

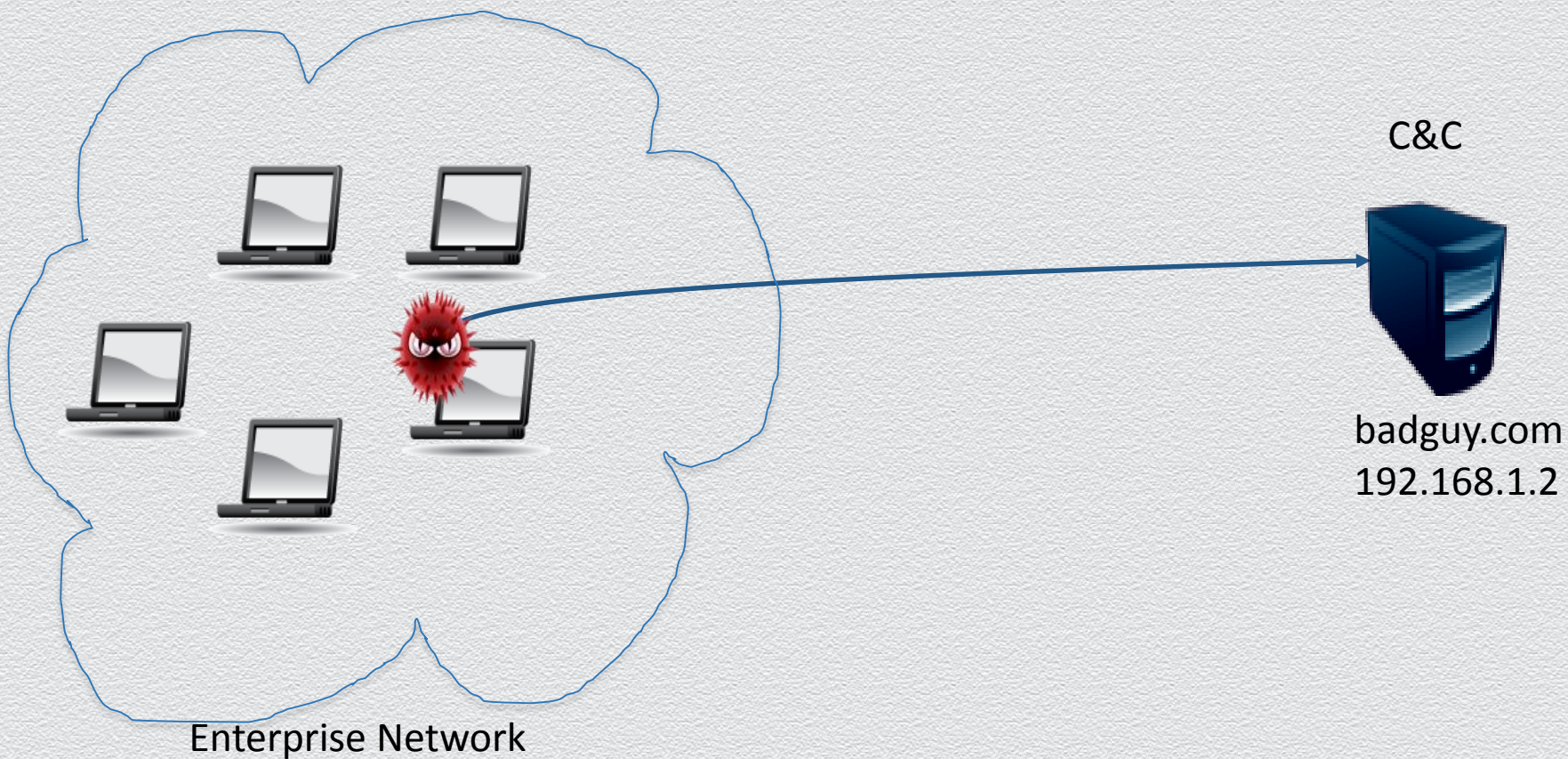
Learning Malware Languages: Fun with Dick and Jane's Malware

SESSION ID: TRM-W06

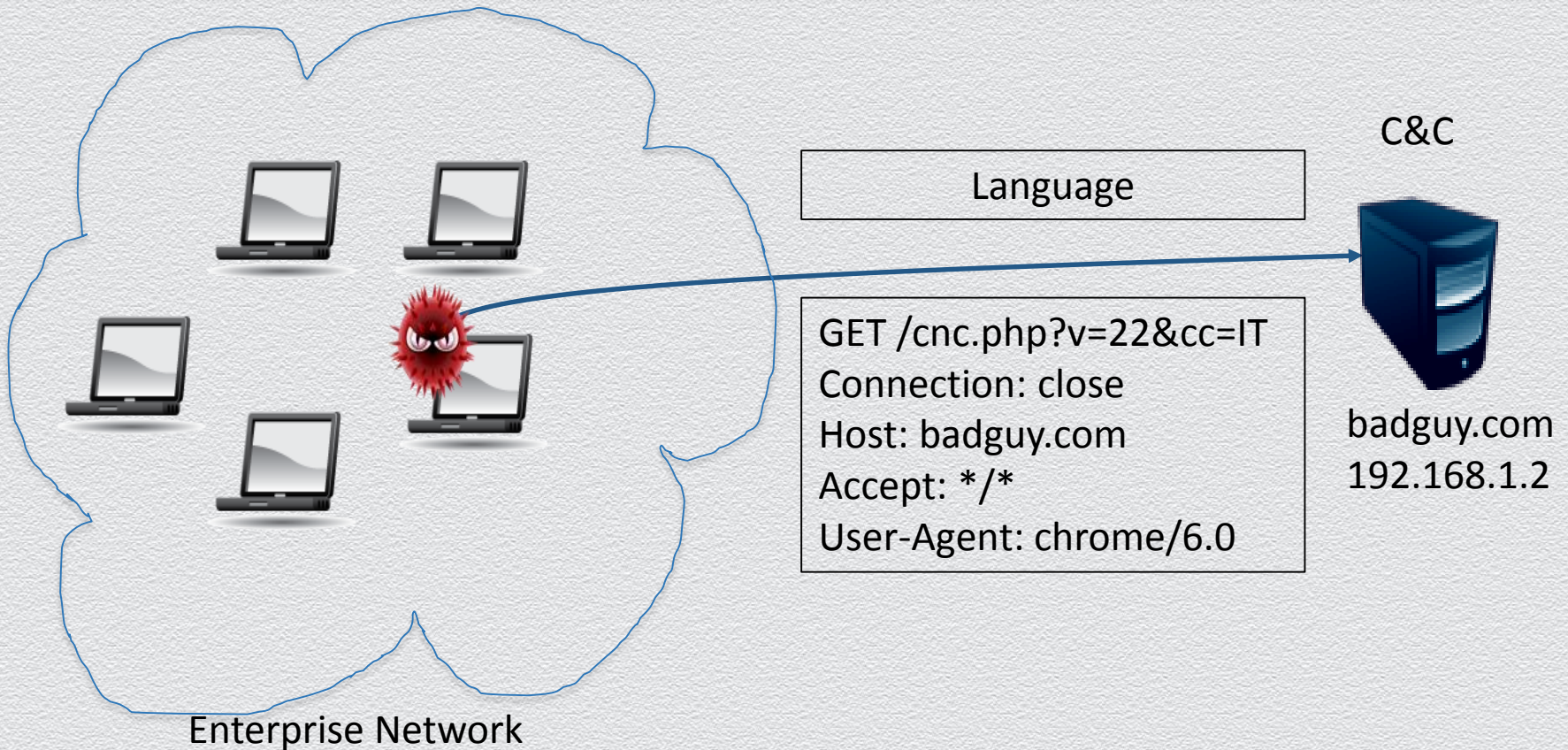
Brian Foster
Chief Technology Officer
Damballa

