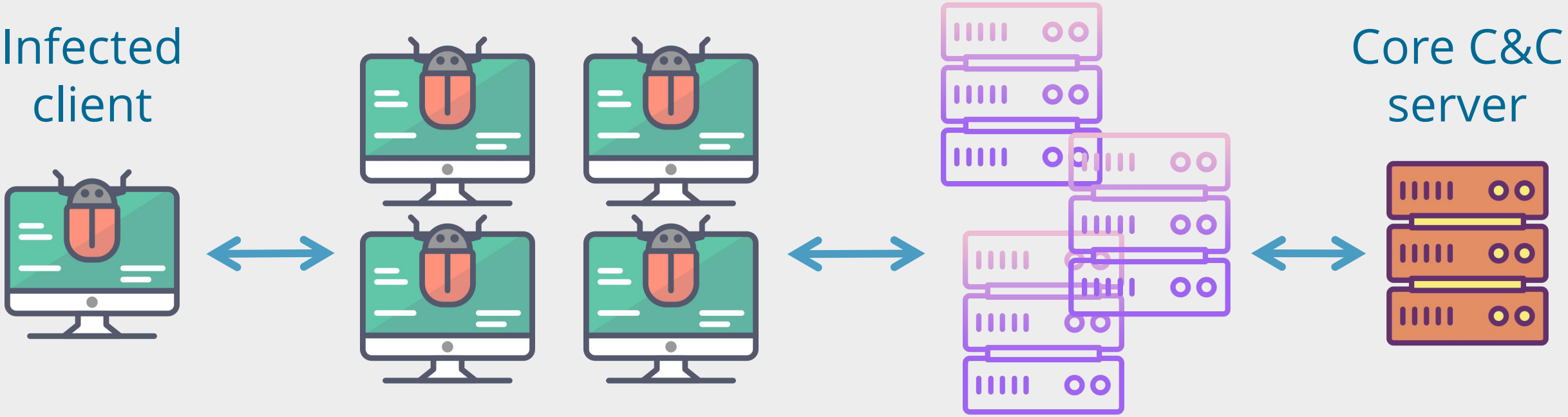
A Practical Approach for Taking Down Avalanche Botnets Under Real-World Constraints

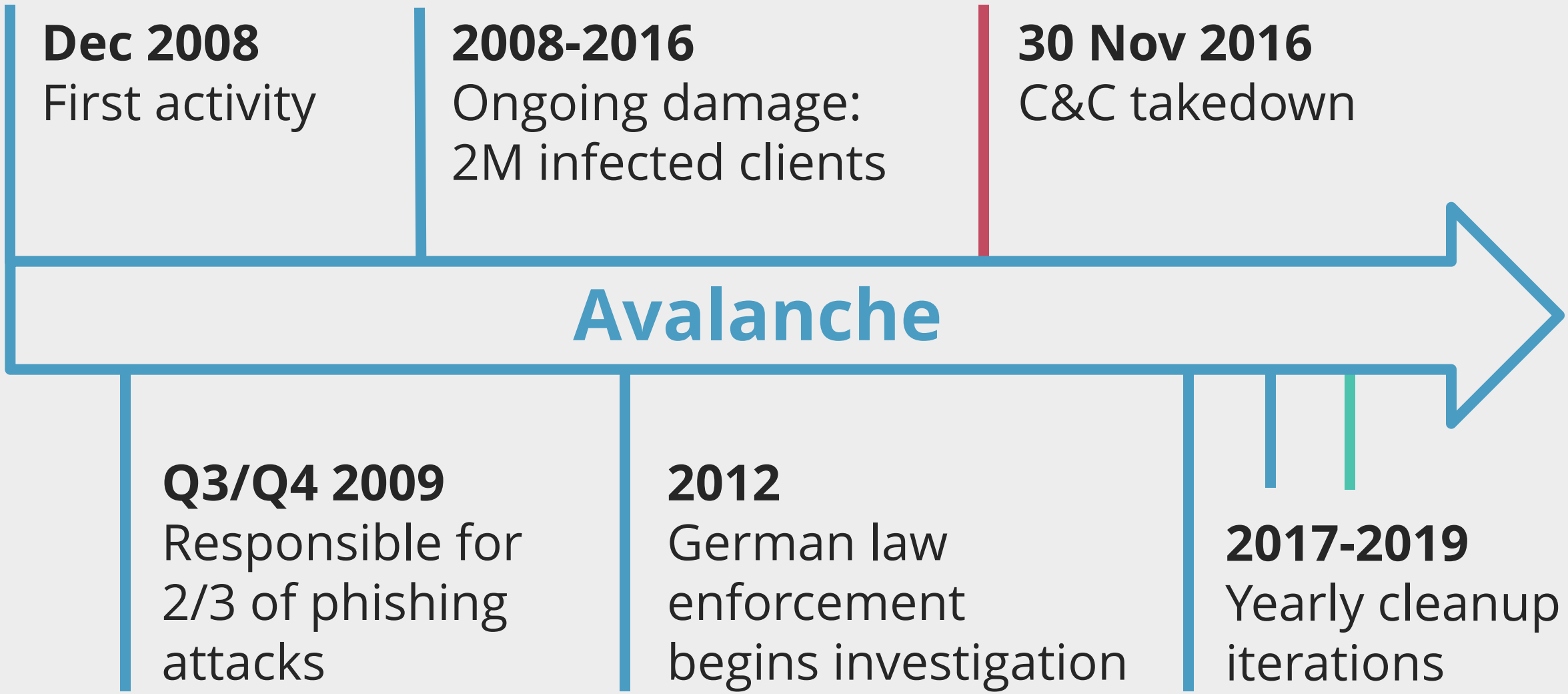
Victor Le Pochat, Tim Van hamme, Sourena Maroofi, Tom Van Goethem, Davy Preuveneers, Andrzej Duda, Wouter Joosen, Maciej Korczyński

