Botnets: detection & mitigation

I. Pustogarov

Measurement and detection Techniques

- Passive techniques
  - Data collected from observations – honeypots
  - Does not interfere with botnet activities
    - Transparent to botmaster
- Active techniques
  - Actively interact with the botnet to understand it
  - Better understanding and measurements
  - Bot activities may be disrupted – detectable by botmaster
  - Researchers may be targeted by botmaster
- Reverse engineering of bots
  - Several anti-reverse engineering techniques are used: obfuscation, encryption, dynamic updating

Passive techniques

Packet inspection

- Packet inspection – through detection signatures
- Match protocol fields or packet payload against pre-defined patterns of bot traffic, such as:
  - Packet with shell-code
  - Communication with known malicious IPs
  - Unrelated/unwanted protocol run by a server
    - A file server that suddenly begins to communicate via IRC
- Implemented in: intrusion detection systems (IDS)
  - HIDS & NIDS
  - Also related tool: intrusion prevention systems (IPS)

Packet inspection – drawbacks

- Full packet inspection is costly and not scalable
- Traffic sampling may not capture many important features
- High false negatives: May not detect anything beyond the signature database; multi-packet payload; encrypted payload
- Dealing with false positives is difficult

Passive technique – flow analysis

- Packet payload is not used
  - More scalable than packet inspection
- Flow record attributes are used
  - source and destination address, port numbers
  - protocol used inside the packets
  - the duration of the session
  - cumulative size and number of transmitted packets
- Goal: identify traffic patterns (normal vs. malicious)
- Example: Cisco NetFlow

Example system

- Traffic Aggregation for Malware Detection (TAMD)
  - Yen and Reiter, DIMVA, 2008
  - http://cs.unc.edu/~reiter/papers/2008/DIMVA.pdf

- Features used:
  - flows that communicate with a common destination that is busier than the average of all destinations
  - those that have a similar payload
  - those flows that belong to hosts with a common OS, as most malware is OS-specific