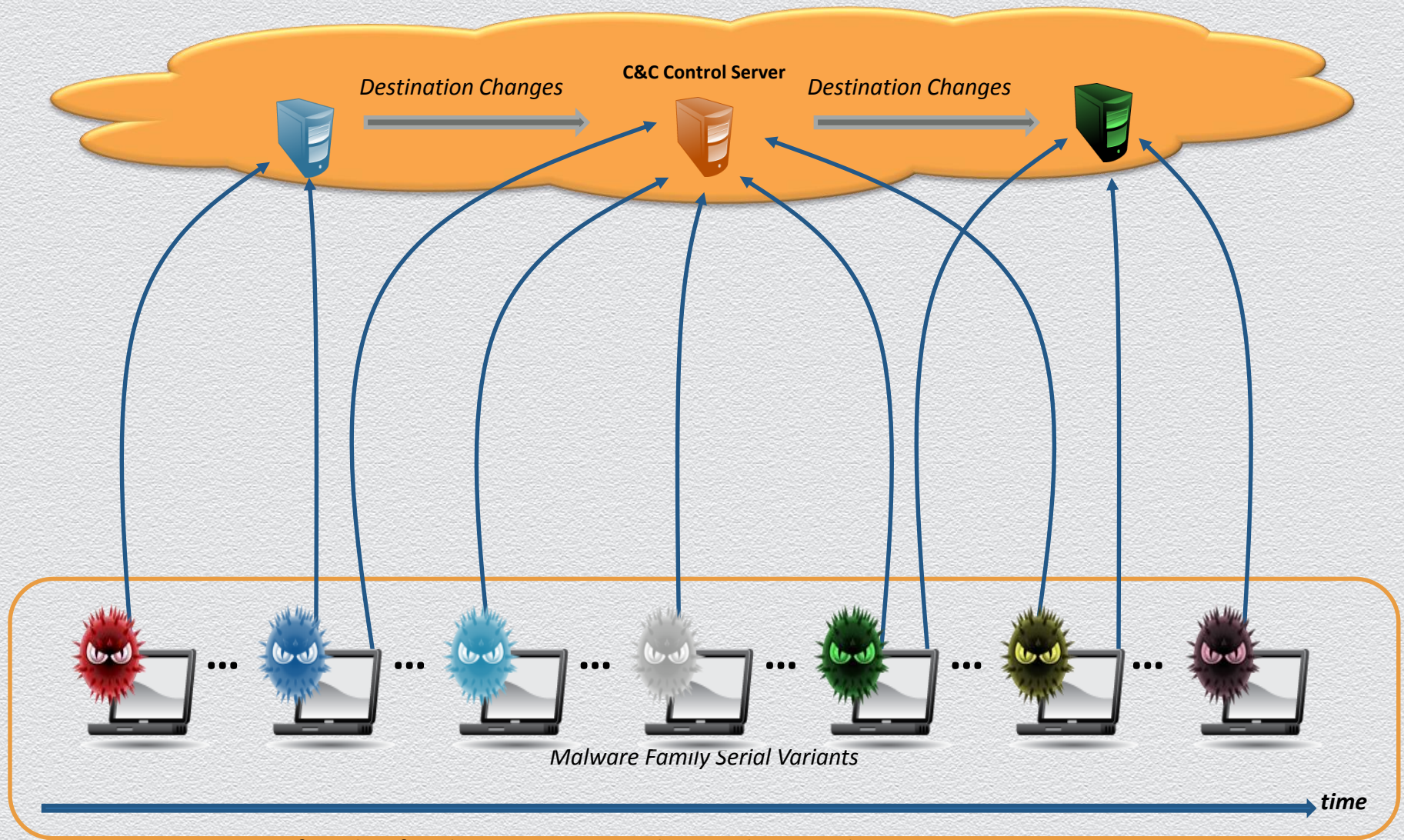


Malware Languages

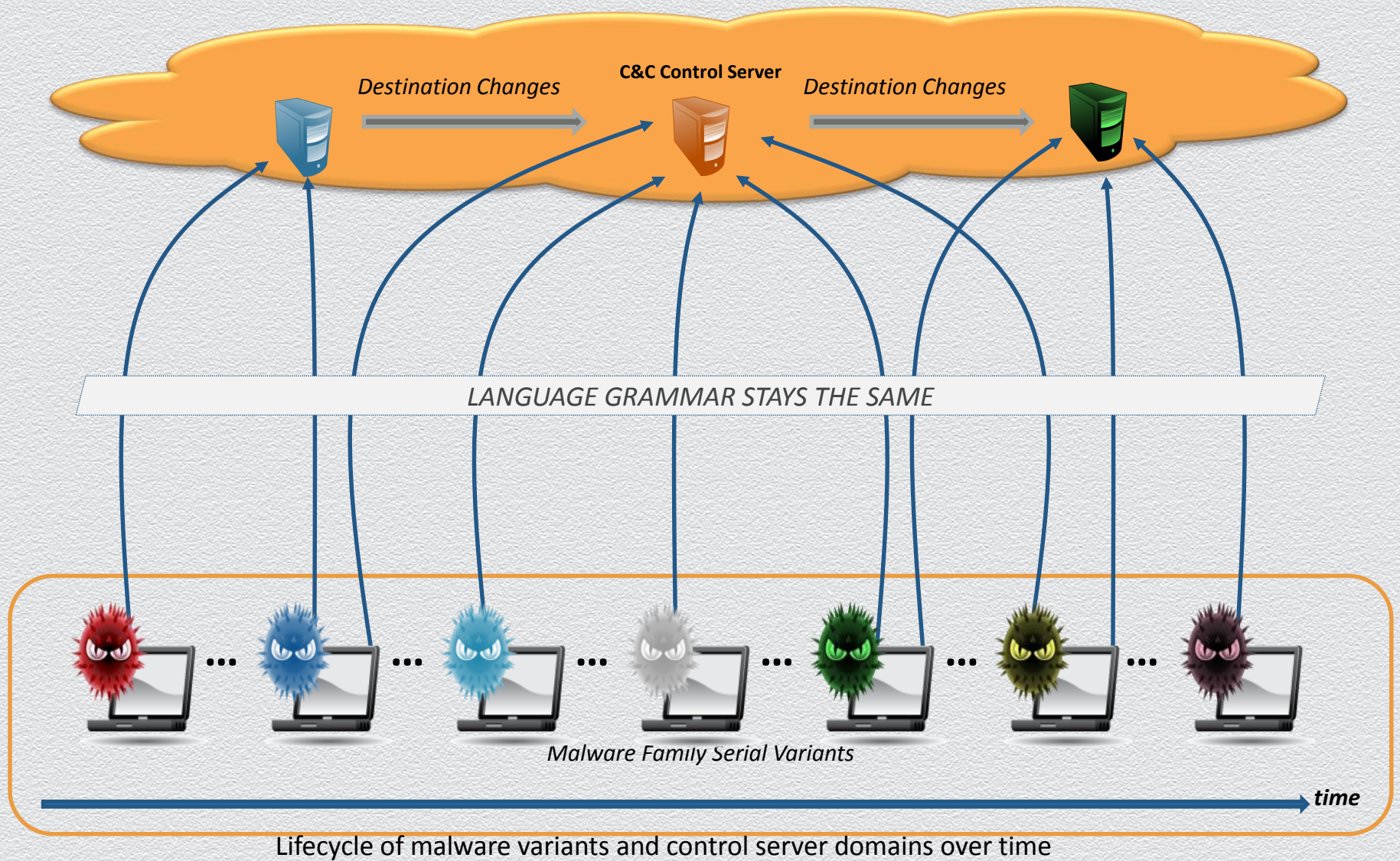


Malware Languages

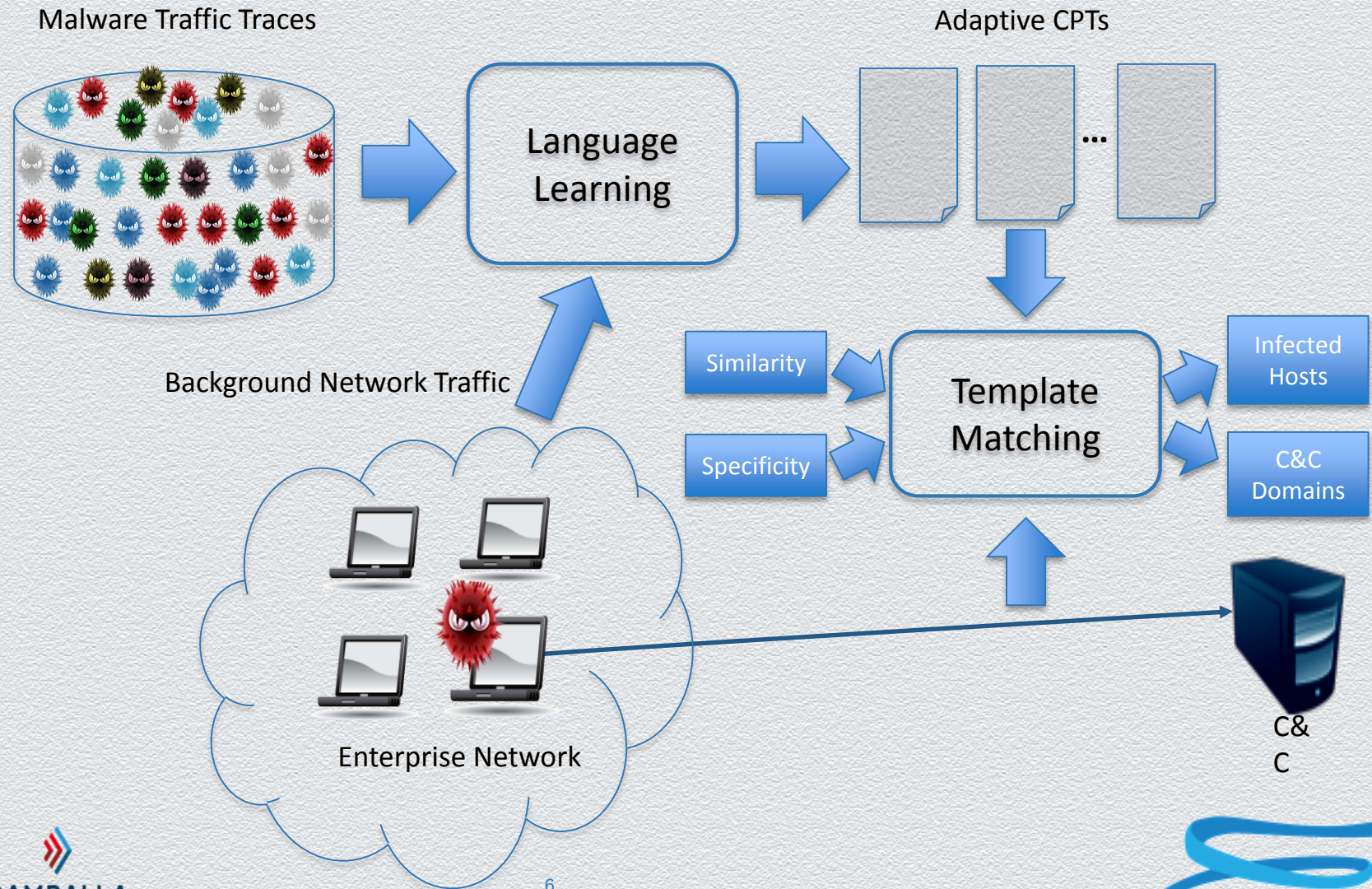




