Dynamical Understanding of DNS Events (DUDE)

H. Van Dyke Parunak, Alex Nickels, Rich Frederiksen
Soar Technology, Inc.

Problem

Advanced Persistent Threat

- Internal machine connects to multiple URLs to (e.g.)
  - Download exploit
  - Persistent C2 channel
  - Exfiltration
- Infection can spread within the enterprise → not all machines show all steps of the threat

Preparation

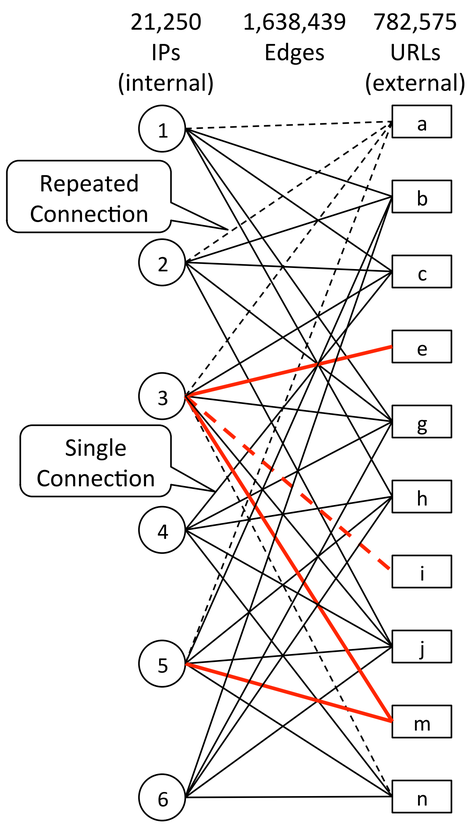
Objectives

- Reduce data size
- Avoid artifacts

Actions

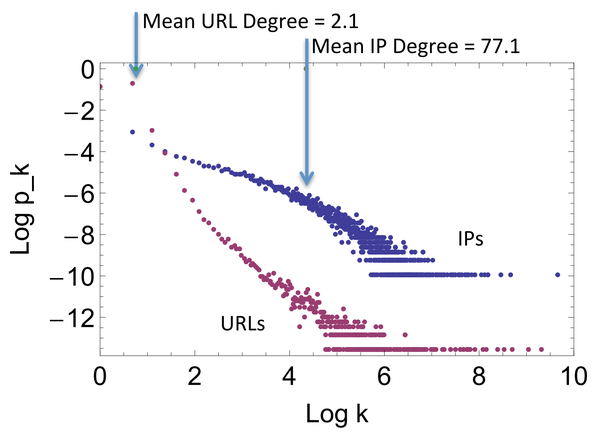
- Round all times to nearest second
- Delete all but one of multiple DNS responses to the same IP for the same URL
- Collapse DNS resolution chains: retain records only from Ips that never respond to a DNS query
- Whitelist URLs from Month 1

Processing: Graph

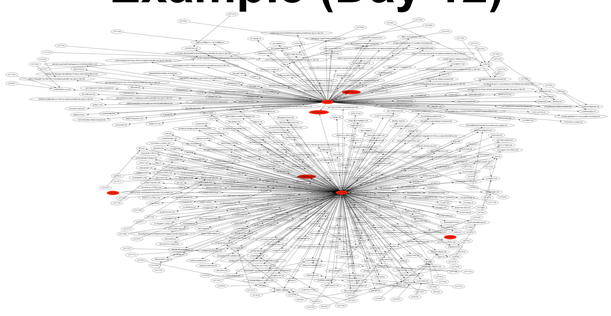


- **All relevant data is local** in the graph
  - Connection times (on edge)
  - Degree of URL
  - Propagation
- **Scale by distributing** across CPUs
- Agents on each processor
  - Implement different heuristics
  - Interact by annotating the graph

→ Giant component has 800,472 nodes, 1500 smaller components total 3353 nodes



Example (Day 12)



Detecting Connections

Look for

- IPs adjacent to rare URLs (≤ 5 IPs)
- Rare URLs adjacent to suspect IPs

Detecting Regular Beaconing

On a link with 10 or more connections to a rare URL (5 or fewer IPs)

- Compute successive differences between connection times
- Subtract their mean
- Compute Fourier transform (off-label use!)
- Drop first coefficient (0)
- Compute ratio of minimum coefficient to maximum coefficient
- Select links with highest ranking ratio

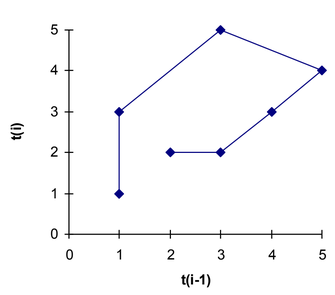
This scores type B (e.g., mine.starving.wad.f8) low. Candidate extension:

