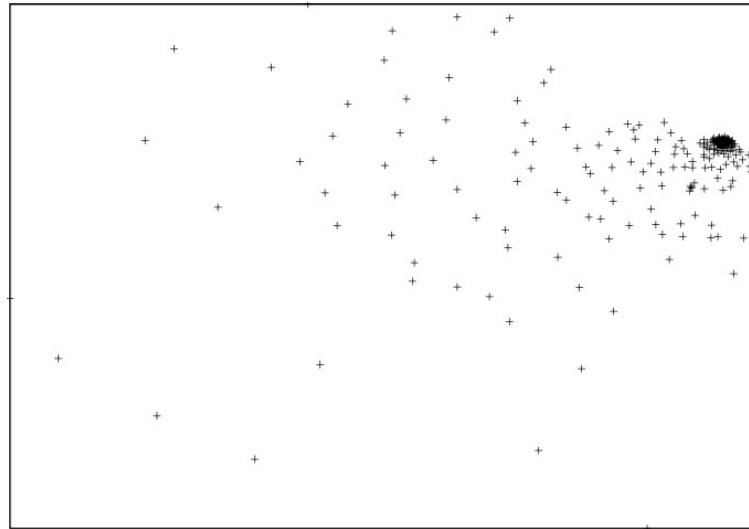


Detection and Identification of BotNets



Dr. Alexander K. Seewald



Early Warning System for BotNets (1)

„Revisiting DarkNets“

- Analyzes traffic sent to unused IPs (DarkNet)
- Completely passive, relies on botnet propagation
- Very efficient! (~ 300,000 IPs analyzed in RT)

How?

- Reference data about known bots / botnets
- Feature construction at several levels
- Machine learning: Traffic patterns of specific bots
- Validation & Test

Early Warning System for BotNets (2)

Demonstrating a three-level (passive) approach

- **Level 1:** Analysis of single packets
Spambot identification system
- **Level 2:** Analysis of network traffic
Analyzing access patterns of spambots
- **Level 3:** Analysis of traffic contents
Correlating spam content and spambot type

Reference Data & Features

Reference Data

- Contributed by Marshal via their TRACE system
- IP, Timestamp, Spambot Type - matched to our logs

Data collection

- Learn spambot type from single(!) packet
- No packets outgoing (not even SYN/ACK)

Features

- ICMP, TCP and UDP traffic indicators
- IP destination port (no source!)
- 2-byte-gram of payload for ICMP/UDP
- 2-byte-gram of TCP options for TCP (no payload)

Matching Reference Data to DarkNet IPs

Static vs. Dynamic IPs

- Dynamic IPs only valid for limited time periods
- 4.38% static IPs, 95.62% dynamic IPs in our data
- Matching to reference data within +/- 1h (~ 5%)

DNSBL overlap

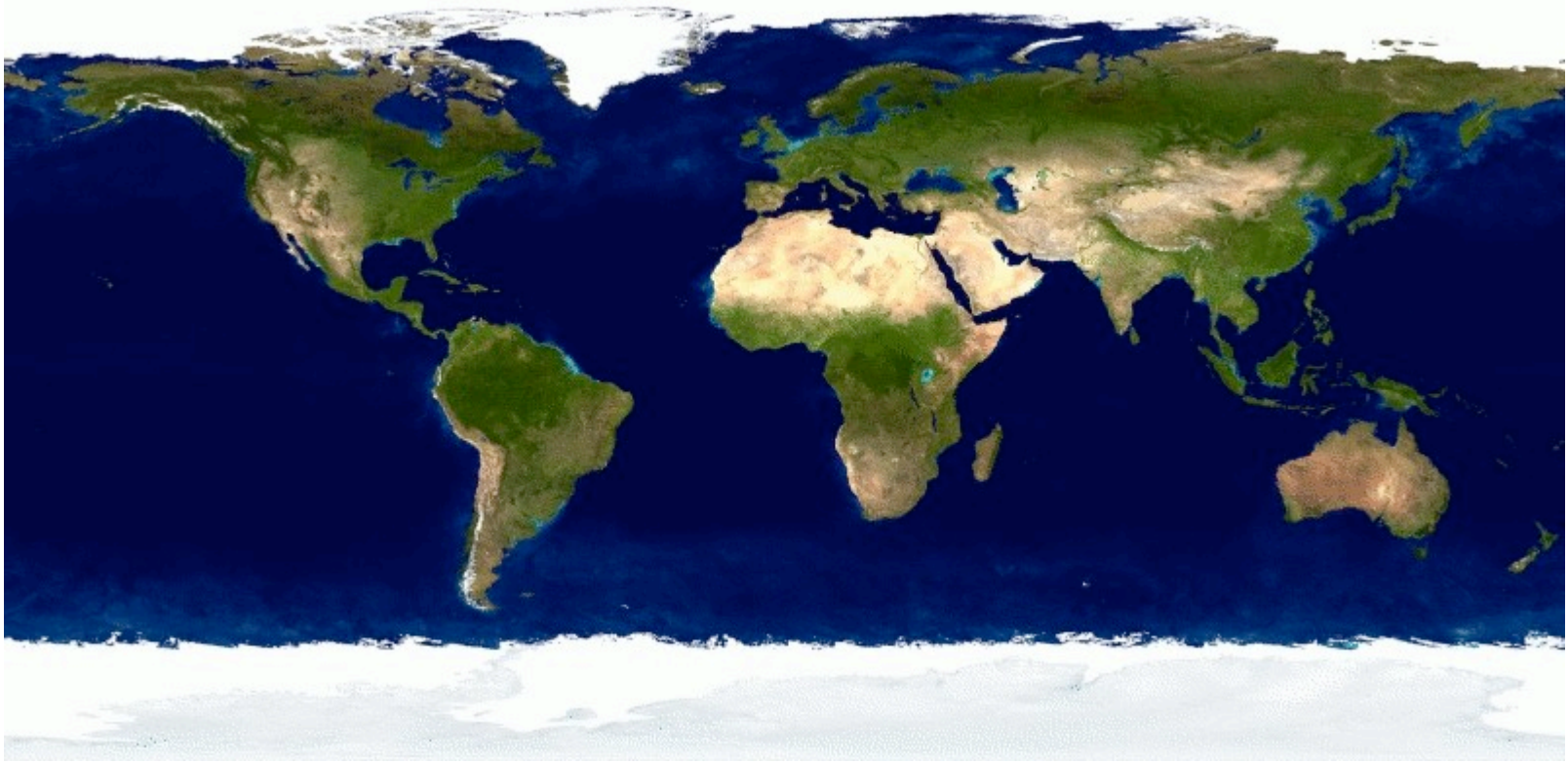
- At time of access to our darknet, only 2.75% of Spambots are listed in the Spamhaus XBL.
- Indicates possible usage of our system as DNSBL

