On the Robustness of Random Walk Algorithms for the Detection of Unstructured P2P Botnets

Dominik Muhs [1] Steffen Haas [2] Thorsten Strufe [1] Mathias Fischer [2]

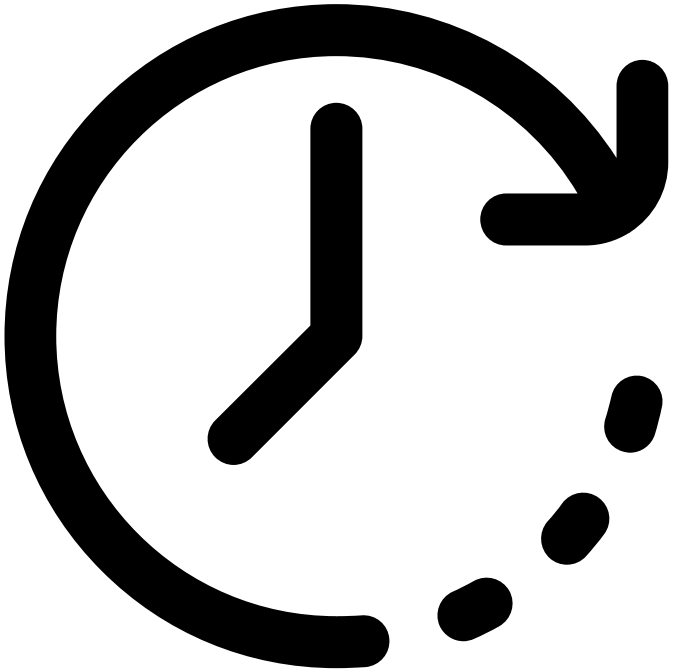
TECHNISCHE UNIVERSITÄT DRESDEN

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

[1] Technische Universität Dresden
Dresden, Germany
first.last@tu-dresden.de

[2] Universität Hamburg
Hamburg, Germany
first.last@informatik.uni-hamburg.de

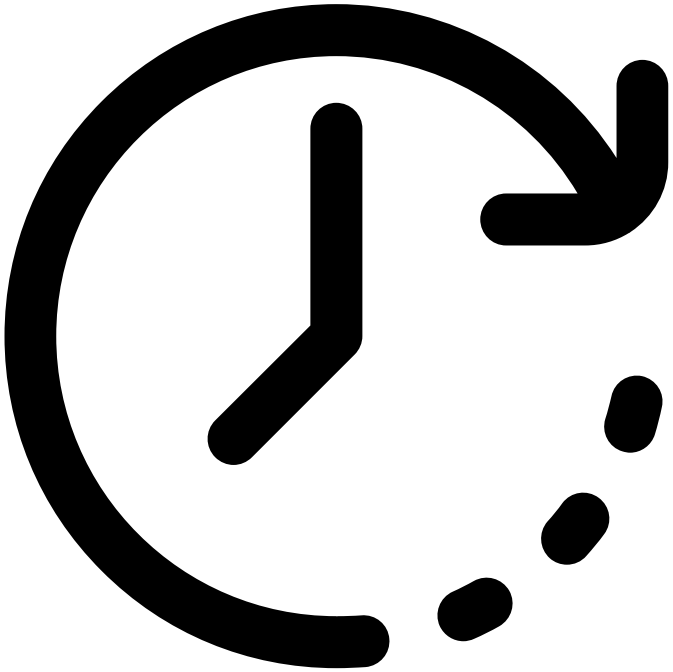
Outline



[7]

Outline

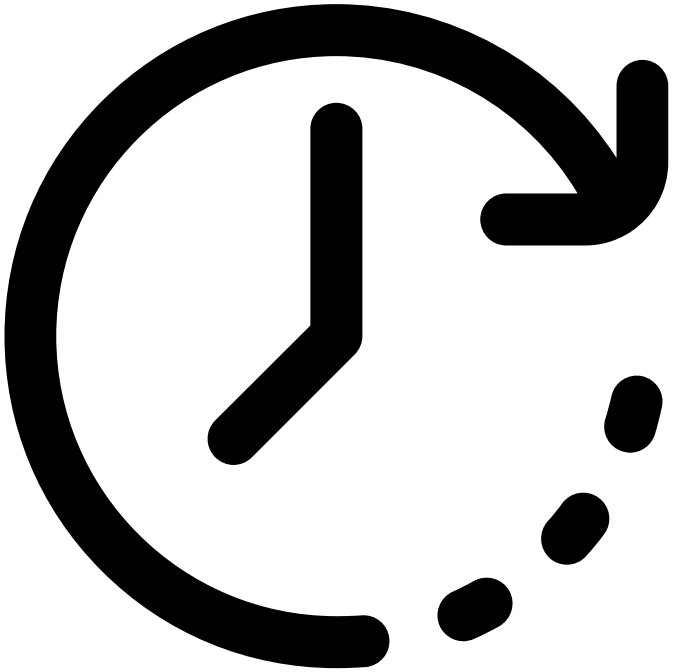
I. Motivation



[7]

Outline

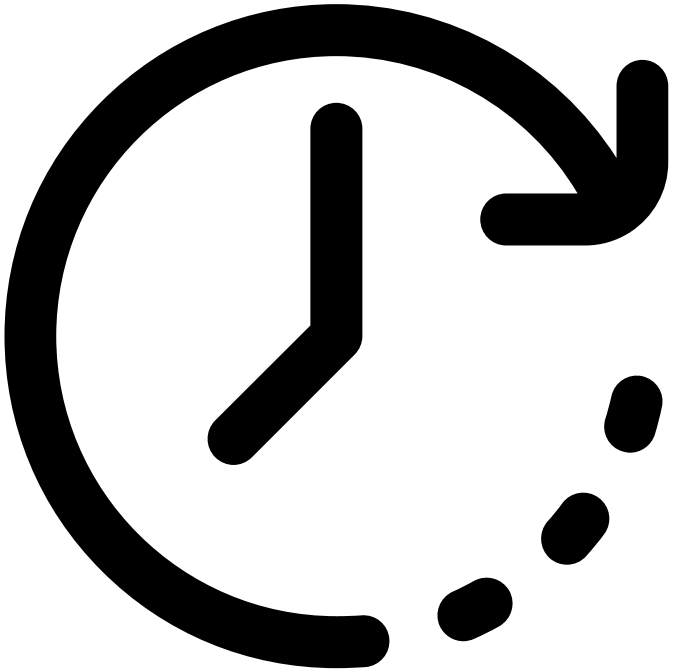
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model



[7]

Outline

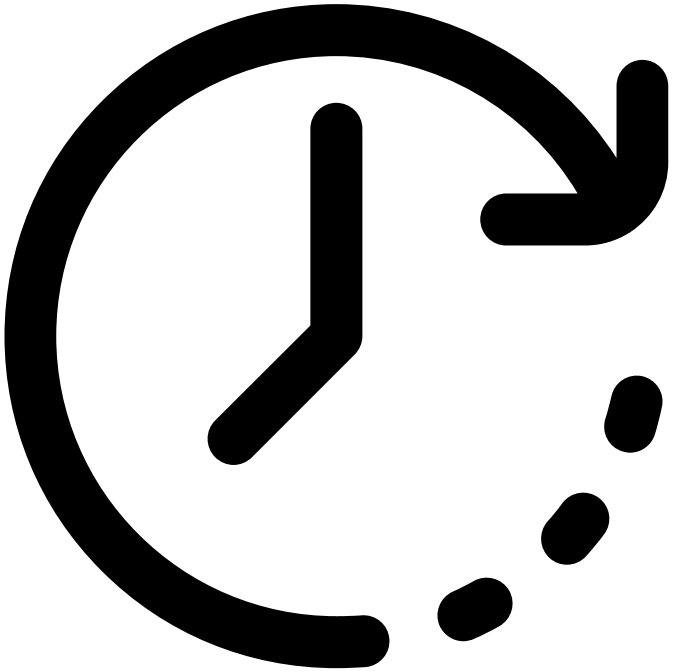
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model
III. Random Walks



[7]

Outline

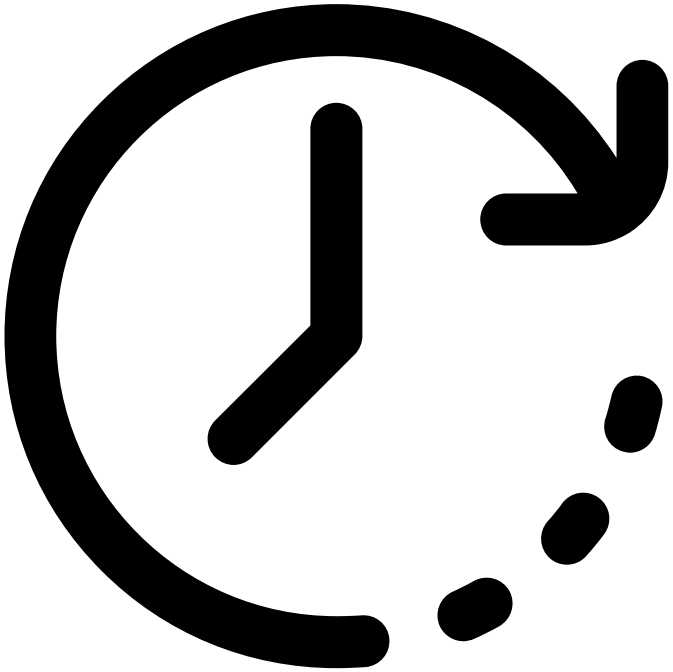
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model
III. Random Walks
IV. Analysis and Detection



[7]

Outline

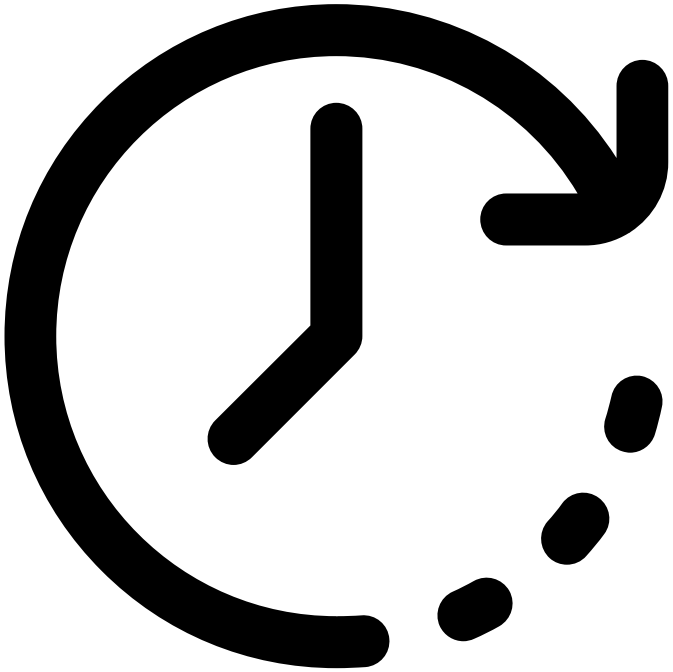
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model
III. Random Walks
IV. Analysis and Detection
V. Limiting Knowledge



[7]

Outline

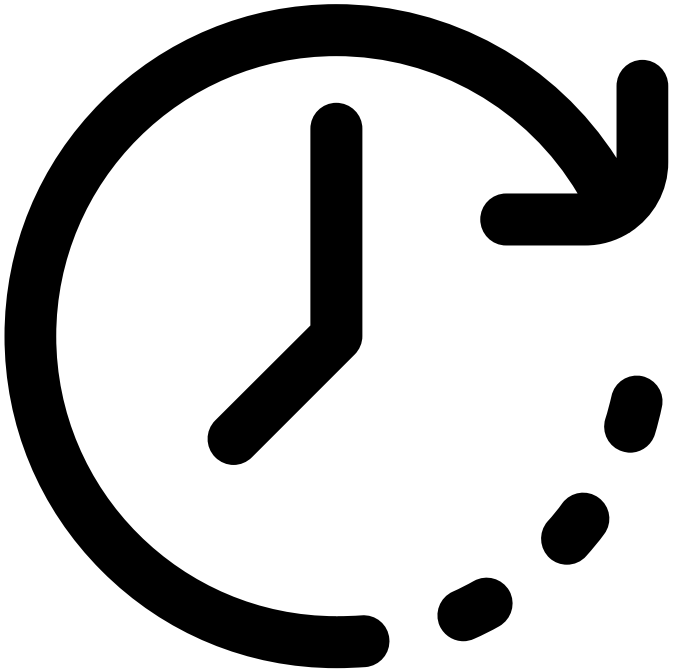
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model
III. Random Walks
IV. Analysis and Detection
V. Limiting Knowledge
VI. Results



[7]

Outline

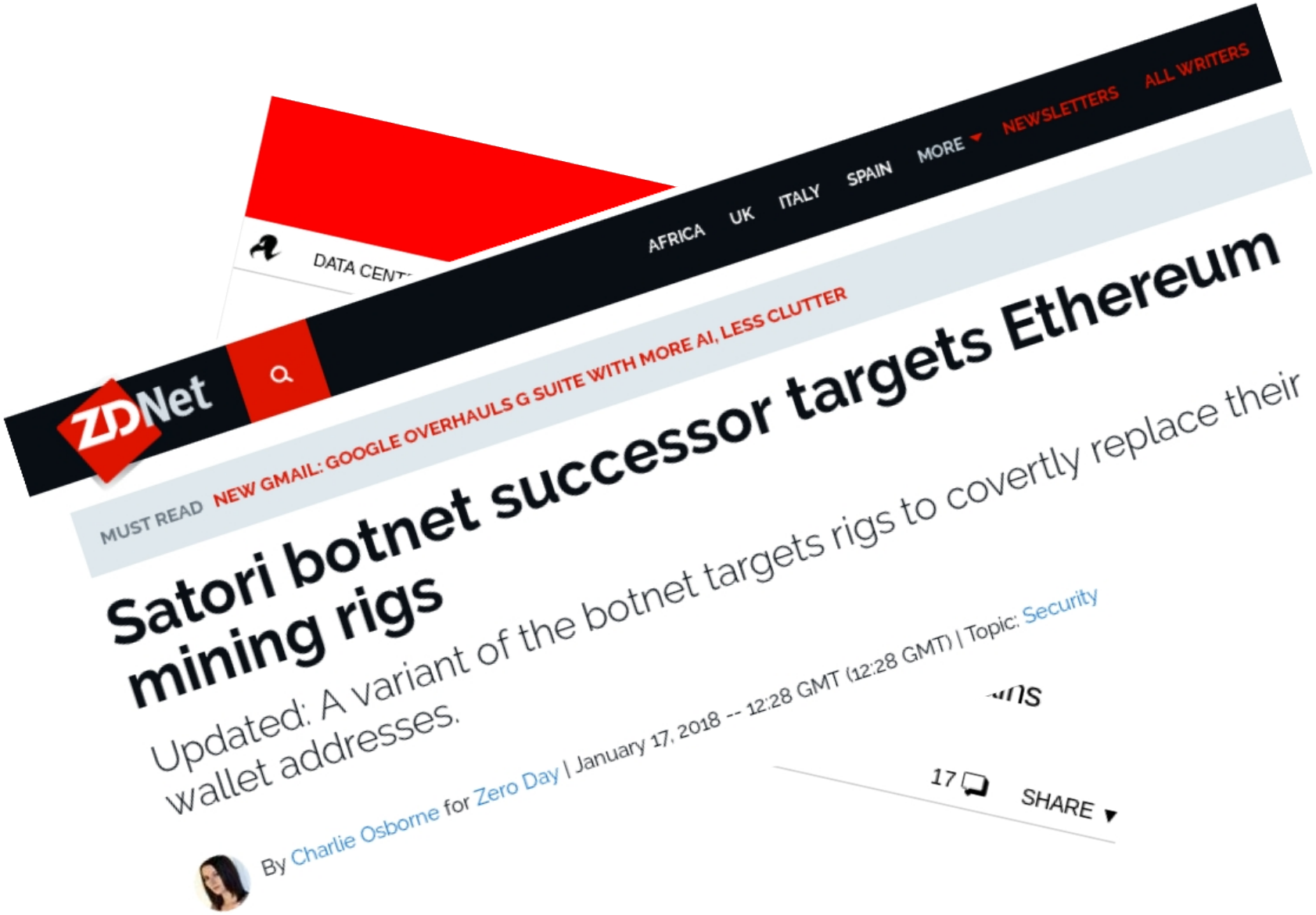
I. Motivation
II. Botnets
    1. Definition
    2. Graph Model
III. Random Walks
IV. Analysis and Detection
V. Limiting Knowledge
VI. Results
VII. Conclusion



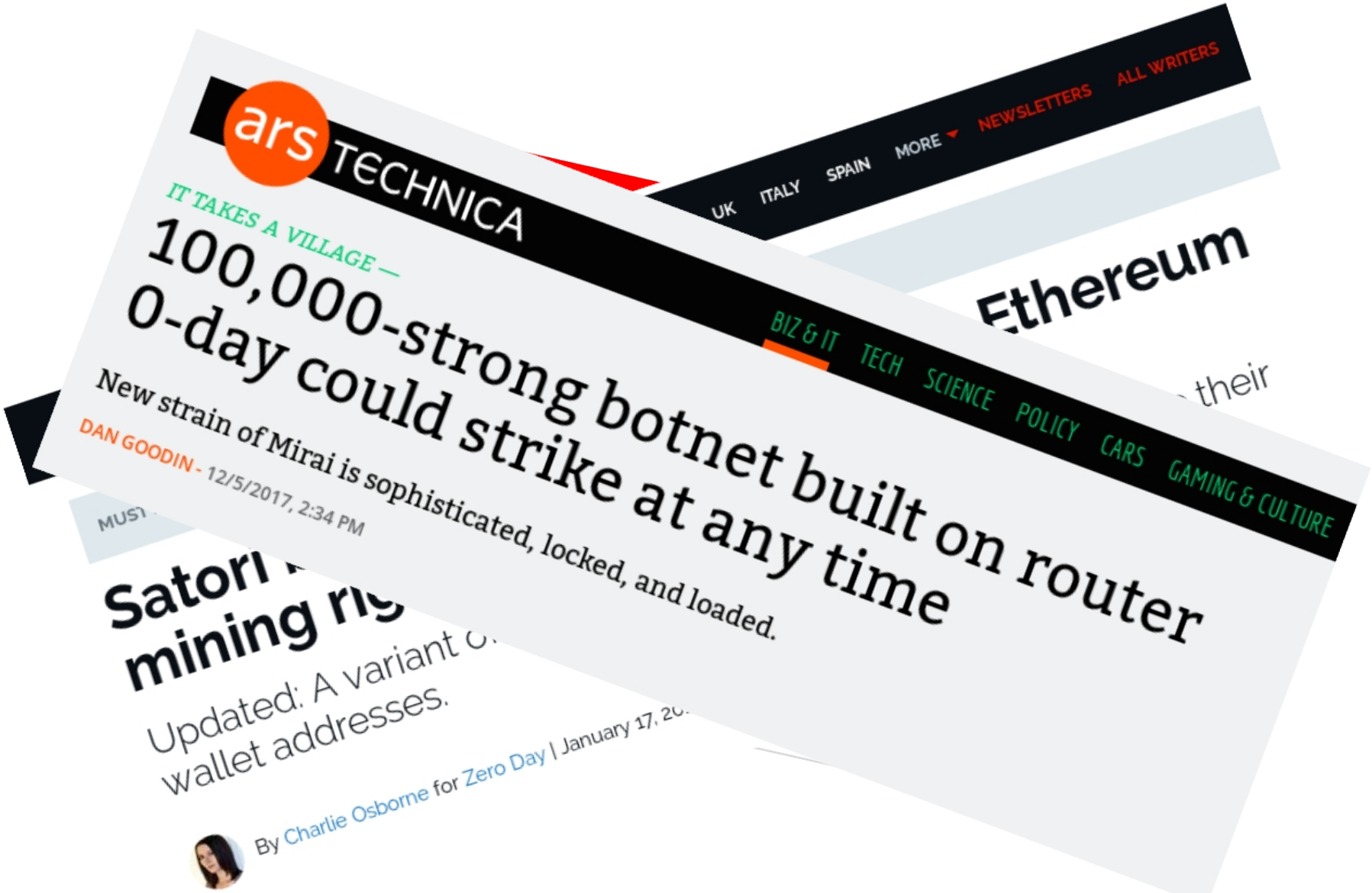
[7]


The Register®
Biting the hand that feeds IT
DATA CENTRE
SOFTWARE
SECURITY
DEVOPS
BUSINESS
PERSONAL TECH
SCIENCE
Security
Mysterious Hajime botnet has pwned 300,000 IoT devices
The Dark Knight of malware's purpose remains unknown
By John Leyden 27 Apr 2017 at 16:02
17
SHARE

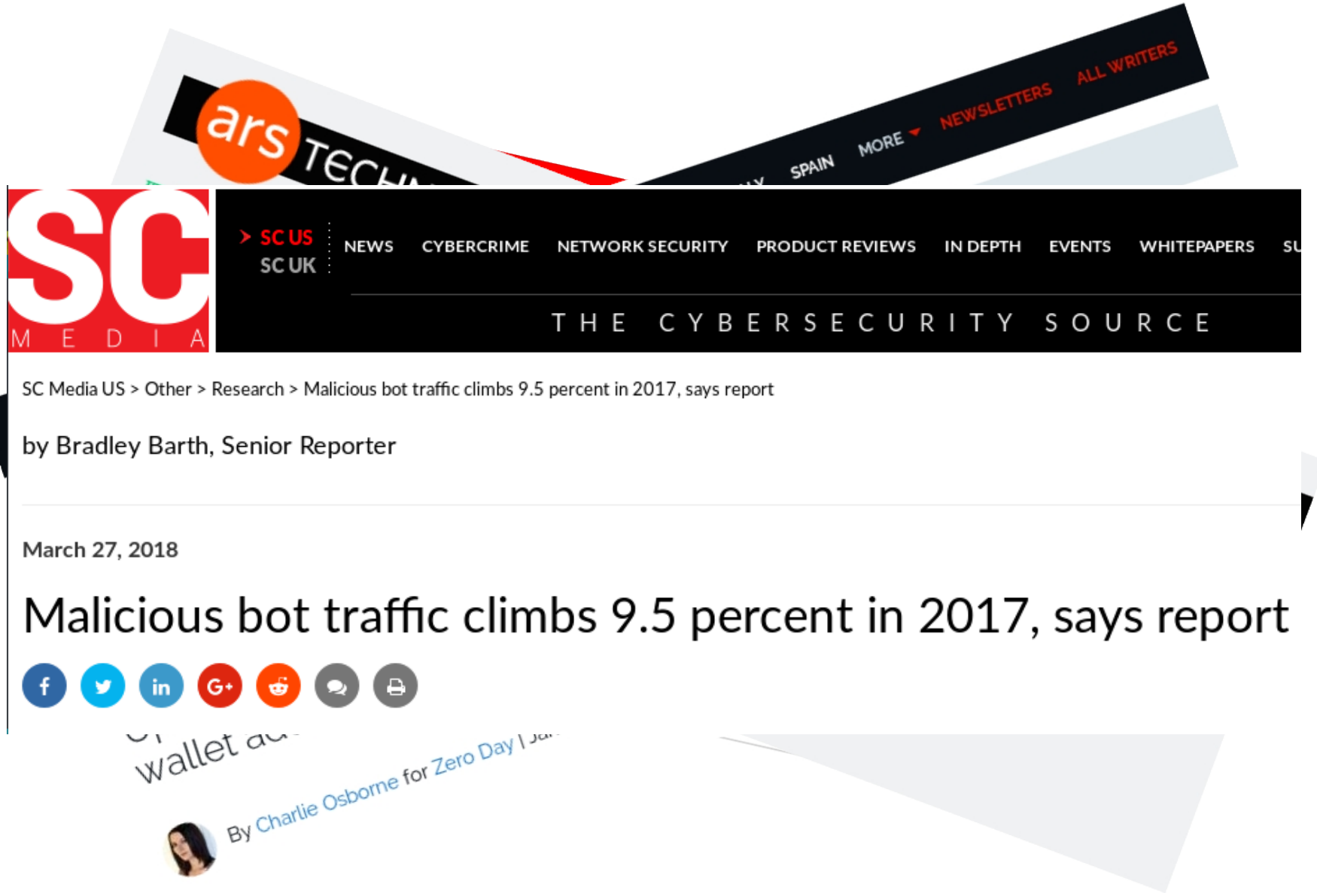
[1]