- Replace long intervals with sequence

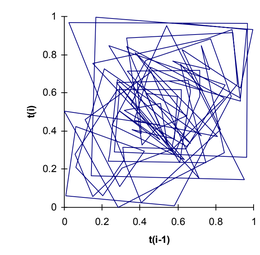
Processing: Dynamics

Exploring: Time Delay Embedding

x=t(i-1)	y=t(i)
1	1
1	3
3	5
5	4
4	3
3	2
2	2



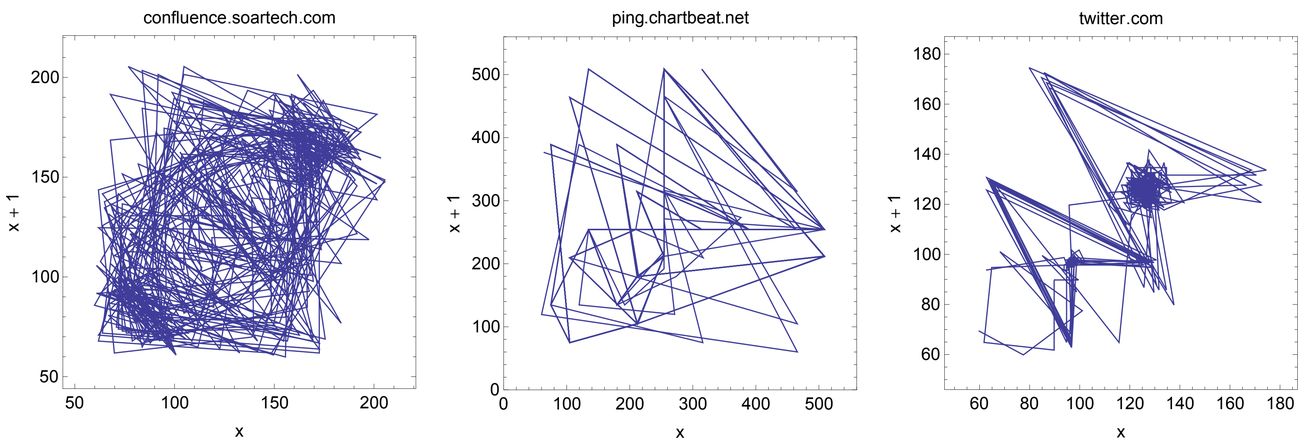
100 Random Points



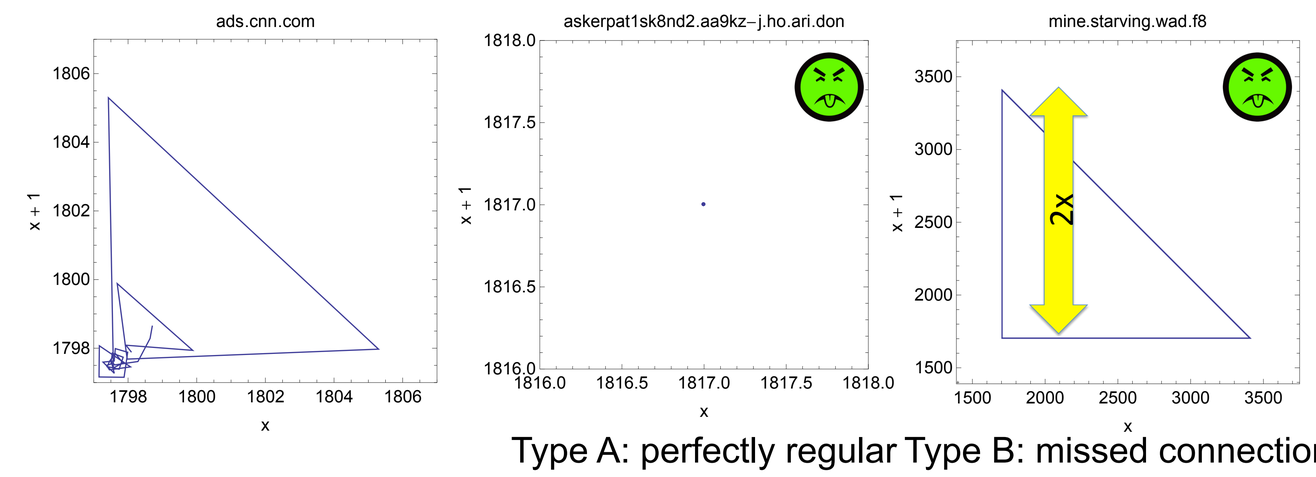
Takens' Theorem (1981): Such plots capture the complete topology of the system trajectory in the underlying (unknown) state space.