Performance

Spambot identification based on single packets

Spambot	Prec.	Rec.	F-m.
Unknown (4)	1,000	1,000	1,000
Unknown (11)	0,947	0,973	0,960
Srizbi (3)	0,847	0,981	0,909
Rustock (7)	1,000	0,263	0,416
Unknown (9)	1,000	0,143	0,250
Hacktool.Spammer (6)	0,500	0,313	0,385
Pushdo (2)	0,000	0,000	0,000
Mega-D (1)	0,000	0,000	0,000

Activity Overview (24h compressed)

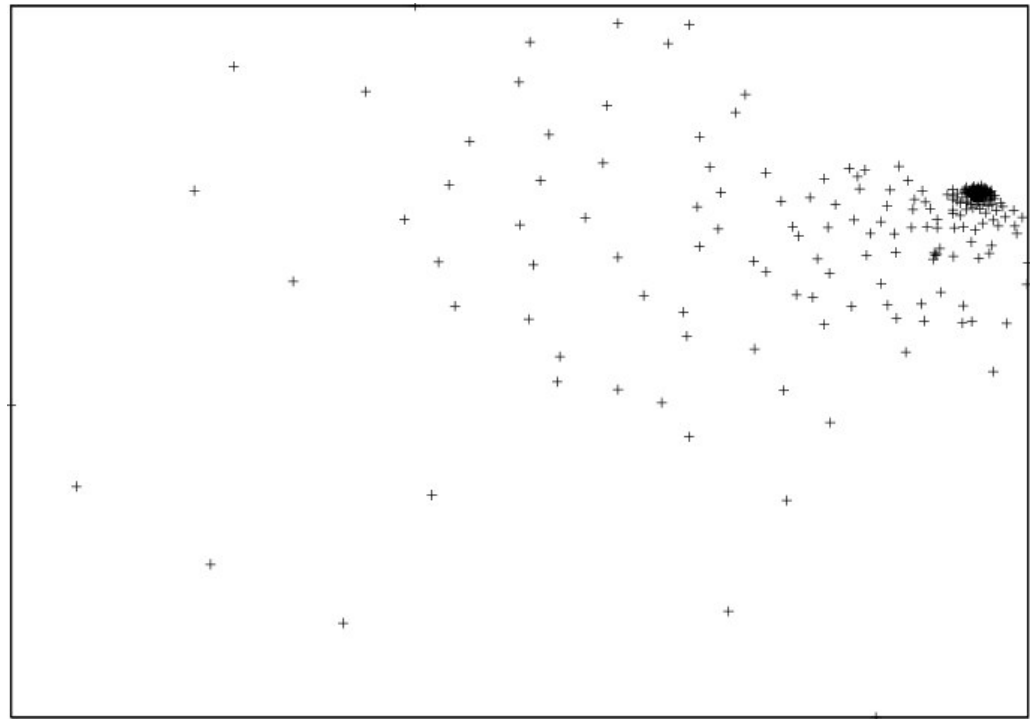


Different colors indicate access by different spambots.
Background: [Visible Earth \(NASA\)](#), IP-localization by
[IP Address Location](#). Reference data by [Marshal Trace](#).

