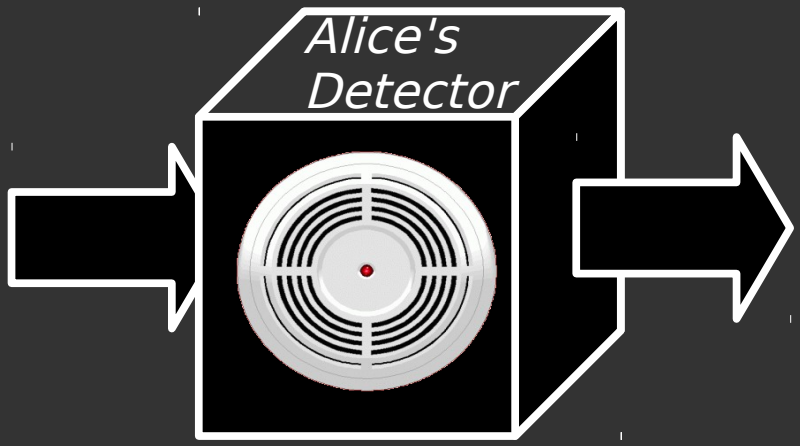
Challenges in Experimenting *with* Botnet Detection Systems

Adam J. Aviv Andreas Haeberlen

University of Pennsylvania

Alice has developed a new botnet detector!!!

What should the evaluation show?



Ideal

Alice deploys her detector live on her local network

- Alice is provided with a list of hosts that are botnet infected

Alice deploys her detector on various other networks

- Academic, Residential, Corporate, etc.

Alice records traces of each deployment

- Improve detector in the lab
- Readily available to other researchers

Realities

Production-ready deployment?

Ground truth of botnet infections?

Deployment on various networks?

Record trace and replay experiment?

Traces available to other researchers?

Taking a Step Back

so reporting performance statistics for the detector is very difficult.

As a result, researchers have relied on synthetic traces generated using an overlay methodology, where botnet traffic is mixed in with benign background traffic. This methodology can lead to a number of pitfalls, which could cause researchers to over- or underestimate the performance of their detector. A larger consequence of this methodology is that the background traces used in experiments are not easily shared due to privacy concerns, complicating basic scientific practices, such as performance comparisons and experimental reproducibility.

We observe that many of these issues are similar to those faced by researchers experimenting with large-scale distributed systems before the advent of PlanetLab. We propose a strawman system similar to PlanetLab that, through collaboration, could ameliorate many of the experimental challenges in botnet detection research. There are several research problems that would need to be addressed before such a system could become a reality, but there is evidence that these problems are solvable.

Acknowledgments

We thank Sean Peisert for shepherding this paper, and Jonathan Smith, Benjamin Pierce, Angelos Keromytis, Micah Sherr, and the anonymous reviewers for their helpful comments. This research was supported in part by ONR Grant N00014-09-1-0770 and by US National Science Foundation grants CNS-1054229 and CNS-1065060.

References

[1] P. Bächer, T. Holz, M. Kötter, and G. Wicherski. Know Your Enemy: Tracking Botnets. Technical report, The Honeynet Project, Aug. 2008.

[2] M. Barbaro and T. Zeller. A face is exposed for AOL searcher no. 4417749. *The New York Times*, Aug. 2006. http://www.nytimes.com/2006/08/09/technology/09aol.html.

[3] V. Berk, G. Bakos, and R. Morris. Designing a framework for active worm detection on global networks. In *1st IEEE International Workshop on Information Assurance (IWIAS)*, Mar. 2003.

[4] CAIDA. http://www.caida.org/.

[5] H. Choi, H. Lee, and H. Kim. BotGAD: detecting botnets by capturing group activities in network traffic. In *4th International ICST Conference on Communication System Software and Middleware (COMSWARE)*, June 2009.

[6] H. Choi, H. Lee, H. Lee, and H. Kim. Botnet Detection by Monitoring Group Activities in DNS Traffic. In *7th IEEE International Conference on Computer and Information Technology (CIT)*, Oct. 2007.

[7] B. Coskun and S. Dietrich. Friends of An Enemy: Identifying Local Members of Peer-to-Peer Botnets Using Mutual Contacts. In *26th Annual Computer Security Applications Conferenec (ACSAC)*, Dec. 2010.

[8] Cyber-TA. http://cyber-ta.org/.

[9] M. Dischinger, A. Haeberlen, I. Beschastnikh, K. P. Gummadi, and S. Saroiu. SatelliteLab: Adding heterogeneity to planetary-scale network testbeds. In *ACM SIGCOMM Conference*, Aug 2008.

[10] D. R. Ellis, J. G. Aiken, K. S. Attwood, and S. D. Tenaglia. A behavioral approach to worm detection. In *ACM Workshop on Rapid Malcode (WORM)*, Oct. 2004.

[11] J. Goebel and T. Holz. Rishi: Identify bot contaminated host by IRC nickname evaluation. In *1st USENIX Workshop on Hot Topics in Understanding Botnets (HotBots)*, Apr. 2007.

[12] B. S. Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna. Your botnet is my botnet: Analysis of a botnet takeover. In *16th ACM conference on Computer and Communications Security (CCS)*, Nov. 2009.

[13] G. Gu, R. Perdisci, J. Zhang, and W. Lee. BotMiner: Clustering analysis of network traffic for protocol- and structure-independent botnet detection. In *17th USENIX Security Symposium*, July 2008.

[14] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee. BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. In *16th USENIX Security Symposium*, Aug. 2007.

[15] G. Gu, J. Zhang, and W. Lee. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. In *16th Network and Distributed System Security Symposium (NDSS)*, Feb. 2008.

[16] T. Holz, M. Steiner, F. Dahl, E. Biersack, and F. Freiling. Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm. In *1st USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, Apr. 2008.

[17] Honey@home. http://www.honeyathome.org/.

[18] M. Jelasity and V. Bilicki. Towards automated detection of peer-to-peer botnets: on the limits of local approaches. In *2nd USENIX Conference on Large-scale Exploits and Emergent Threats (LEET)*, Apr. 2009.

[19] C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, and S. Savage. The Heisenbot Uncertainty Problem: Challenges in Separating Bots from Chaff. In *1st USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, Apr. 2008.

[20] A. Karasaridis, B. Rexroad, and D. Hoeflin. Wide-scale botnet detection and characterization. In *1st USENIX Workshop on Hot Topics in Understanding Botnets (HotBots)*, Apr. 2007.

[21] D. Kotz and T. Henderson. CRAWDAD: A community resource for archiving wireless data at Dartmouth. http://crawdad.cs.dartmouth.edu/.

[22] C. P. Lee. *Framework for Botnet Emulation and Analysis*. PhD thesis, Georgia Institute of Technology, Atlanta, Georgia, May 2009.

[23] C. Livadas, R. Walsh, D. Lapsley, and W. T. Strayer. Using Machine Learning Techniques to Identify Botnet Traffic. In *31st Annual IEEE Conference on Local Computer Networks (LCN)*, Nov. 2006.

[24] W. Lu, G. Rammidi, and A. A. Ghorbani. Clustering botnet communication traffic based on n-gram feature selection. *Computer Communications*, 34(3):502–514, Mar. 2011.

[25] W. Lu, M. Tavallaee, and A. A. Ghorbani. Automatic discovery of botnet communities on large-scale communication networks. In *4th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, Mar. 2009.

[26] A. G. Miklas, S. Saroiu, A. Wolman, and A. D. Brown. Bunker: a privacy-oriented platform for network tracing. In *6th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Apr. 2009.

[27] J. C. Mogul and M. Arlitt. SC2D: An alternative to trace anonymization. In *SIGCOMM Workshop on Mining Network Data (MineNet)*, Sept. 2006.

[28] S. Nagaraja, P. Mittal, C. Y. Hong, M. Caesar, and N. Borisov. BotGrep: Finding P2P Bots with Structured Graph Analysis. In *USENIX Security Symposium*, Aug. 2010.

[29] A. Narayanan and V. Shmatikov. Robust deanonymization of large sparse datasets. In *29th IEEE Symposium on Security and Privacy*, May 2008.

[30] L. Peterson, A. Bavier, M. E. Fiuczynski, and S. Muir. Experiences building PlanetLab. In *7th Symposium on Operating Systems Design and Implementation (OSDI)*, Nov. 2006.

[31] Protected Repository for the Defense of Infrastructure Against Cyber Threats. http://www.predict.org.

[32] H. Pucha, Y. C. Hu, and Z. M. Mao. On the impact of research network based testbeds on wide-area experiments. In *6th ACM SIGCOMM Conference on Internet Measurement (IMC)*, Oct. 2006.