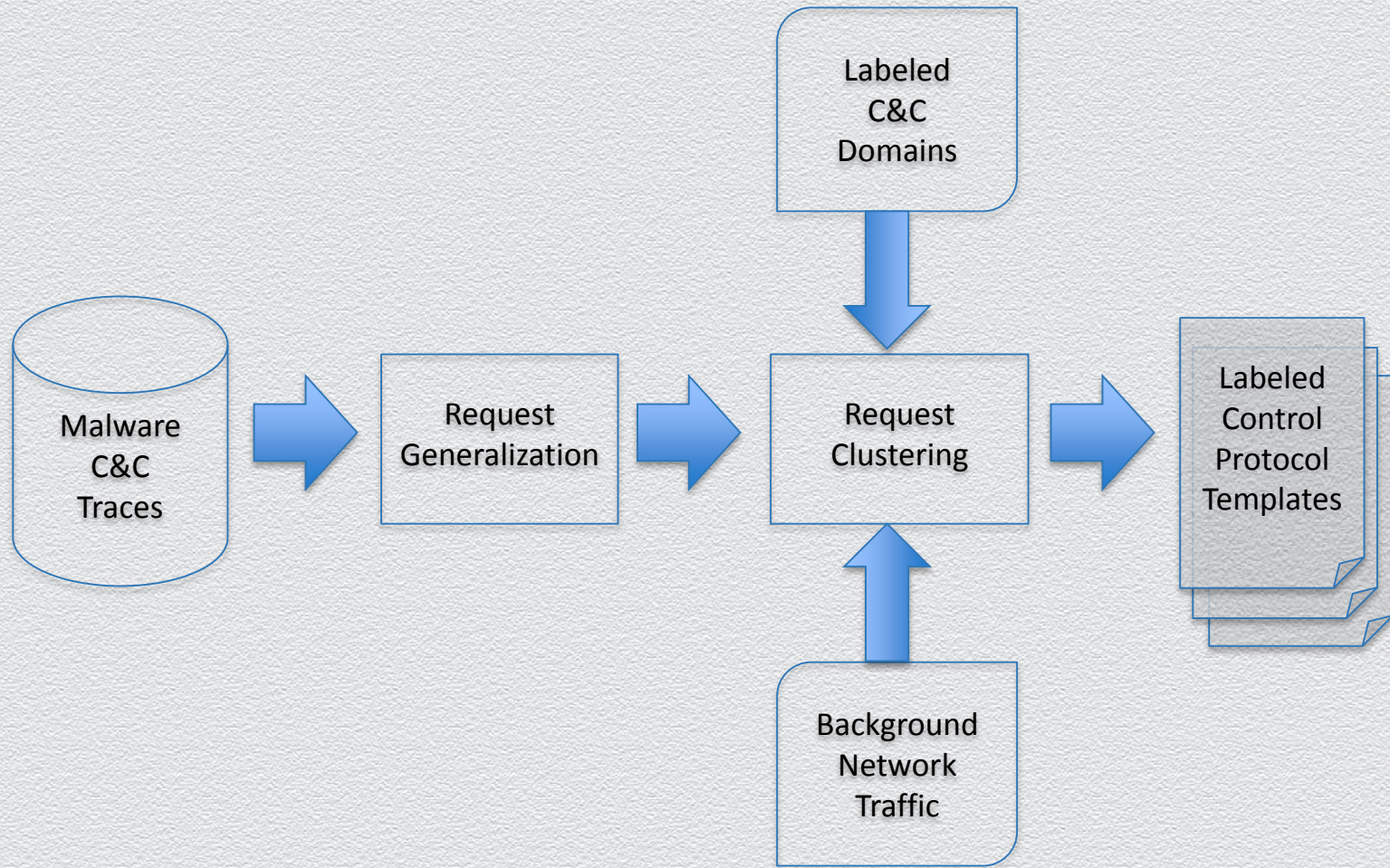
Lifecycle of malware variants and control server domains over time



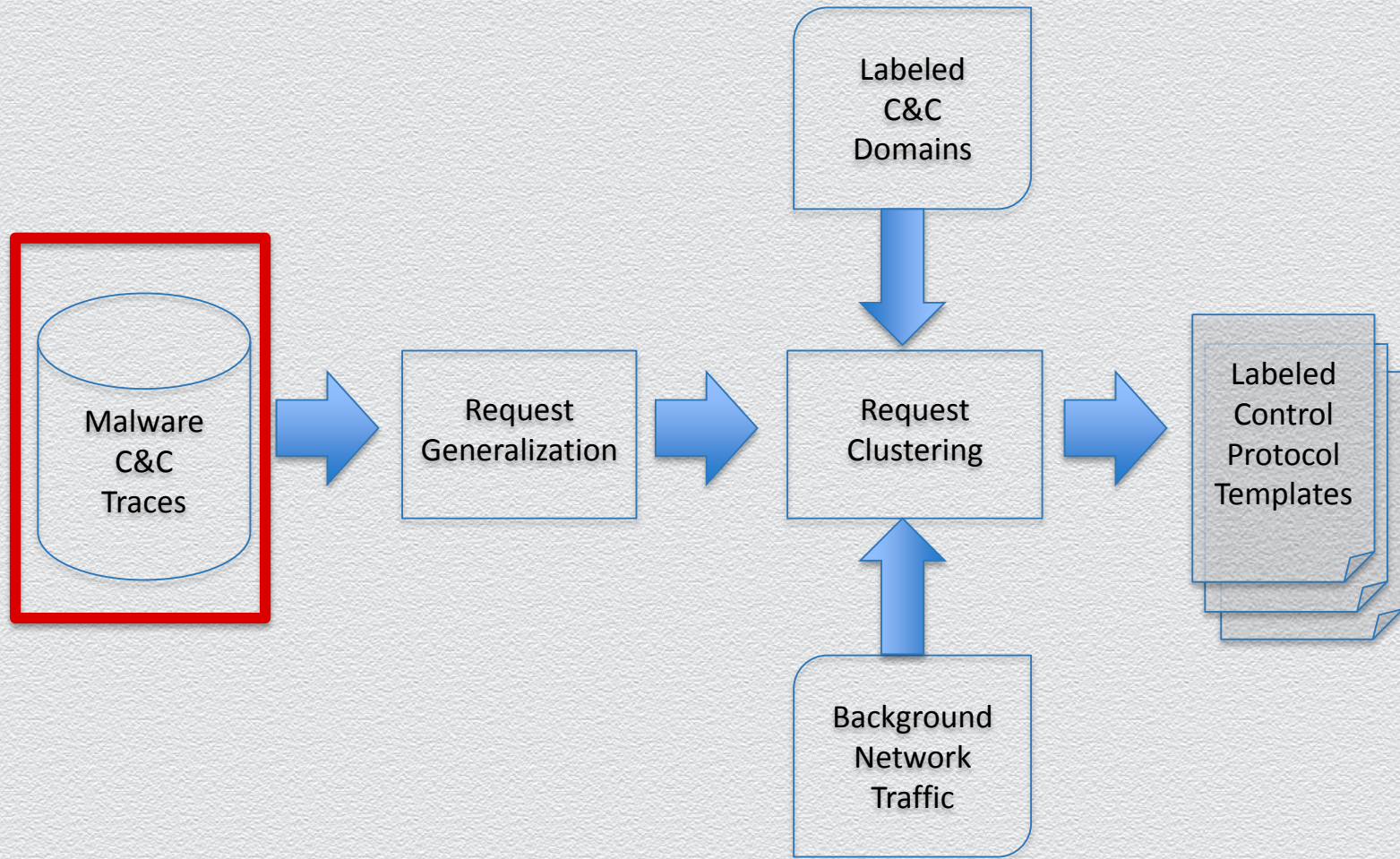
Learning & Recognizing Malware Languages