AFRICA UK ITALY SPAIN MORE NEWSLETTERS ALL WRITERS
ZDNet
MUST READ NEW GMAIL: GOOGLE OVERHAULS G SUITE WITH MORE AI, LESS CLUTTER
Satori botnet successor targets Ethereum mining rigs
Updated: A variant of the botnet targets rigs to covertly replace their wallet addresses.
By Charlie Osborne for Zero Day | January 17, 2018 -- 12:28 GMT (12:28 GMT) | Topic: Security
17
SHARE

[2]


UK ITALY SPAIN MORE NEWSLETTERS ALL WRITERS
ars TECHNICA
BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE
IT TAKES A VILLAGE —
100,000-strong botnet built on router 0-day could strike at any time
New strain of Mirai is sophisticated, locked, and loaded.
DAN GOODIN - 12/5/2017, 2:34 PM
Ethereum
their
mining
Updated: A variant
wallet addresses.
By Charlie Osborne for Zero Day | January 17, 20

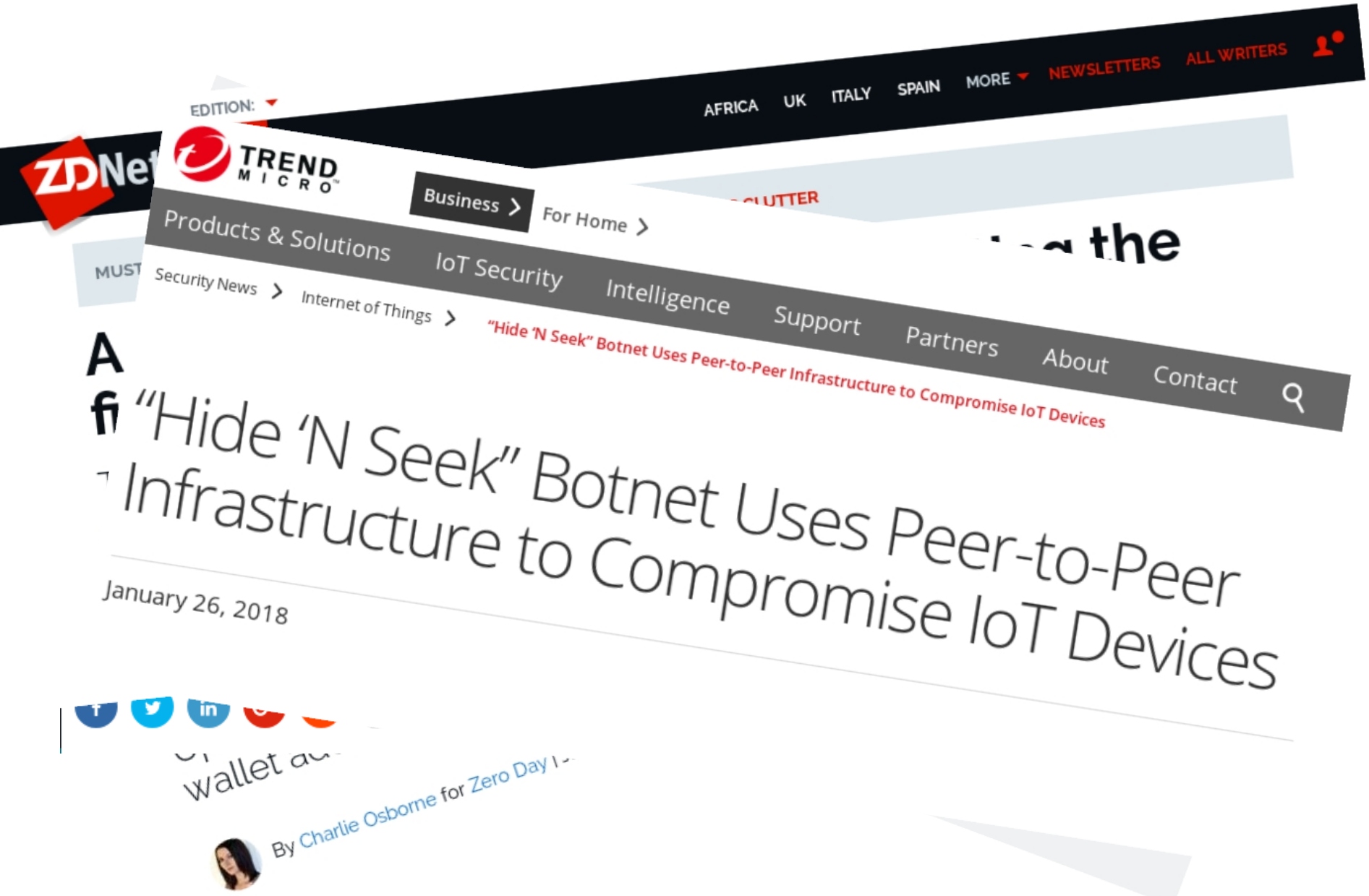
[3]


ars TECH
SPAIN
MORE
NEWSLETTERS
ALL WRITERS
SC MEDIA
SC US
SC UK
NEWS
CYBERCRIME
NETWORK SECURITY
PRODUCT REVIEWS
IN DEPTH
EVENTS
WHITEPAPERS
THE CYBERSECURITY SOURCE
SC Media US > Other > Research > Malicious bot traffic climbs 9.5 percent in 2017, says report
by Bradley Barth, Senior Reporter
March 27, 2018
Malicious bot traffic climbs 9.5 percent in 2017, says report
wallet
By Charlie Osborne for Zero Day

[4]

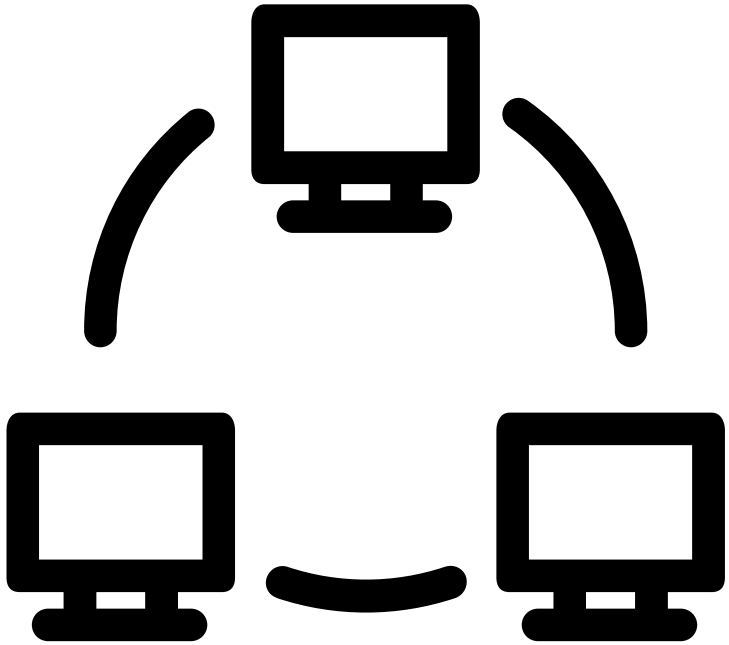

EDITION:
AFRICA
UK
ITALY
SPAIN
MORE
NEWSLETTERS
ALL WRITERS
ZDNet
MUST READ NEW GMAIL: GOOGLE OVERHAULS G SUITE WITH MORE AI, LESS CLUTTER
A new Mirai-style botnet is targeting the financial sector
The researchers say it's the largest attack since the Mirai-powered cyberattack in October 2016 that took down large swathes of the Western internet.
By Zack Whittaker for Zero Day | April 5, 2018 -- 13:00 GMT (14:00 BST) | Topic: Security
percent in 2017, says report
wallet
By Charlie Osborne for Zero Day

[5]


EDITION:
AFRICA UK ITALY SPAIN MORE NEWSLETTERS ALL WRITERS
TREND MICRO
Business
For Home
Products & Solutions
IoT Security
Intelligence
Support
Partners
About
Contact
Security News
Internet of Things
"Hide 'N Seek" Botnet Uses Peer-to-Peer Infrastructure to Compromise IoT Devices
"Hide 'N Seek" Botnet Uses Peer-to-Peer Infrastructure to Compromise IoT Devices
January 26, 2018
By Charlie Osborne for Zero Day

[6]

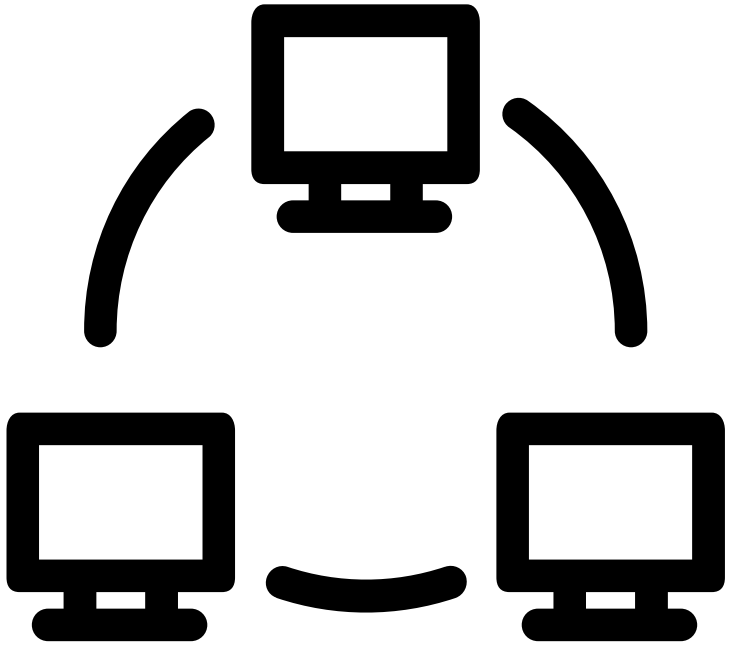
What are Botnets?



[9]

What are Botnets?

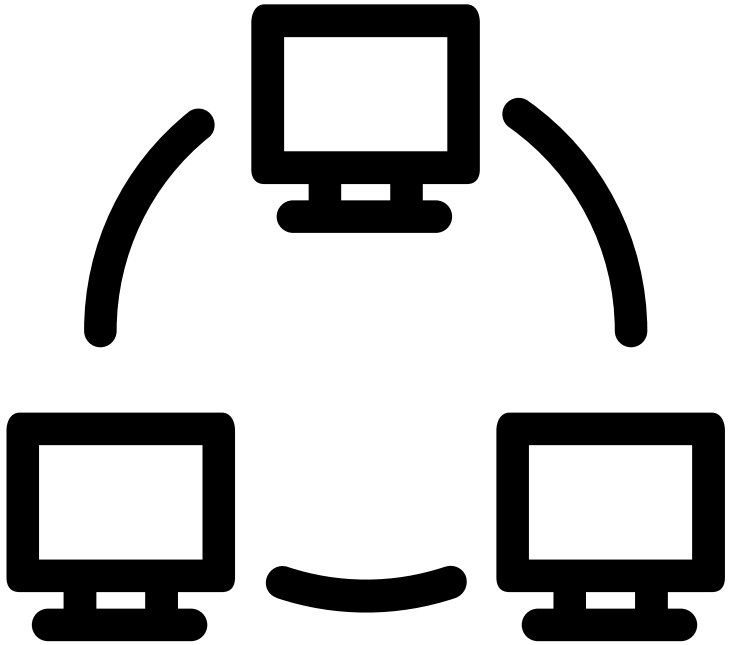
- Device collection



[9]

What are Botnets?

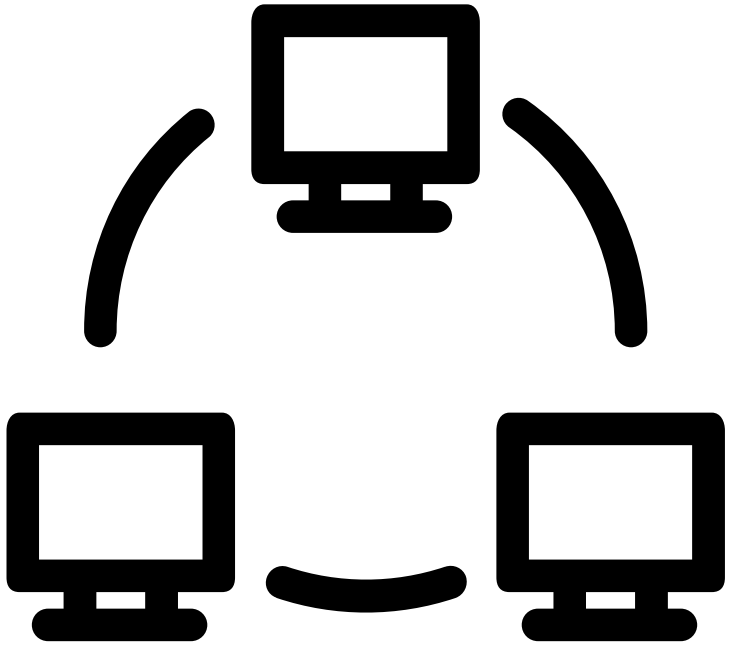
- Device collection
- Internet-connected



[9]

What are Botnets?

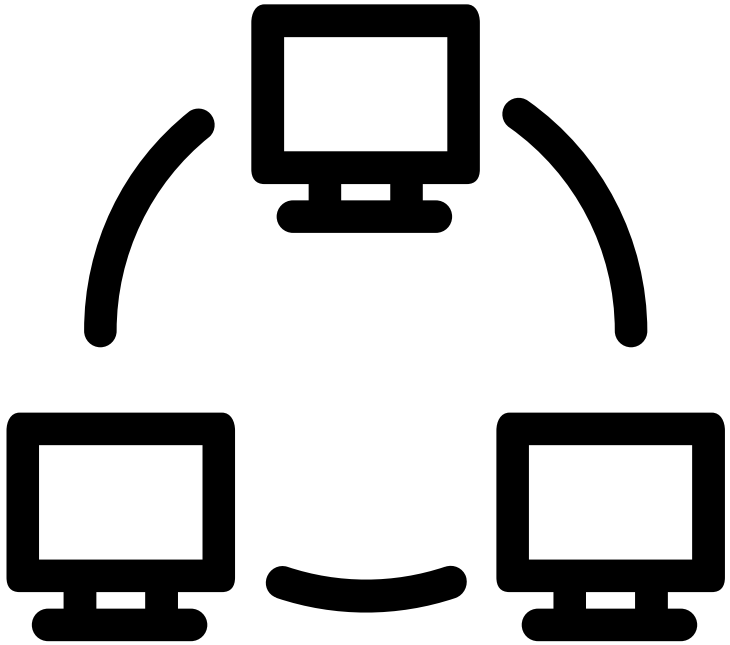
- Device collection
- Internet-connected
- Malware-infected



[9]

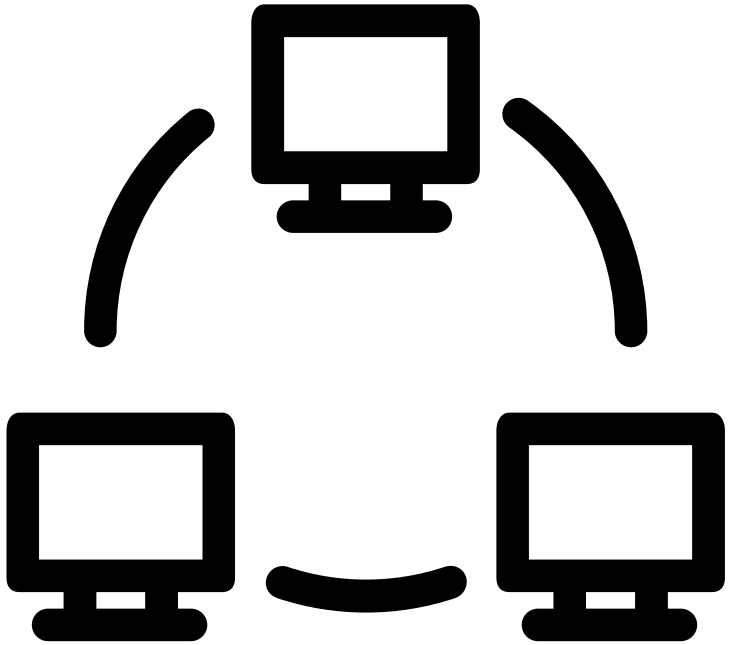
What are Botnets?

- Device collection
- Internet-connected
- Malware-infected
- Remotely controlled (usually centralized)



[9]

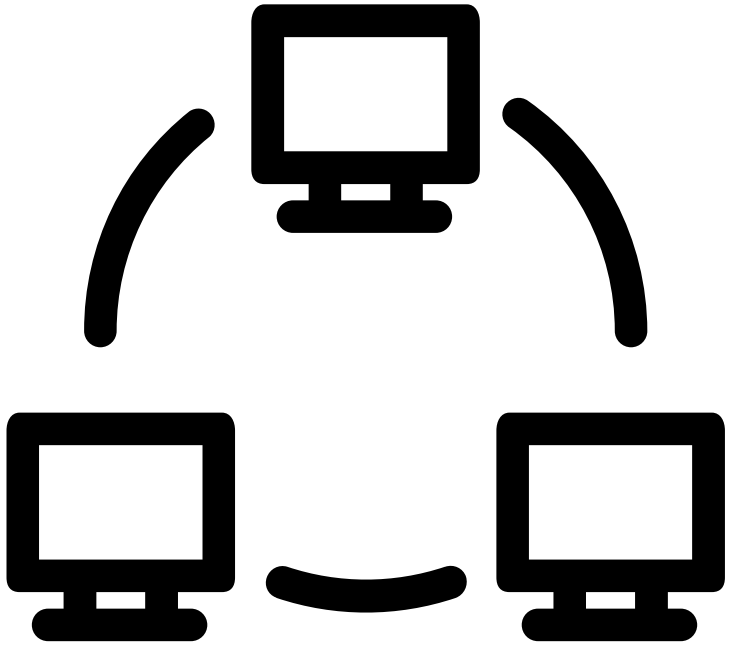
Why are Botnets bad?



[9]

Why are Botnets bad?

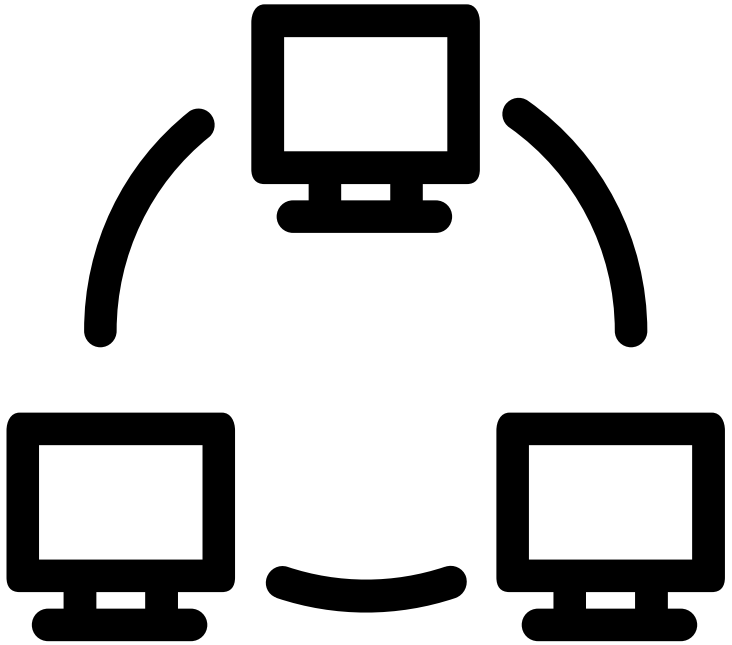
- Clickfraud



[9]

Why are Botnets bad?

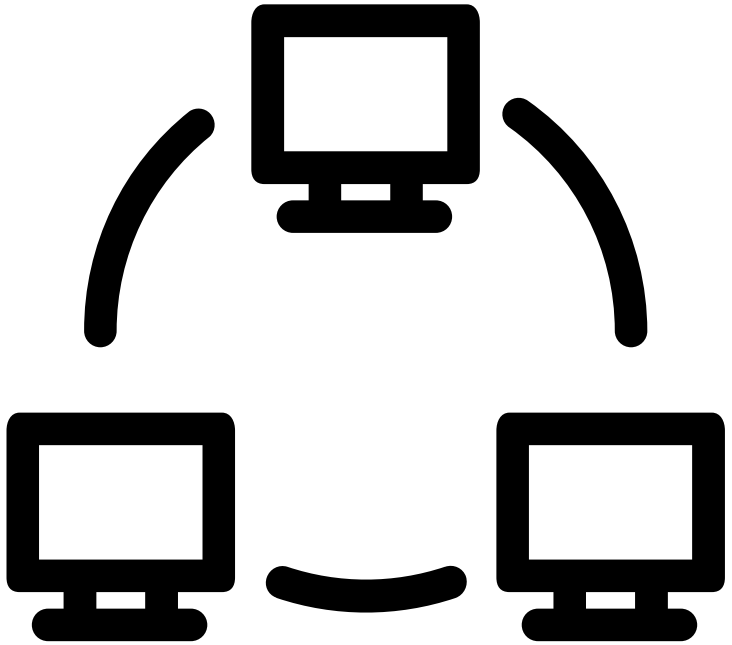
- Clickfraud
- Spam



[9]

Why are Botnets bad?

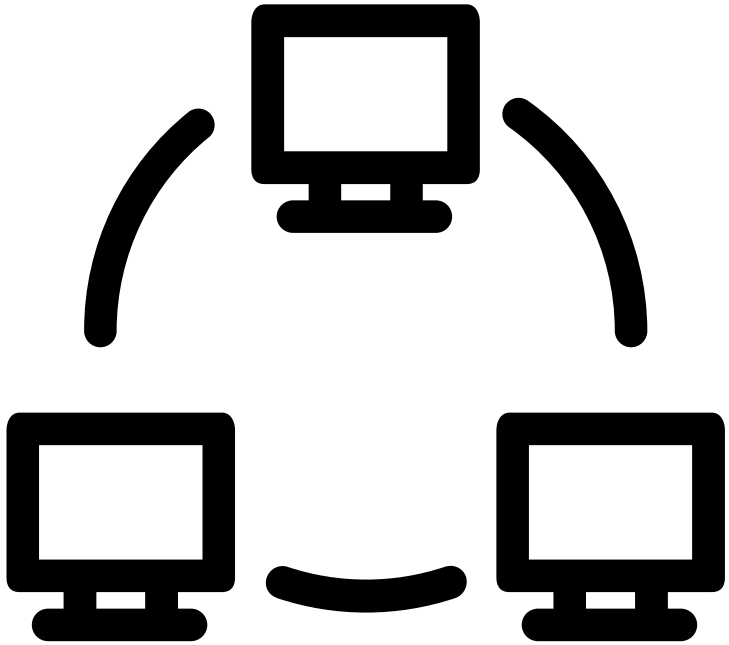
- Clickfraud
- Spam
- DDoS attacks



[9]

Why are Botnets bad?