[33] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis. My botnet is bigger than yours (maybe, better than yours): why size estimates remain challenging. In *1st USENIX Workshop on Hot Topics in Understanding Botnets (HotBots)*, Apr. 2007.

[34] A. Ramachandran, N. Feamster, and D. Dagon. Revealing botnet membership using DNSBL counter-intelligence. In *2nd USENIX Conference on Steps to Reducing Unwanted Traffic on the Internet (SRUTI)*, July 2006.

[35] J. Reed, A. J. Aviv, D. Wagner, A. Haeberlen, B. C. Pierce, and J. M. Smith. Differential privacy for collaborative security. In *3rd European Workshop on System Security (EuroSec)*, Apr. 2010.

[36] K. Rieck, G. Schwenk, T. Limmer, T. Holz, and P. Laskov. Botzilla: Detecting the "phoning home" of malicious software. In *25th ACM Symposium on Applied Computing (SAC)*, Mar. 2010.

[37] Shadowserver. http://shadowserver.org/.

[38] N. Spring, L. Peterson, A. Bavier, and V. Pai. Using PlanetLab for network research: myths, realities, and best practices. *SIGOPS Operating Systems Review*, 40(1):17–24, Jan. 2006.

[39] E. Stinson and J. C. Mitchell. Towards systematic evaluation of the evadability of bot/botnet detection methods. In *2nd USENIX Workshop on Offensive Technologies (WOOT)*, July 2008.

[40] B. Stone-Gross, T. Holz, G. Stringhini, and G. Vigna. The underground economy of Spam: A botmaster's perspective of coordinating large-scale Spam campaigns. In *4th USENIX Workshop on Large-Scale Exploits and Emerging Threats (LEET)*, Mar. 2011.

[41] W. T. Strayer, R. Walsh, C. Livads, and D. Lapsley. Detecting Botnets with Tight Command and Control. In *31st IEEE Conference on Local Computer Networks (LCN)*, Nov. 2006.

[42] K. V. Vishwanath and A. Vahdat. Swing: Realistic and responsive network traffic generation. *IEEE/ACM Transactions on Networking*, 17(3):712–725, June 2009.

[43] VX Heavens. http://vx.netlux.org/.

[44] M. C. Weigle, P. Aduthi, F. Hernández-Campos, K. Jeffay, and F. D. Smith. Tmix: a tool for generating realistic tcp application workloads in ns-2. *SIGCOMM Computer Communication Review*, 36:65–76, July 2006.

[45] P. Wurzinger, L. Bilge, T. Holz, J. Goebel, C. Kruegel, and E. Kirda. Automatically Generating Models for Botnet Detection. In *14th European Symposium on Research in Computer Security (ESORICS)*. Sept. 2009.

[46] T.-F. Yen and M. K. Reiter. Traffic aggregation for malware detection. In *5th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, July 2008.

[47] T.-F. Yen and M. K. Reiter. Are Your Hosts Trading or Plotting? Telling P2P File-Sharing and Bots Apart. In *30th International Conference on Distributed Computing Systems (ICDCS)*, June 2010.

[48] N. Zeldovich, S. Boyd-Wickizer, and D. Mazières. Securing distributed systems with information flow control. In *5th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Apr. 2008.

[49] U. Zhang, X. Luo, R. Perdisci, G. Gu, W. Lee, and N. Feamster. Boosting the scalability of botnet detection using adaptive traffic sampling. In *6th ACM Symposium on Information, Computer and Communcations Security (ASIACCS)*, Mar. 2011.

[50] Y. Zhang and V. Paxson. Detecting stepping stones. In *9th USENIX Security Symposium*, Aug. 2000.

Many Challenges

- Multiple Administrative Domains
- Network Heterogeneity
- Multimorbidity
- Privacy
- Controlled Environments
- Artifact Overfitting
- Botnet Overfitting
- Focus on Academic Networks
- Scale
- Mixing Artifacts
- False Postives & Negatives
- Repeatability
- Comparability
- Lack of Verification

privacy

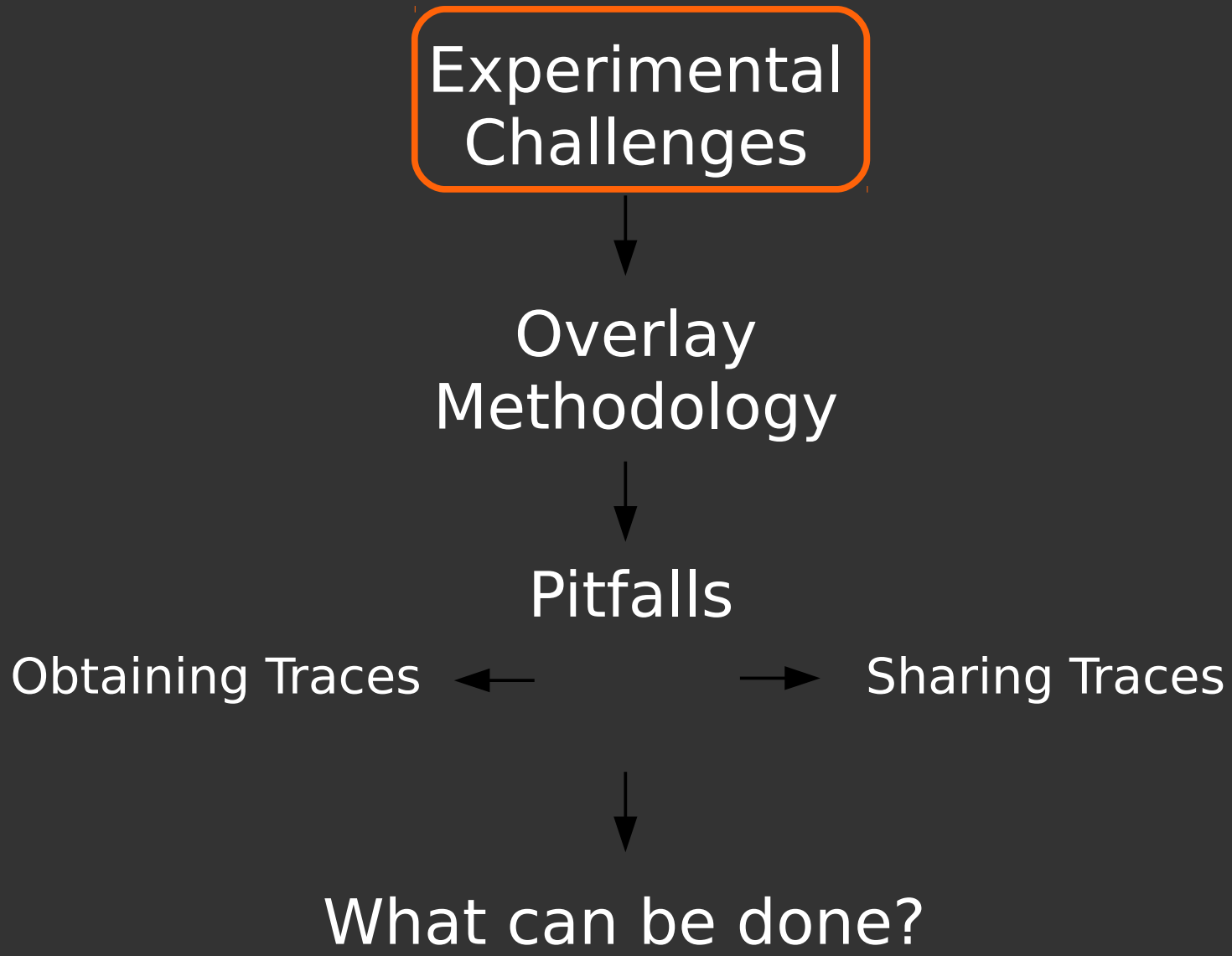
We have to worry about privacy, but the botnet authors don't!

Can we do better together?

Discussion/Topics/Questions

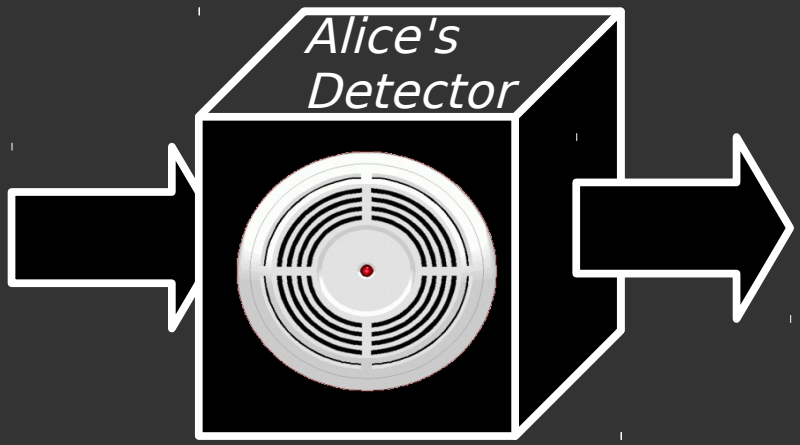
- Experimental Ideals vs. Realities
  - Not just botnet detectors ...
- Raw Materials of the Experiment
  - Sharing and Obtaining Traces
  - Botnet and Background Traces
- Can we do better via collaboration?