Level 2: Access patterns (1)

Multiple accesses from same IP

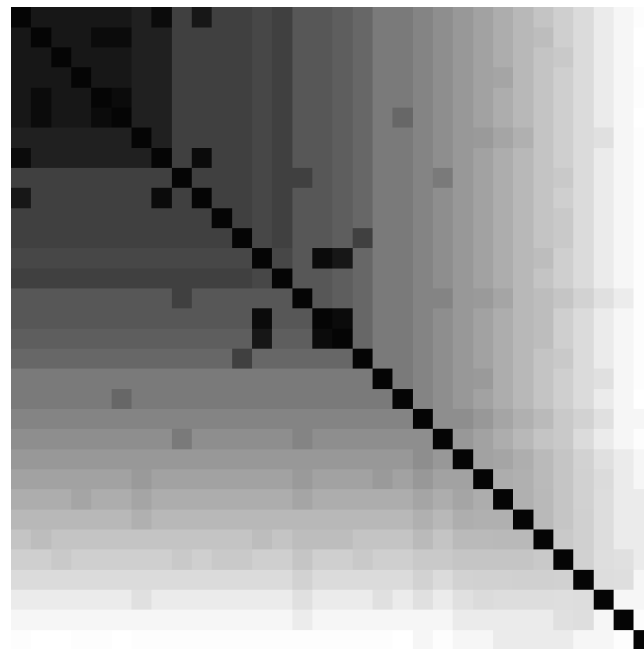
- Pattern type: random, consecutive, stepping, ...
- Some variability in access patterns (edit distance)



Level 2: Access patterns (2)

Patterns vs. Spambot types

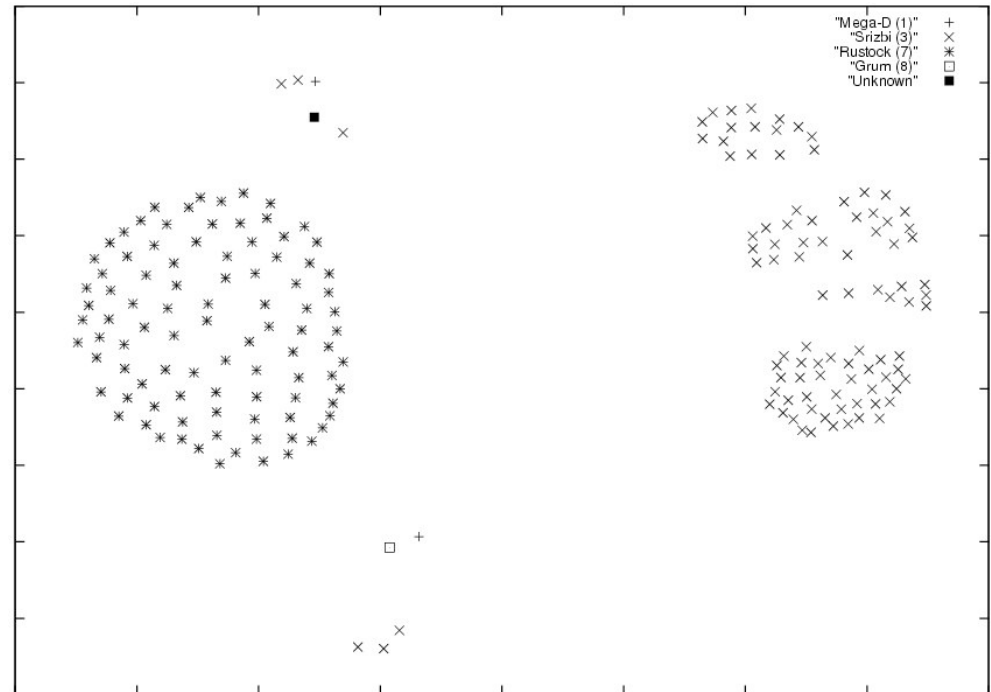
- No spambot-specific patterns – all patterns are shown by at least three different spambots.
- Possibly hinting at underlying control systems?



Level 3: Spam content vs. spambot type

Matching Spambot reference data to spam contents

- Allows to find correlations between spambot type and the contents of a spam mail.
- Different spambots send out different spam mails.



Outlook & Future Work

We have...

- Server with capacity up to 300,000 darknet IPs
- Working prototype for Spambot identification
- Lots of ideas ;-)

We are lacking...

- Sufficient reference data, especially for BotNets
- A large darknet (1.5 million IPs would suffice ;-)
- Minor funding (network traffic / server bills etc.)

Thanks for your attention!