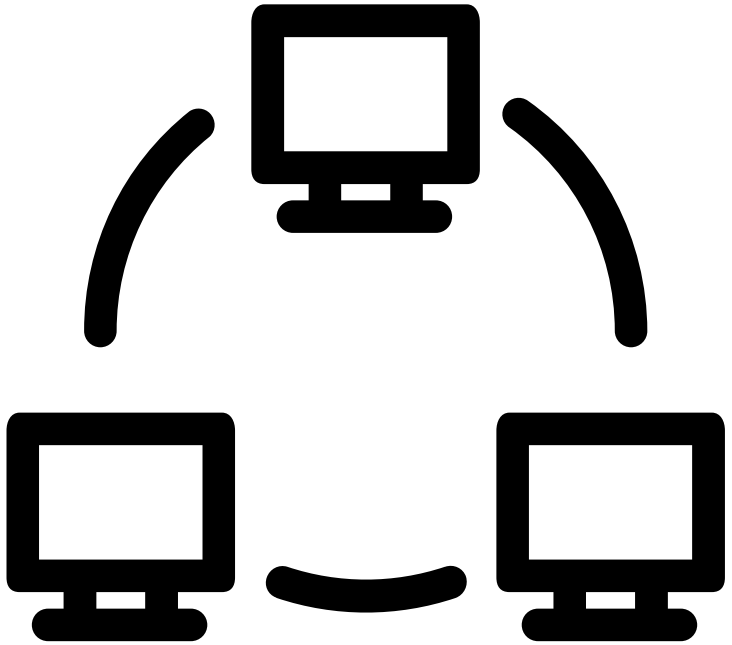
- Clickfraud
- Spam
- DDoS attacks
- Cryptocurrency mining



[9]

Why are Botnets bad?

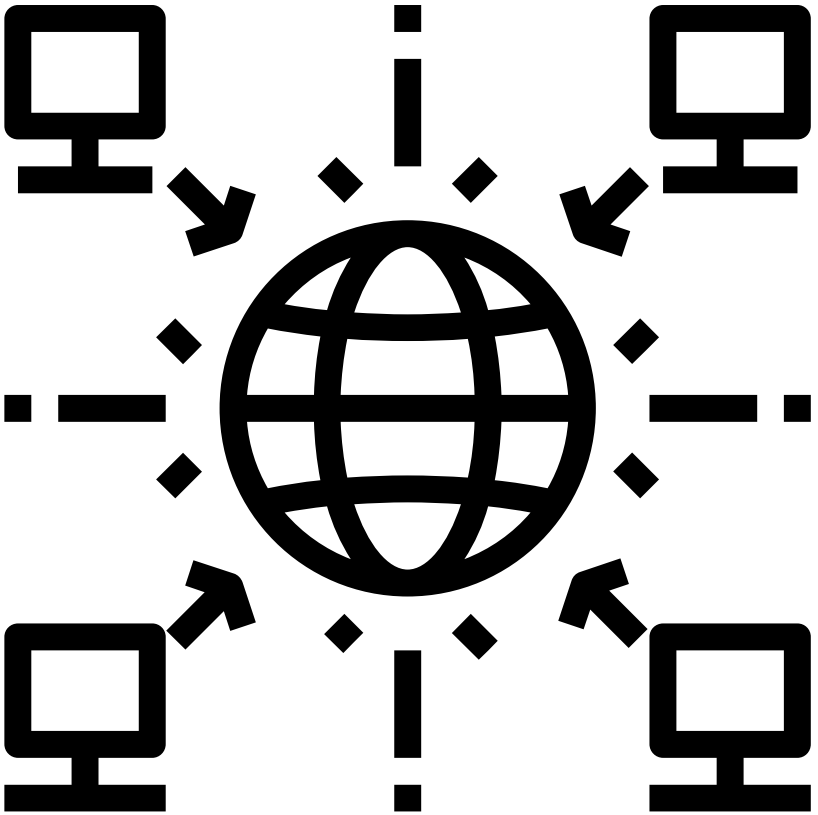
- Clickfraud
- Spam
- DDoS attacks
- Cryptocurrency mining
- Intellectual property theft



[9]

Topological Categories

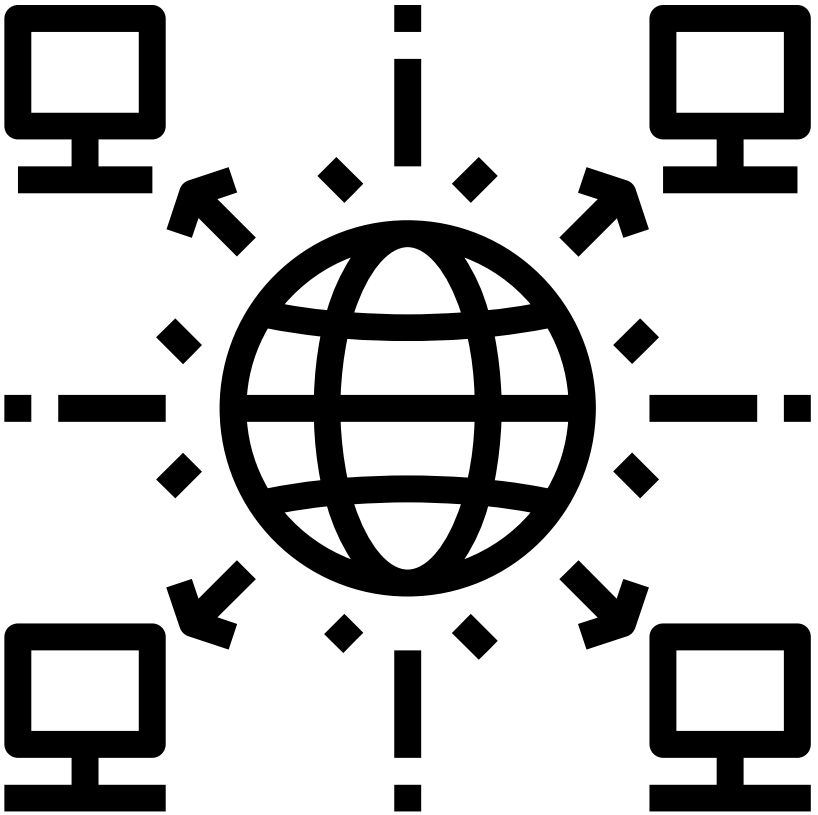
- Centralized



[8]

Topological Categories

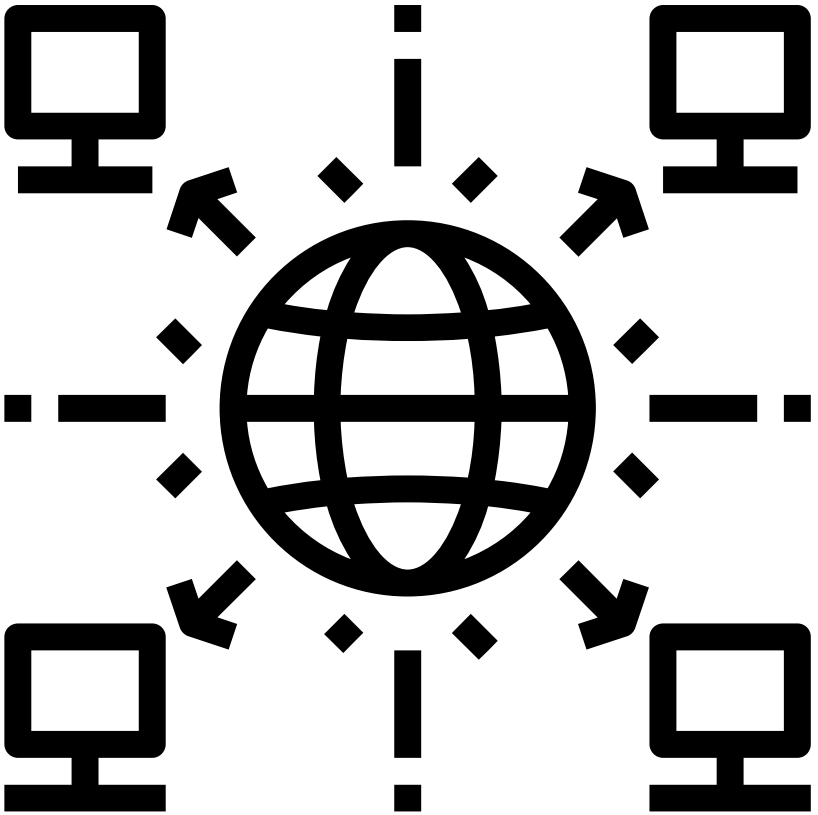
- Centralized
- Decentralized



[8]

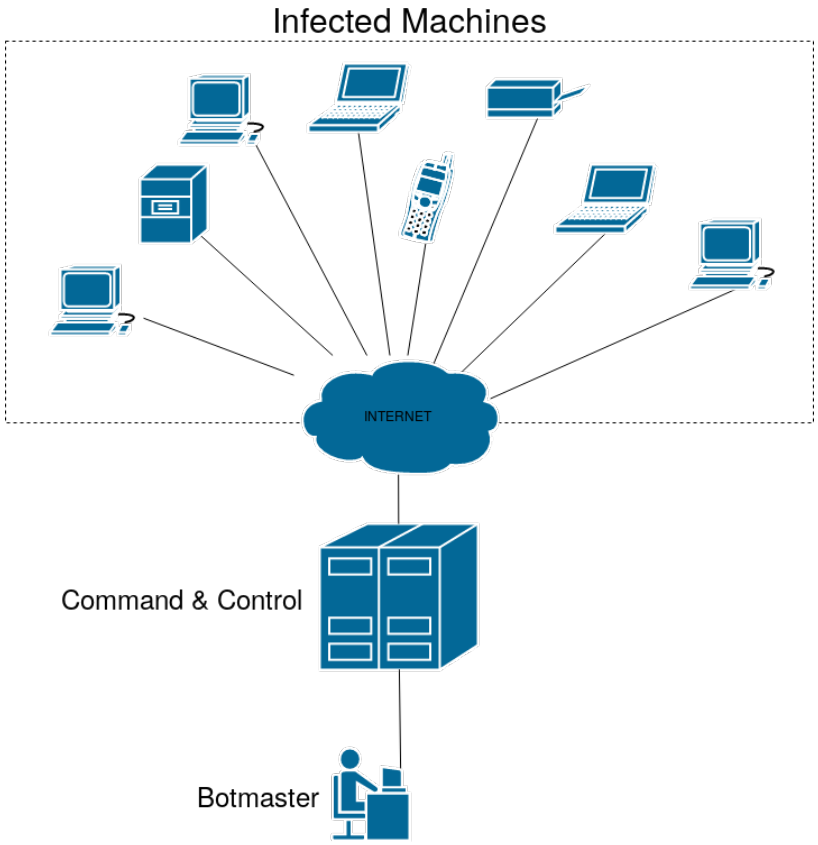
Topological Categories

- Centralized
- Decentralized
  - Structured
  - Unstructured



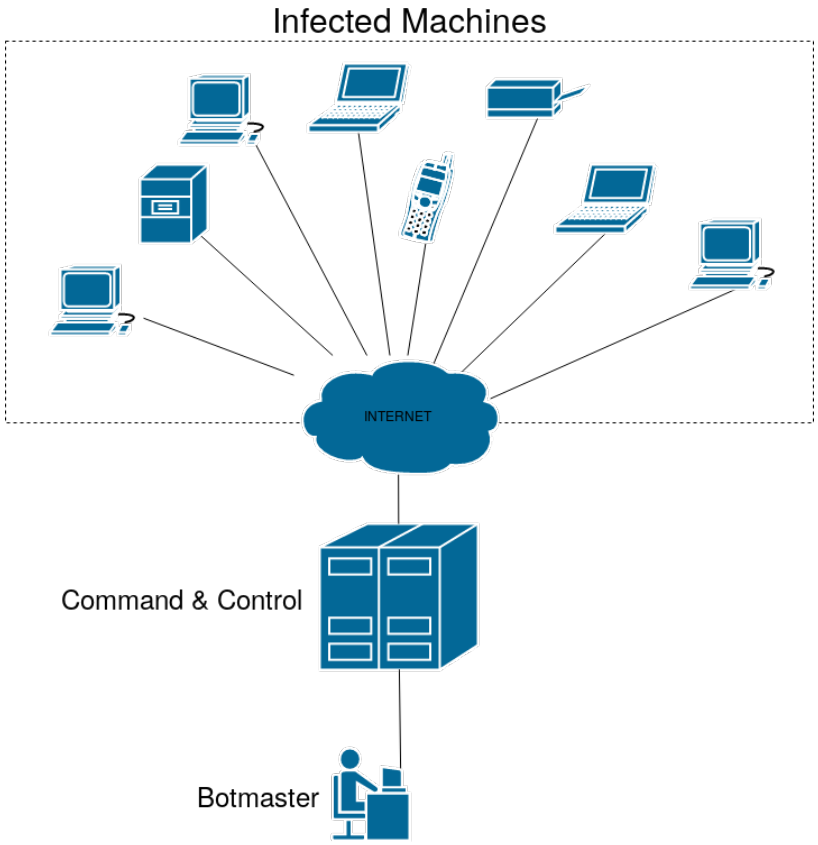
[8]