Presentation Outline



Alice has developed a new botnet detector!!!

What should the evaluation show?



Ideal vs. Reality

Alice deploys her detector live on her local network

- Alice is provided with list of hosts that are botnet infected

Production-ready deployment?

Ground truth of botnet infections?

Alice deploys her detector on other various networks

- Corporate, Residential, Corporate, etc.

Deployment on various networks?

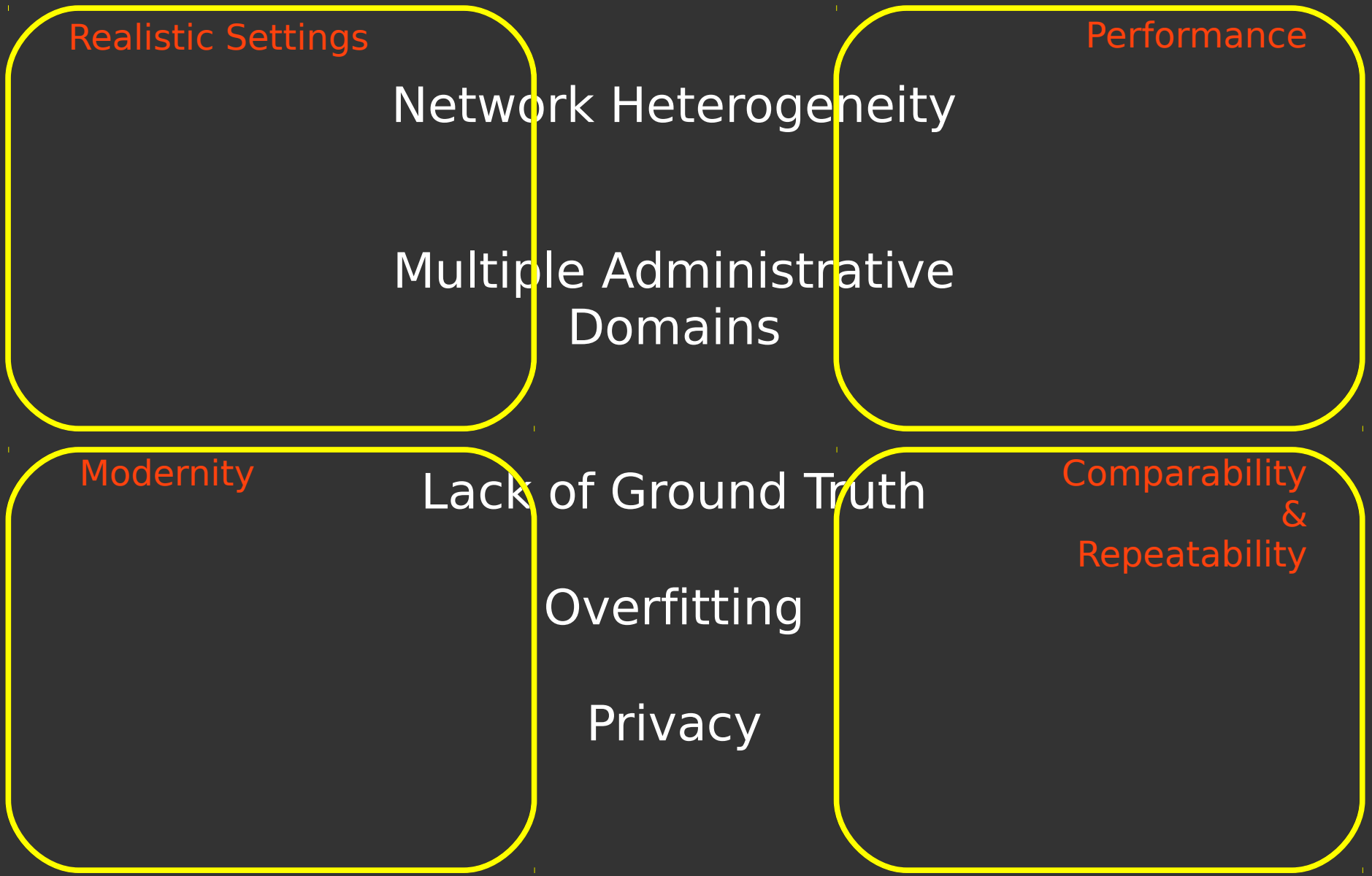
Alice records traces of each deployment

- Improve detector in the lab
- Readily available to other researchers

Record trace and replay experiment?

Traces available to other researchers?

