RIPs}}{\text{\# of RIPs}}$$

Domain Clustering

- Group domains that are related to each other
  - Hierarchical clustering algorithm
  - Similarity measure based on resolved IPs

$$sim(\alpha, \beta) = \frac{|R_\alpha \cap R_\beta|}{|R_\alpha \cup R_\beta|} \cdot \frac{1}{1 + e^{\gamma - \min(|R_\alpha|, |R_\beta|)}}$$

Jaccard Index

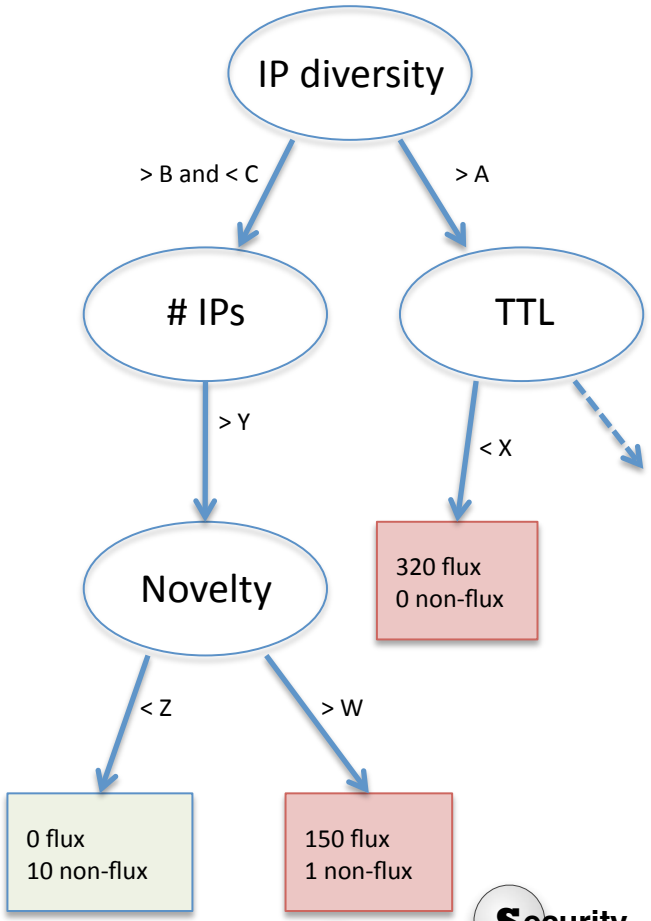


oparle.com
ns1.chokode.com
ns2.chokode.com
ns3.chokode.com
ns4.chokode.com
ns5.chokode.com
ns6.chokode.com

free-pass.porn-4-free-here.ru
free-pass.allhotpornhere.ru
free-pass.all-porn-access-free.ru

Supervised Classifier

- Input: Clusters of domains
  - Clusters are translated into *feature vectors*
- Supervised Training:
  - Need labeled data (ground truth)
  - We built a web interface to facilitate semi-manual labeling
- Output: new (unlabeled) clusters are labeled as either *flux* or *non-flux*



Statistical Features

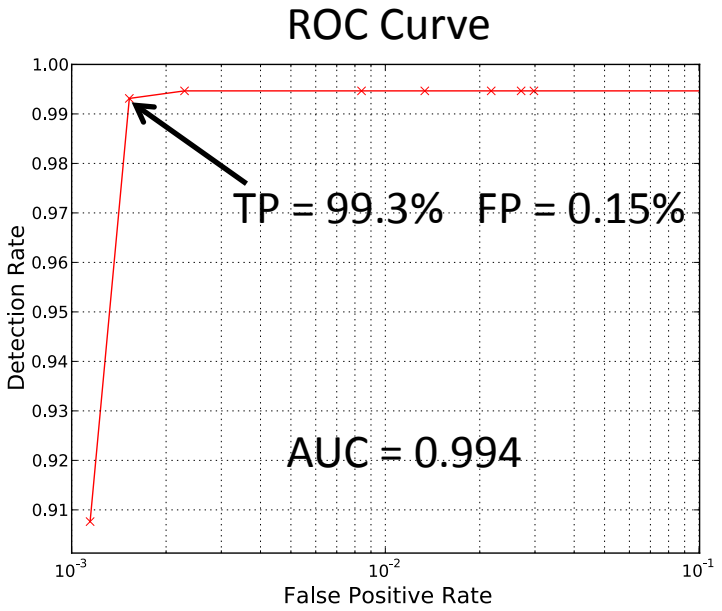
- Measurements on each domain cluster
  - ϕ_1 – # of IPs in RIPs set
  - ϕ_2 – # of Domains
  - ϕ_3 – avg(TTL)
  - ϕ_4 – # domains that have recently pointed to any of RIPs
  - ϕ_5 – Entropy of /16 prefixes
  - – …