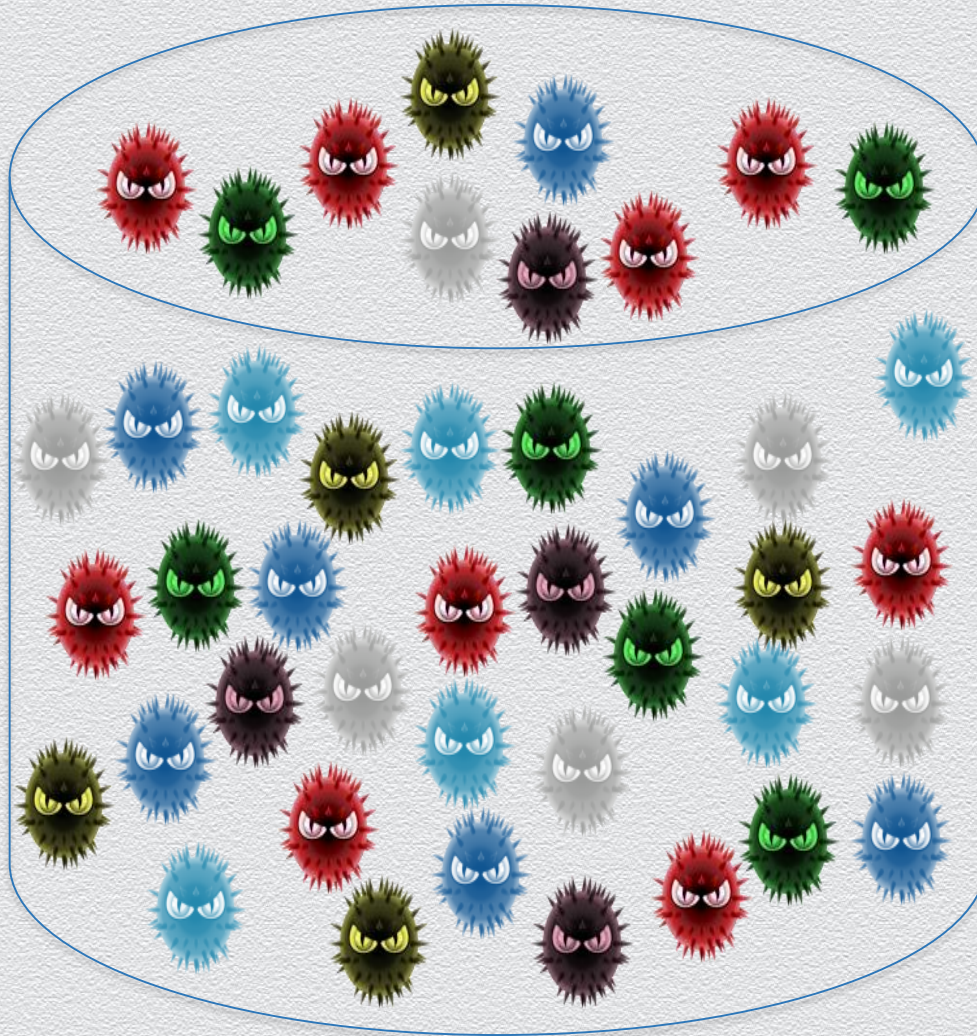
Learning Malware Languages



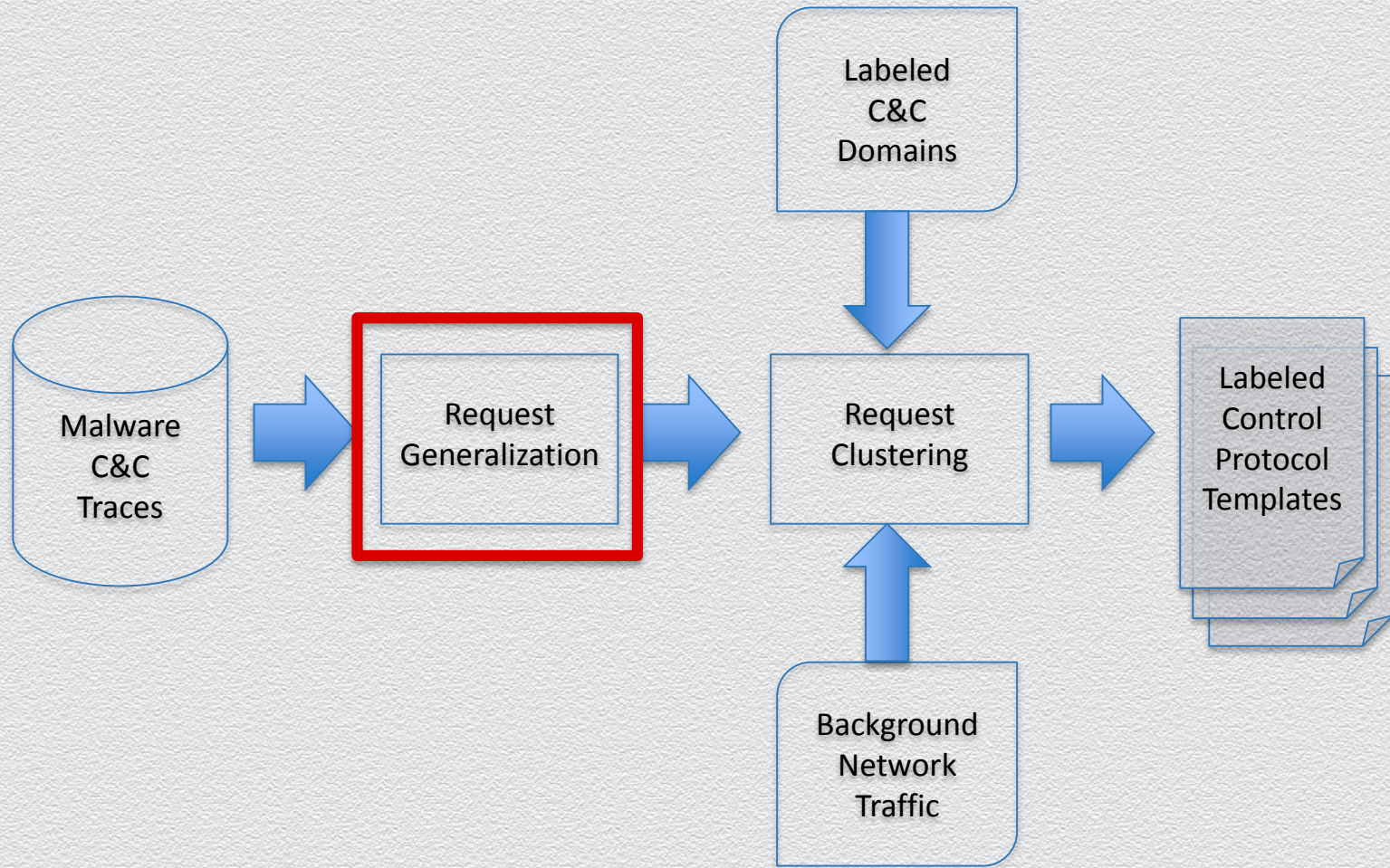
Learning Malware Languages



C:/Malware/Run



Learning Malware Languages



Request Generalization

Request 1:

GET /Ym90bmq=/cnc.php?v=220&cc=IT

Host: www.bot.net

User-Agent: 680e4a9a

Request 2:

GET /bWFsd2F=/cnc.php?v=139&cc=US

Host: www.malwa.re

User-Agent: dae4a661

Request Generalization

Request 1:

GET /Ym90bmQ=/cnc.php?v=220&cc=IT
Host: www.bot.net
User-Agent: 680e4a9a

Request 2:

GET /bWFsd2F=/cnc.php?v=139&cc=US
Host: www.malwa.re
User-Agent: dae4a661

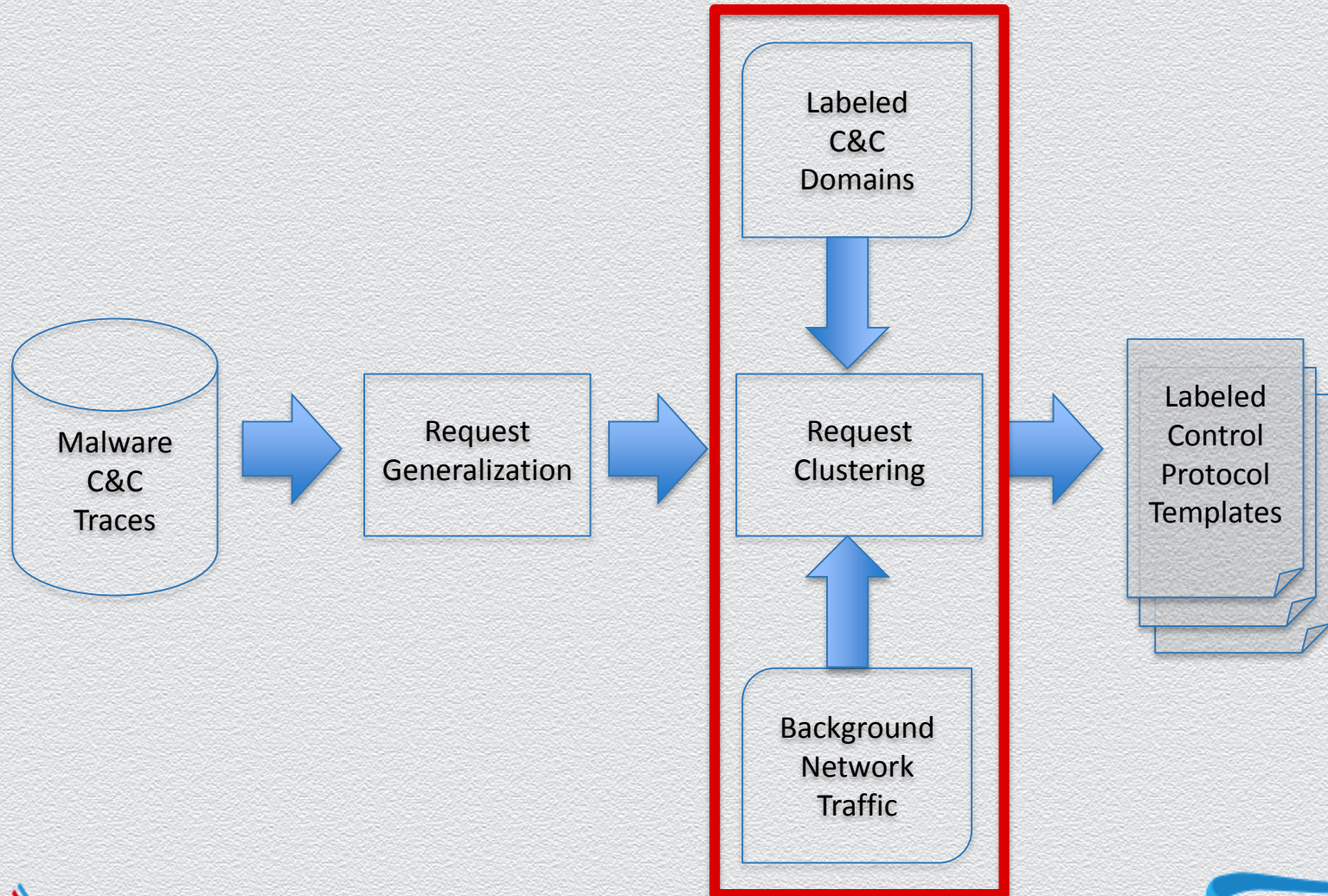
Generalized Request 1:

GET /<base64,8>/cnc.php?v=<int,3>&cc=<str,2>
Host: www.bot.net
User-Agent: <hex,8>