NDSS 2020, 25 February 2020

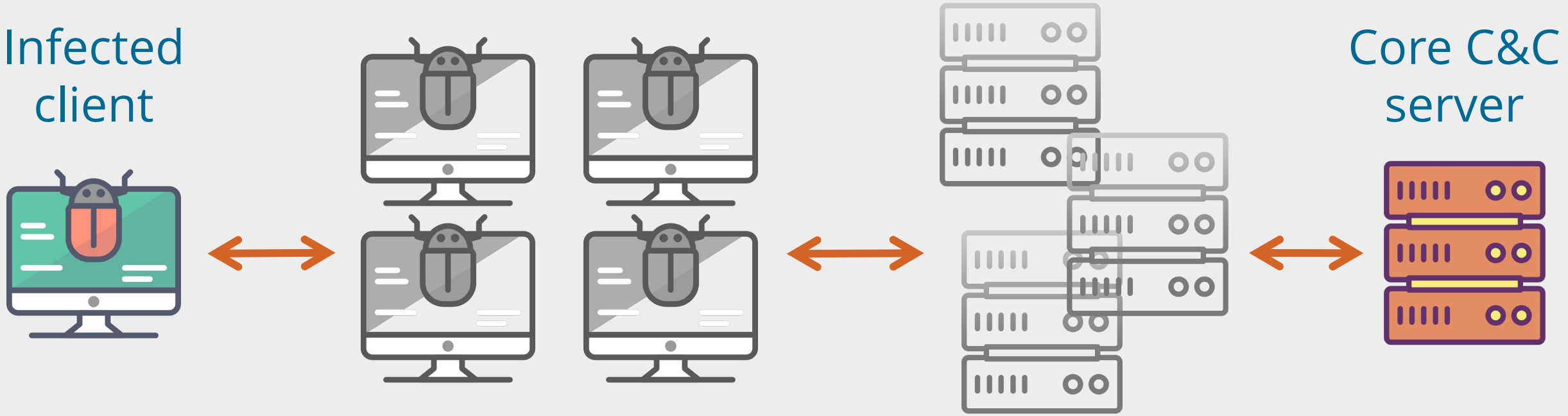
A Practical Approach for **Taking Down Avalanche Botnets** Under Real-World Constraints

Avalanche operated an advanced infrastructure




Dec 2008
First activity
2008-2016
Ongoing damage:
2M infected clients
30 Nov 2016
C&C takedown
Avalanche
Q3/Q4 2009
Responsible for
2/3 of phishing
attacks
2012
German law
enforcement
begins investigation
2017-2019
Yearly cleanup
iterations