Other passive techniques

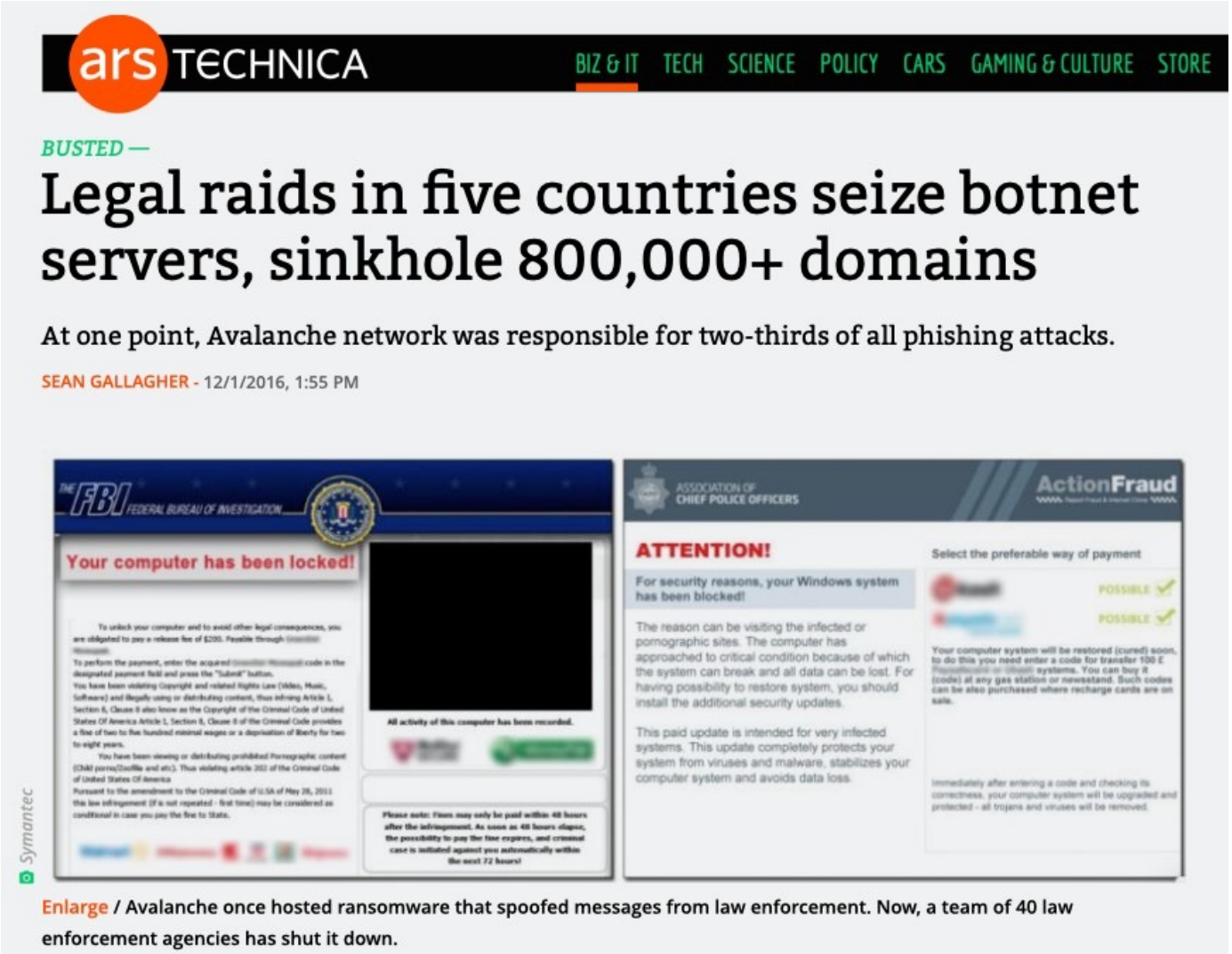
- DNS-based detection/measurement
  - From DNS queries to malicious domains
- Analysis of spam records
  - Indirect technique
  - Spam content, SMTP conversations, email header fields
  - Spam mails are generated from a template
    - How to distinguish the template
- Honeypots