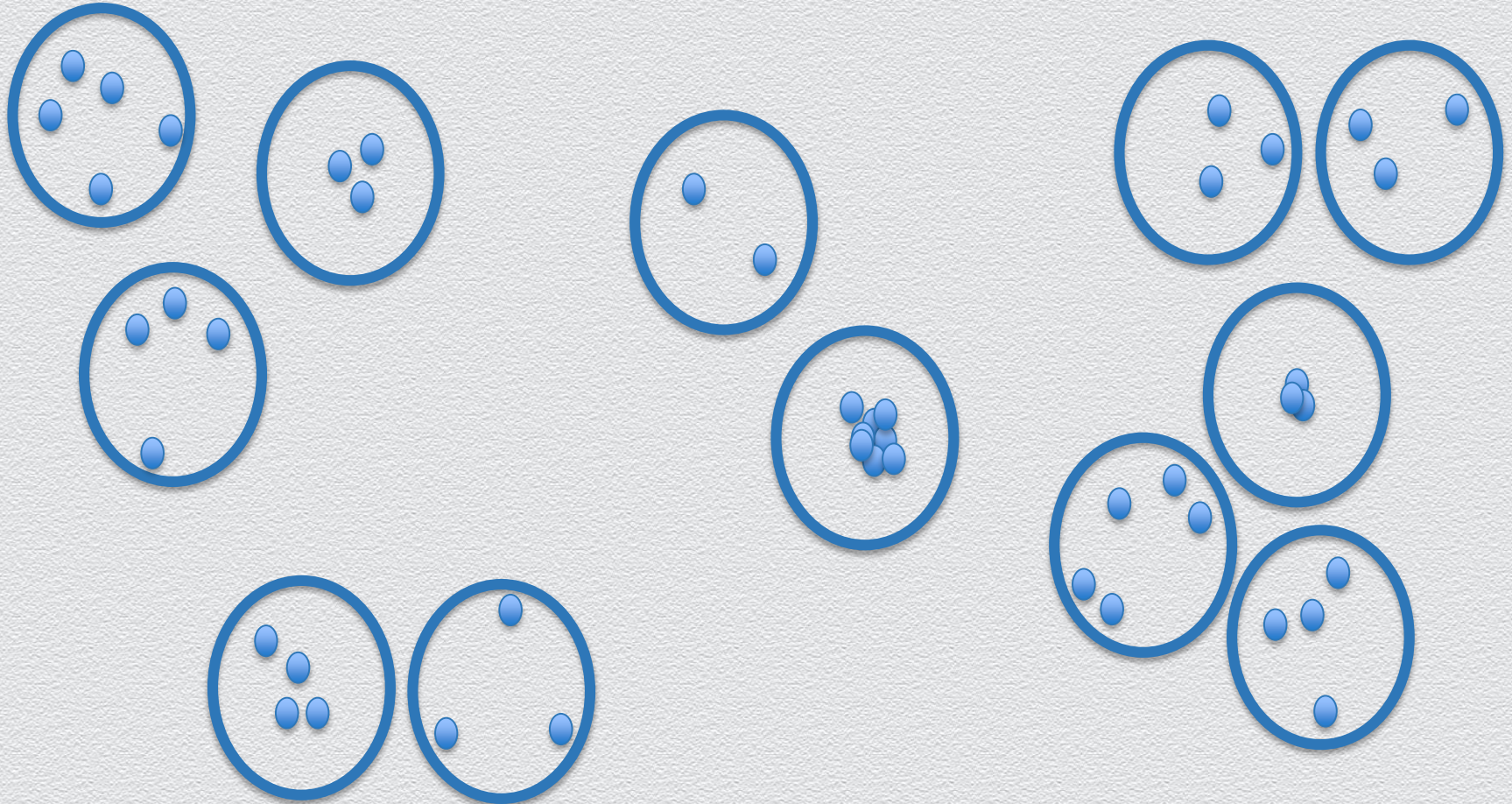
Generalized Request 2:

GET /<base64,8>/cnc.php?v=<int,3>&cc=<str,2>
Host: www.malwa.re
User-Agent: <hex,8>