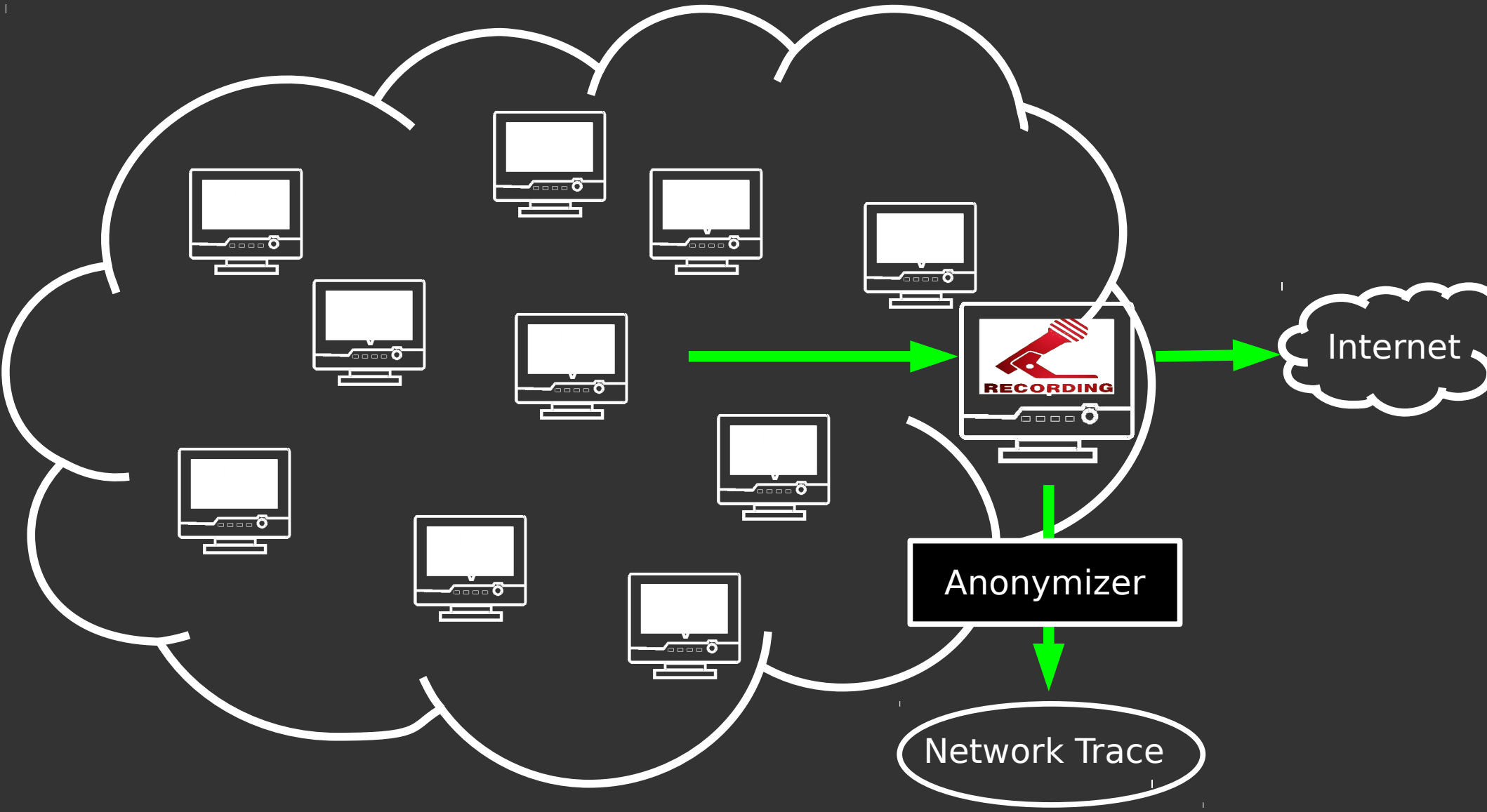
Evaluation Realities



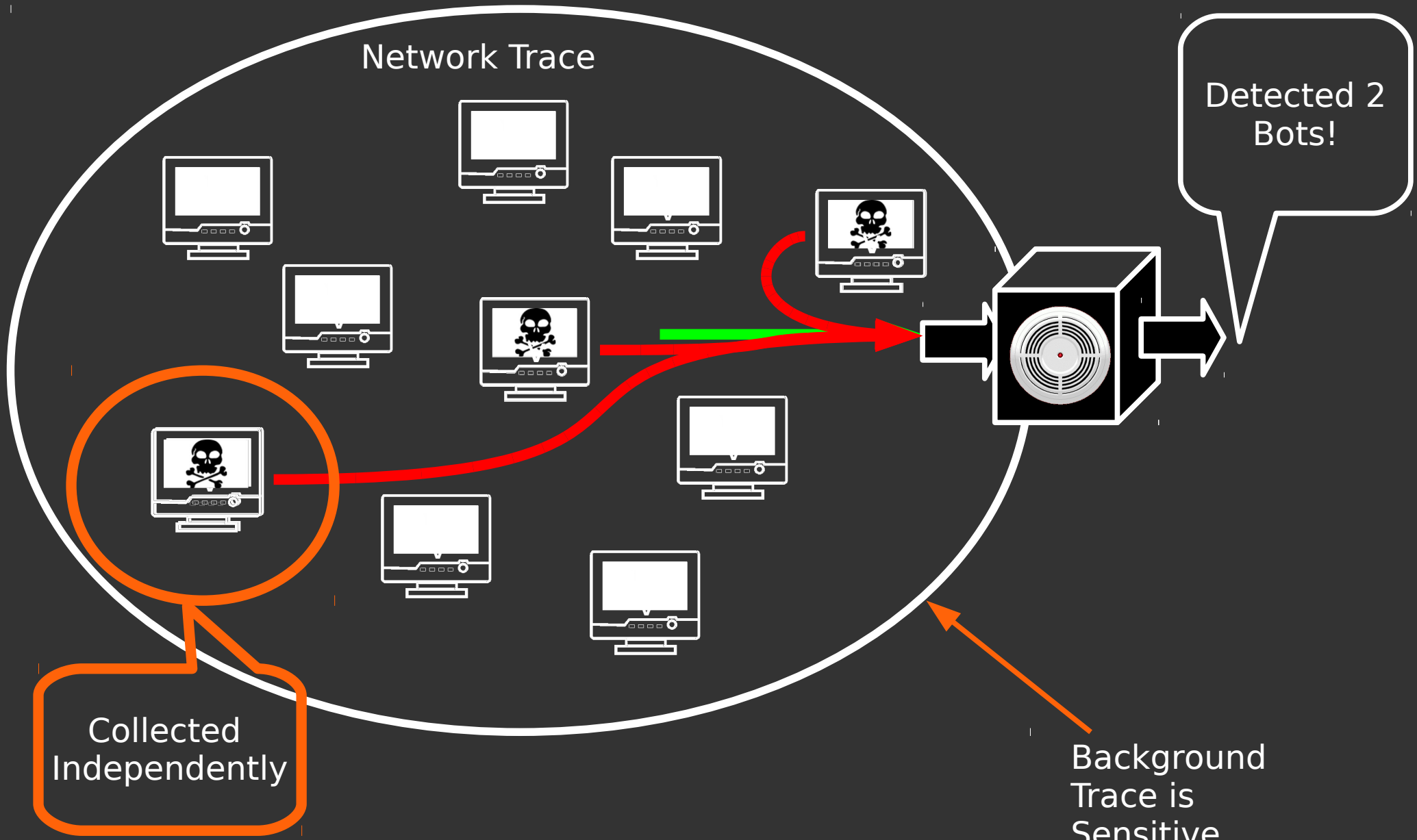
Pitfalls


Experimental Challenges
Overlay Methodology
Pitfalls
Obtaining Traces
Sharing Traces
What can be done?