Centralized Botnets



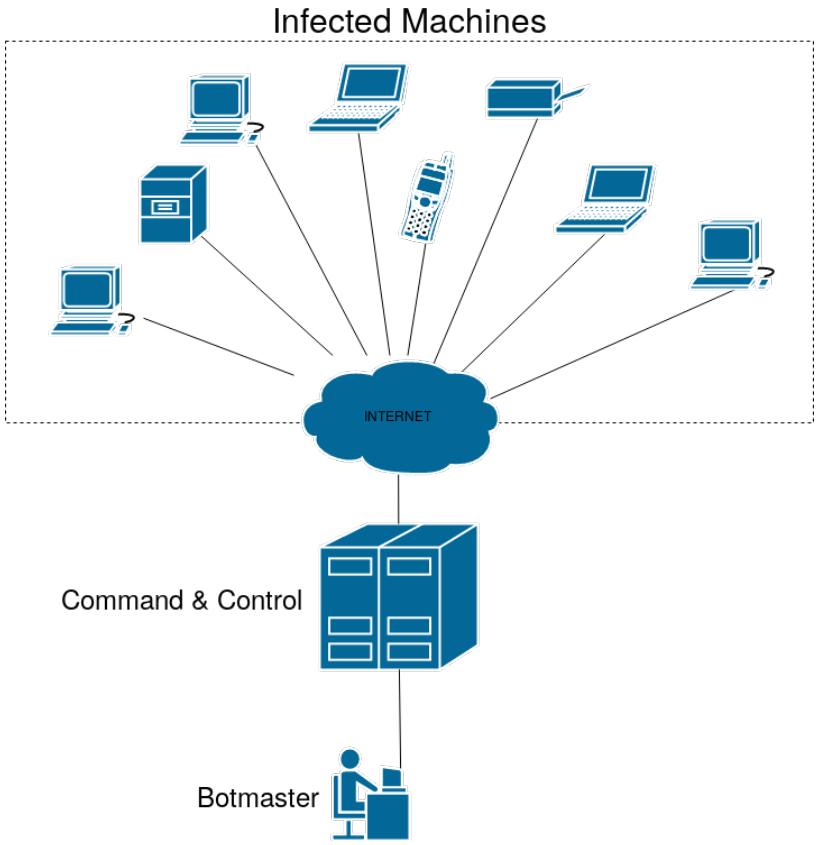
Centralized Botnets

- Central C2 server



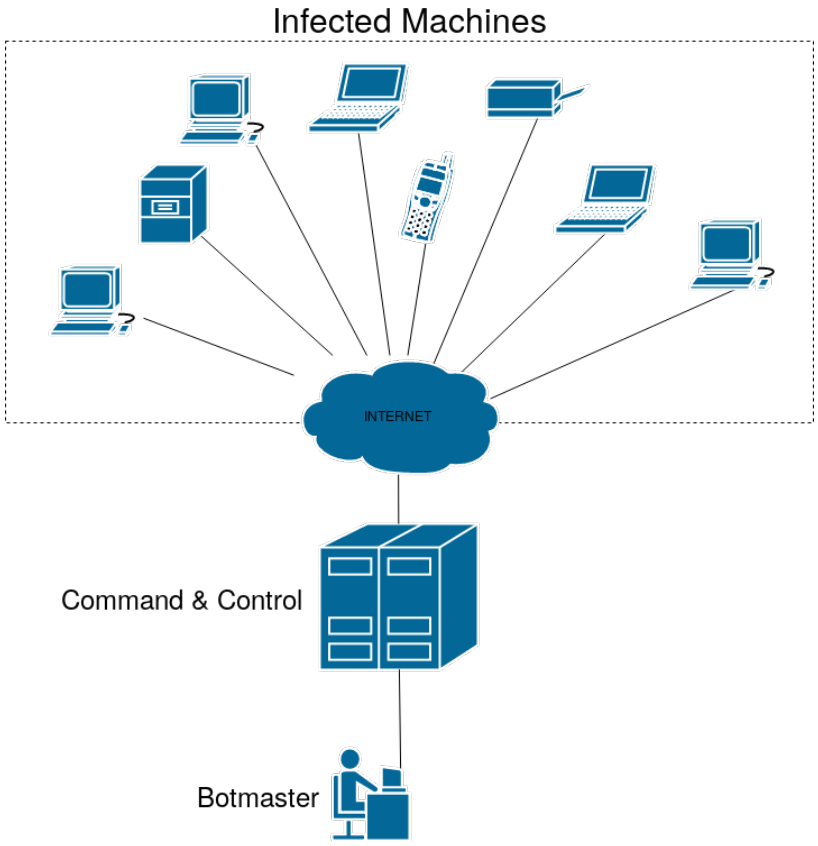
Centralized Botnets

- Central C2 server
- Star topology



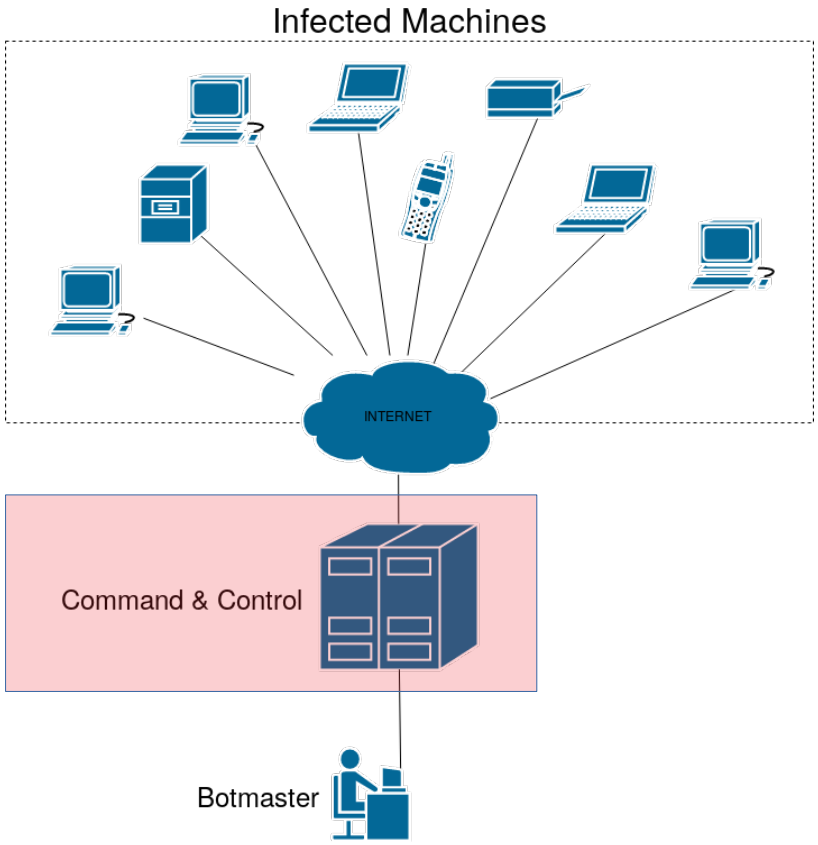
Centralized Botnets

- Central C2 server
- Star topology
- IRC/HTTP/…

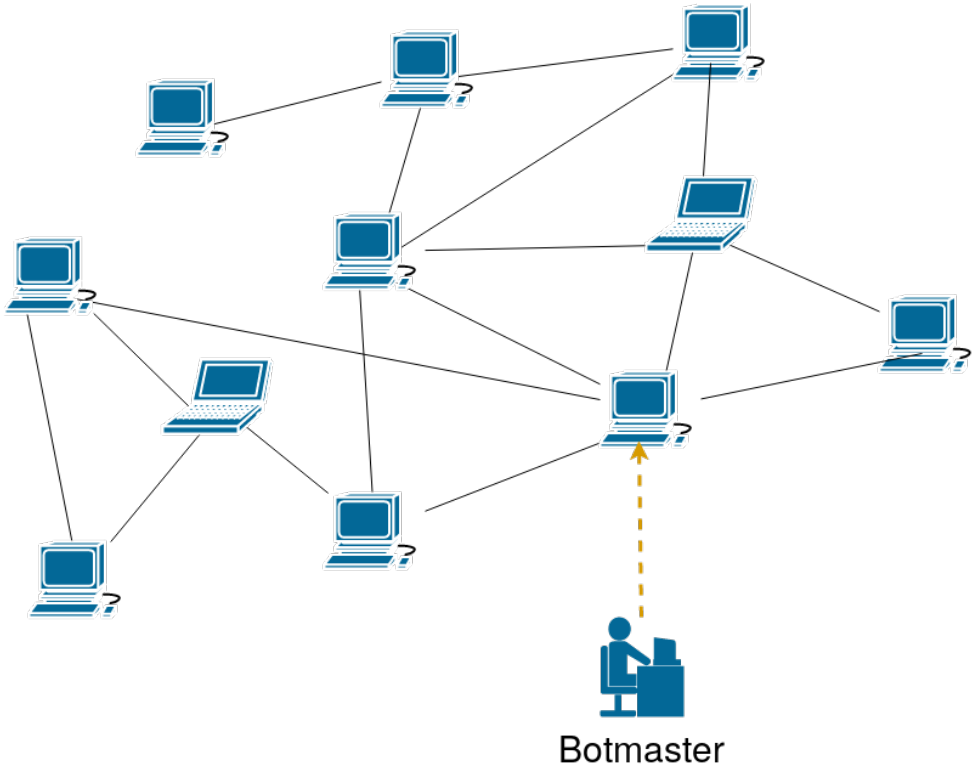


Centralized Botnets

- Central C2 server
- Star topology
- IRC/HTTP/…
- Single point of failure

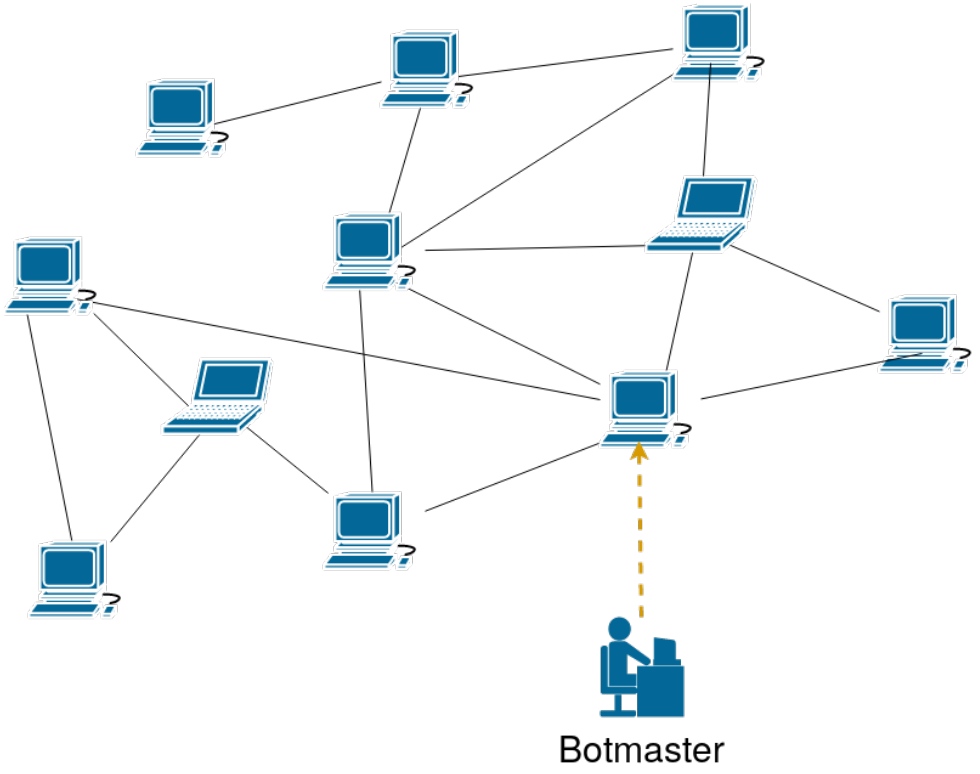


Structured P2P Botnets



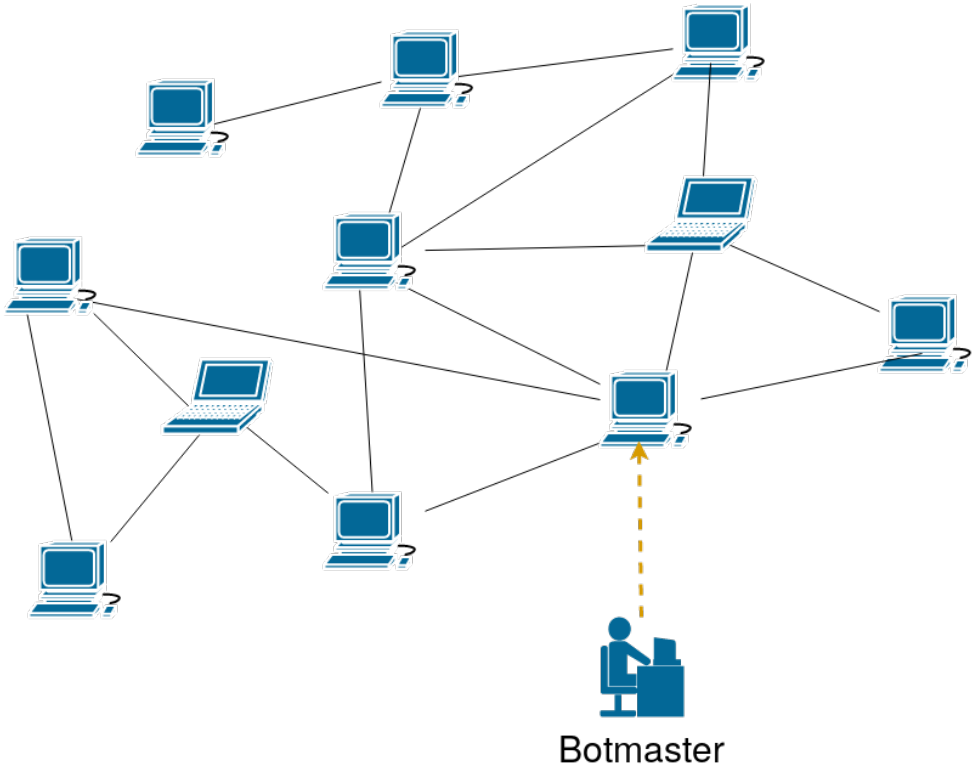
Structured P2P Botnets

- No C2 server



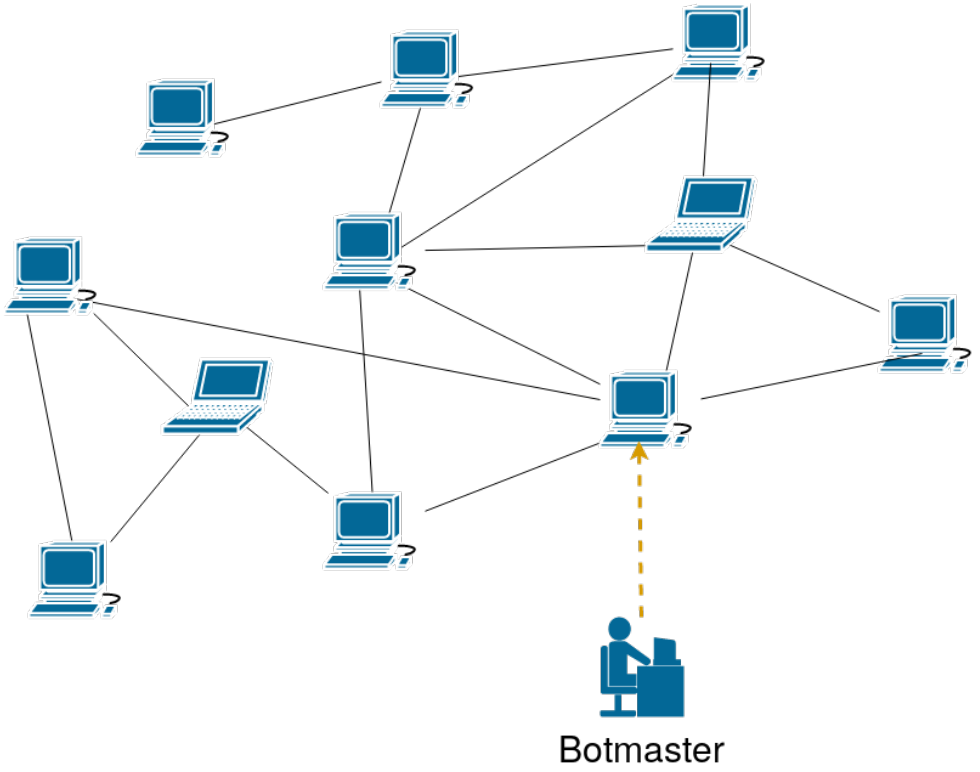
Structured P2P Botnets

- No C2 server
- Hard to take down



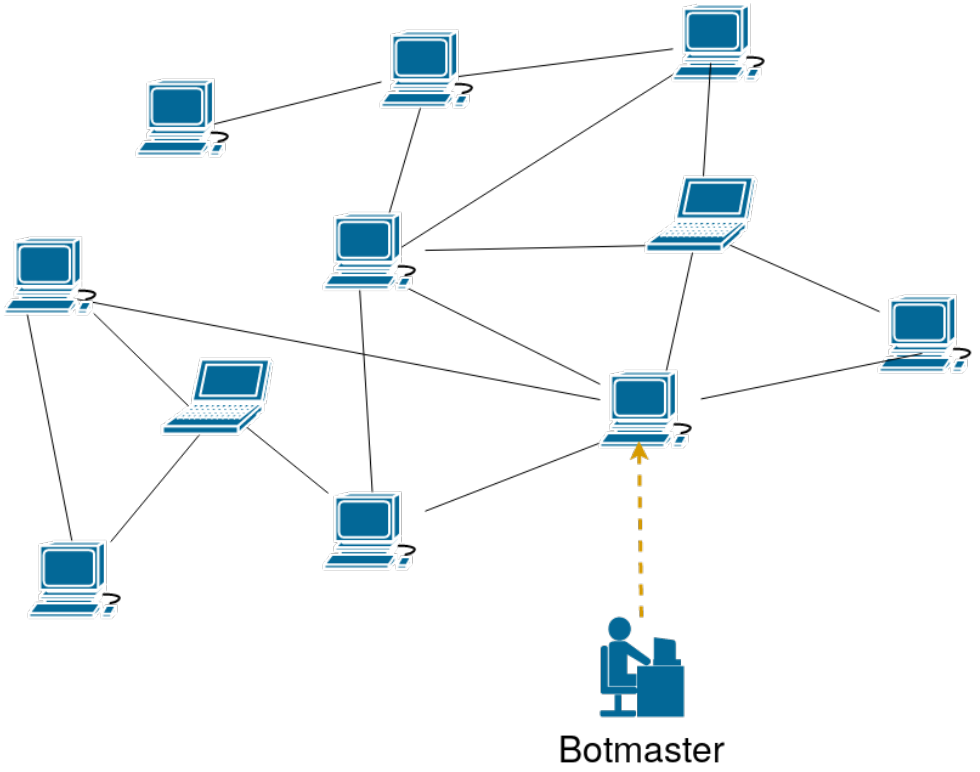
Structured P2P Botnets

- No C2 server
- Hard to take down
- Specific rule set

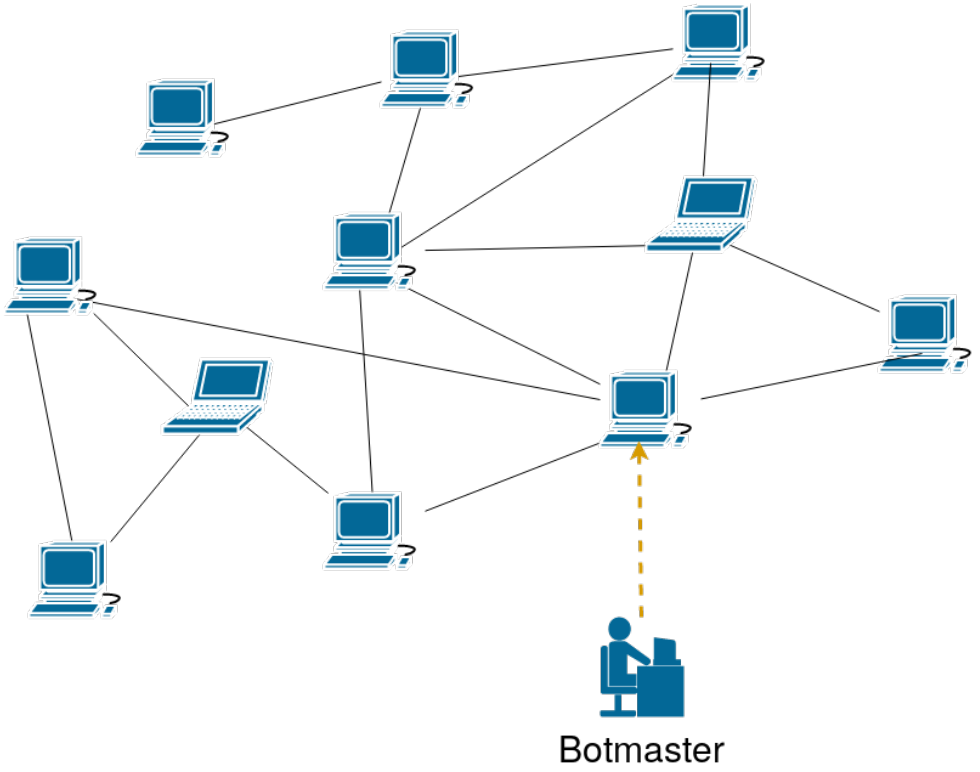


Structured P2P Botnets

- No C2 server
- Hard to take down
- Specific rule set
- Kademlia, Chord

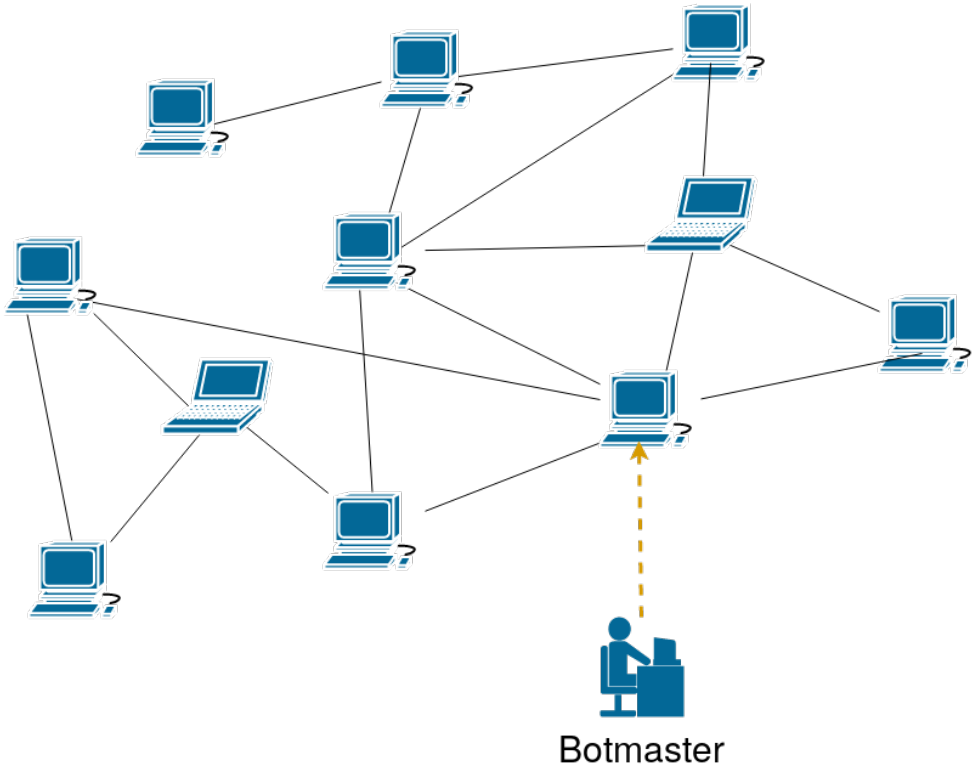


Unstructured P2P Botnets



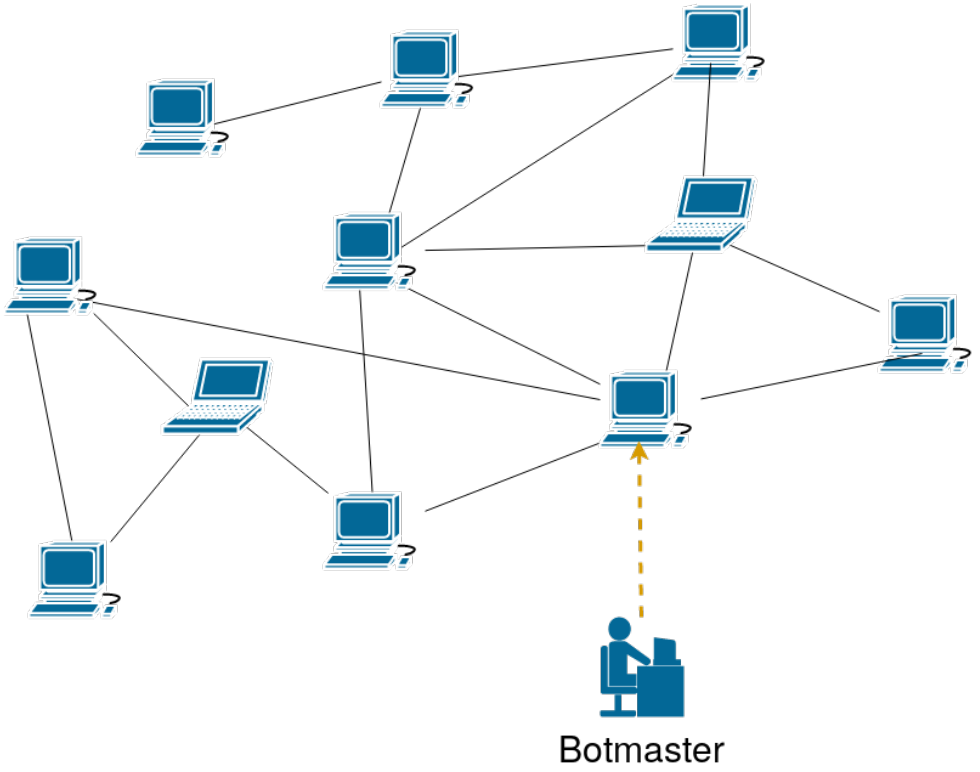
Unstructured P2P Botnets

- Randomized



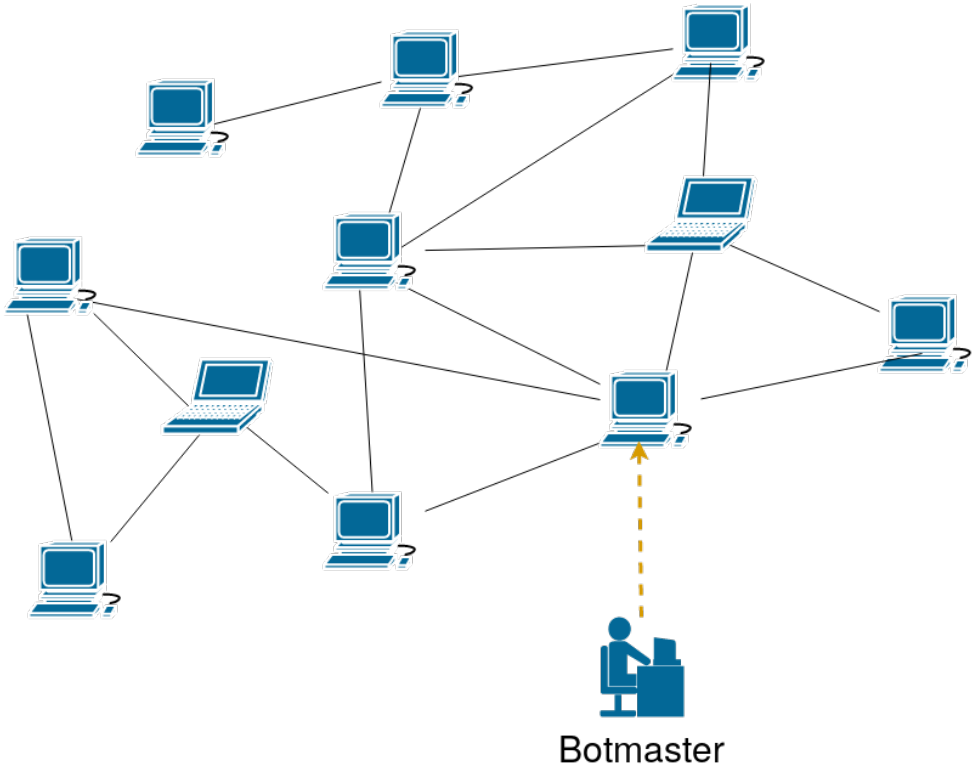
Unstructured P2P Botnets

- Randomized
- Evade topological matching

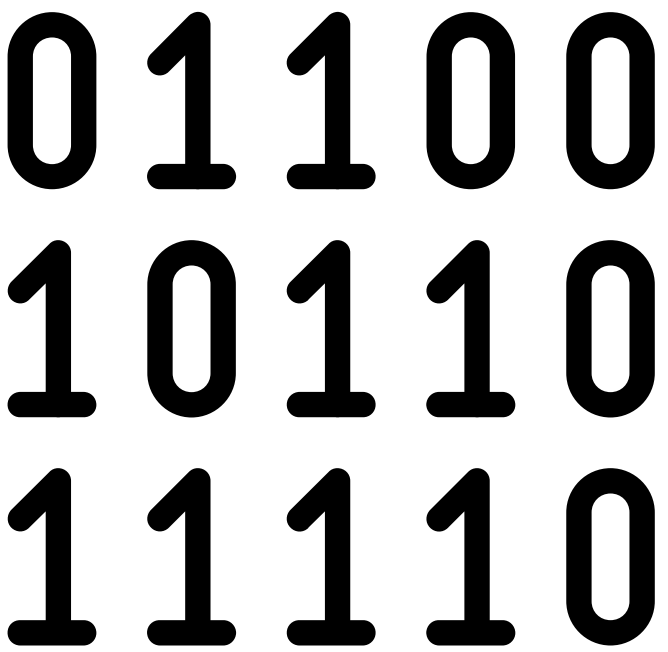


Unstructured P2P Botnets

- Randomized
- Evade topological matching
- Statistical methods necessary



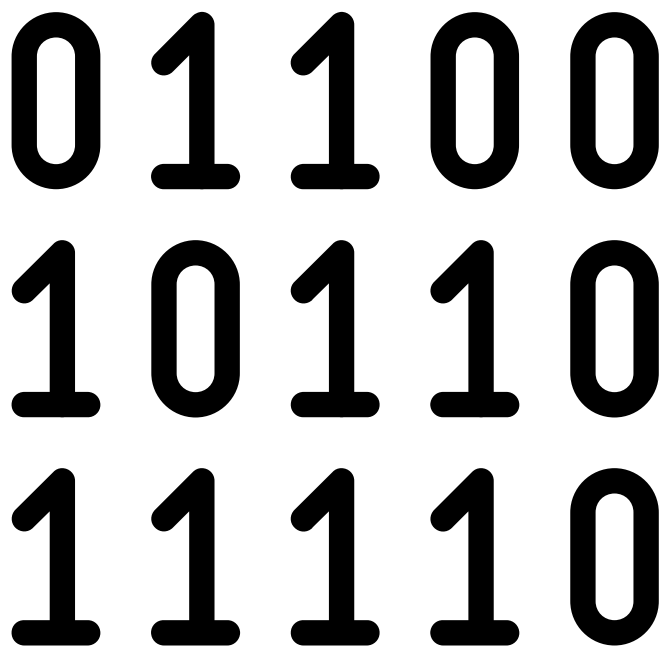
Existing Approaches



[7]

Existing Approaches

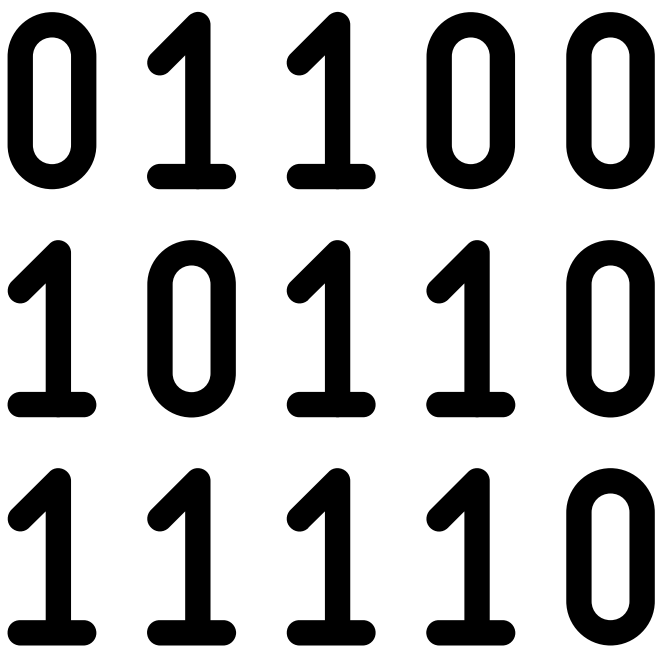
- Leverage graph models



[7]

Existing Approaches

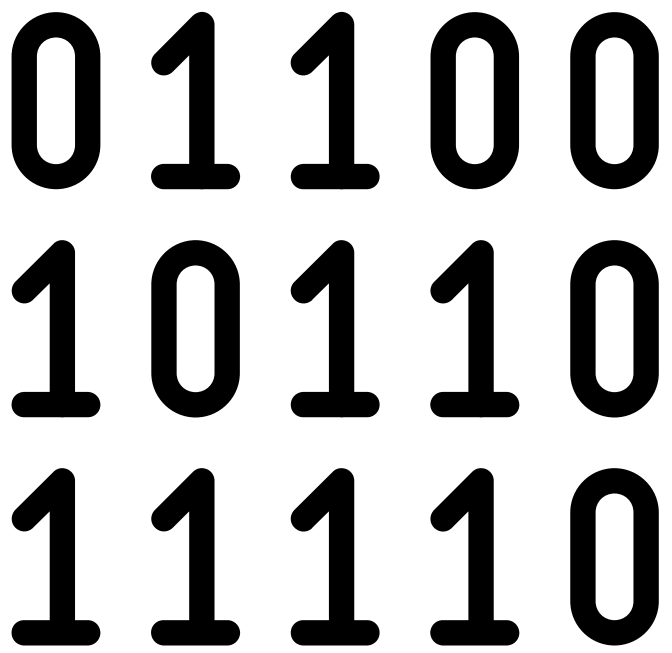
- Leverage graph models
- … and random walks



[7]