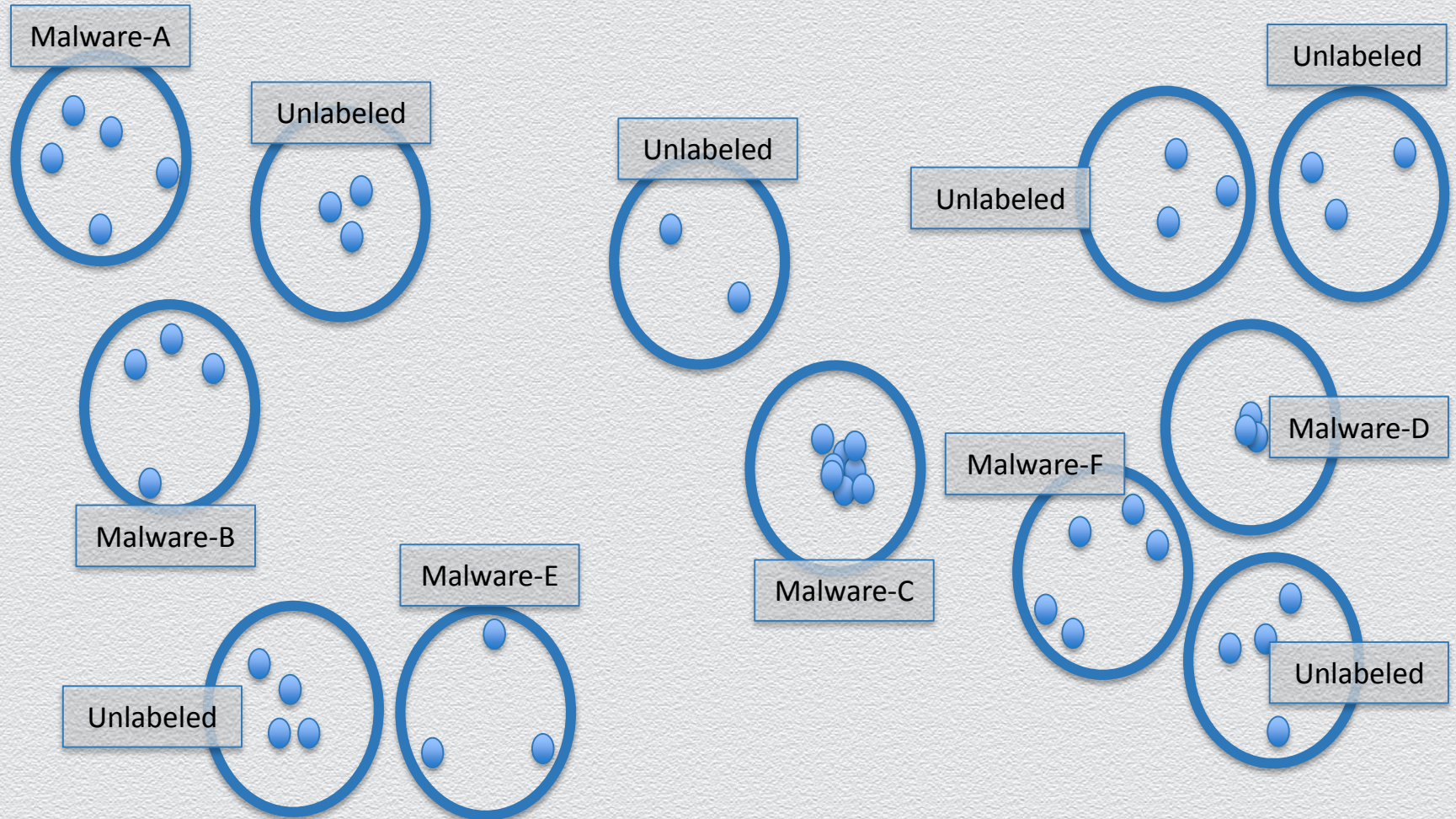
Learning Malware Languages



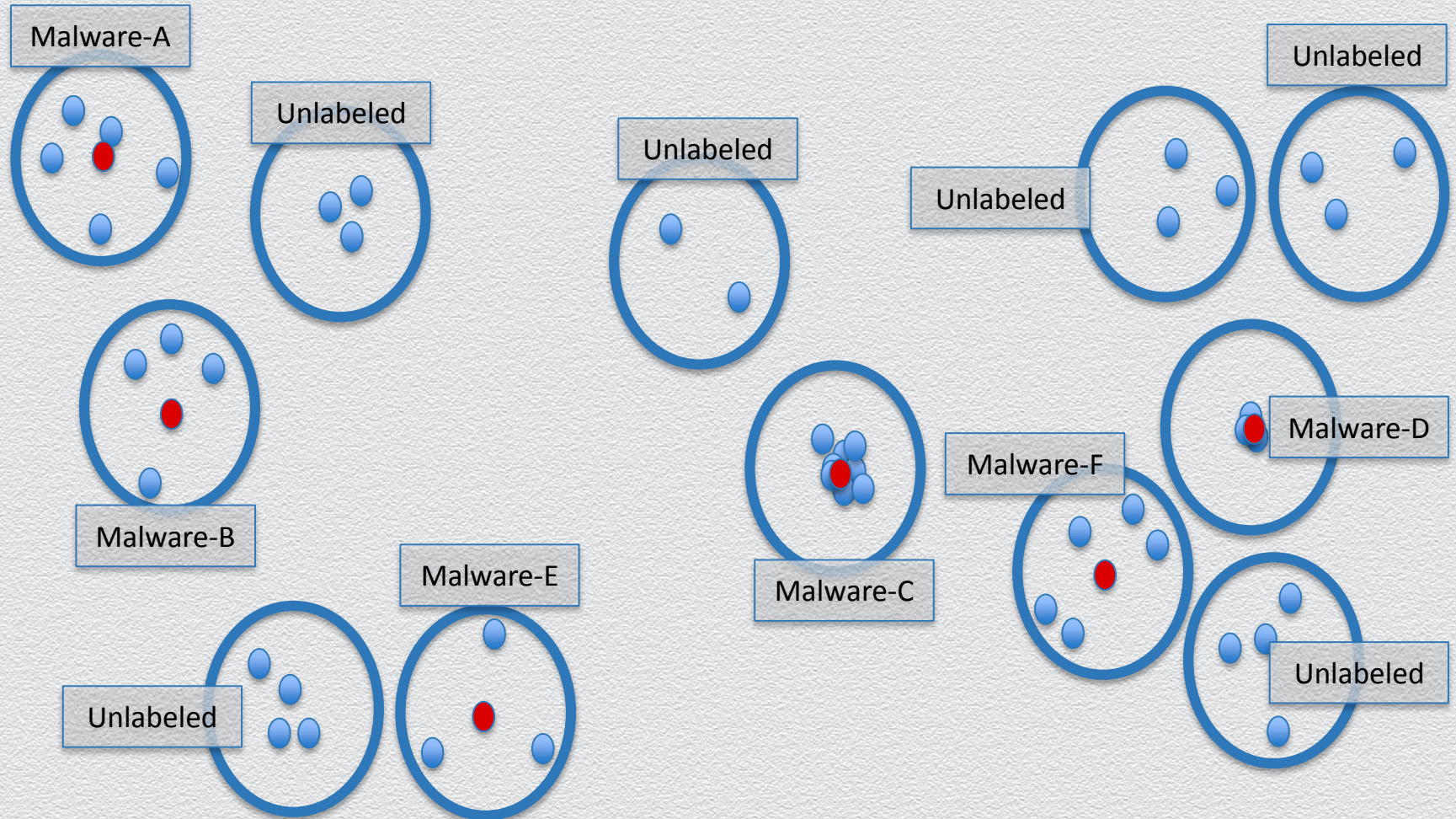
Request Clustering



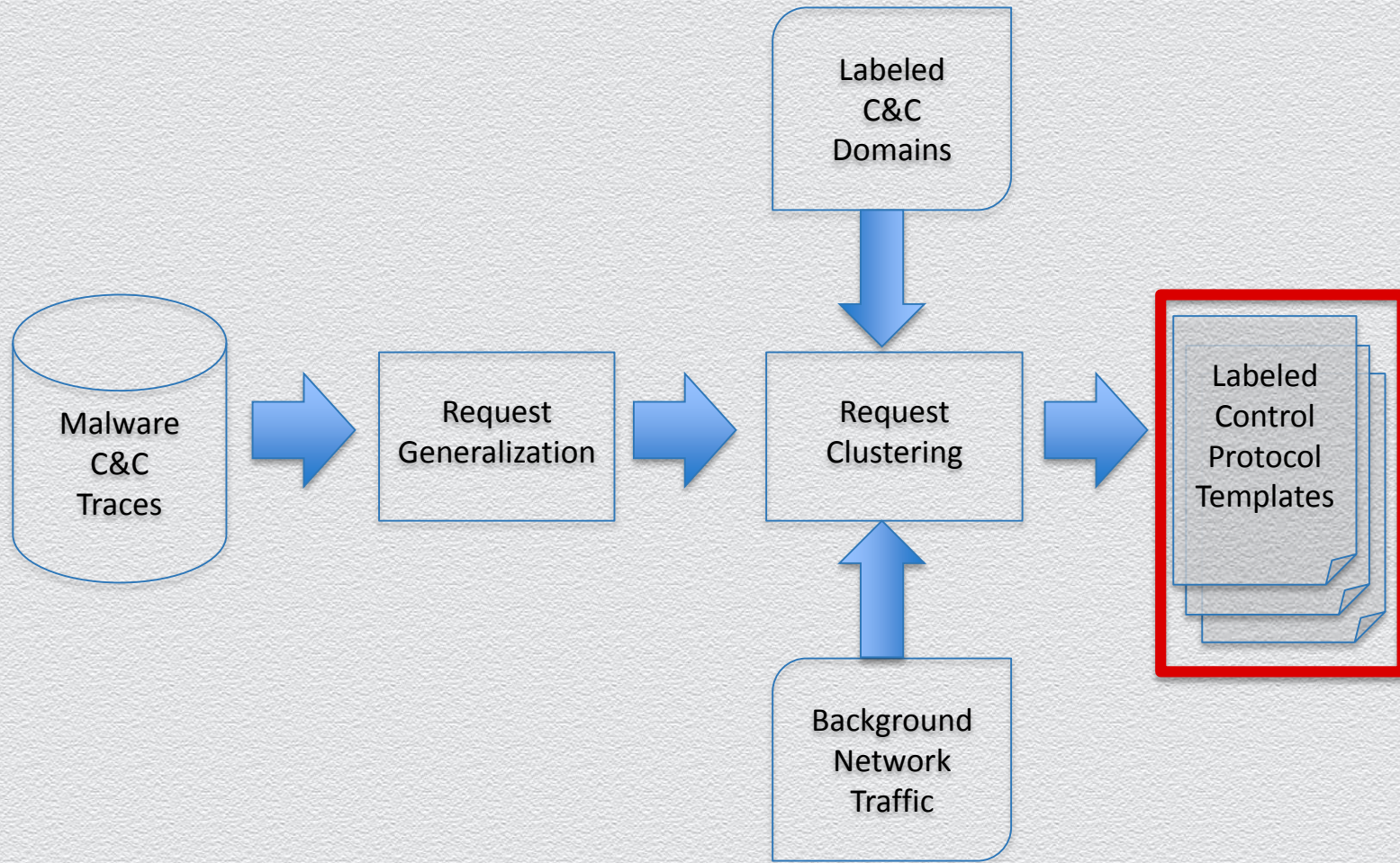
Request Clustering



Request Clustering



Learning Malware Languages

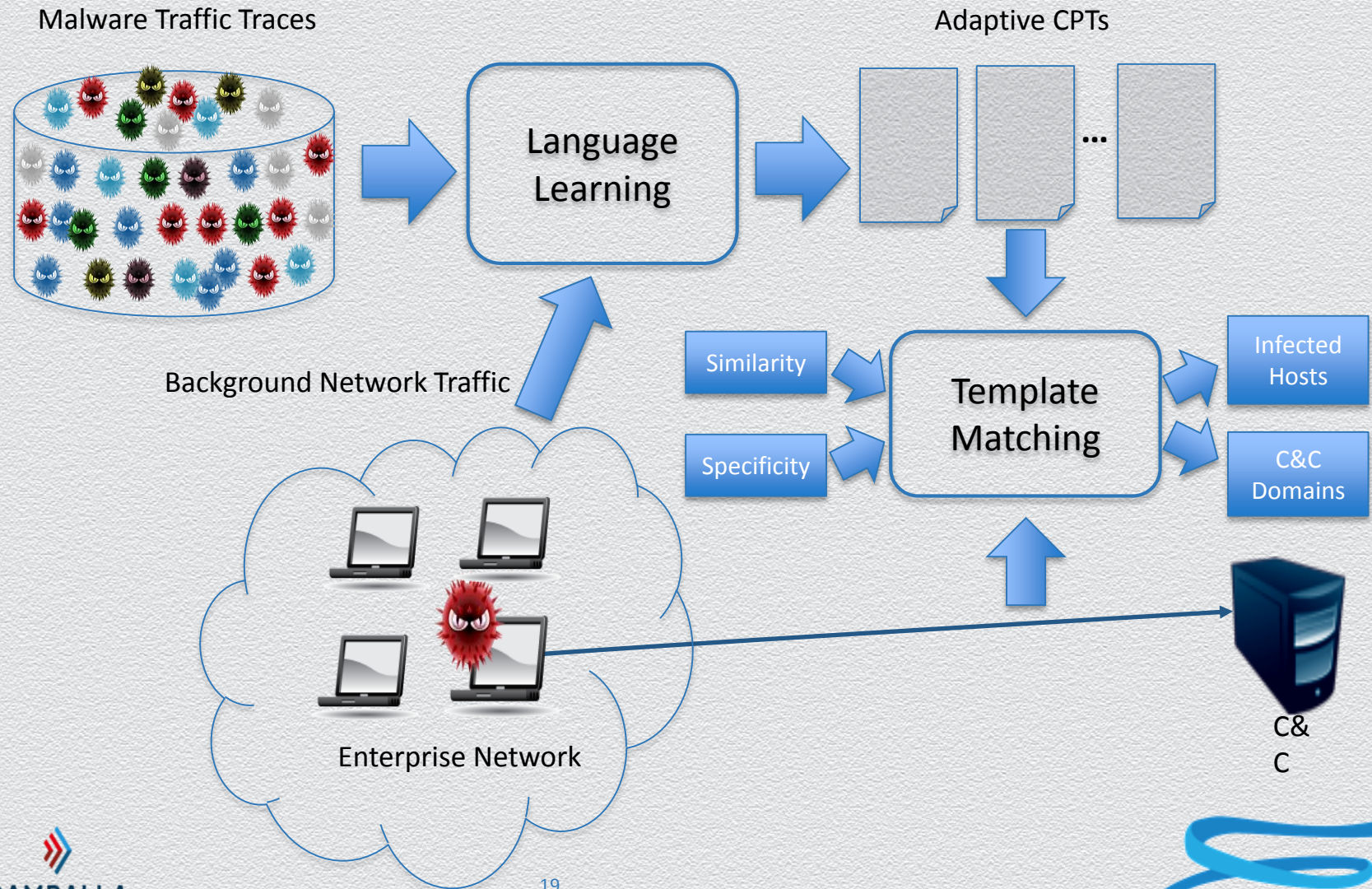


Labeled Control Protocol Template

- ◆ τ_1) **Median URL path:** /<base64,8>/cnc.php.
- ◆ τ_2) **URL query component:** {v=<int,3>, cc=str, 2}.
- ◆ τ_3) **User-Agent:** <hex,8>.
- ◆ τ_4) **HTTP Headers:** {(Host,12)}.
- ◆ τ_5) **DST nets:** {172.16.8.0/24, 10.10.10.0/24, 192.168.4.0/24}.