Overlay Methodology



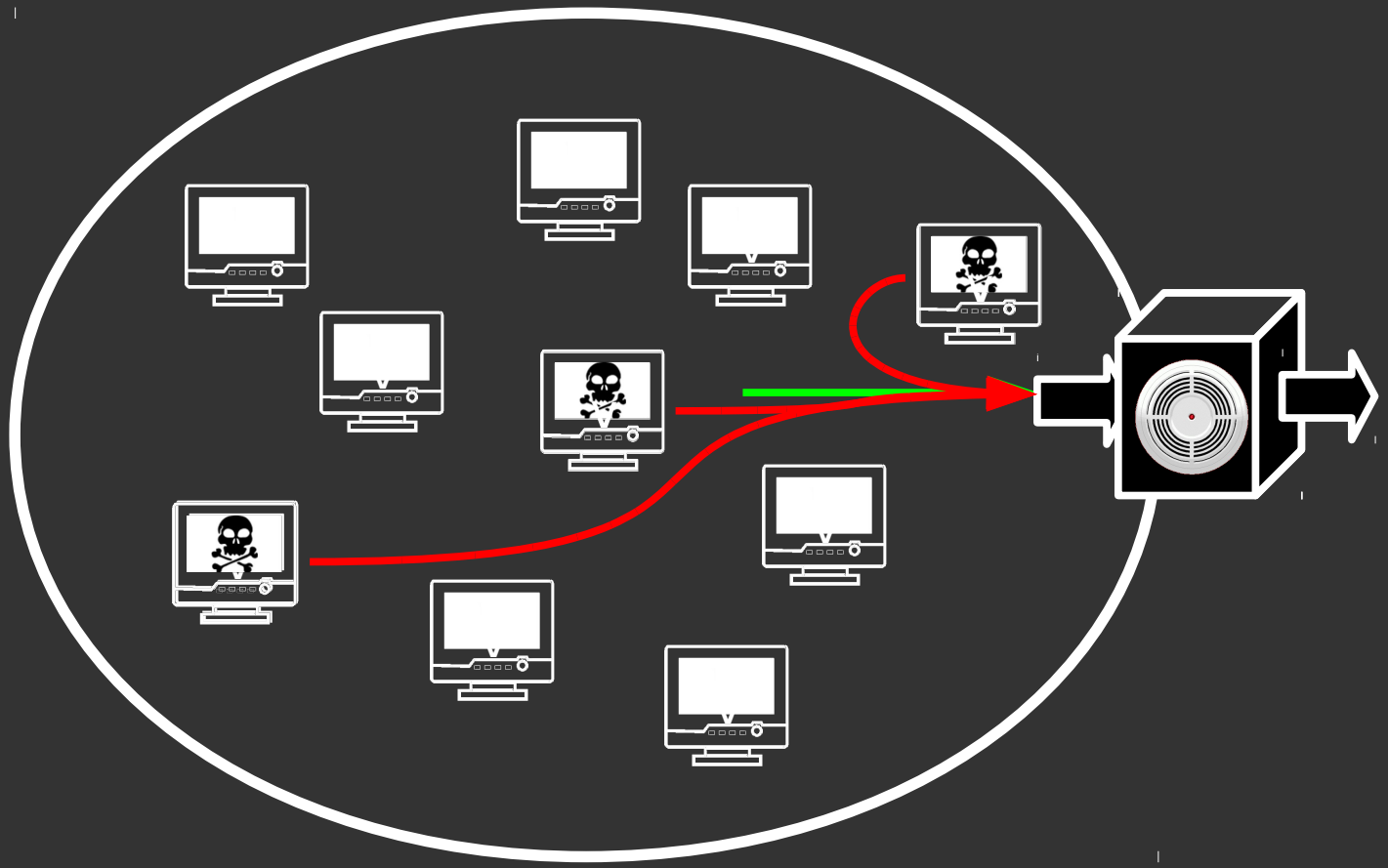
Replay and Evaluate



Prevalence in the Literature

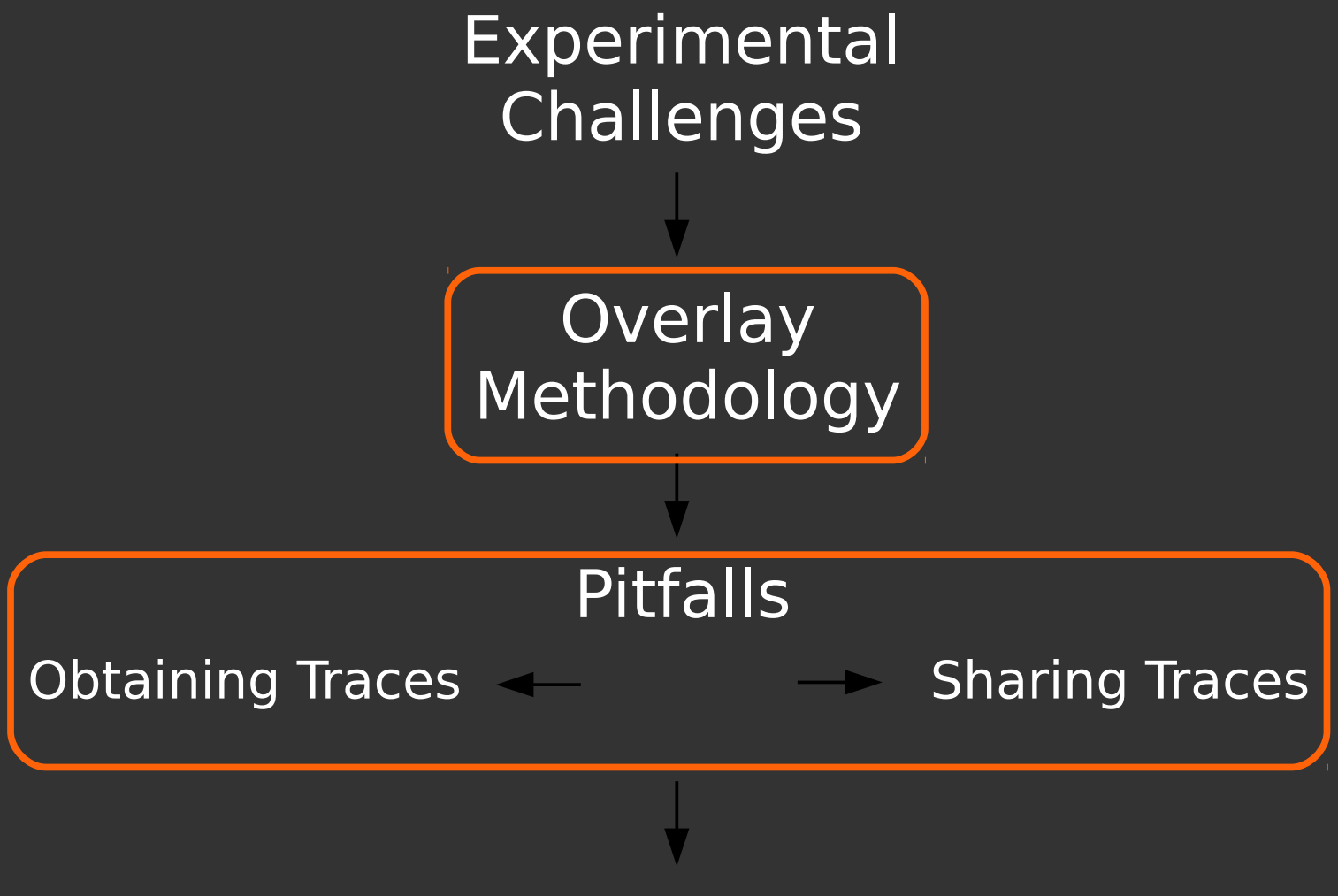
* See paper for references.

Advantages of Overlay Methodology



Ground Truth

Pitfalls

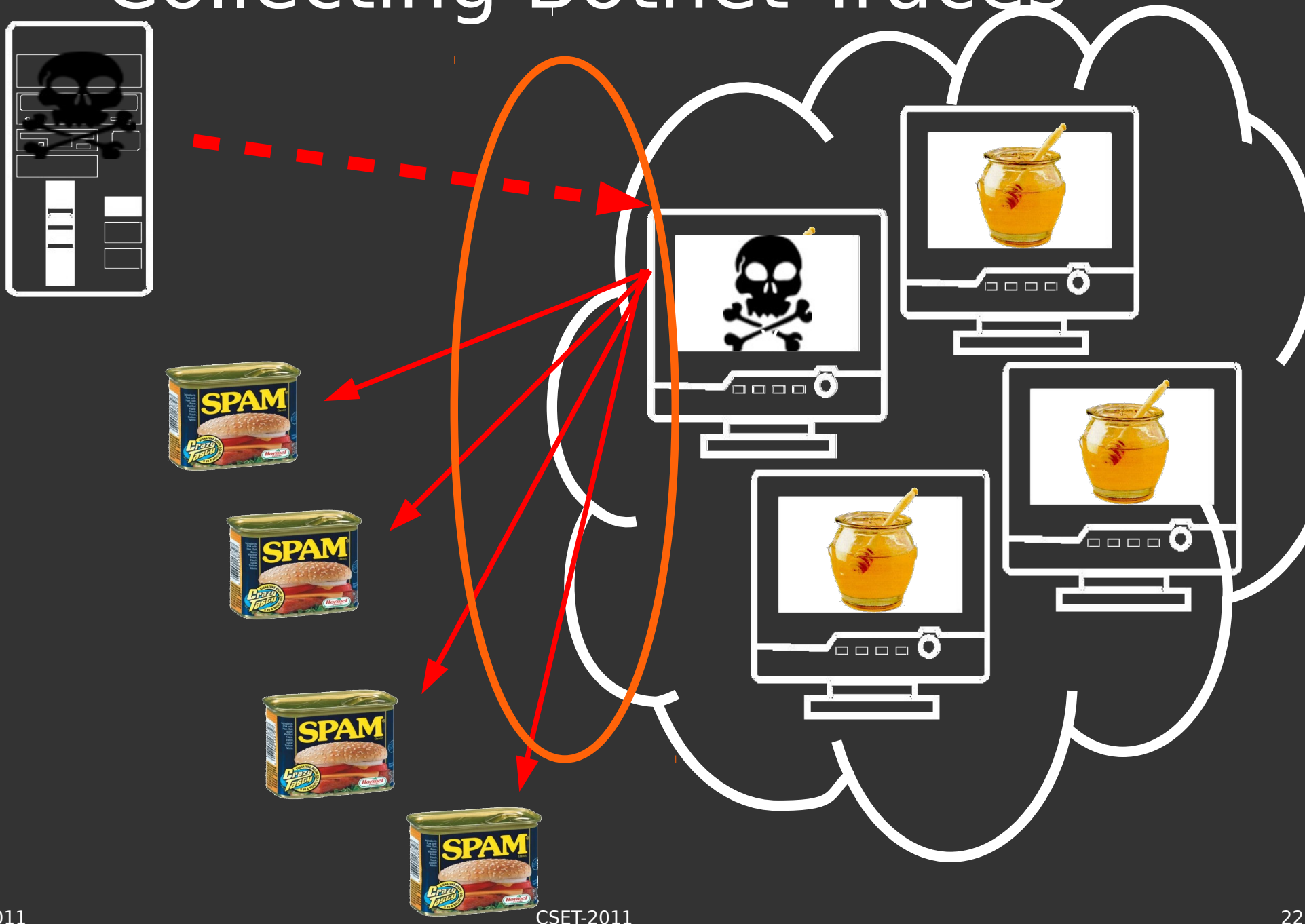

Experimental Challenges
Overlay Methodology
Pitfalls
Obtaining Traces
Sharing Traces
What can be done?