Existing Approaches

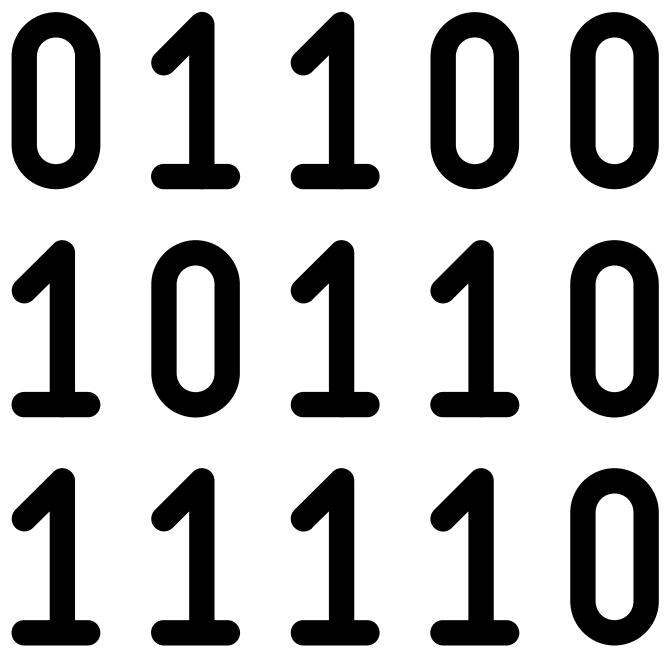
- Leverage graph models
- … and random walks



[7]

Existing Approaches

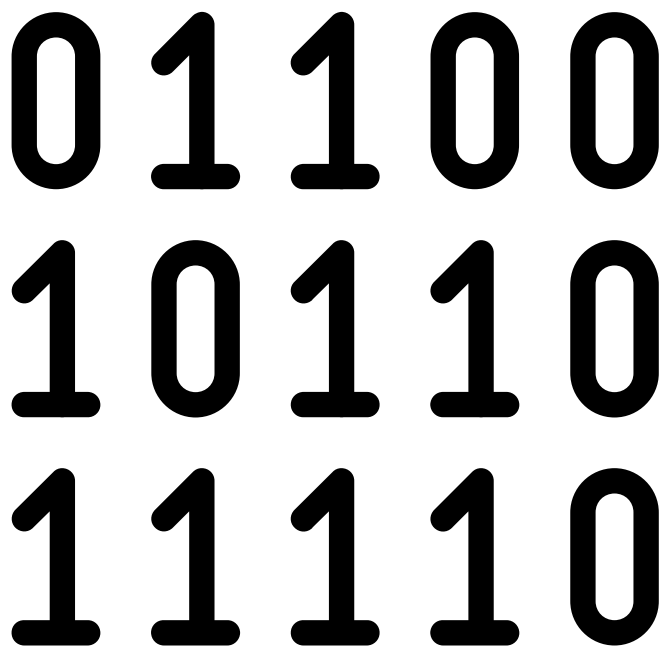
- Leverage graph models
- … and random walks
- Focus on structured botnets [10, 11, 12]



[7]

Existing Approaches

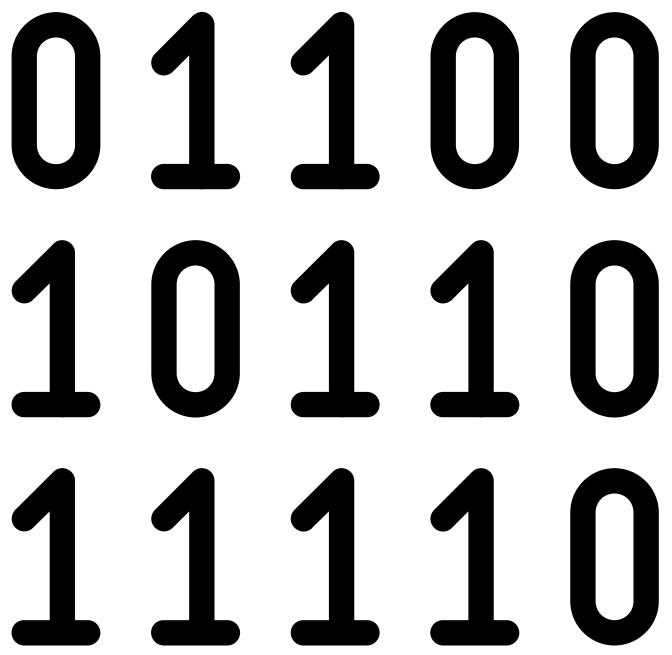
- Leverage graph models
- … and random walks
- Focus on structured botnets [10, 11, 12]
- Do not use open technologies



[7]

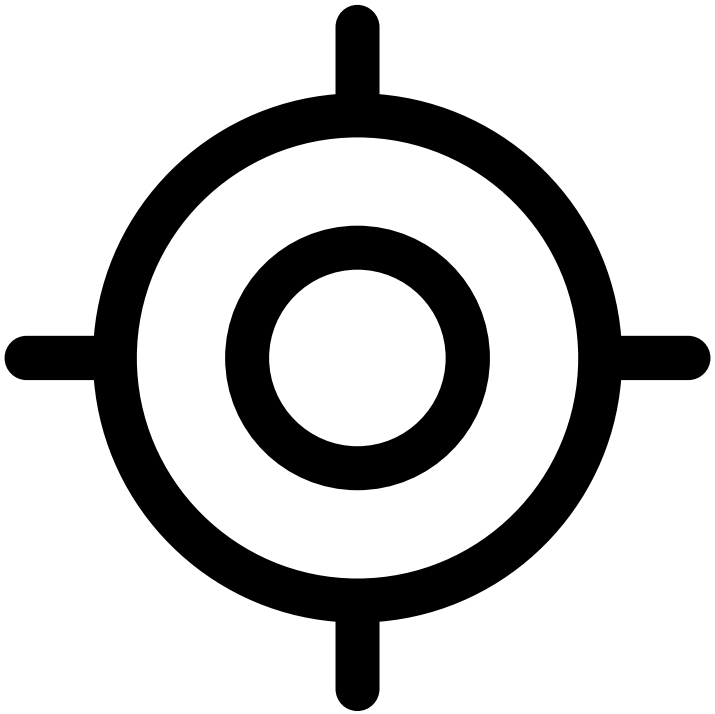
Existing Approaches

- Leverage graph models
- … and random walks
- Focus on structured botnets [10, 11, 12]
- Do not use open technologies
- Often assume complete knowledge on botnet communication



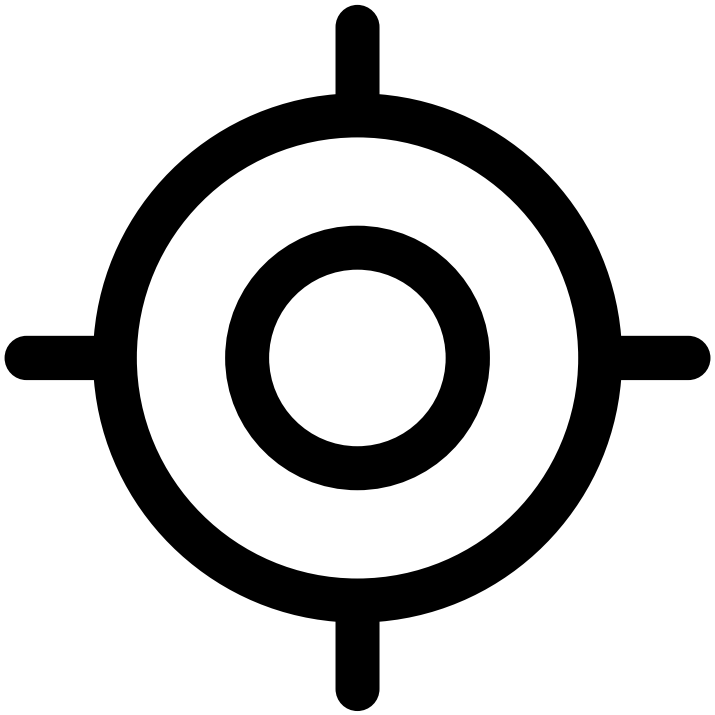
[7]

Our Approach



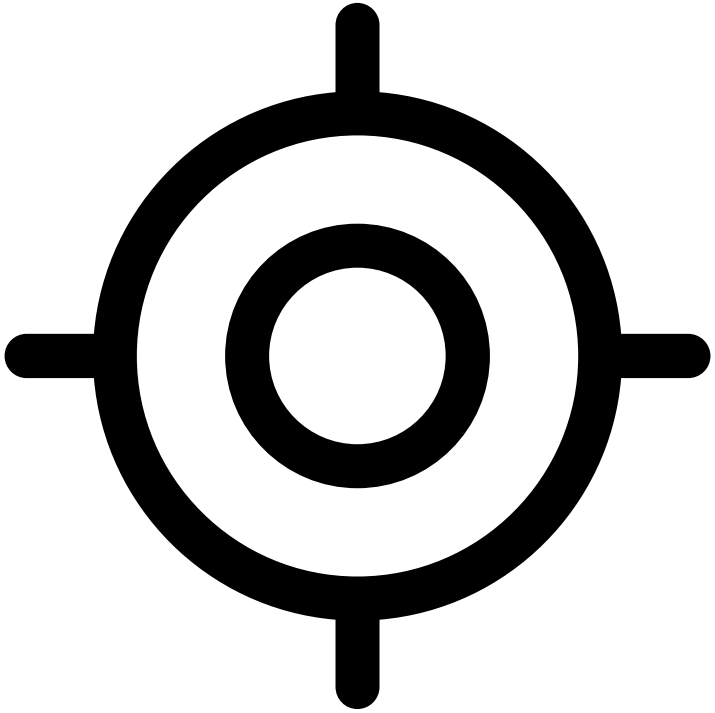
Our Approach

- Leverages random walks



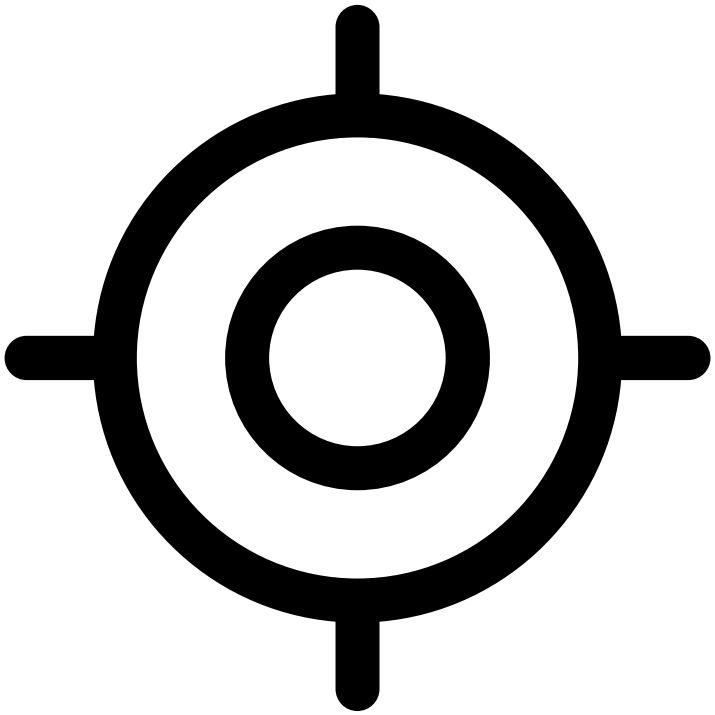
Our Approach

- Leverages random walks
- Uses open-source technologies



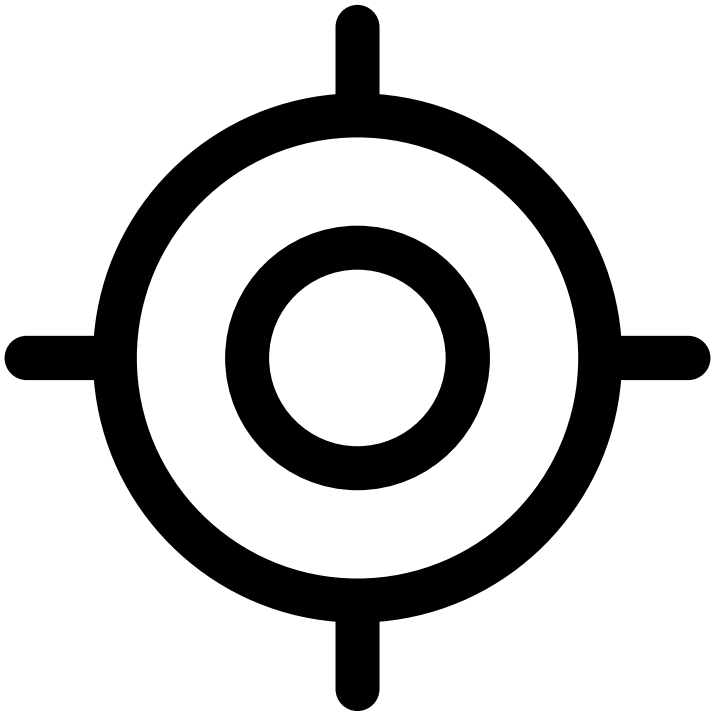
Our Approach

- Leverages random walks
- Uses open-source technologies
- Tested on unstructured botnets



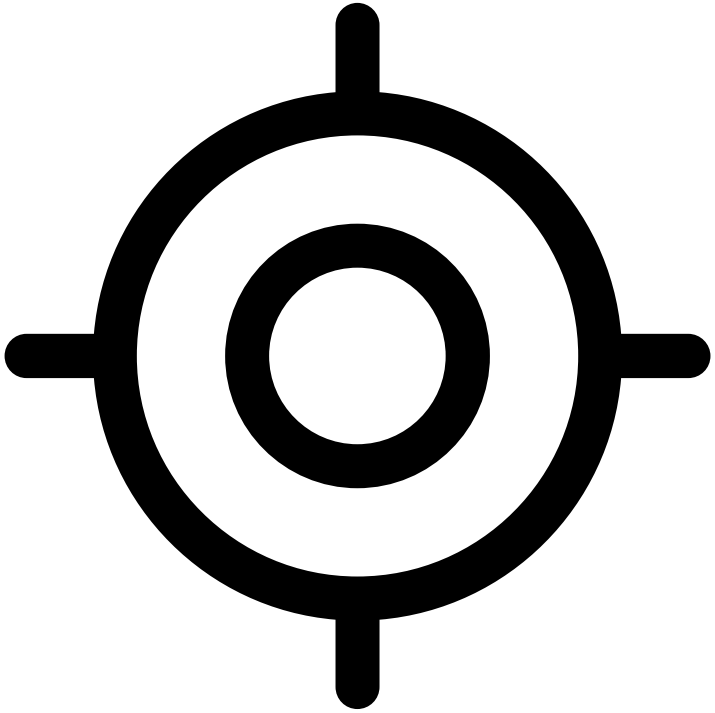
Our Approach

- Leverages random walks
- Uses open-source technologies
- Tested on unstructured botnets
- Precise when information is limited

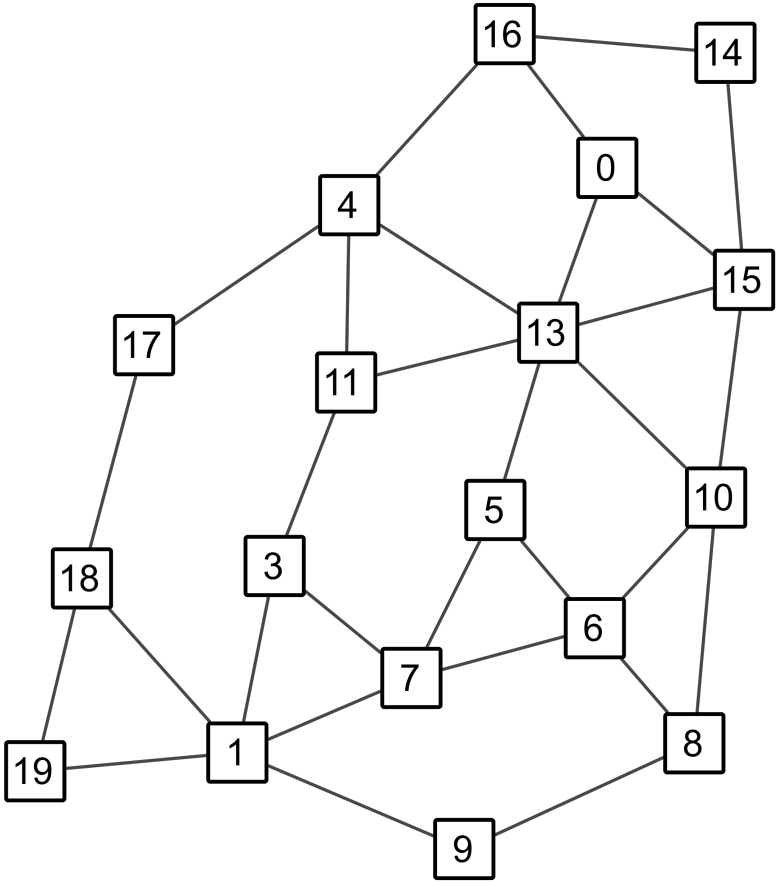


Our Approach

- Leverages random walks
- Uses open-source technologies
- Tested on unstructured botnets
- Precise when information is limited
- Can be combined with other approaches