Active techniques

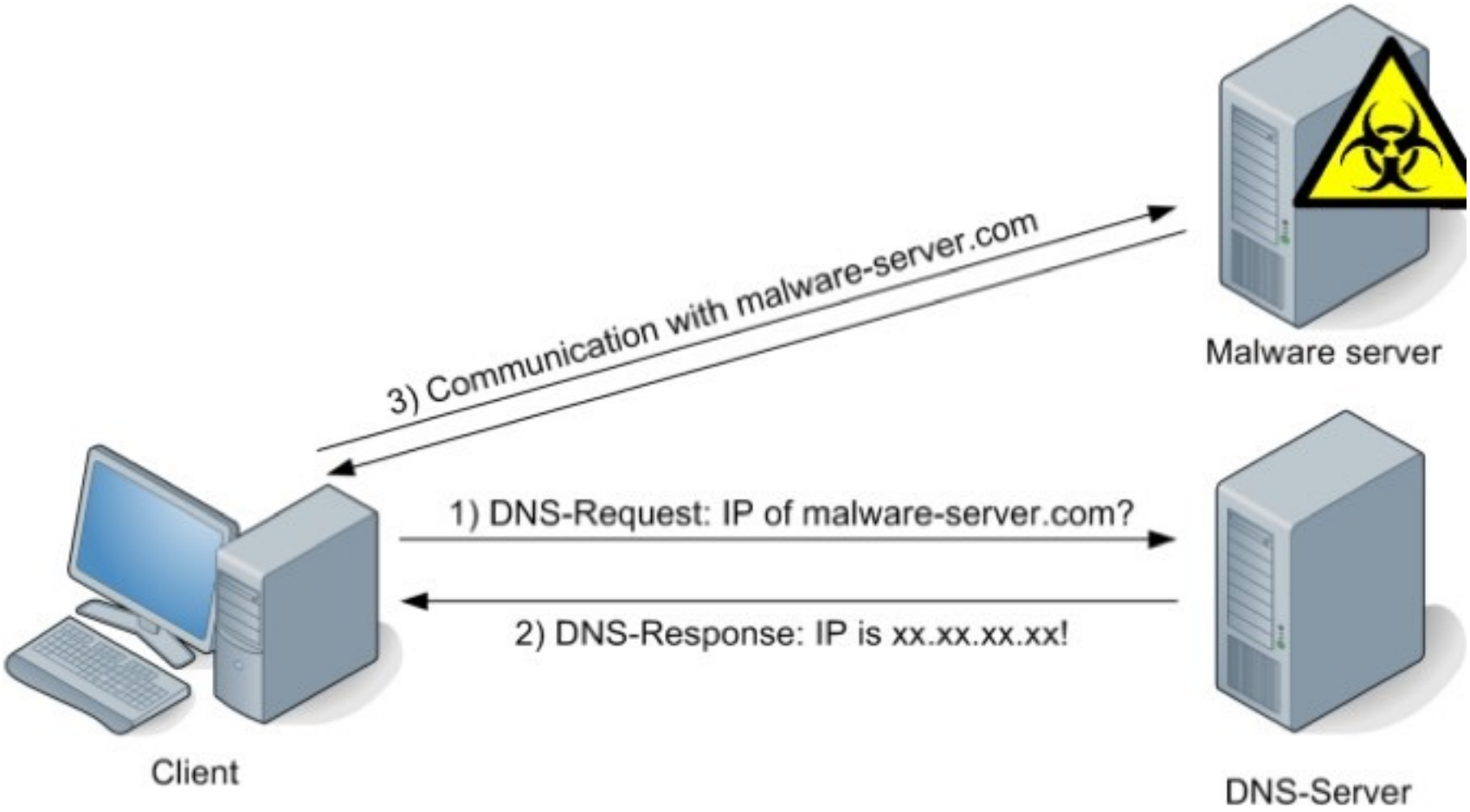
Sinkholing

- Redirecting or dropping traffic destined to a C&C server, malware distribution server, or attack server
- Sinkholing provides a view of the botnet's live population
- Very effective against: botnets, phishing attacks, ad fraud, ransomware etc.
- Example: Stuxtnet measurement by Symantec, WannaCry ransomware kill-switch, Microsoft used this many times (see: https://www.wired.com/story/microsoft-russia-fancy-bear-hackers-sinkhole-phishing/)
- Running a sinkhole may get tricky
  - What to do with sensitive data (ID theft, government data)