Key hunch: "nature writes straight with crooked lines." Dynamics that emerge from legitimate software used by many interacting machines should be qualitatively more complex than beaconing from a trojan.
→ Base time series: delays between successive connections

Known, Benign Examples



Regular Connections



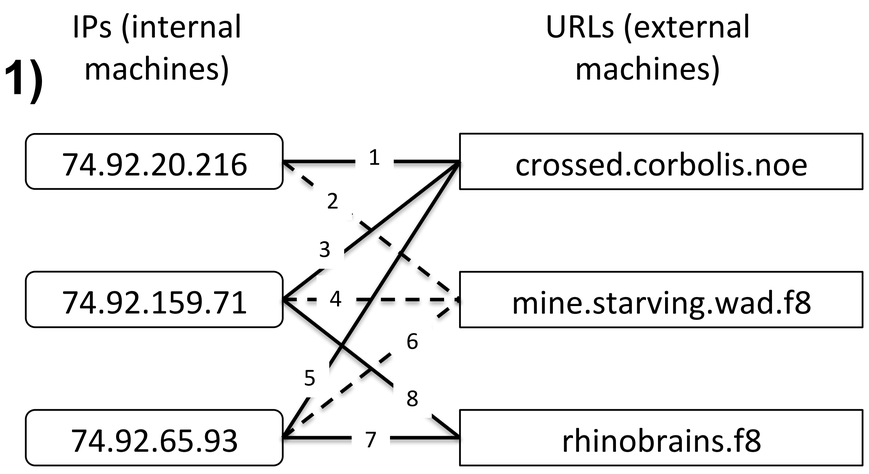
Type A: perfectly regular Type B: missed connection

Performance

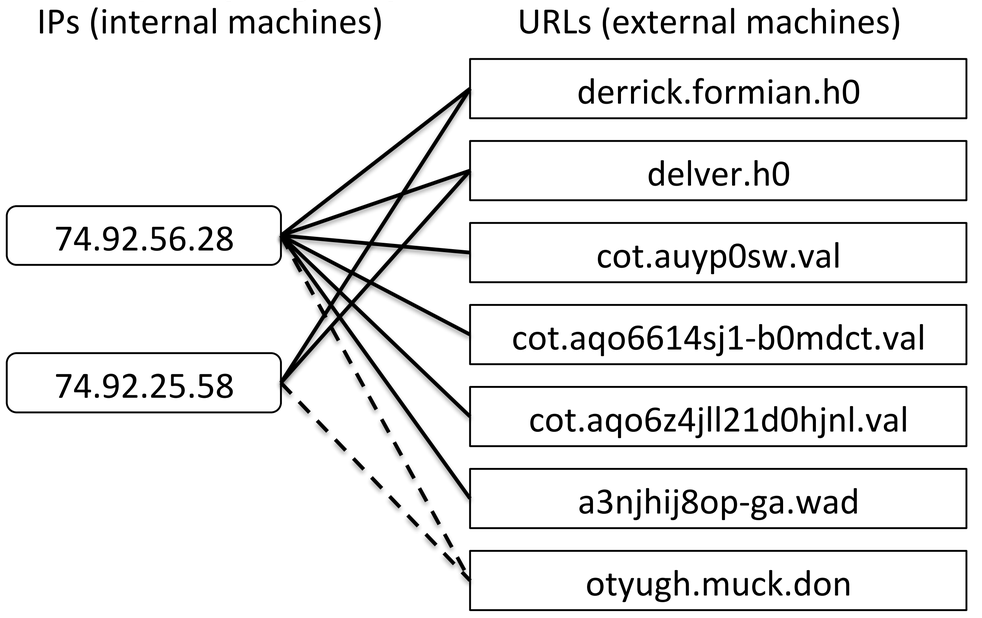
Summary

			Documented Attacks: Yes: IP	Yes: Predecessors	Yes: Beacons	Yes: Successors	No: IP	No: Predecessors	No: Beacons	No: Successors
DUDE	Hit	Without Hints	15	10	15	3	11	325[1]	21	249[2]
		With Hints	14	10			0	0		
	Miss		5	17			?			
Total			34	55			?			

Example (Day 11)



Example (Day 22)



[1] These are associated with only 9 beacons. # of predecessors/beacon = {264, 31, 18, 6, 2, 2, 2, 1, 1}
[2] These are associated with only 6 beacons. # of successors/beacon = {227, 12, 7, 3, 1, 1}. The beacon with 227 successors is the same as the one with 264 prececessors.