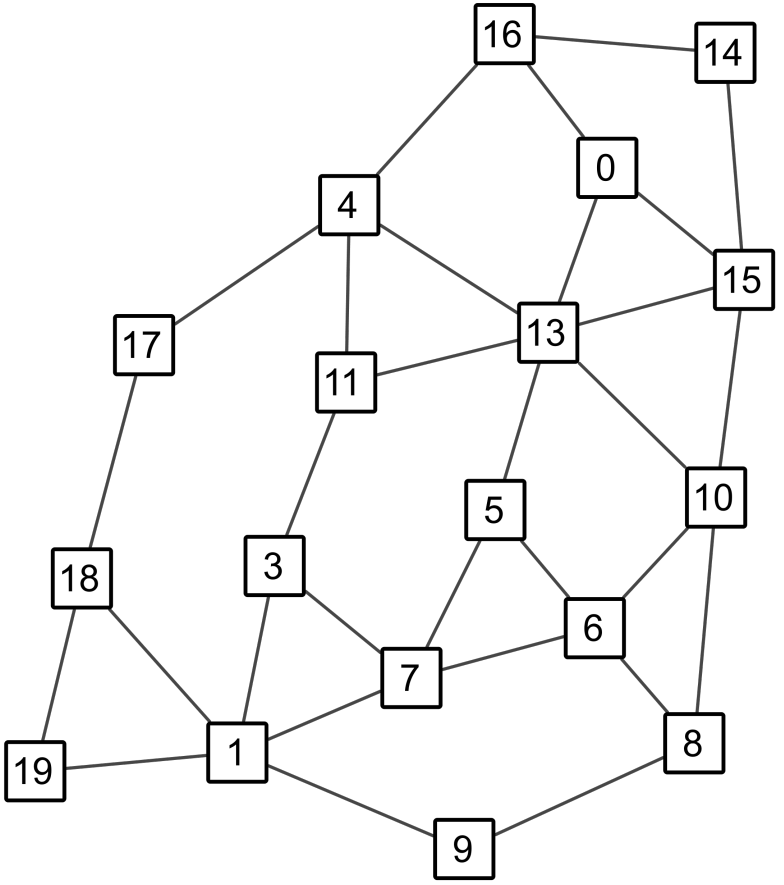


Communication Graph



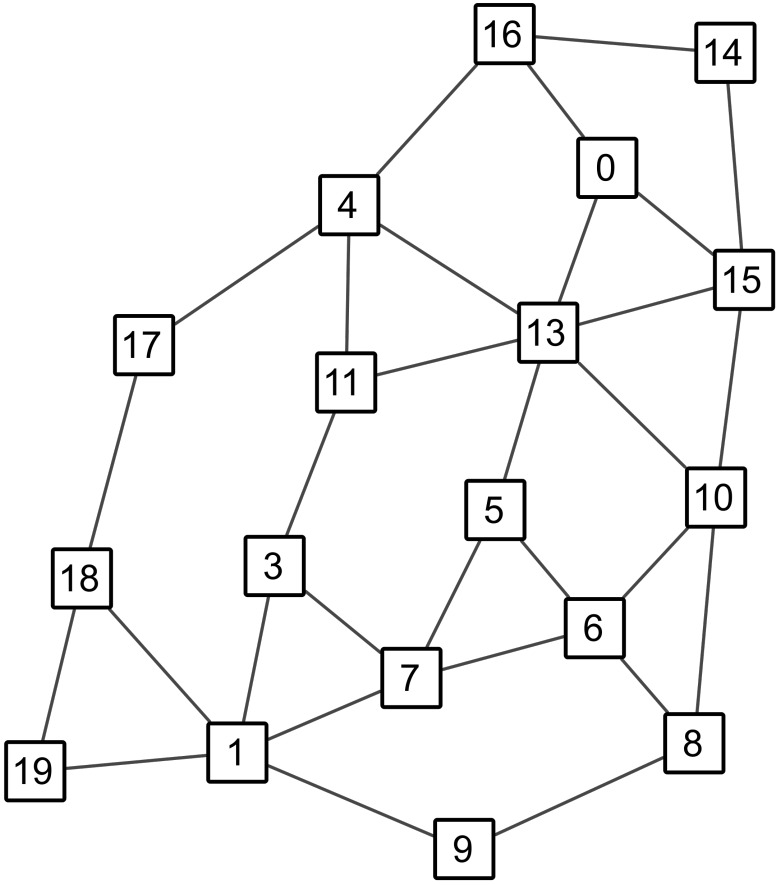
Communication Graph

- No payload data needed



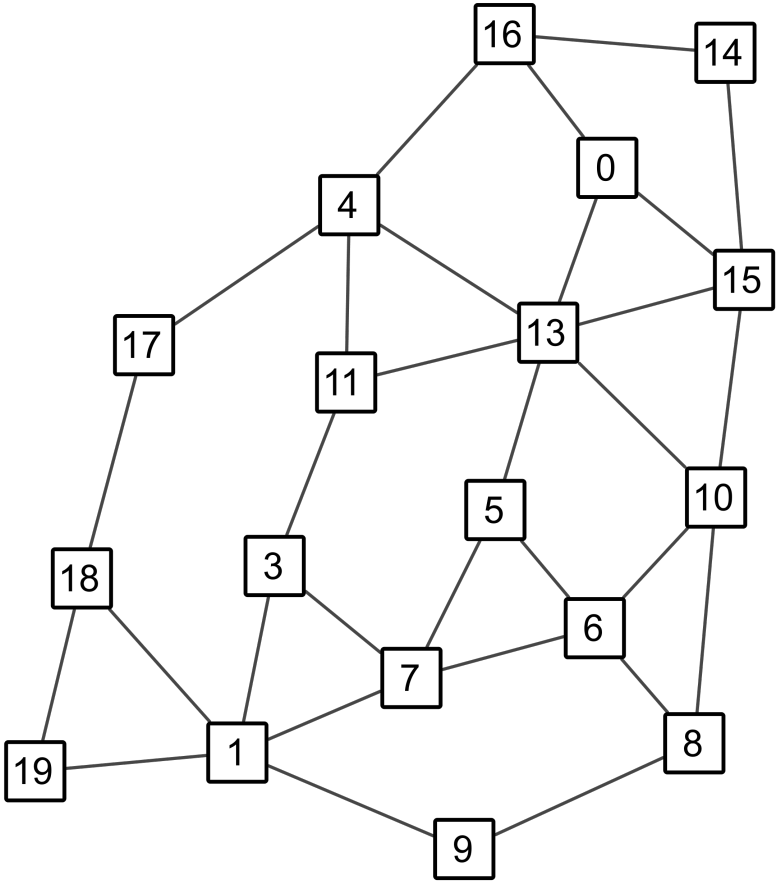
Communication Graph

- No payload data needed
- Network operator's view



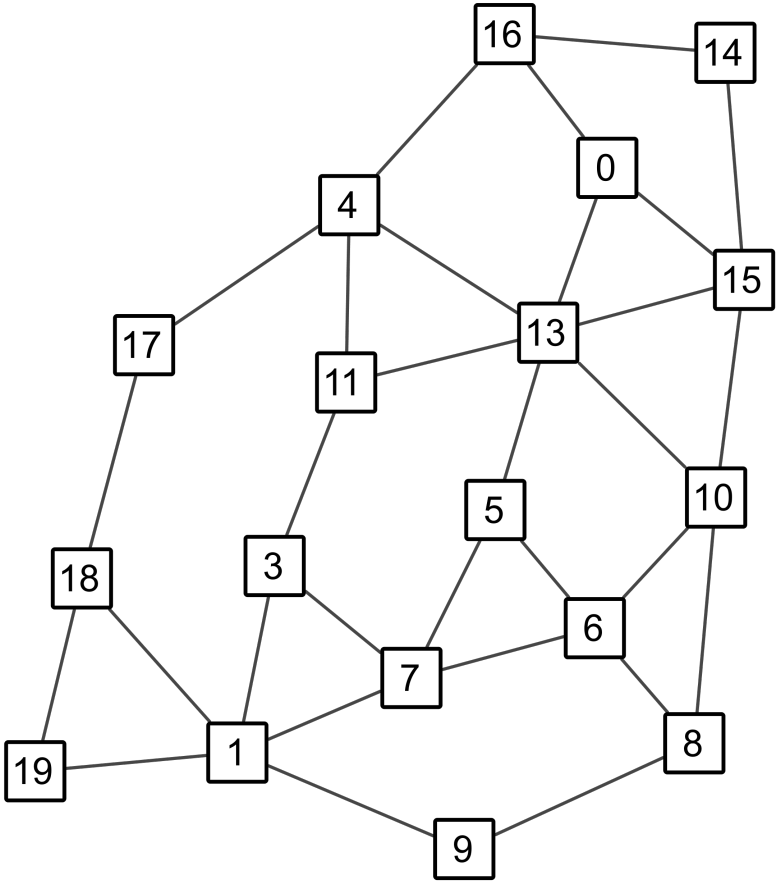
Communication Graph

- No payload data needed
- Network operator's view
- Aggregated NetFlow data



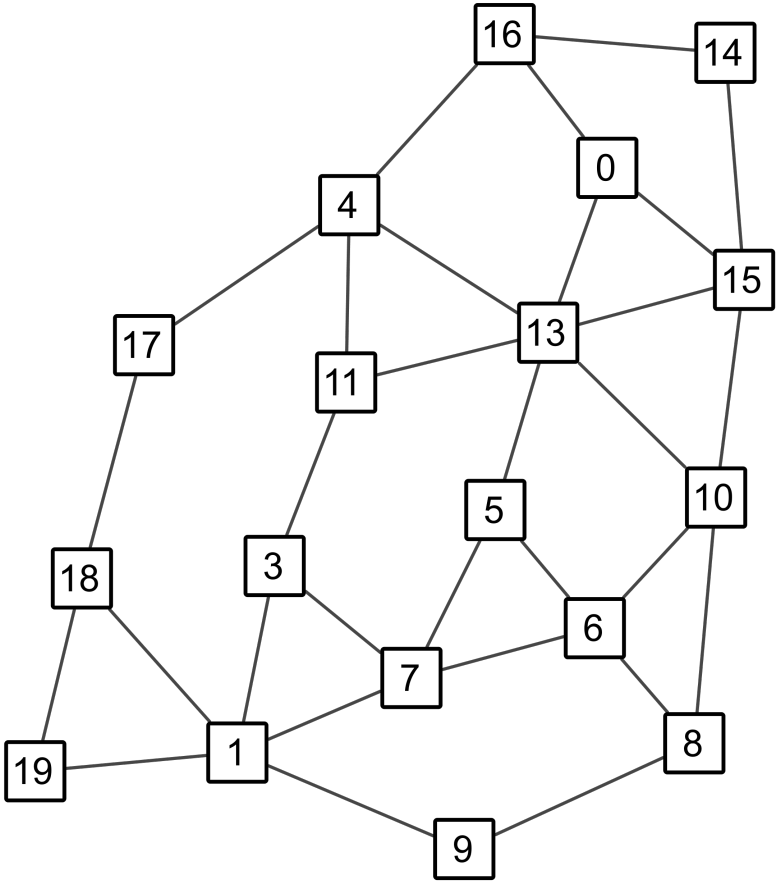
Communication Graph

- No payload data needed
- Network operator's view
- Aggregated NetFlow data
- Idea: extract well-connected subgraph

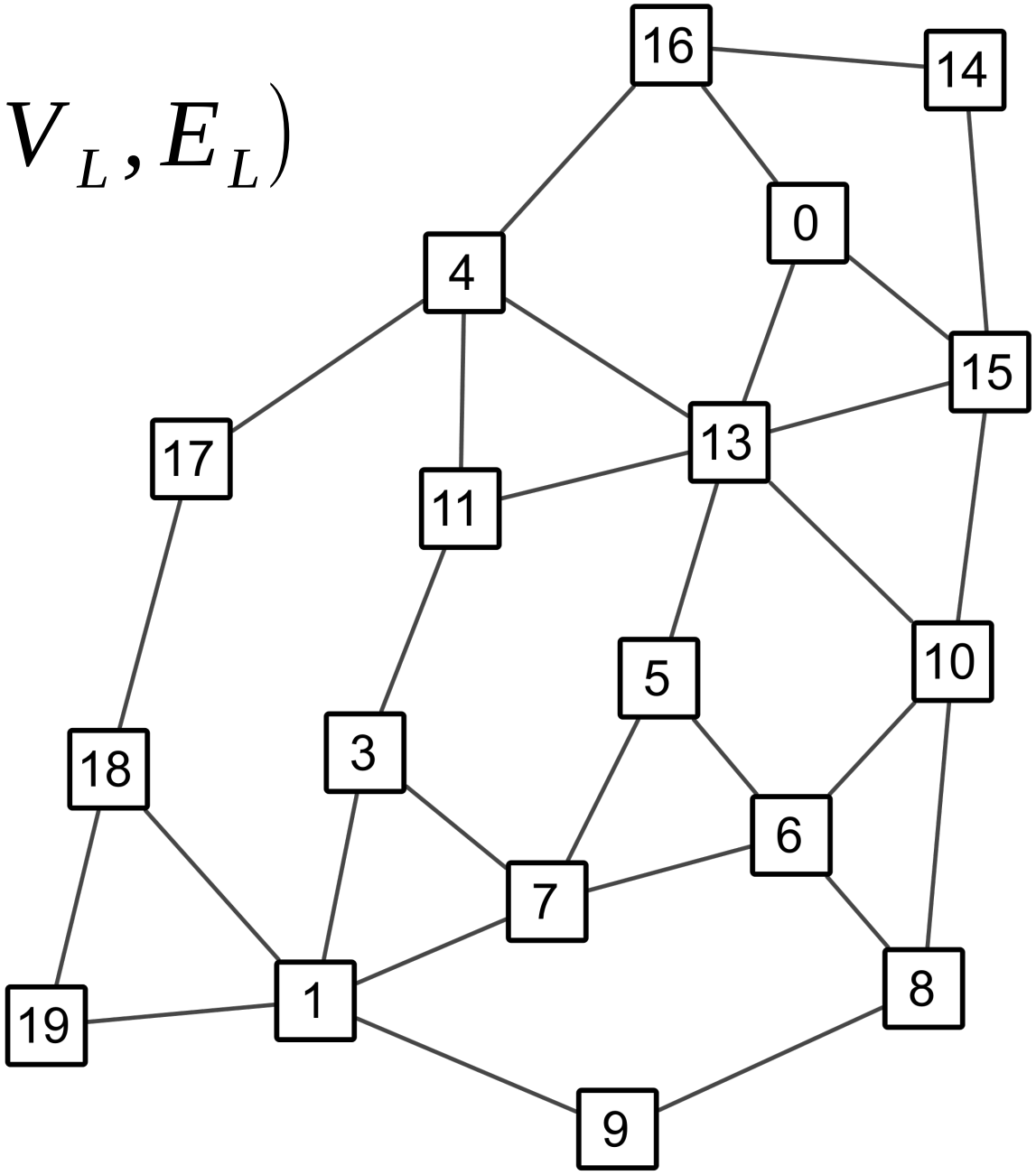


Communication Graph

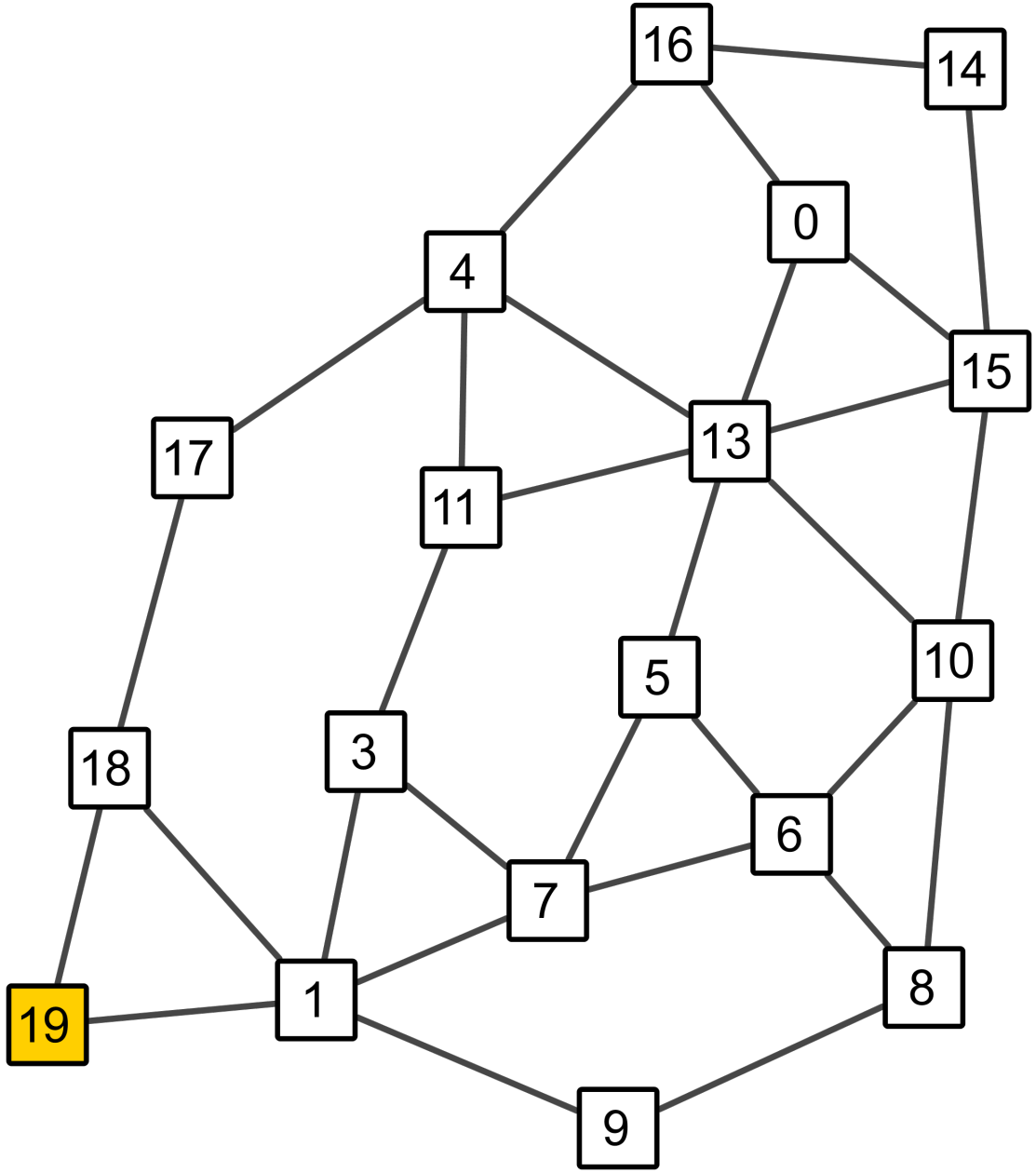
- No payload data needed
- Network operator's view
- Aggregated NetFlow data
- Idea: extract well-connected subgraph
- Approach: Random Walks