$$\phi_5 = \frac{-\sum_x p(x) \cdot \log_2 p(x)}{\log_2(\phi_1)}$$

Overall we measure **13 statistical features**

Cross-Validation

- Labeled Dataset
  - semi-manual labeling process
  - If no clear-cut decision, exclude cluster to minimize training *noise*
  - 1,337 clusters labeled as flux
    - 100,644 distinct 2LDs (113,580 FQDs)
  - 5,708 labeled as non-flux
    - 2,116 distinct 2LDs (59,215 FQDs)



Live Traffic Evaluation

- 5 months of operational deployment
  - 4,084 domain clusters labeled as *flux*
    - 1,743 2LDs (63,442 FQDs)
  - 3,633 domain clusters labeled as *non-flux*
    - 227,667 2LDs (264,550 FQDs)
  - Threshold K = 30 distinct IPs
    - Clusters with less than 30 resolved IPs are discarded

- Measure four different quantities
  - False Positives
  - False Negatives
  - True Positives
  - True Negatives
- Separately measured due to many *unknown* domains that cannot be easily verified as either flux or not

Ground Truth

- A domain cluster C may fall into three categories:
  - (1) TPs: C includes domains and/or IPs that are known to be related to a flux network
  - (2) FPs: C does not represent a flux network, and may instead represent a CDN or other legitimate services
  - (3) NAs: the true nature of C is *unknown*, that is no prior information exists on this cluster in any public (or even private) security data sources.

- (1) top flux domains from abuse.ch (**KFD**) + domains from public malware domain blacklists (**KMD**)
- (2) *consistently top* 100k Alexa (**ATD**), >300k domains from Yahoo DMOZ (**YDD**), list of known CDN domains (**CDN**)

Live Results Summary

domain1.com
domain2.com
domain3.com
domain4.com
domain5.com
domain6.com

TPs

- 24/75 KFD (50 domains not visible in SIE)
- 179/10,447 KMD (Note: most malware domains are not flux domains)
- 525 + 595 ***new*** flux domains using KFD and KMD as "seed", respectively (*guilty by association*)

FPs

- 2/57,910 2LDs in ADT (pool.ntp.org, qyq3606.meibu.com)
  pool.ntp.org appeared only briefly (filtered at the source by SIE for ch.204?)
- 0 in CDN and 0 from YDD

FNs

- Domains consistently classified as *non-flux*
  1 from KFD: discountpharmacyhealth.net
  30 from KMD

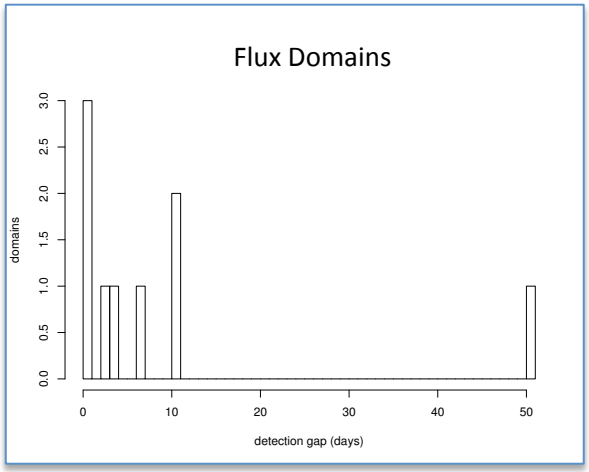
TNs

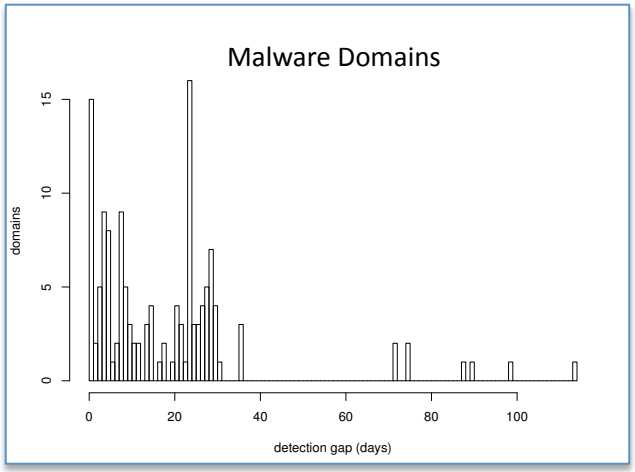
171 2LDs in ATD+YDD+CDN
227,667 2LDs remain *unknown*