Obtaining Traces

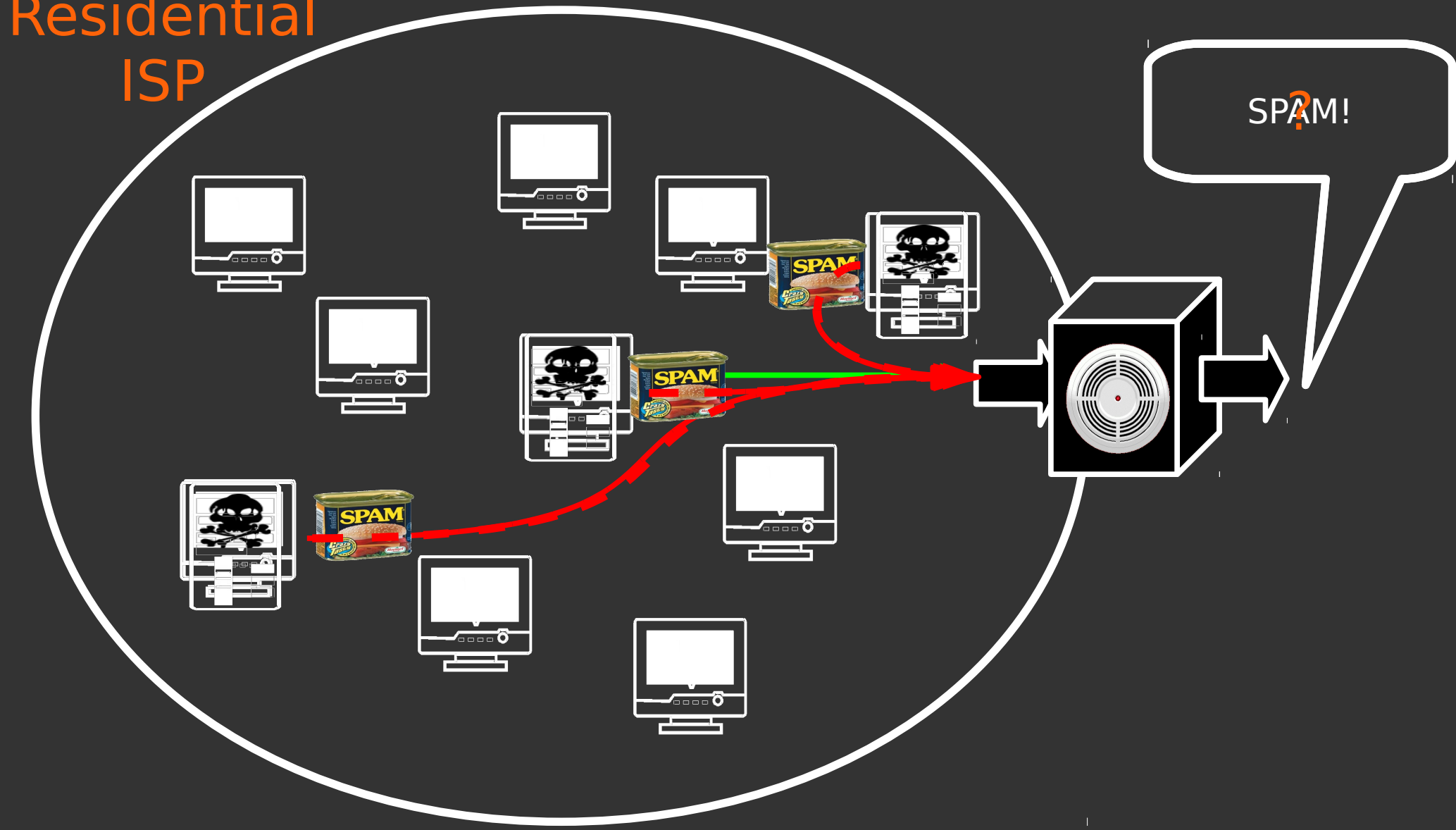
Realism

Merging of Botnet and Background trace should be realistic

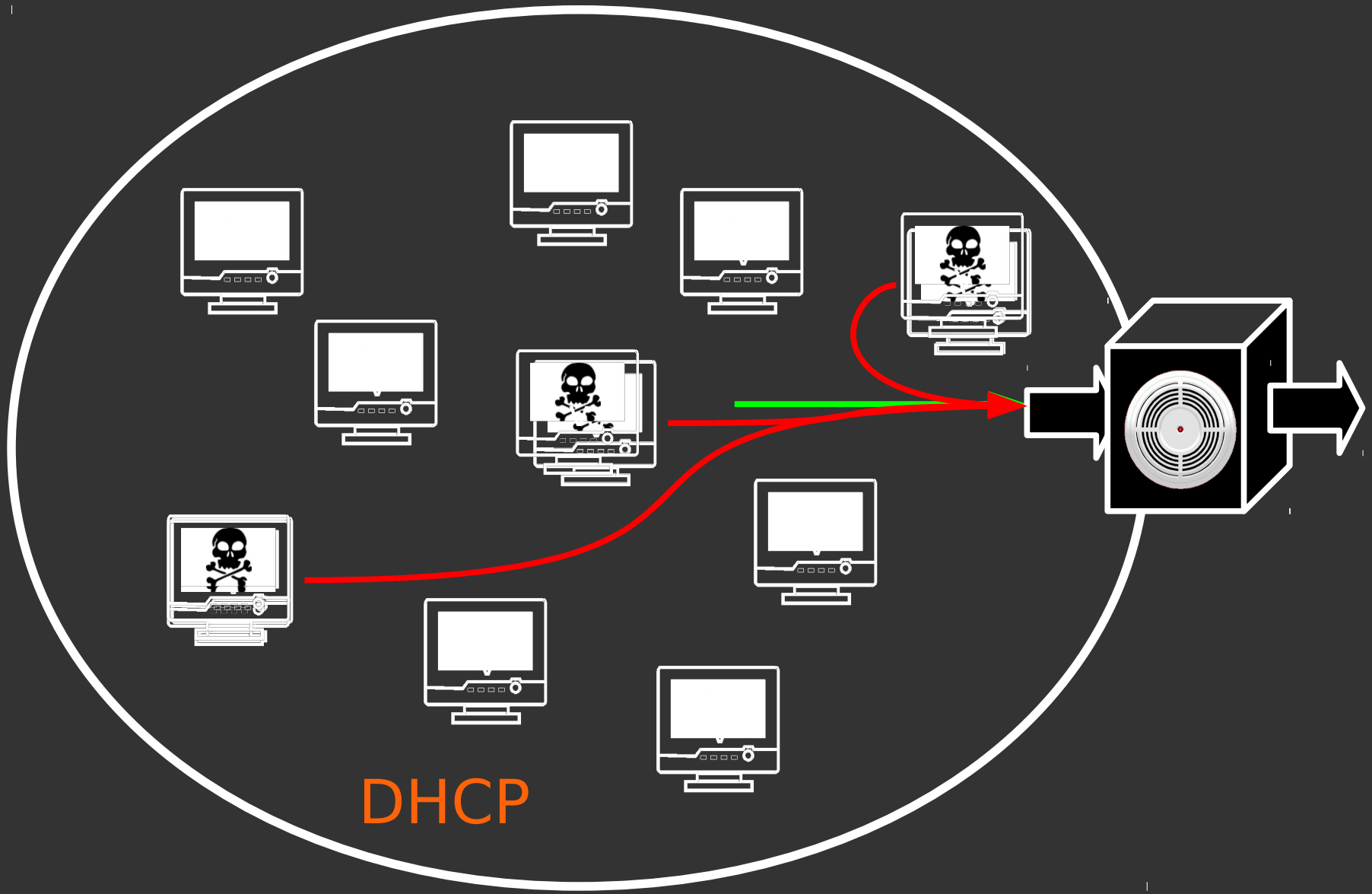
Collecting Botnet Traces



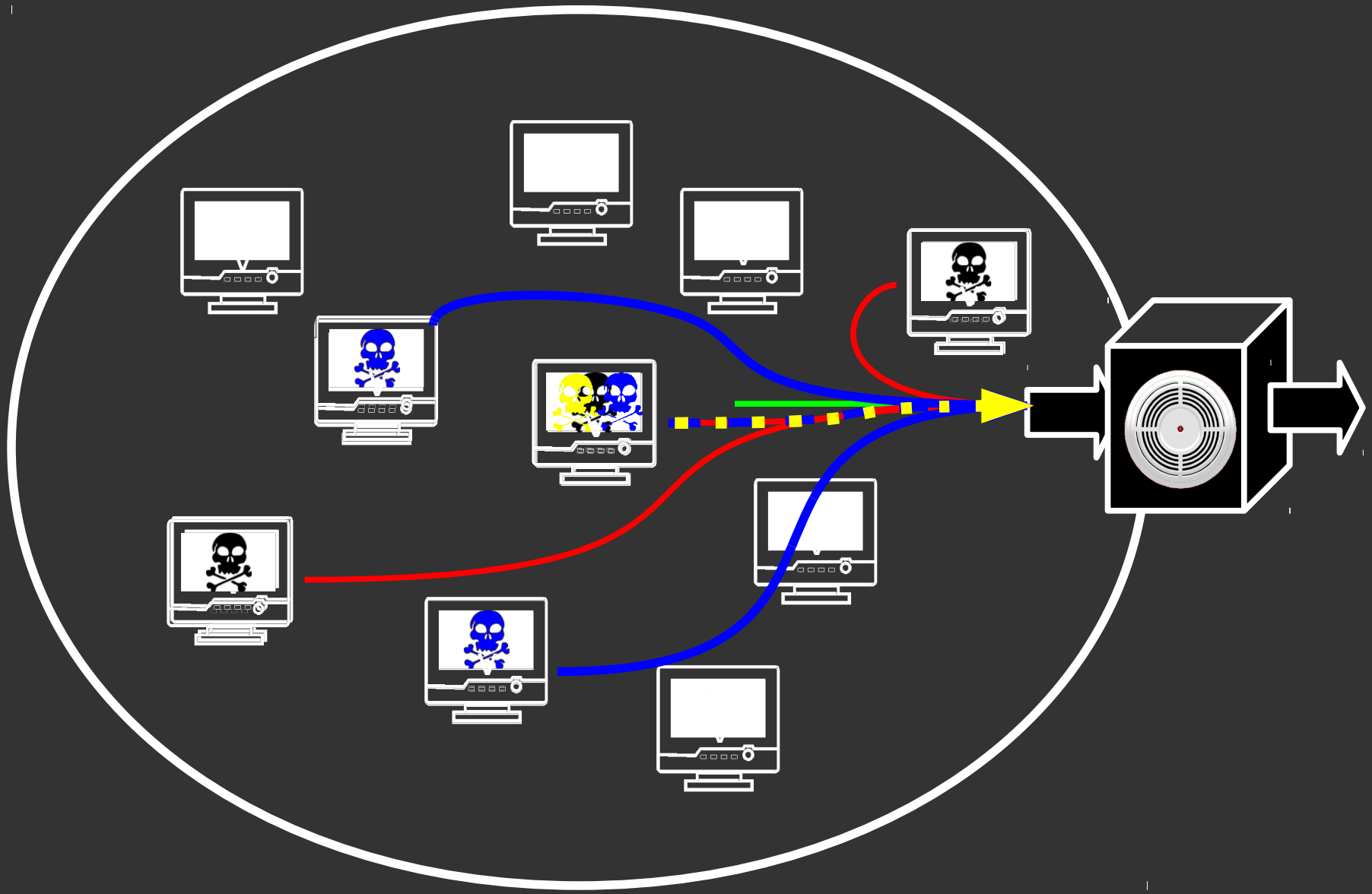
Realistic Embedding



Mixing Artifacts



Multimorbidity



Obtaining Traces

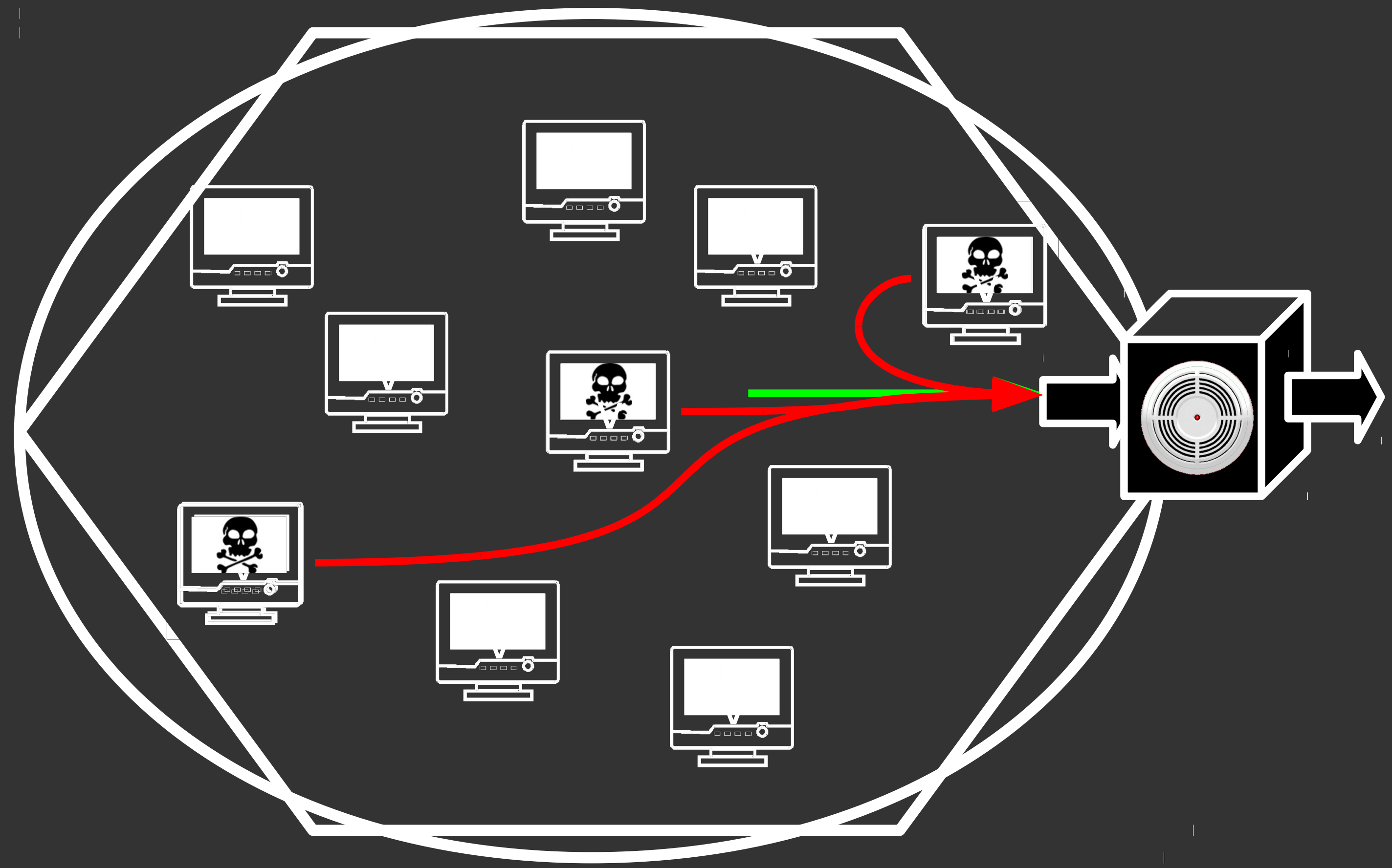
Realism

Merging of Botnet and Background trace should be realistic