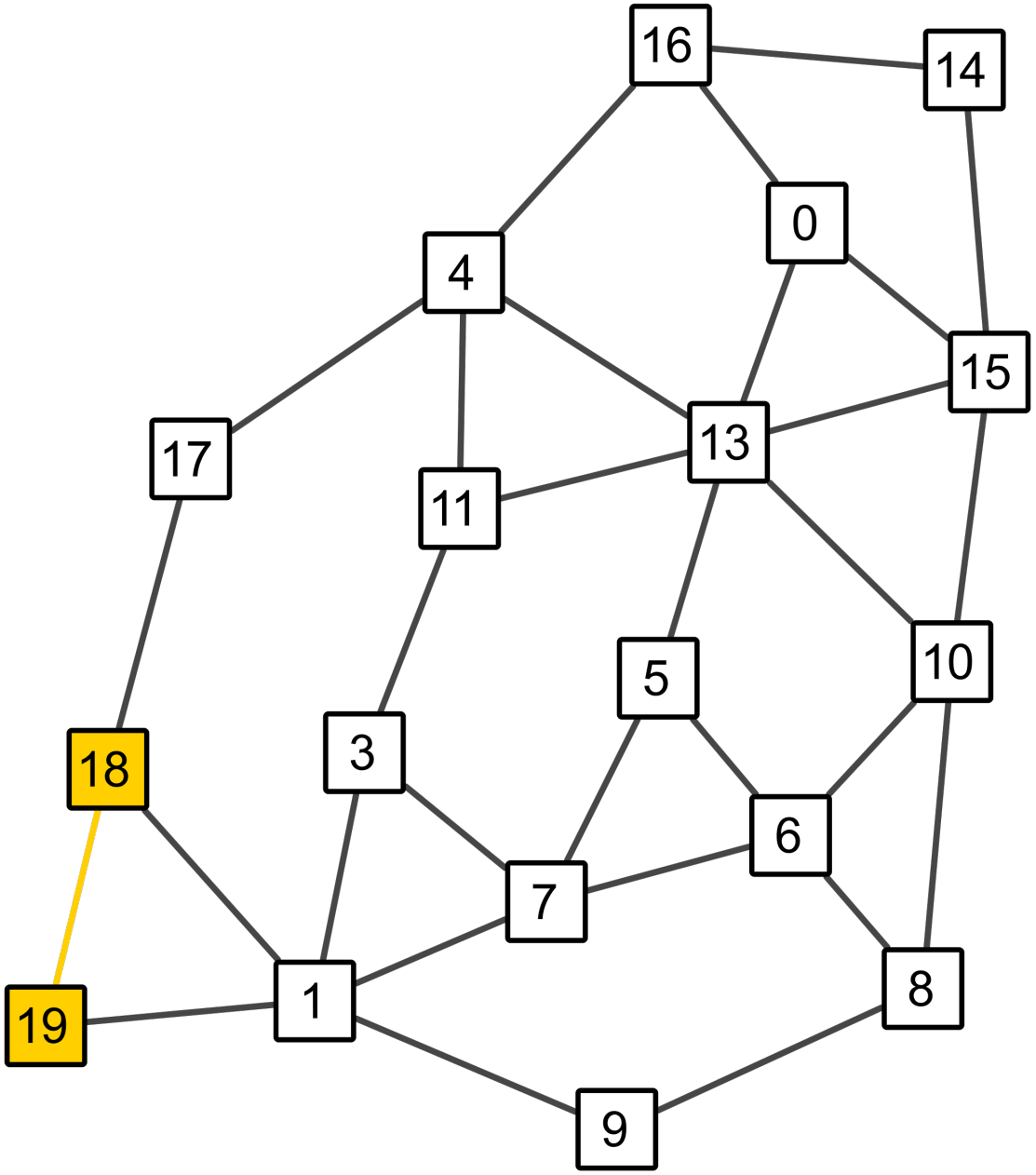
$$G_L = (V_L, E_L)$$



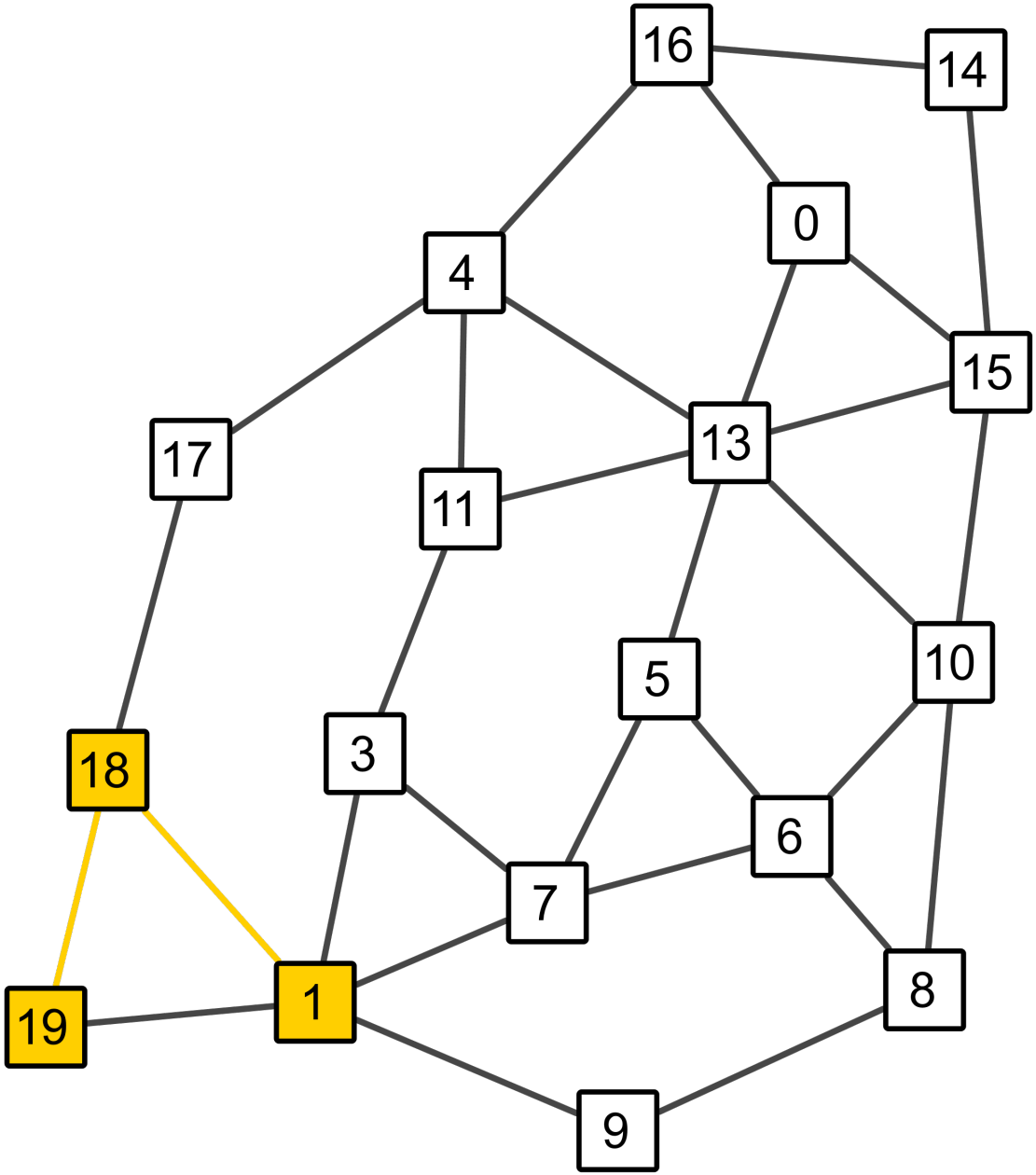
$k=0$



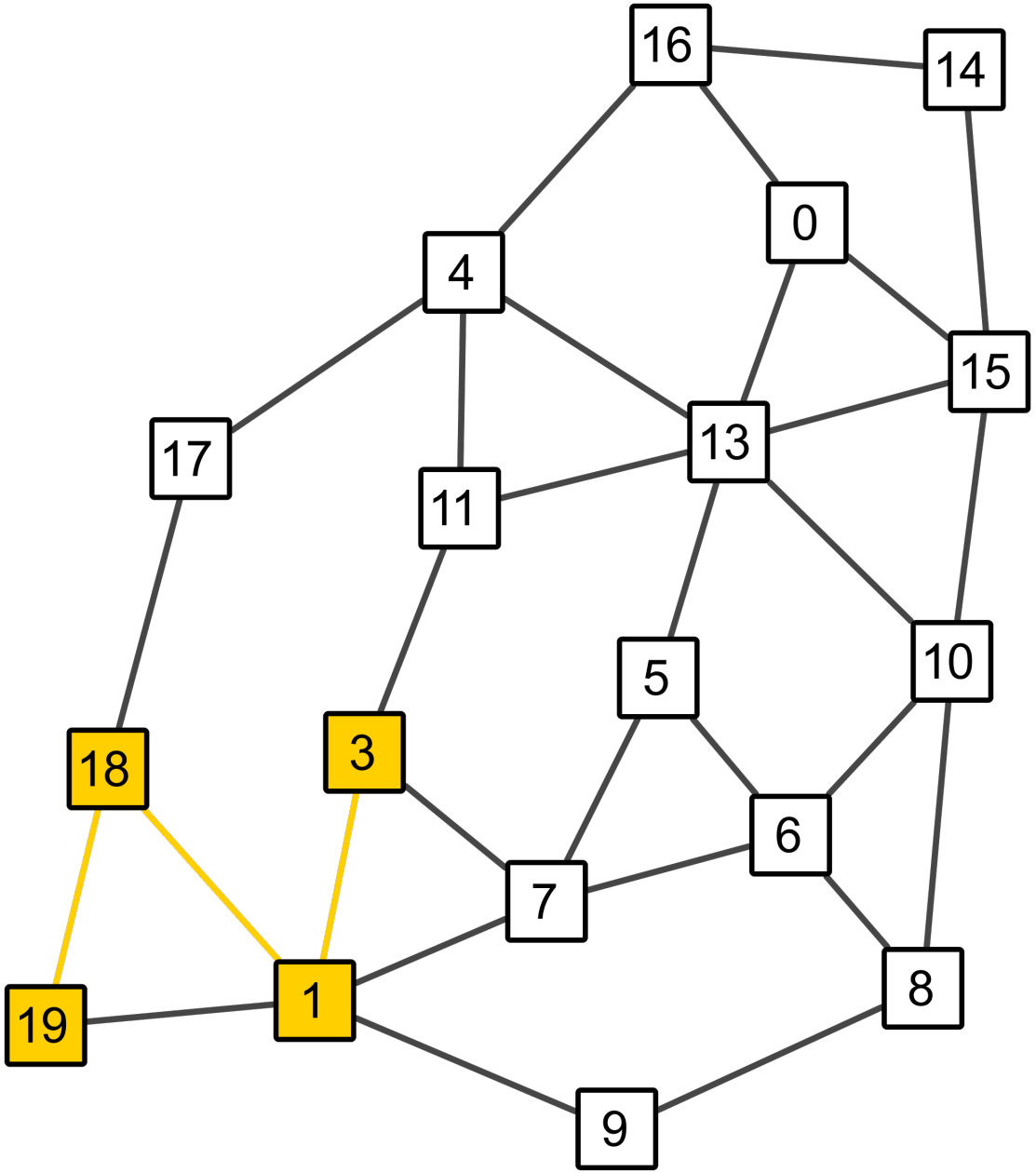
$k=1$



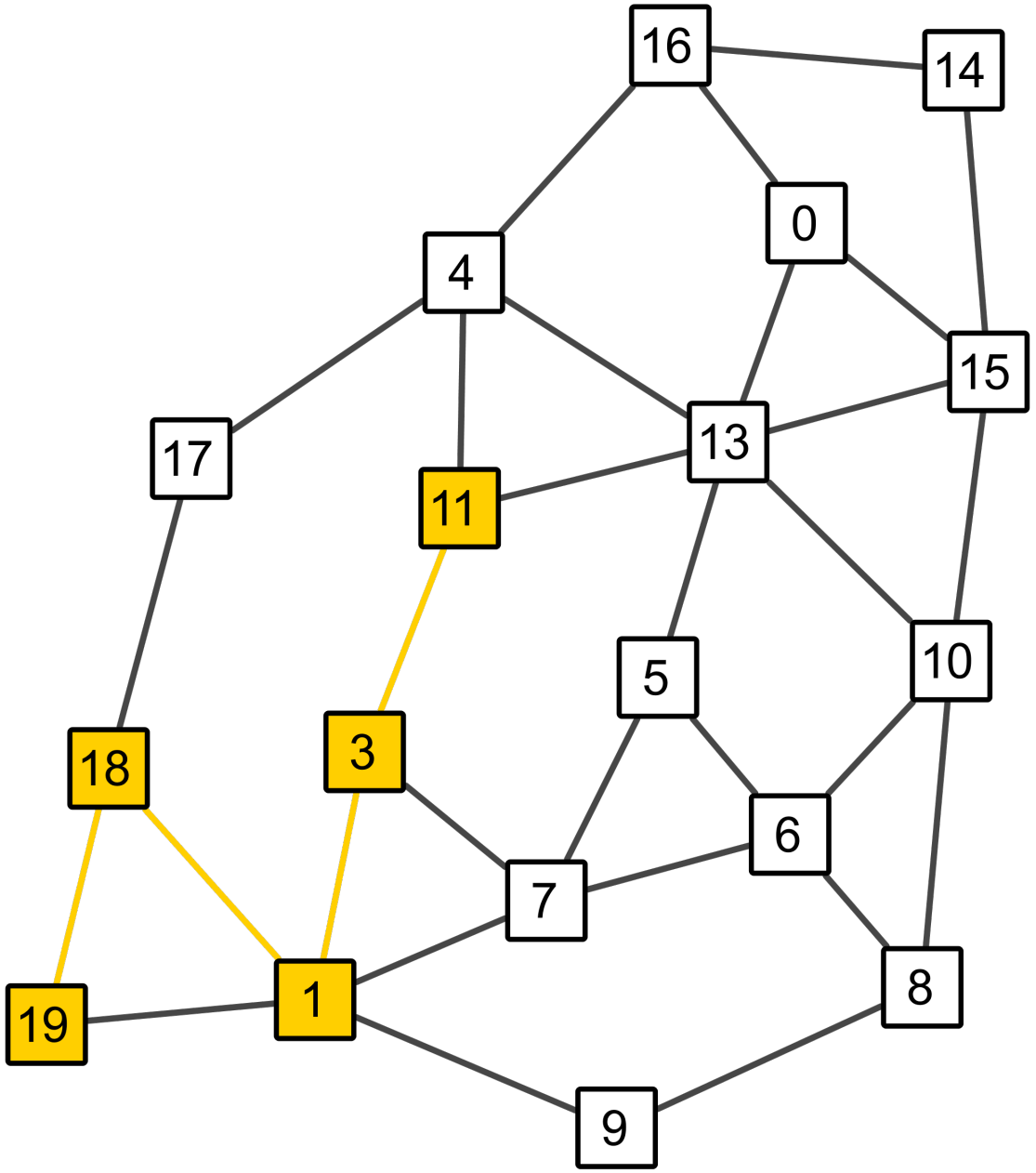
$k=2$



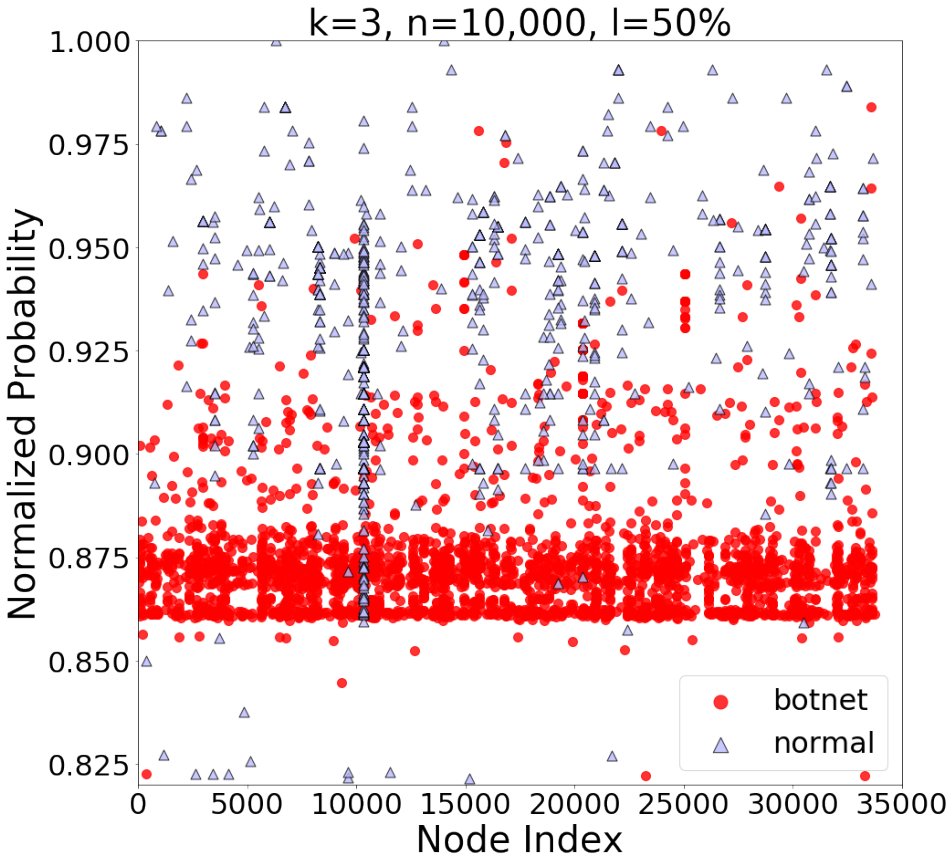
$k=3$



$k = 4$

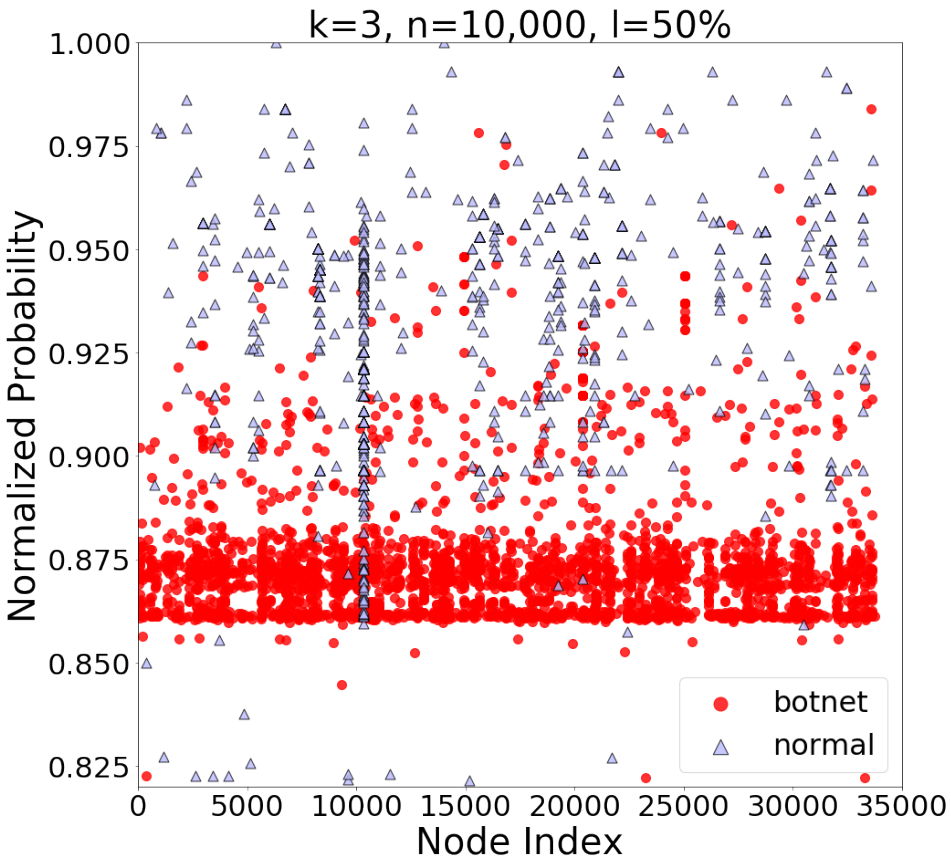


Probability Distribution



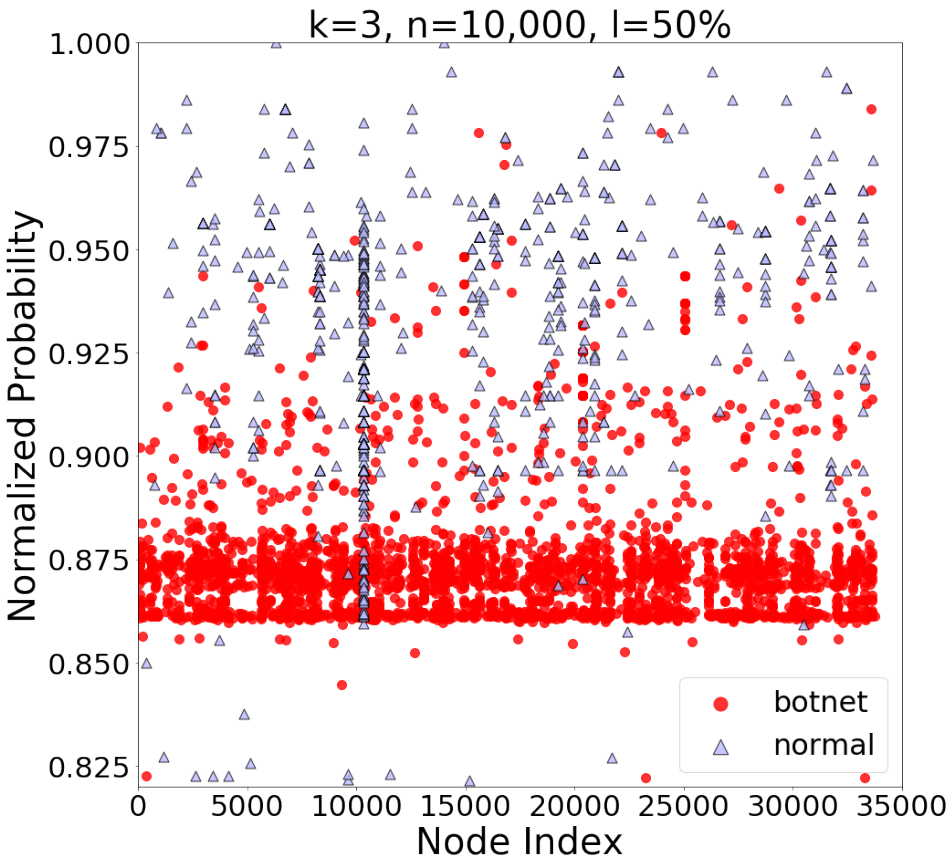
Probability Distribution

- n=10,000 walks



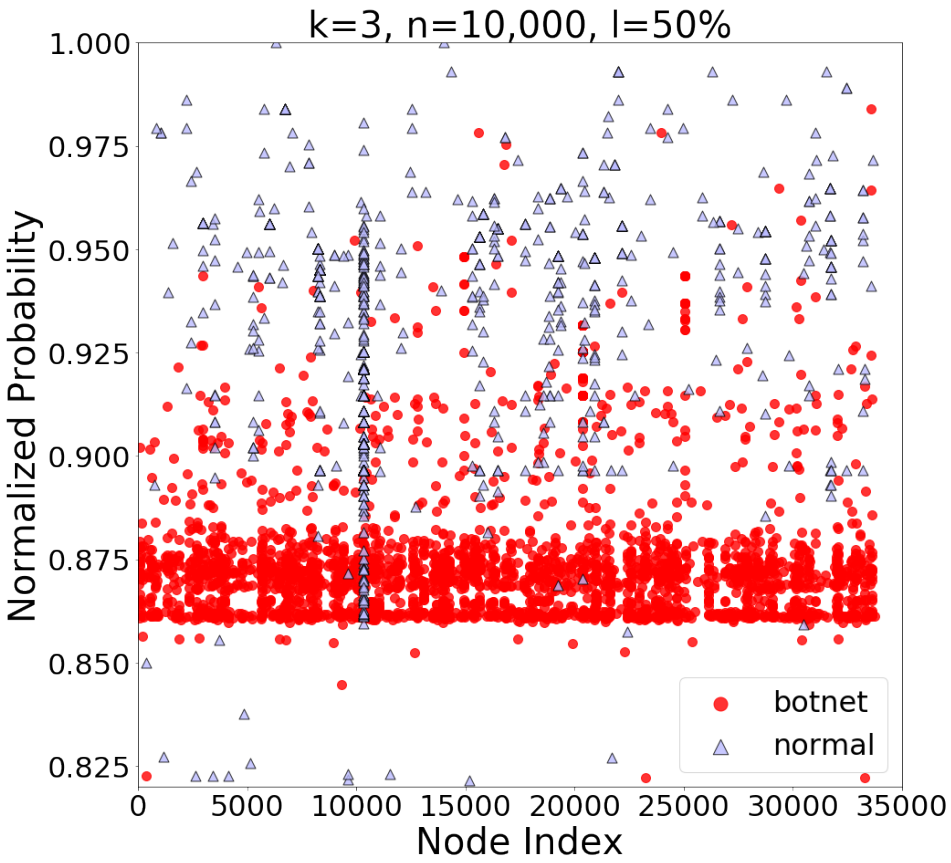
Probability Distribution

- $n=10{,}000$ walks
- Of length $k=3$



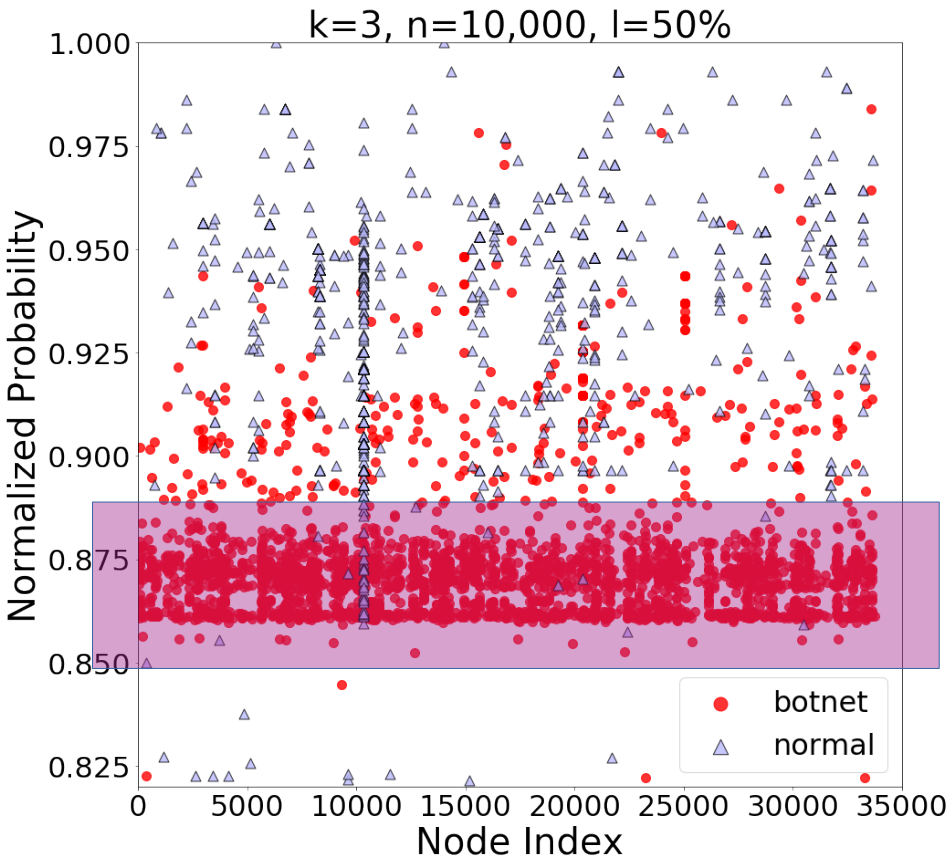
Probability Distribution

- $n=10{,}000$ walks
- Of length $k=3$
- With loss $l=0.5$



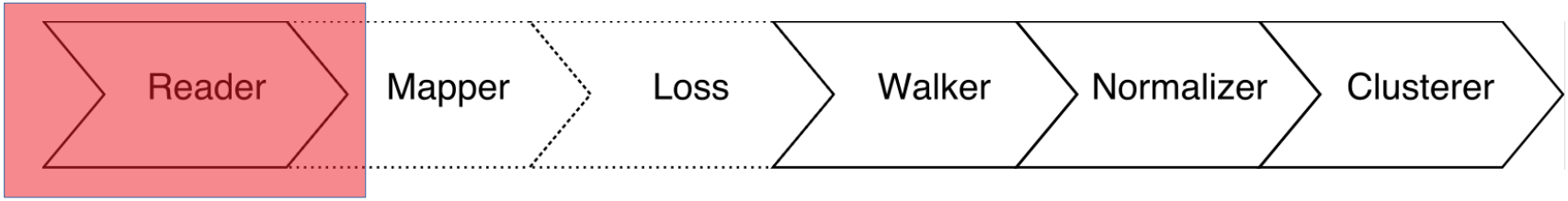
Probability Distribution

- $n=10{,}000$ walks
- Of length $k=3$
- With loss $l=0.5$
- Fast-mixing artifact



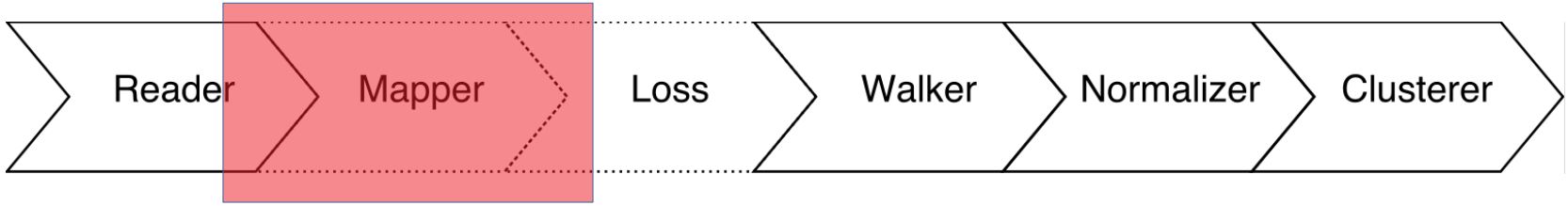
The Analysis Pipeline

The Analysis Pipeline



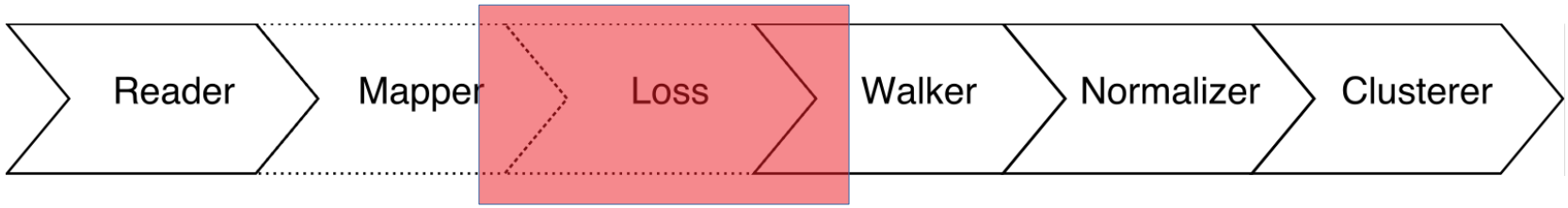
- Aggregate NetFlow data (Python 3.6, `networkx`)

The Analysis Pipeline



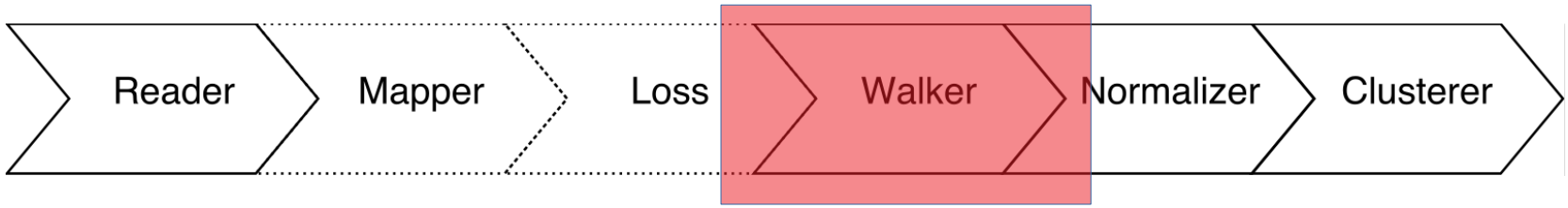
- Aggregate NetFlow data (Python 3.6, `networkx`)
- Evaluation steps:
  - Botnet node mapping

The Analysis Pipeline



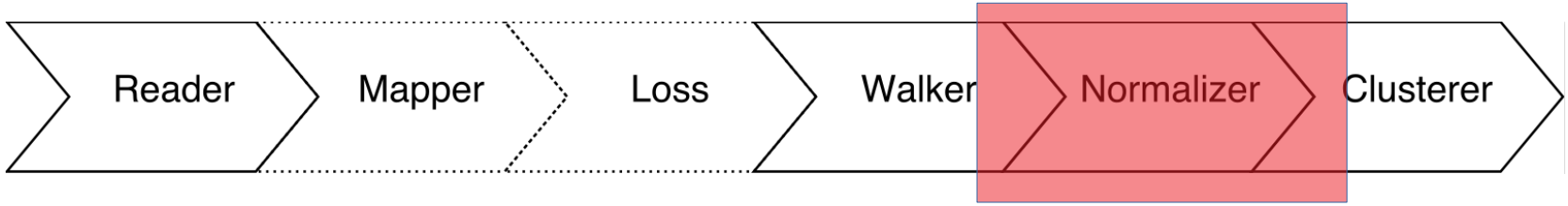
- Aggregate NetFlow data (Python 3.6, `networkx`)
- Evaluation steps:
  - Botnet node mapping
  - Apply loss functions

The Analysis Pipeline



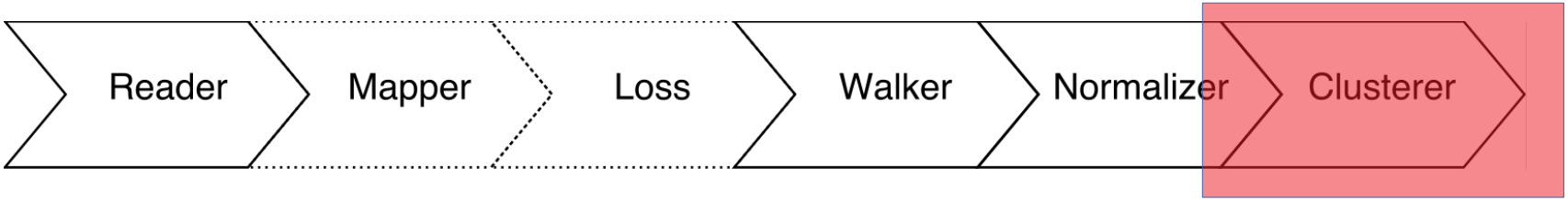
- Aggregate NetFlow data (Python 3.6, `networkx`)
- Evaluation steps:
  - Botnet node mapping
  - Apply loss functions
- Execute random walks (`numpy`)

The Analysis Pipeline



- Aggregate NetFlow data (Python 3.6, `networkx`)
- Evaluation steps:
  - Botnet node mapping
  - Apply loss functions
- Execute random walks (`numpy`)
- Normalize resulting probability distribution

The Analysis Pipeline



- Aggregate NetFlow data (Python 3.6, `networkx`)
- Evaluation steps:
  - Botnet node mapping
  - Apply loss functions
- Execute random walks (`numpy`)
- Normalize resulting probability distribution
- Cluster walk destinations (DBSCAN)

The Test Dataset

	Carrier Graphs		Botnet Graphs	
	TW07	CTU11	ZA24	SA25
Network Diameter	7	12	5	5
Number of Nodes	66 408	38 130	4805	1422
Average Node Degree	2.103	2.062	187.415	416.769
Number of Edges	139 628	78 626	734 010	592 646
Average Path Length	3.959	2.808	2.163	1.776
Average Clustering Coefficient	7×10^{-4}	3×10^{-3}	0.327	0.605