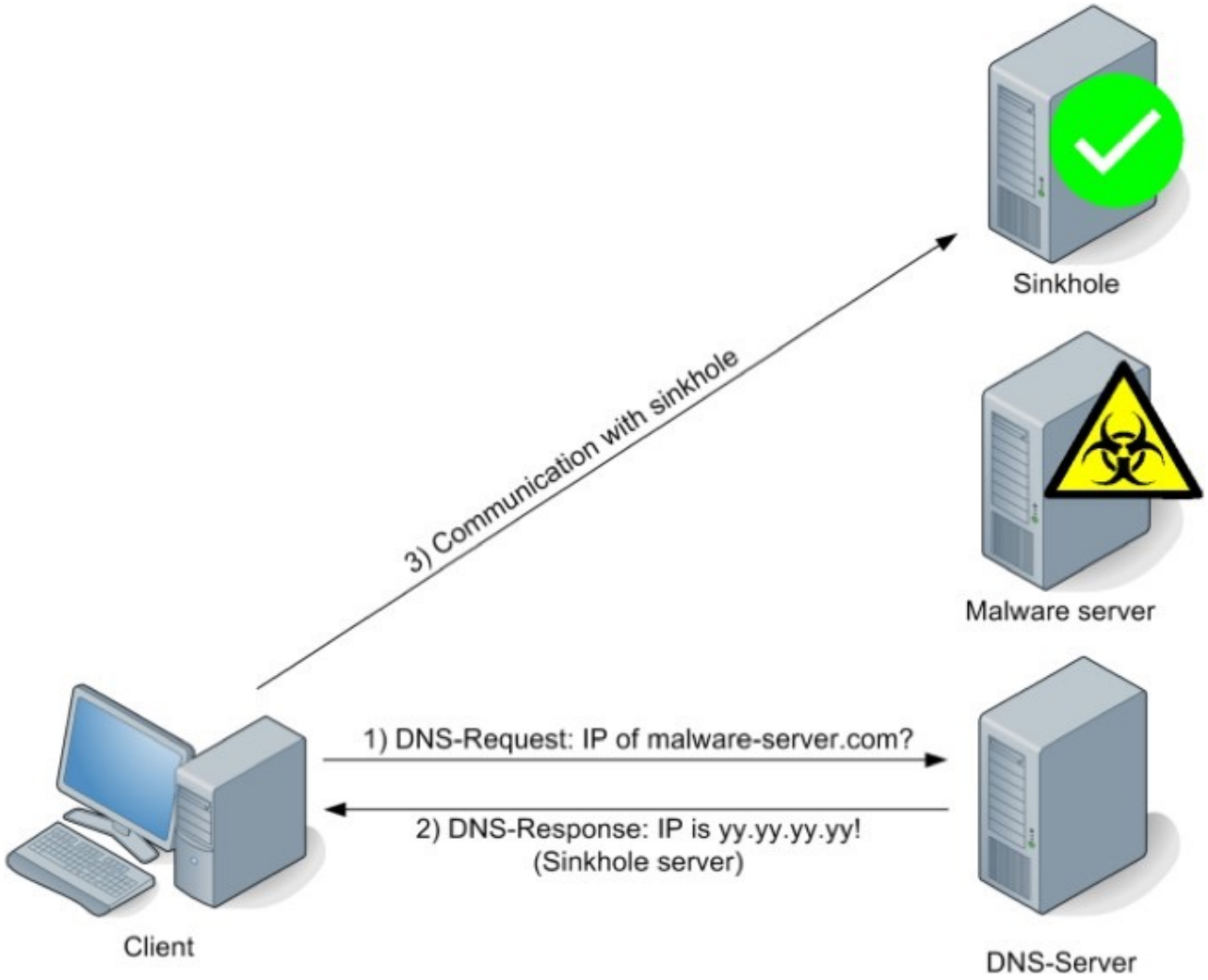
A large-scale operation from 2016



Sinkholing – before redirection



Sinkholing – after redirection



Sinkholing of WannaCry

- Reverse engineering of the binary revealed a special domain checked by the ransomware code: iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com
- The malware continues its operation as long as this gibberish domain remained unregistered
- The domain was bought ($10.69) by MalwareTech and sinkholed
- The IP addresses of infected machines were shared with corresponding companies