Representativeness

Reflect diversity in network scenarios

Focus on Academic Networks



Corporate Business

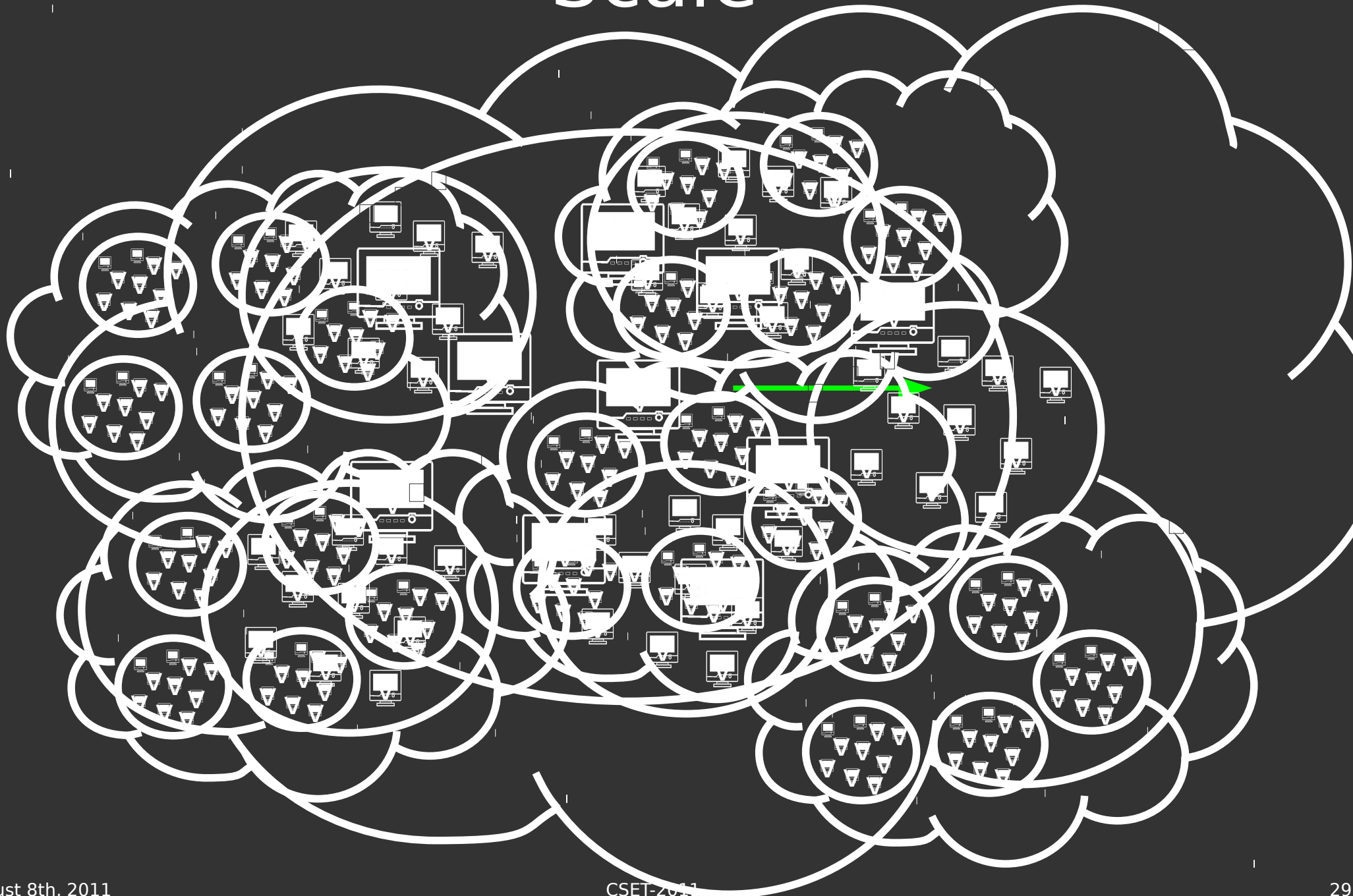
State Universities

Prevalence in the Literature

	Academic Traces	At Least One Other Trace
Overlay Methodology	[13] [49] [15] [36] [46] [47] [41] [23] [6] [7]	[28] [25] [24] [14]
Other Methodology	[36] [11] [5]	[20] [14] [45]

* See paper for references.