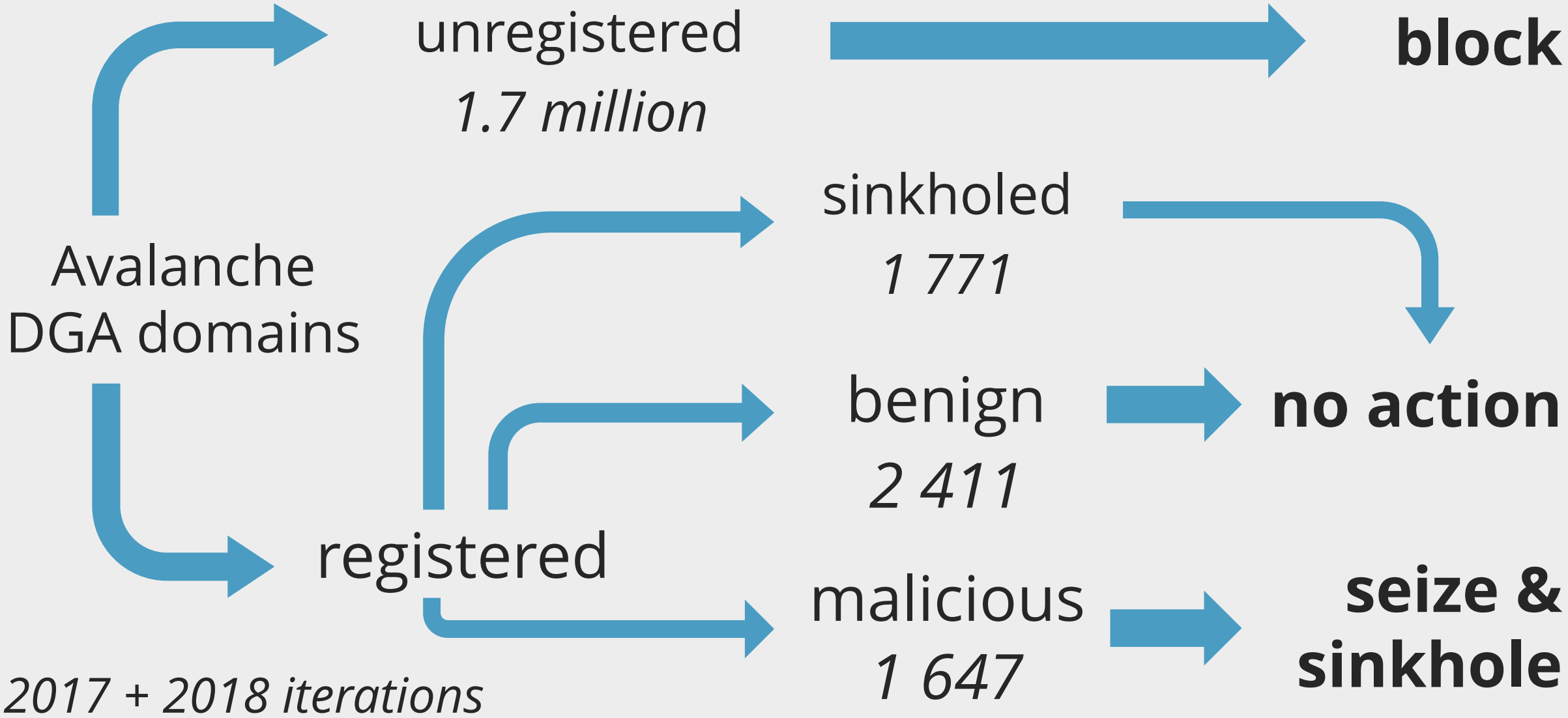
Avalanche operated an advanced infrastructure