Infiltration

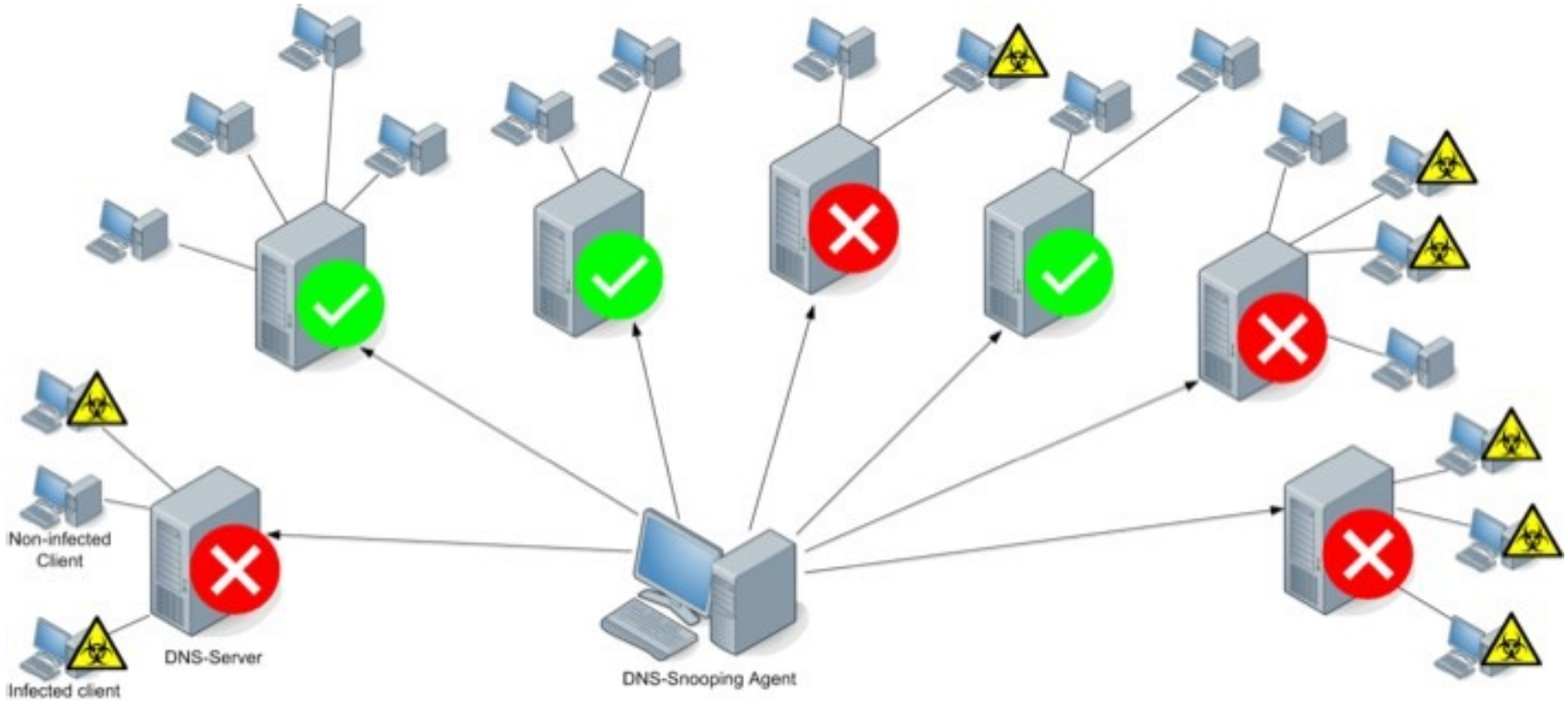
- Software-based
  - Learning from inside
  - Take control of a botnet (protocol reverse engineering, exploiting "vulnerabilities")
- Hardware-based
  - Access to the ISP that's hosting C&C servers
  - Can monitor all traffic to/from servers

DNS cache-snooping

- DNS caching
  - DNS query for an unknown domain: the DNS server will forward the query towards the responsible authoritative name server
  - Store the resulting data record in a local cache
    - Future queries will be served from this cached entry
  - Caching increases the performance of a name server

- Caching for detection/measurement:
  - check indirectly if a target domain has been queried through a specific domain server by testing if a cached answer is stored