Scale



Obtaining Traces

Realism

- Merging of Botnet and Background trace should be realistic

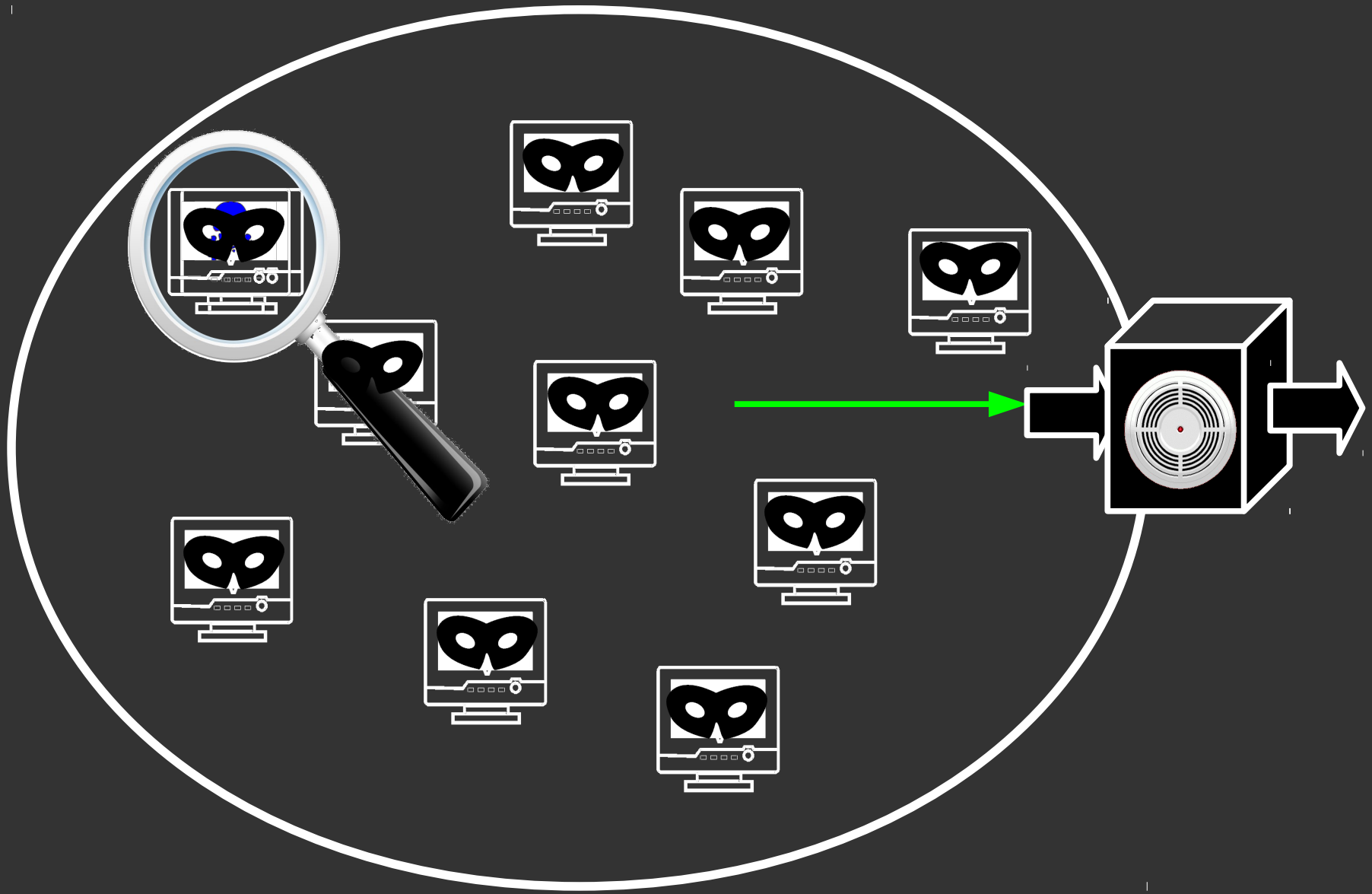
Representativeness

- Reflect diversity in network scenarios

Performance

- False postives and negatives

Lack of Verification

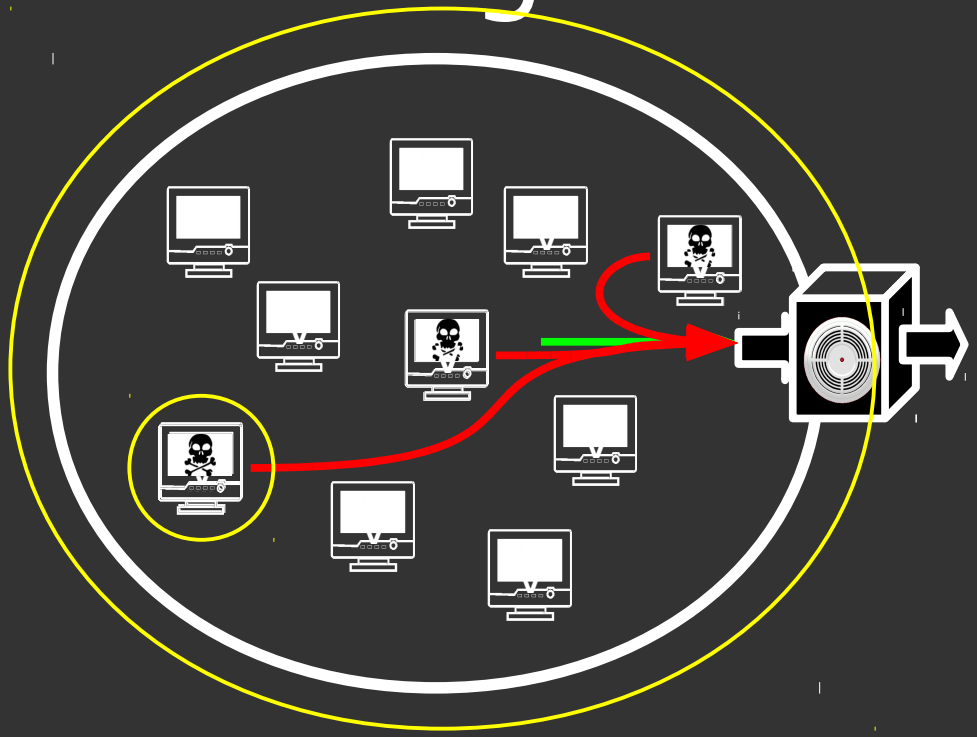


Example From the Literature
TaMD

“We suspect that the reason not every bot in the botnet was detected is due to the randomness in our choice of selected internal hosts to which the malware traffic was assigned, such that a selected internal host that was also contacting other suspicious subnets (not relevant to the botnet) is likely to bias the dimension reduction and clustering algorithm.”

privacy

Sharing Traces



Is the experiment independently repeatable?

Can we do apples to apples comparison?

What can be done?

Observations

Much of these challenges stem from difficulties in *sharing and obtaining* realistic data sets.

Similar to problems faced by researchers studying large scale distributed systems

---> PlanetLab

Can we do better together?

A PlanetLab for Botnet Detection?

Strawman

Distributed Evaluation

- PlanetLab-like nodes on participating networks
- Cannot communicate network traces outside of network

Researchers Deploy Detector Code on Nodes

- Reports are reviewed and declassified by sys-admins
- Researcher can test and debug on local node

Incentives

- Sys-Admins gain access to bleeding edge detectors, for FREE!
- Researchers gain insight into usefulness of reports or “ground truth”

Address Challenges

Realistic Settings

Network Heterogeneity

Multiple Administrative Domains

Performance

Lack of Ground Truth

Modernity

Overfitting

Comparability & Repeatability

Privacy

Huge Deployment Challenges

Privacy

Accountability

Conclusions

Taking a step back

- Literature Review
- Ideal is hard

Ideal vs. Reality

- Privacy!
- *Sharing* and *Obtaining* realistic traces

Overlay Methodology

- And, its pitfalls

Can we do better together?

- PlanetLab for Botnet detectors?

Backup