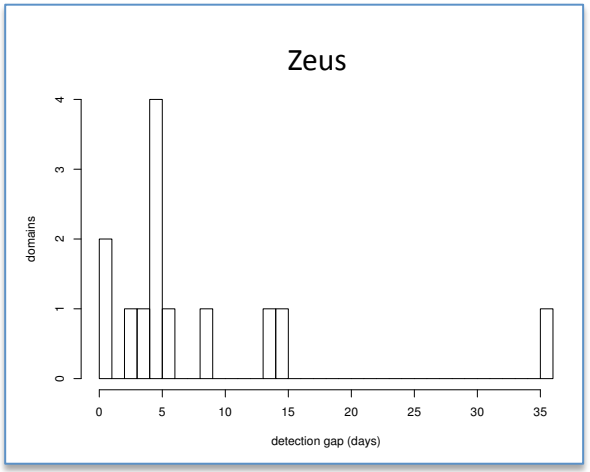
Flux domains	**1,743** 2LDs (**63,442** FQDs)
Flux agent IPs	**317,203** distinct IP addresses (on average 3,265 distinct IPs per day)
Previously unknown flux 2LDs	**995** through a "domain-based" analysis, and **1,030** through an "IP-based analysis" (using *guilty-by-association*)
Early-detection results	**64.5%** of malicious 2LDs detected earlier than other state-of-the-art tools (131 2LDs out of 203)
Previously unknown flux agent IPs	**62%** of flux agents tested against `abuse.ch` DNSBL service

Early Detection

- 9/24 KFD detected earlier than *abuse.ch*
- 125/179 KMD detected earlier than appeared in BLs
- 13/21 Zeus flux domains detected earlier than BLs

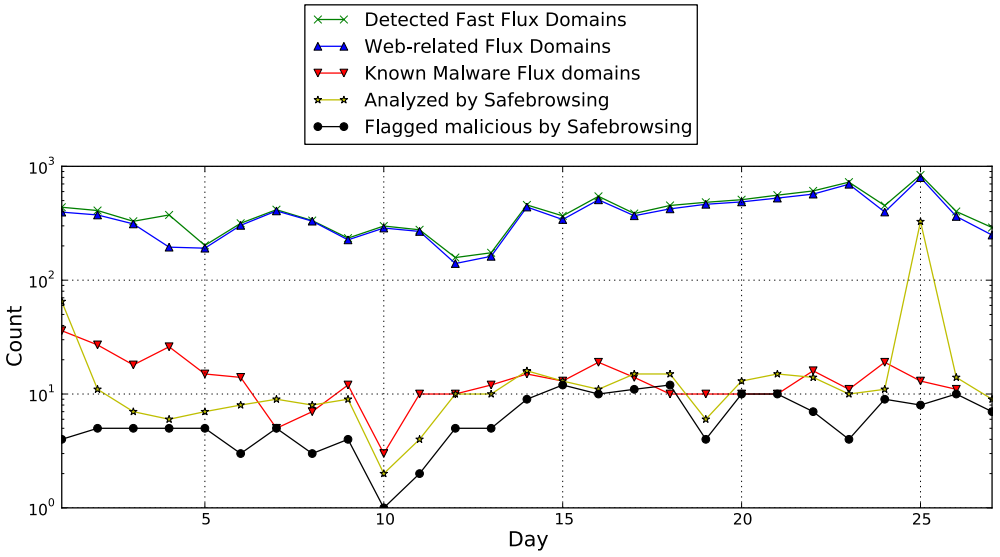






SafeBrowsing

- Take flux domains and
  - Check if port 80 is open
  - Check for valid HTTP response/content
  - Vet against SafeBrowsing (SB) and malware BLs
    - Most missed by SB are rogue pharmacies, porn-related sites
    - SB only reports known phishing and malware sites



Thank You!

Acknowledgements

- ✓ ISC/SIE (especially Robert)
- ✓ Igino Corona (U. Cagliari)
- ✓ David Dagon (GaTech)
- ✓ Wenke Lee (GaTech)
- ✓ Giorgio Giacinto (U. Cagliari)

Source Code Available!

http://code.google.com/p/fluxbuster

Sponsor

- ✓ NSF SDCISec Program

Early Detection of Malicious Flux Networks via Large-Scale Passive DNS Traffic Analysis
IEEE Transactions on Dependable and Secure Computing, 9(5), Sept.-Oct. 2012, pp. 714-726.
http://roberto.perdisci.com/publications/publication-files/FluxBuster-TDSC.pdf