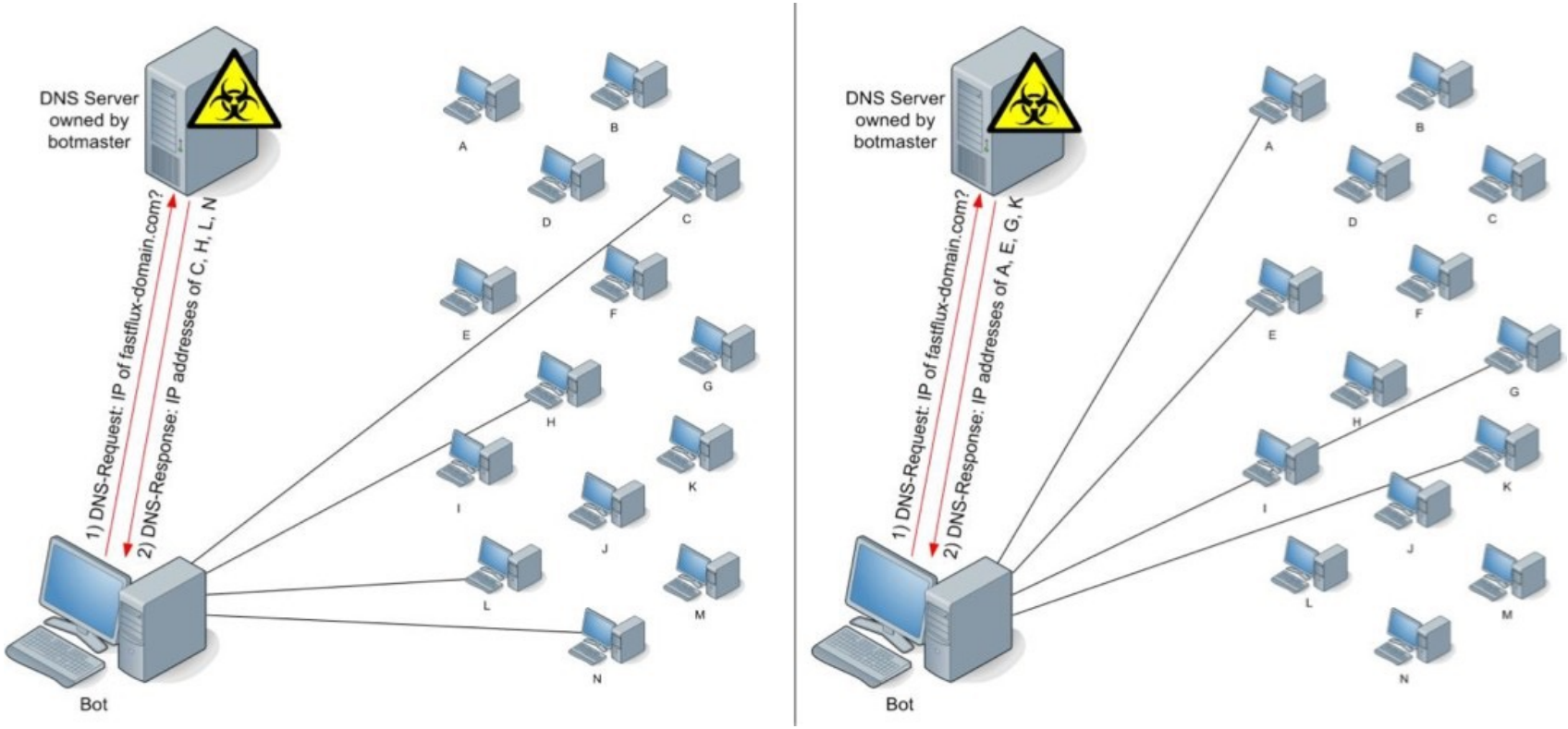
Cache snooping – overview



Tracking of fast-flux networks

- Use their own features against them!
  - Domains with short TTL
  - IP addresses widely vary across queries
  - IPs spread out across many ISPs
    - no topographical relationship – unlike real domains that use similar features