Malware Family: {Trojan-A, BotFamily-1}

Learning & Recognizing Malware Languages



Detecting Infections

- ◆ Similarity
 - ◆ Measures likeness
 - ◆ Weighted average
 - ◆ Match Threshold
- ◆ Specificity
 - ◆ Measures uniqueness
 - ◆ Dynamic weighting
 - ◆ Self-tuning

Input: req, CPT

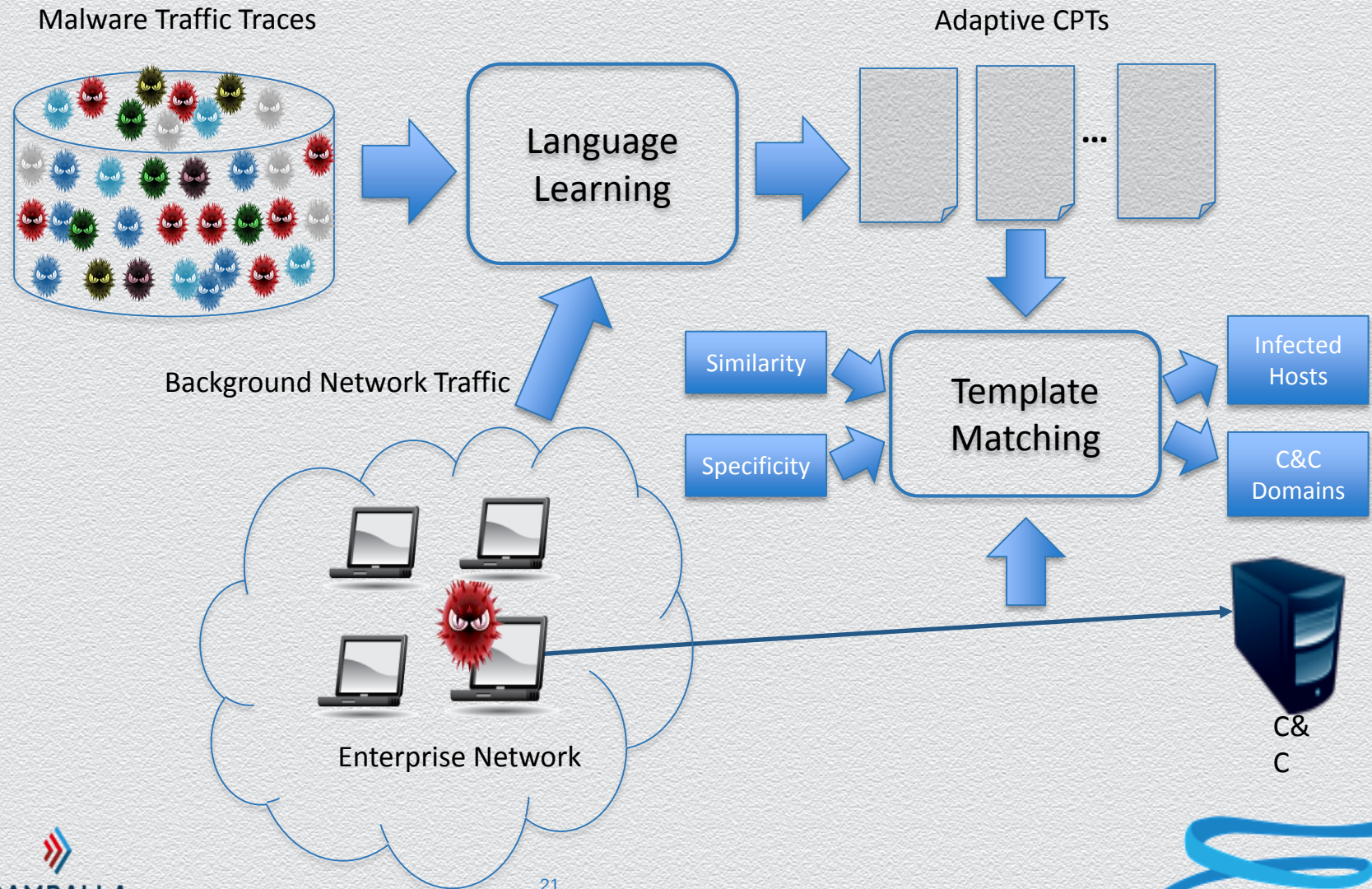
Similarity: $s(\text{req}_i, \text{CPT}_i)$,
for each component i

Specificity: $\sigma(\text{req}_i, \text{CPT}_i)$,
for each component i

Match-Score: $f(\text{sim}, \text{spec})$

If Match-Score $> \Theta$:
return C&C Request

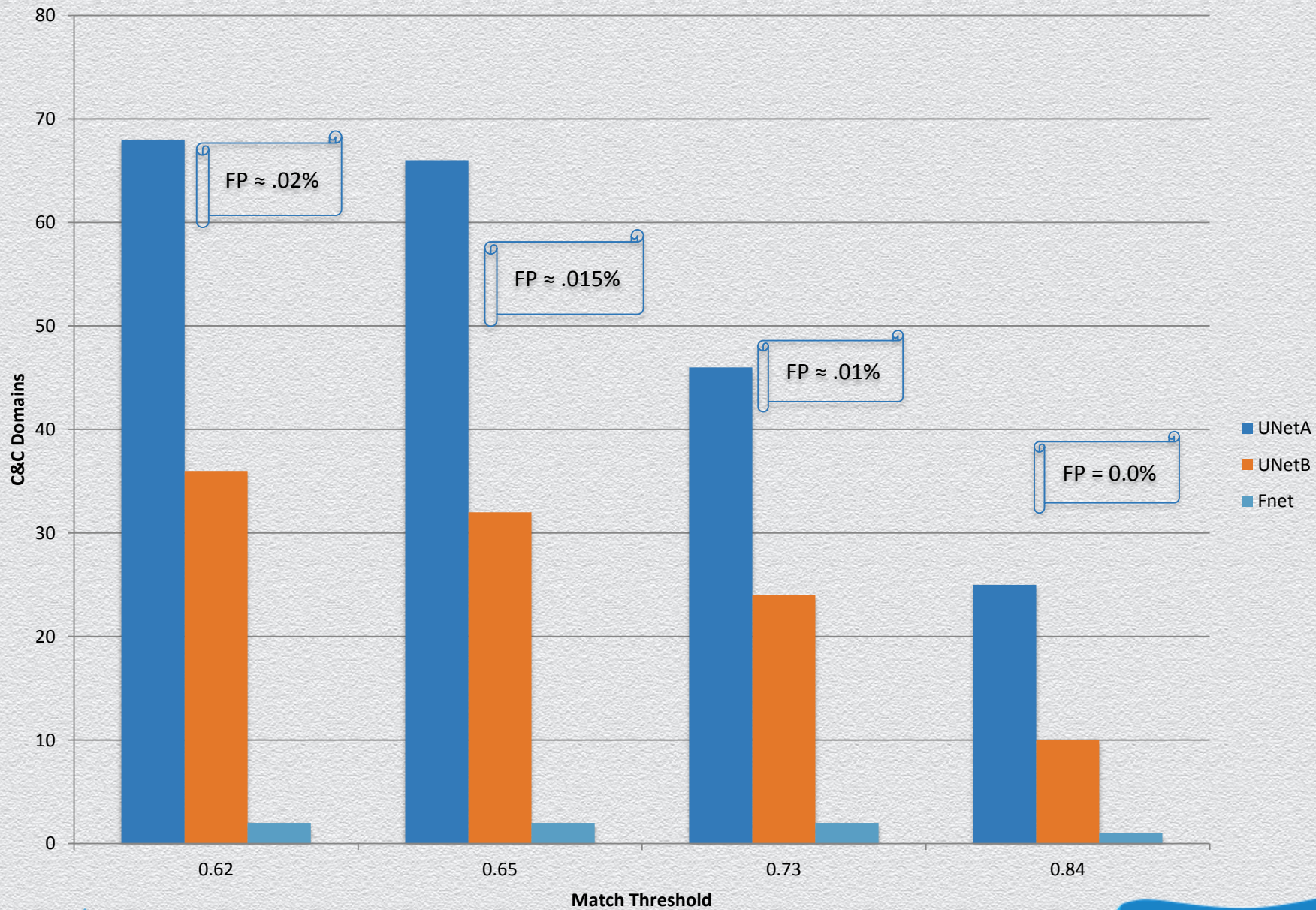
Learning & Recognizing Malware Languages