Domain Generation Algorithms

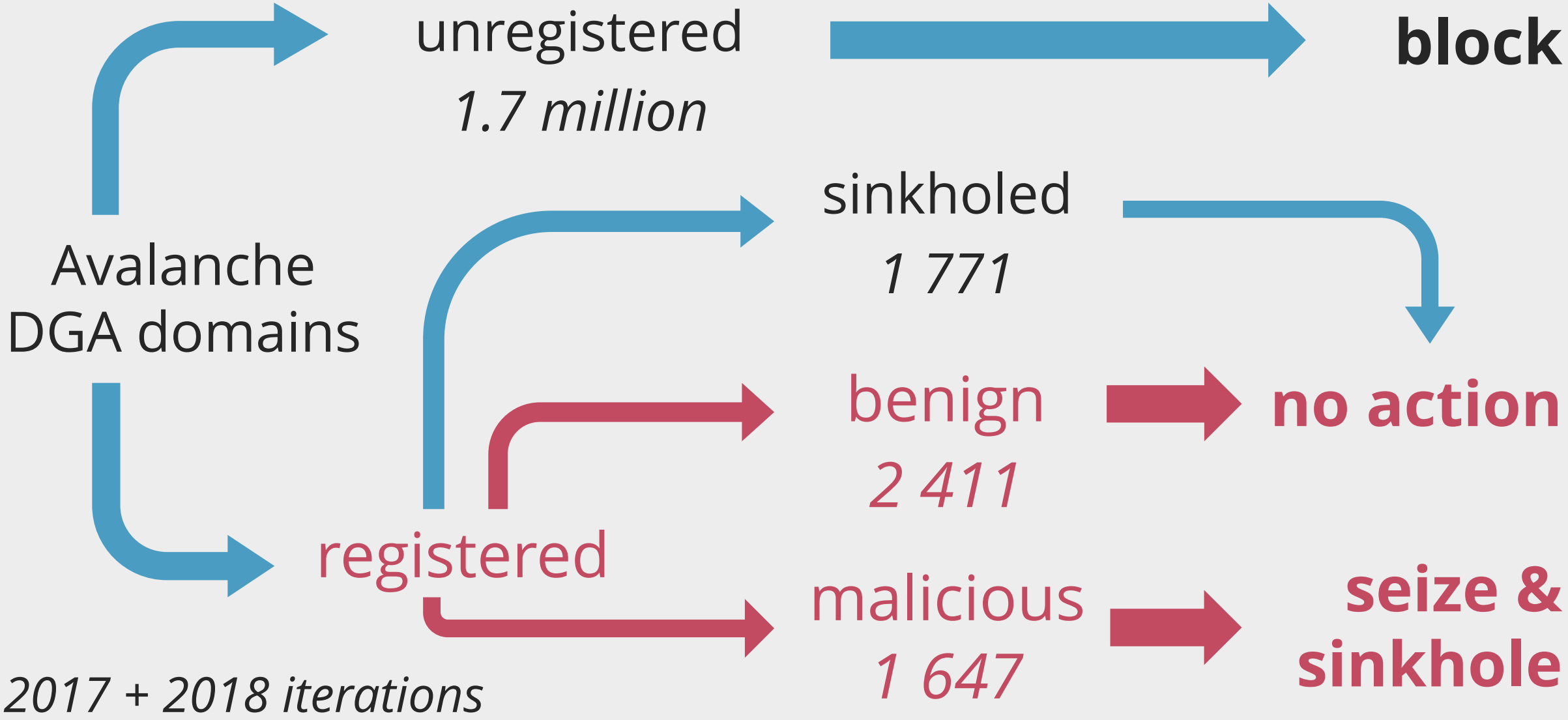
`0a85rcbe2wb5n5f.com` `researchmadness.com` `arbres.com`

Law enforcement has to classify registered domains



2017 + 2018 iterations

Law enforcement has to classify registered domains



2017 + 2018 iterations

We evaluate on a **real-world** takedown: Avalanche

- Design a more **automated** approach
  to reduce extensive **manual classification effort**
- and assist in making **accurate** decisions
  - Take down a *benign* domain: service interruption
  - *Not* take down a *malicious* domain: botnet can respawn
- leveraging *(limited)* **real-world ground truth**
  - synthetic data sets may not be representative [Küh14, LeP19]

A Practical Approach for Taking Down Avalanche Botnets Under **Real-World Constraints**

Constraints affect available indicators

Individual patterns

in contrast to

bulk registration
[Hao16, Spo19]

bulk lexical patterns
[Woo16, Sch18]

Proactive analysis

in contrast to

presence/detection of malicious activity
[Bil11, Ant12]

No active connections

in contrast to

active collection of web content
[Khe14]

A **Practical Approach** for Taking Down Avalanche Botnets Under Real-World Constraints

Our experimental protocol mimics real takedowns

- Enrich with **comprehensive** feature sets (*within constraints*)
- Collect **historical data** as of iteration *(if possible)*

- Some domains have **missing data**
- Classify **all** domains using **ensemble** model

Data set	*Missing*
WHOIS	14.6%
Passive DNS	8.7%
Active DNS	19.5%

train on	test on	accuracy	F_1 score	Effort saved

	train on	*test on*	*accuracy*	*F_1 score*	*Effort saved*
Base	2017	2018	84.3%	73.4%	100%

- Concept drift

	train on	*test on*	*accuracy*	*F_1 score*	*Effort saved*
Base	2017	2018	84.3%	73.4%	100%
Extended A priori	2017 + 15% of 2018	Remaining 85% of 2018	86.4%	78.6%	85.0%

- Hybrid model: Human oracle

	train on	*test on*	*accuracy*	*F_1 score*	*Effort saved*
Base	2017	2018	84.3%	73.4%	100%
Extended A priori	2017 + 15% of 2018	Remaining 85% of 2018	86.4%	78.6%	85.0%
Base A posteriori	2017	2018	97.3%	95.3%	70.3%
Extended A posteriori	2017 + 15% of 2018	Remaining 85% of 2018	97.6%	95.8%	66.2%
				*2019:	76.9%