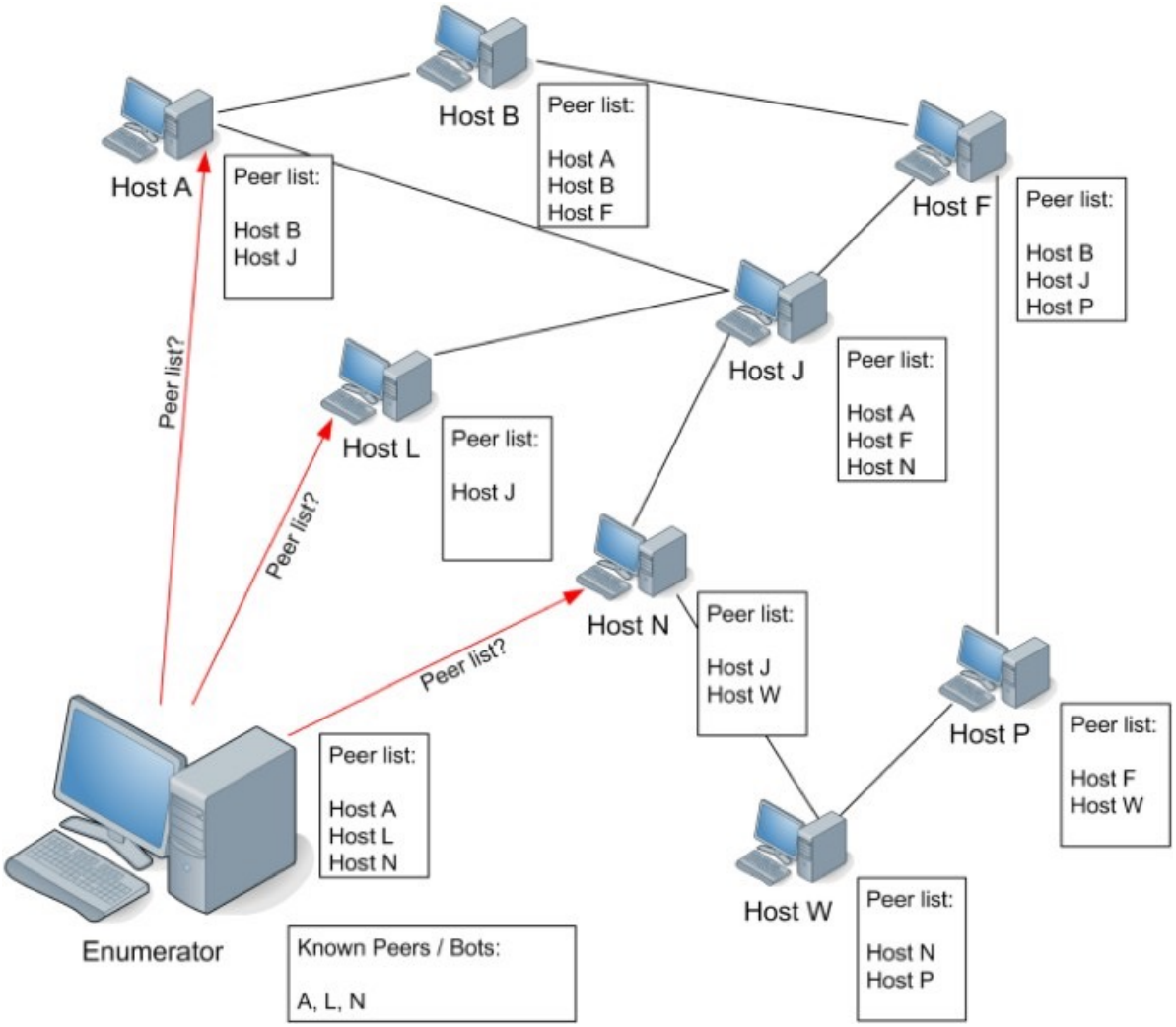
Fast-flux networks – different IPs are returned



Tracking P2P botnets

- Robust against tracking – only few peers are known to each bot
  - No central server to contact/track
  - Used often: Storm, Waledac, Conficker
- Can still be measured
  - Recursive request of peer list (if available/supported by the protocol used)

Recursive peer request



Recursive peer request (cont.)