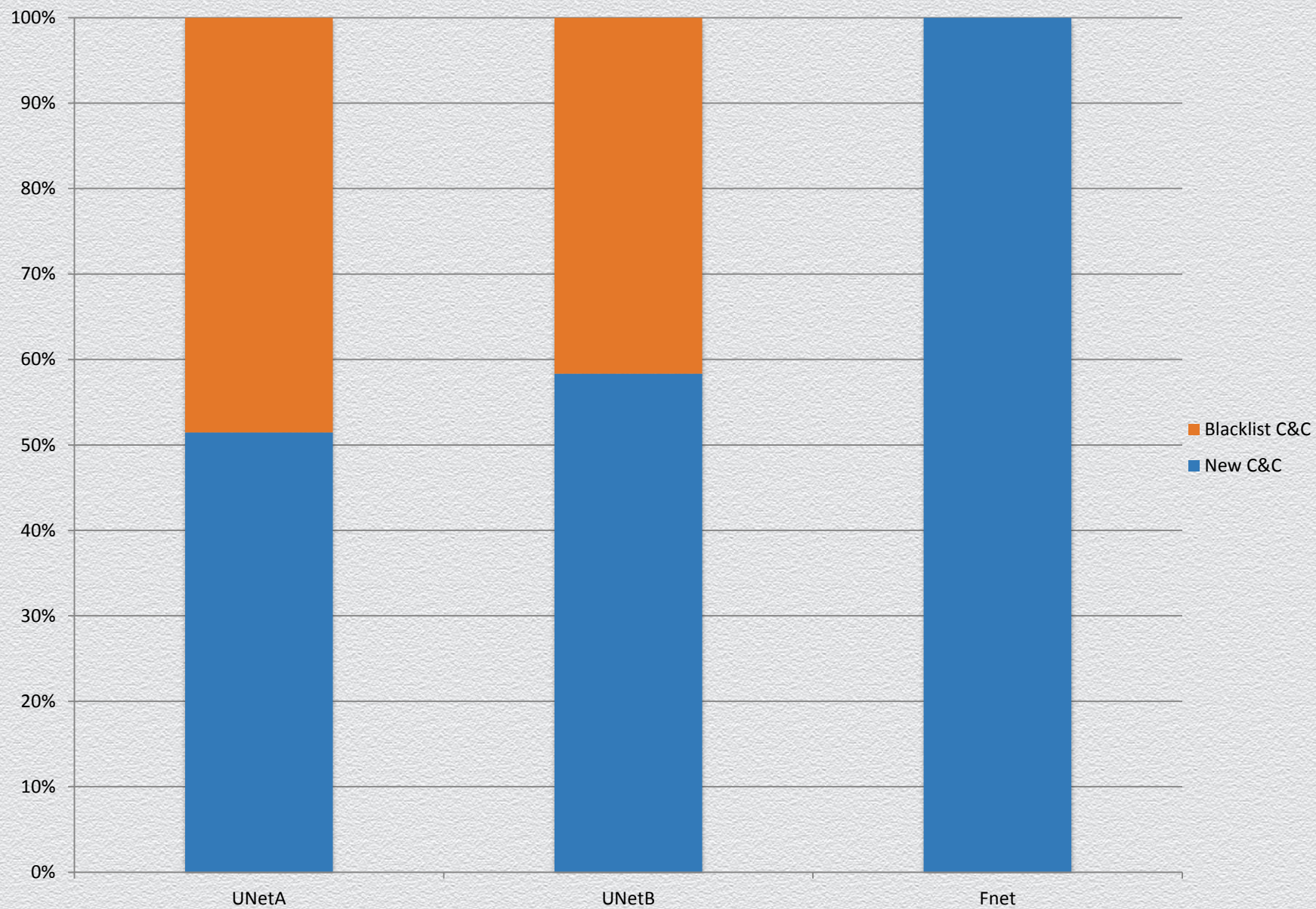


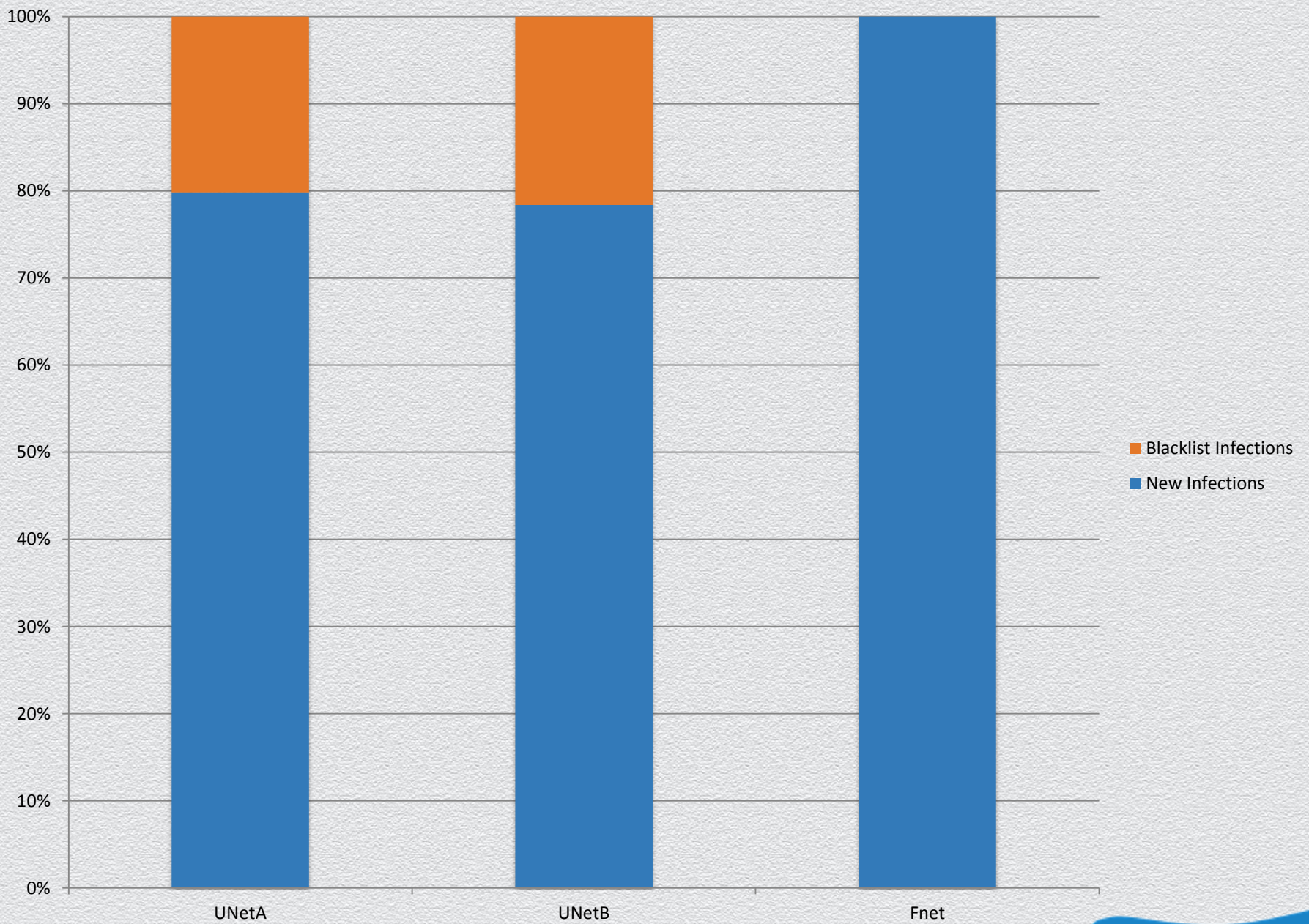
Evaluation Deployment Networks

	UNetA	UNetB	FNet
Distinct Src IPs	7,893	27,340	7,091
HTTP Requests	34,871,003	66,298,395	58,019,718
Distinct Domains	149,481	238,014	113,778

- ◆ Evaluation ran for two weeks.
- ◆ CPTs updated daily beginning two weeks prior to evaluation.



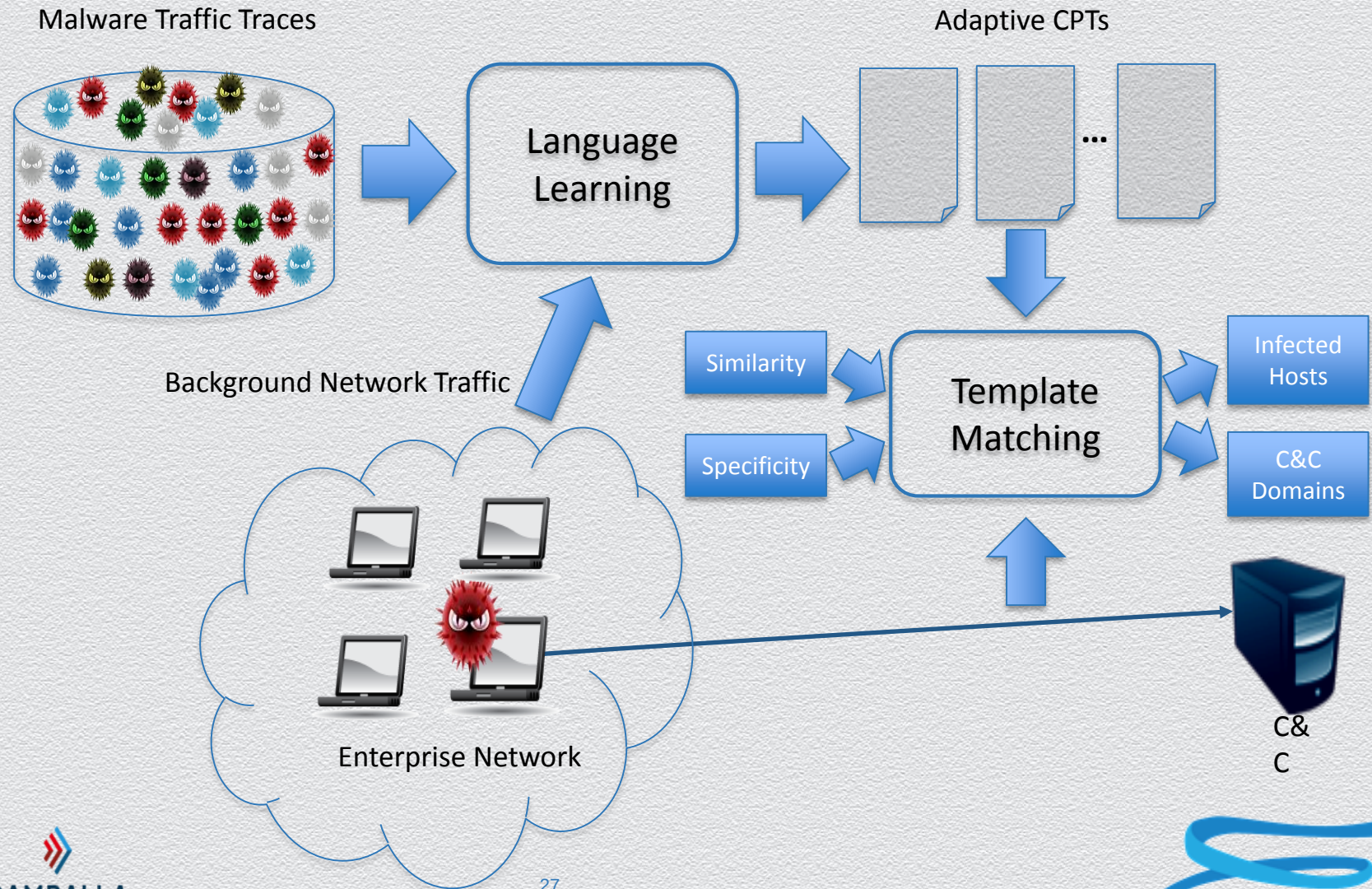




ISP Deployment

- ◆ Deployed the **65 newly discovered domain C&C domains** on **6 ISP networks** for 1 week.
- ◆ Counted the number of distinct source IP addresses contacting the domains daily.
- ◆ Identified **25,584** new potential malware infections.

Learning & Recognizing Malware Languages



Thank You!

Brian.foster@Damballa.com

+1.310.514.7485