We analyze influences on our model

	Set	*Feature*
1	WHOIS	**Time** between creation…
2	WHOIS	**Time** between creation…
3	Passive DNS	**Time** between first seen…
4	Passive DNS	**Time** between first and…
5	WHOIS	**Time** between creation…
6	WHOIS	Renewal of domain …
7	Active DNS	**Days** DNS record seen …
8	WHOIS	Renewal of domain …
9	Active DNS	**Time** between first seen…
10	Joint	Number of pages found…

› Important **time-based** features are **hard to evade**

We analyze influences on our model

	Set	*Feature*
1	**WHOIS**	Time between creation…
2	**WHOIS**	Time between creation…
3	Passive DNS	Time between first seen…
4	Passive DNS	Time between first and…
5	**WHOIS**	Time between creation…
6	**WHOIS**	Renewal of domain …
7	Active DNS	Days DNS record seen …
8	**WHOIS**	Renewal of domain …
9	Active DNS	Time between first seen…
10	Joint	Number of pages found…

- Important **time-based** features are **hard to evade**
- **Data availability** affects **performance**
  - Some redundancy exists

A Practical Approach for Taking Down Avalanche Botnets Under Real-World Constraints

We evaluate on a **real-world** takedown: Avalanche

- Automating **classification** of registered DGA domains
- Real-world setting yields **unique opportunity**
  but also imposes **constraints**
- **Hybrid** model: synergy between model and analyst
- **Insights** for real-world takedowns

A Practical Approach for Taking Down Avalanche Botnets Under Real-World Constraints

victor.lepochat@kuleuven.be @VictorLePochat

KU LEUVEN DistriNet UGA Université Grenoble Alpes LIG fwo

References

[Wai17] R. Wainwright and F. J. Cilluffo, “Responding to cybercrime at scale: Operation Avalanche - a case study,” Europol; Center for Cyber and Homeland Security, The George Washington University, Issue Brief 2017-03, Mar. 2017. [Online]. Available: https://cchs.gwu.edu/sites/g/files/zaxdzs2371/f/Responding%20to%20Cybercrime%20at%20Scale%20FINAL.pdf

[Küh14] M. Kührer, C. Rossow, and T. Holz, “Paint it black: Evaluating the effectiveness of malware blacklists,” in 17th International Symposium on Research in Attacks, Intrusions and Defenses, ser. RAID ’14, 2014, pp. 1–21.

[LeP19] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, “Tranco: A research-oriented top sites ranking hardened against manipulation,” in 26th Annual Network and Distributed System Security Symposium, ser. NDSS ’19, 2019.

[Hao16] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, “PREDATOR: Proactive recognition and elimination of domain abuse at time-of-registration,” in 2016 ACM SIGSAC Conference on Computer and Communications Security, ser. CCS ’16, 2016, pp. 1568–1579.

[Spo19] J. Spooren, T. Vissers, P. Janssen, W. Joosen, and L. Desmet, “Premadoma: An operational solution for DNS registries to prevent malicious domain registrations,” in 35th Annual Computer Security Applications Conference, ser. ACSAC ’19, 2019, pp. 557–567.

[Woo16] J. Woodbridge, H. S. Anderson, A. Ahuja, and D. Grant, “Predicting Domain Generation Algorithms with Long Short-Term Memory Networks,” Nov. 2016, arXiv:1611.00791

[Sch18] S. Schüppen, D. Teubert, P. Herrmann, and U. Meyer, “FANCI : Feature-based automated NXDomain classification and intelligence,” in 27th USENIX Security Symposium, ser. USENIX Security ’18, 2018, pp. 1165–1181.

[Bil11] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, “EXPOSURE: Finding malicious domains using passive DNS analysis,” in 18th Annual Network and Distributed System Security Symposium, ser. NDSS ’11, 2011.

[Ant12] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, and D. Dagon, “From throw-away traffic to bots: Detecting the rise of DGA-based malware,” in 21st USENIX Security Symposium, ser. USENIX Security ’12, 2012, pp. 491–506.

[Khe14] N. Kheir, F. Tran, P. Caron, and N. Deschamps, “Mentor: Positive DNS reputation to skim-off benign domains in botnet C&C blacklists,” in 29th IFIP International Information Security and Privacy Conference, ser. SEC ’14, 2014, pp. 1–14.