The Test Dataset

	Carrier Graphs		Botnet Graphs	
	TW07	CTU11	ZA24	SA25
Network Diameter	7	12	5	5
Number of Nodes	66 408	38 130	4805	1422
Average Node Degree	2.103	2.062	187.415	416.769
Number of Edges	139 628	78 626	734 010	592 646
Average Path Length	3.959	2.808	2.163	1.776
Average Clustering Coefficient	7×10^{-4}	3×10^{-3}	0.327	0.605

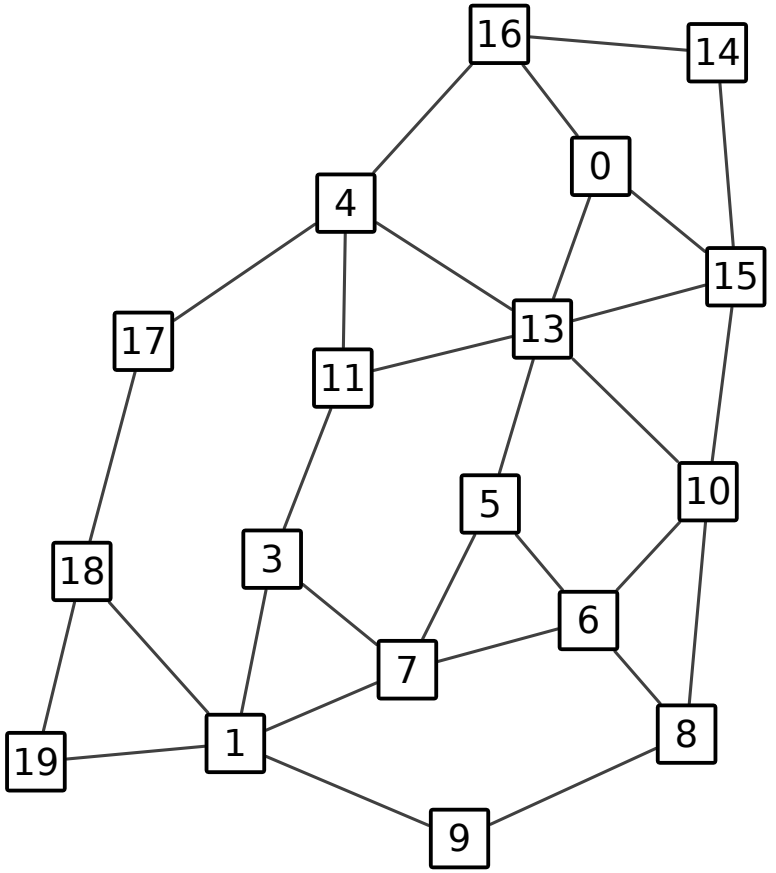
- CTU11 from Czech Technical University

The Test Dataset

	Carrier Graphs		Botnet Graphs	
	TW07	CTU11	ZA24	SA25
Network Diameter	7	12	5	5
Number of Nodes	66 408	38 130	4805	1422
Average Node Degree	2.103	2.062	187.415	416.769
Number of Edges	139 628	78 626	734 010	592 646
Average Path Length	3.959	2.808	2.163	1.776
Average Clustering Coefficient	7×10^{-4}	3×10^{-3}	0.327	0.605

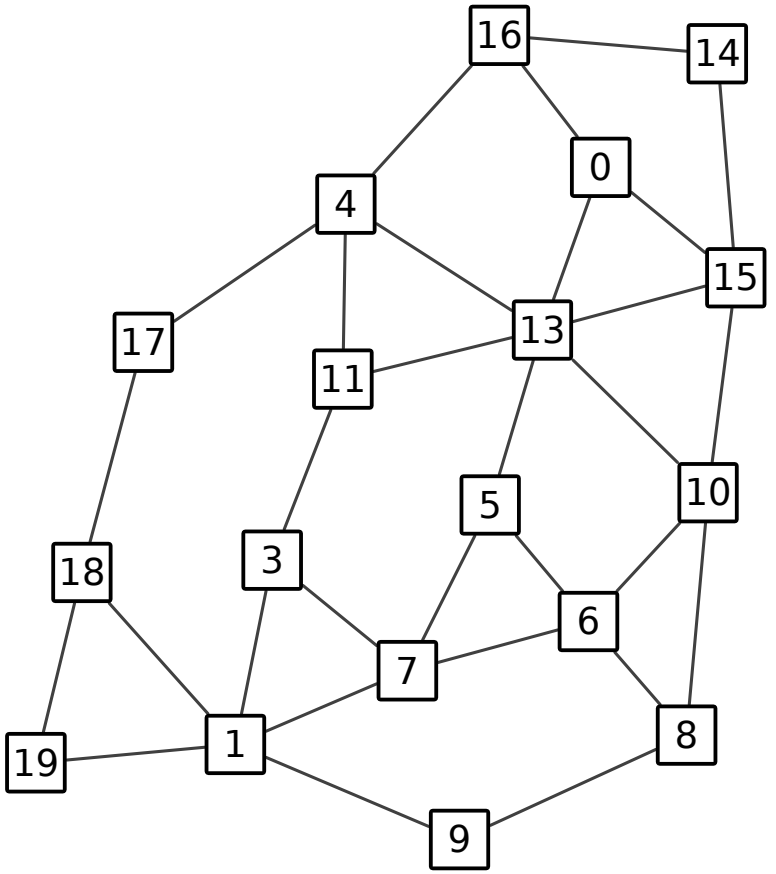
- CTU11 from Czech Technical University
- ZA24 ZeroAccess communication graph

Loss Strategies



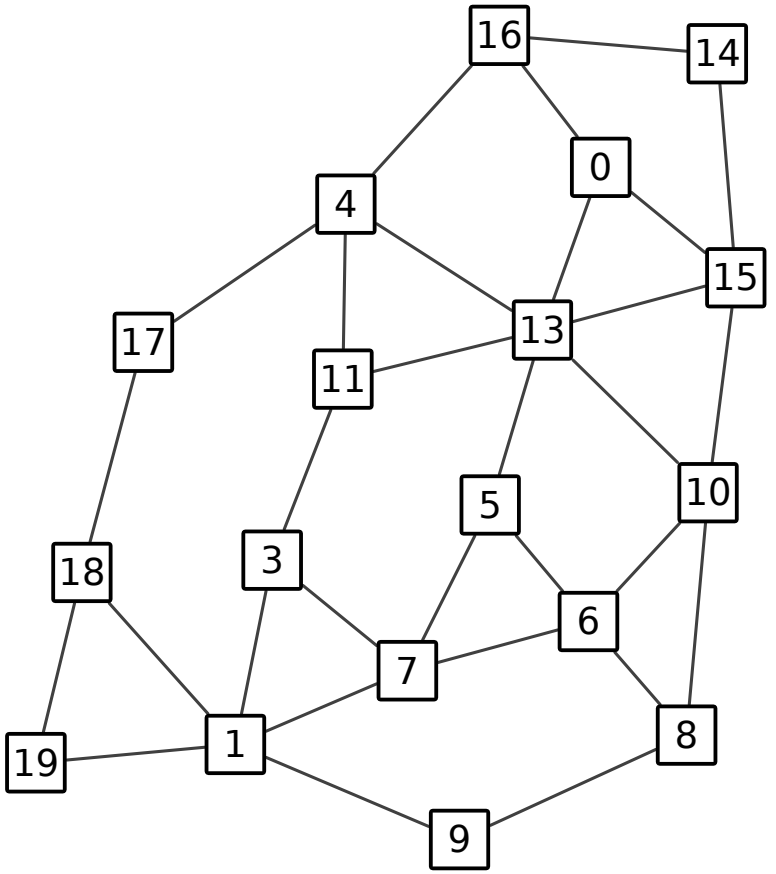
Loss Strategies

- Other approaches do not evaluate limited network view



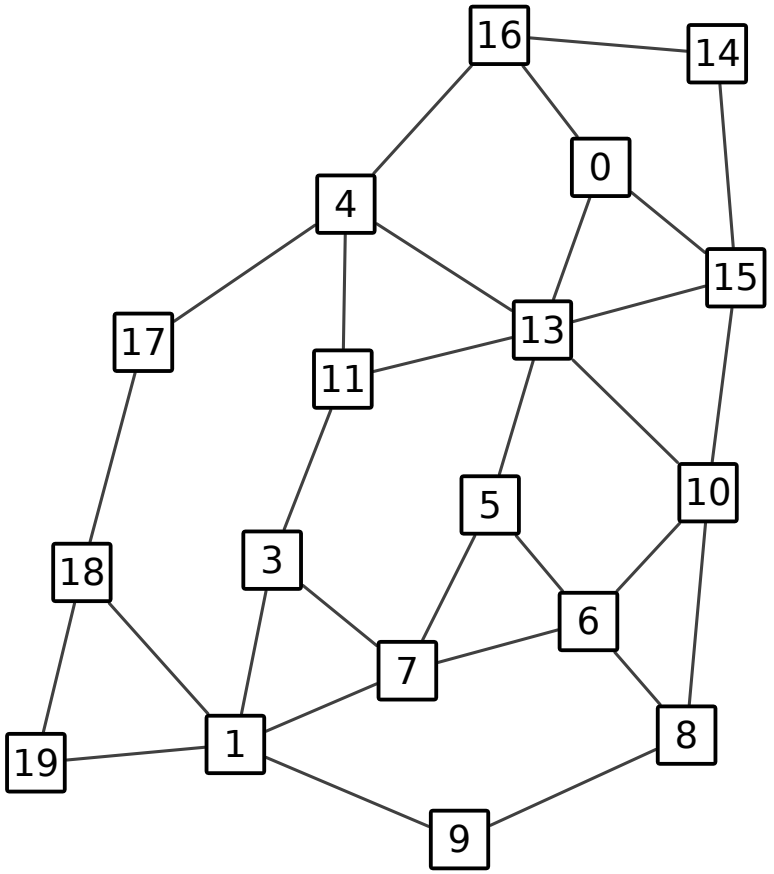
Loss Strategies

- Other approaches do not evaluate limited network view
- Unrealistic assumptions:
  - All communication relationships captured



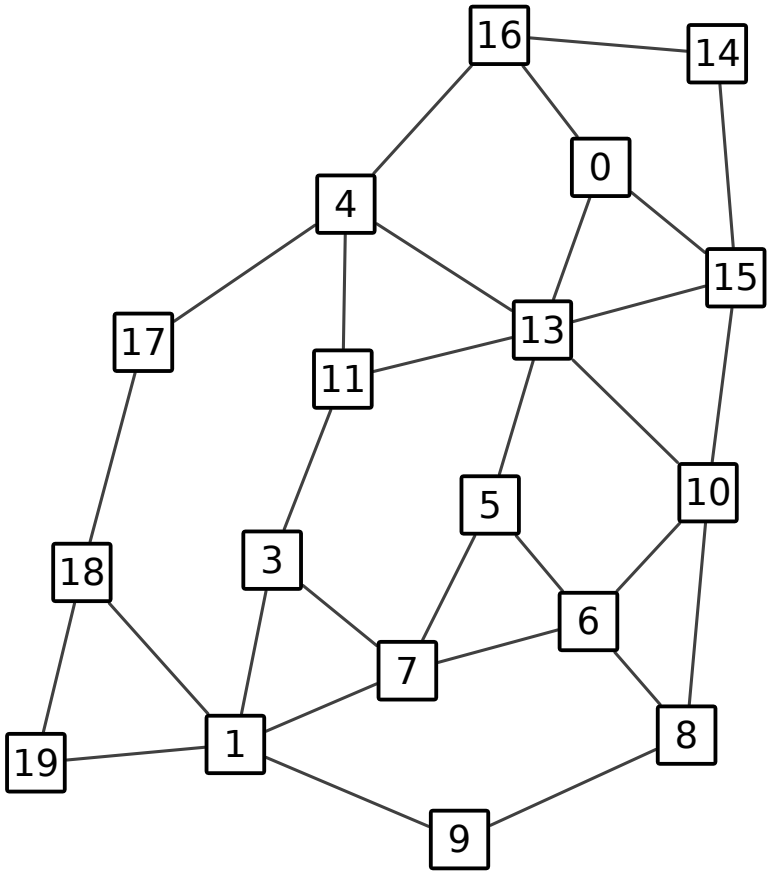
Loss Strategies

- Other approaches do not evaluate limited network view
- Unrealistic assumptions:
  - All communication relationships captured
  - Complete botnet in known network

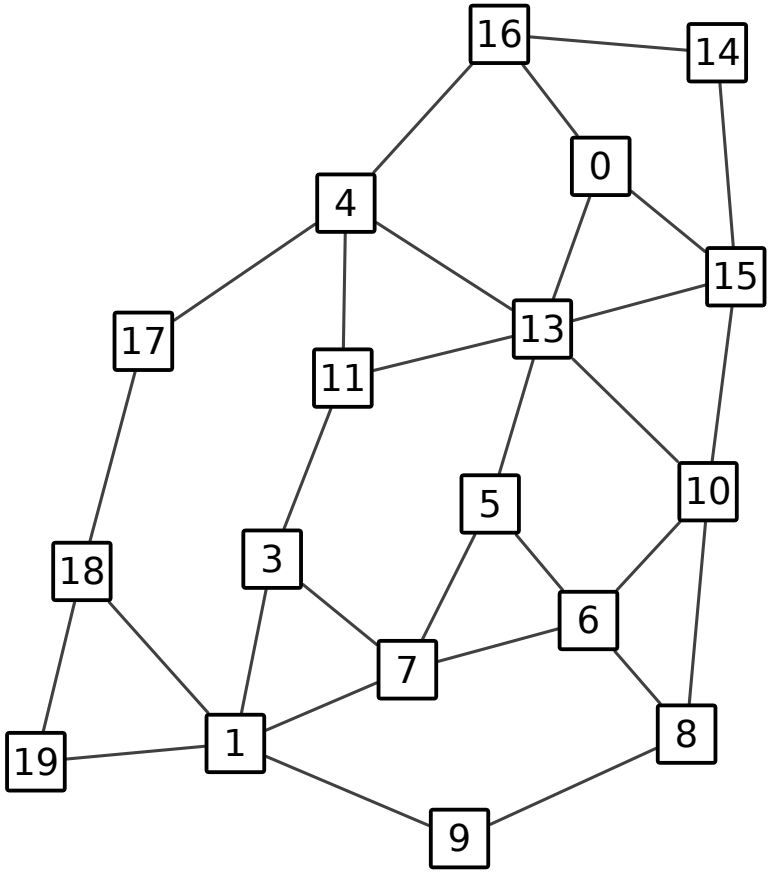


Loss Strategies

- Other approaches do not evaluate limited network view
- Unrealistic assumptions:
  - All communication relationships captured
  - Complete botnet in known network
- Solution: Simulate loss on communication graph

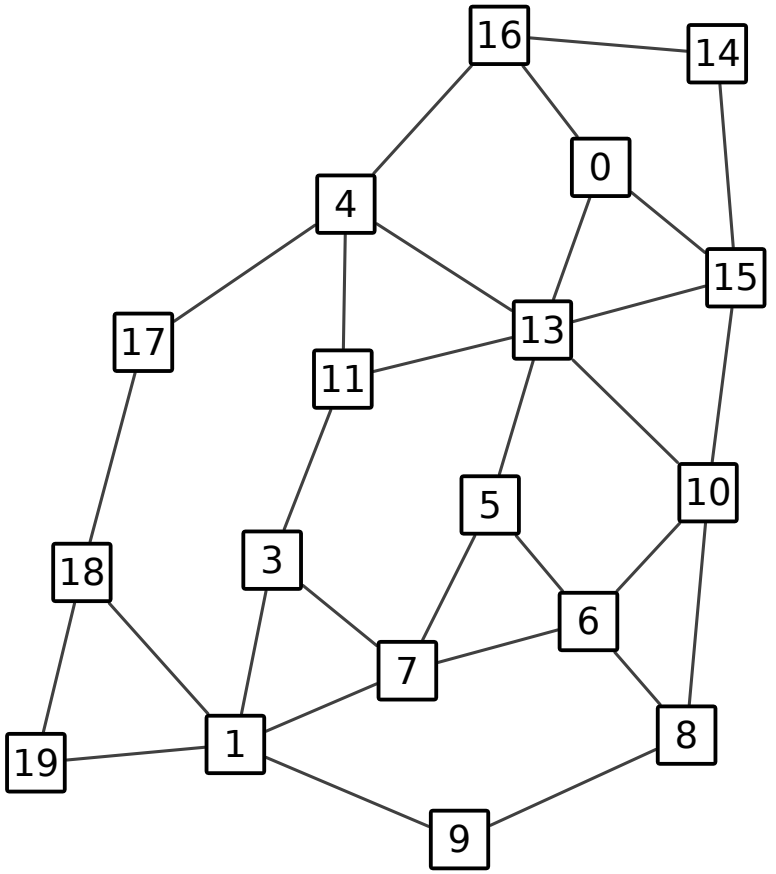


Random Botnet Edge Deletion



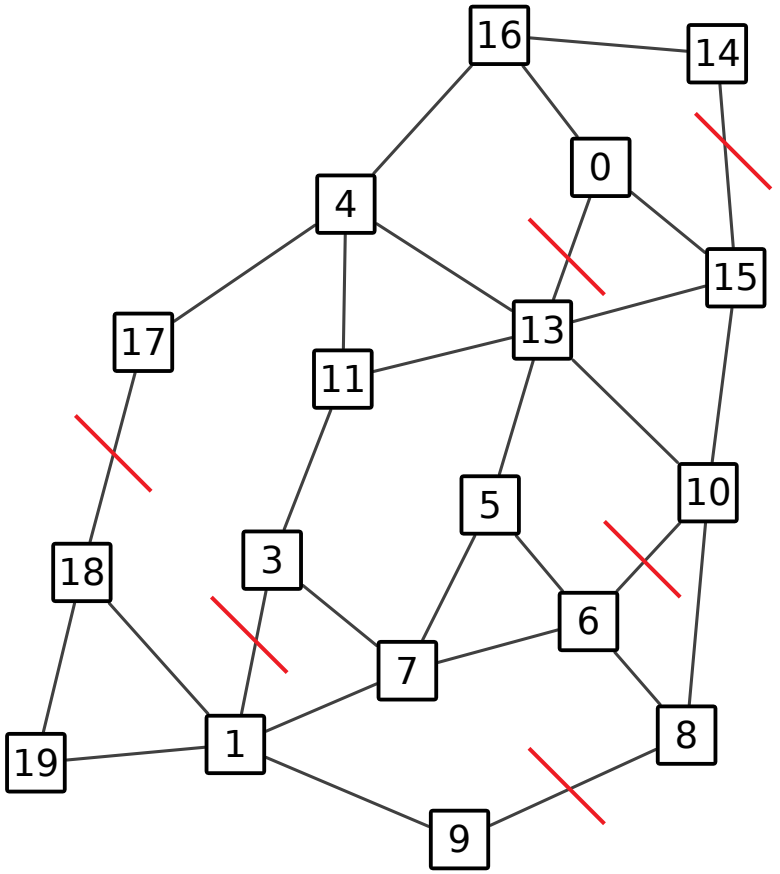
Random Botnet Edge Deletion

- Random subset of botnet edges



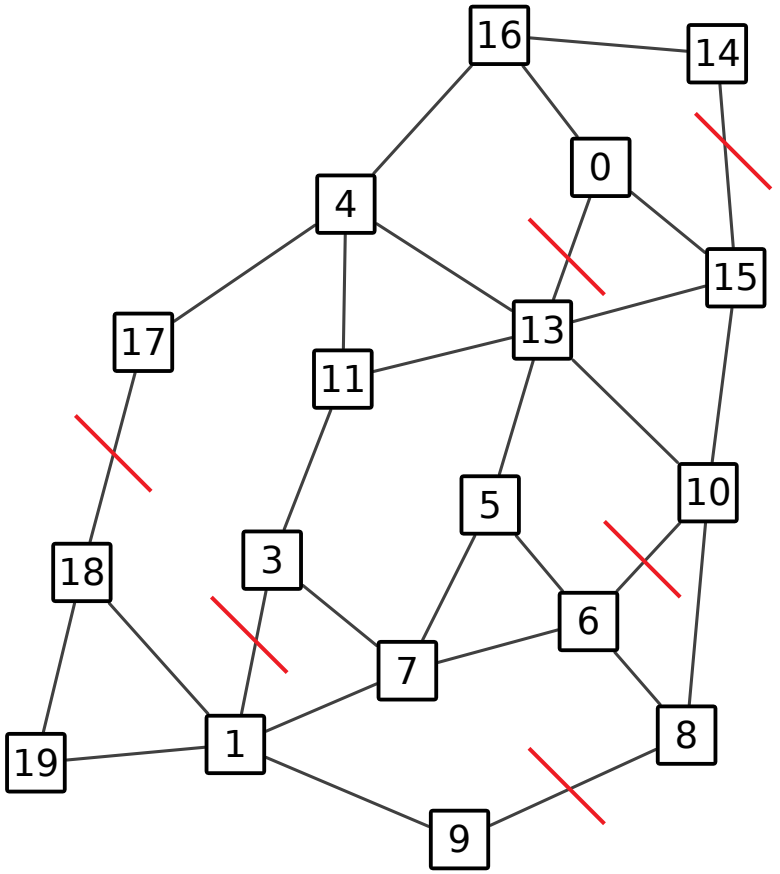
Random Botnet Edge Deletion

- Random subset of botnet edges



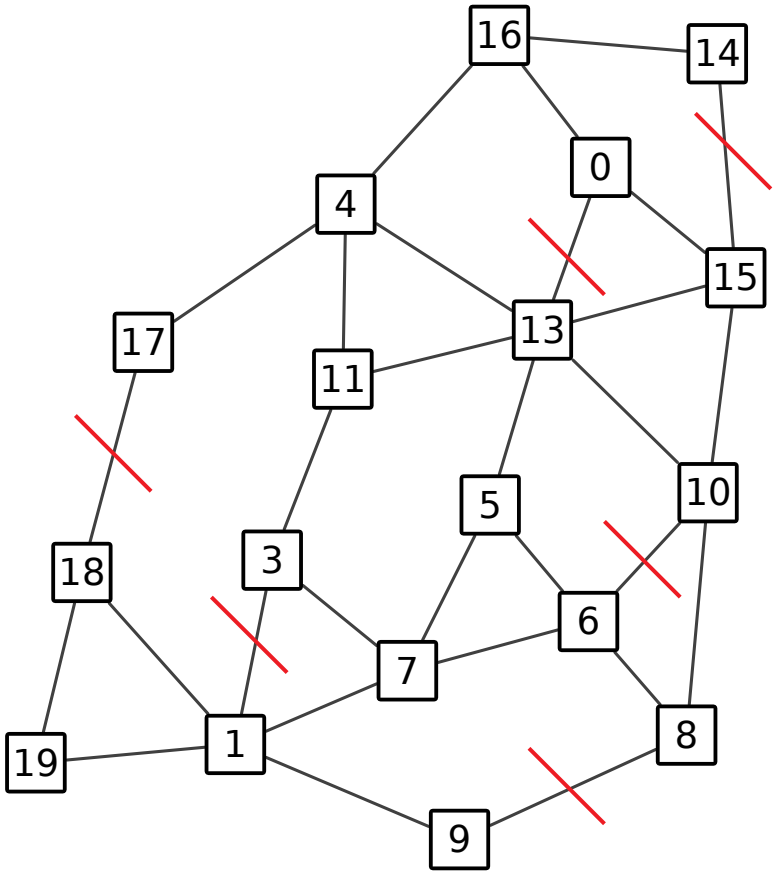
Random Botnet Edge Deletion

- Random subset of botnet edges
- Out-of-view connections



Random Botnet Edge Deletion

- Random subset of botnet edges
- Out-of-view connections
- ISP-related loss (e.g. 1:256 sampling)



RBED Robustness

RBED Robustness

- Random Botnet Edge Deletion

RBED Robustness

- Random Botnet Edge Deletion

$$\text{Precision} = \frac{tp}{tp + fp}$$

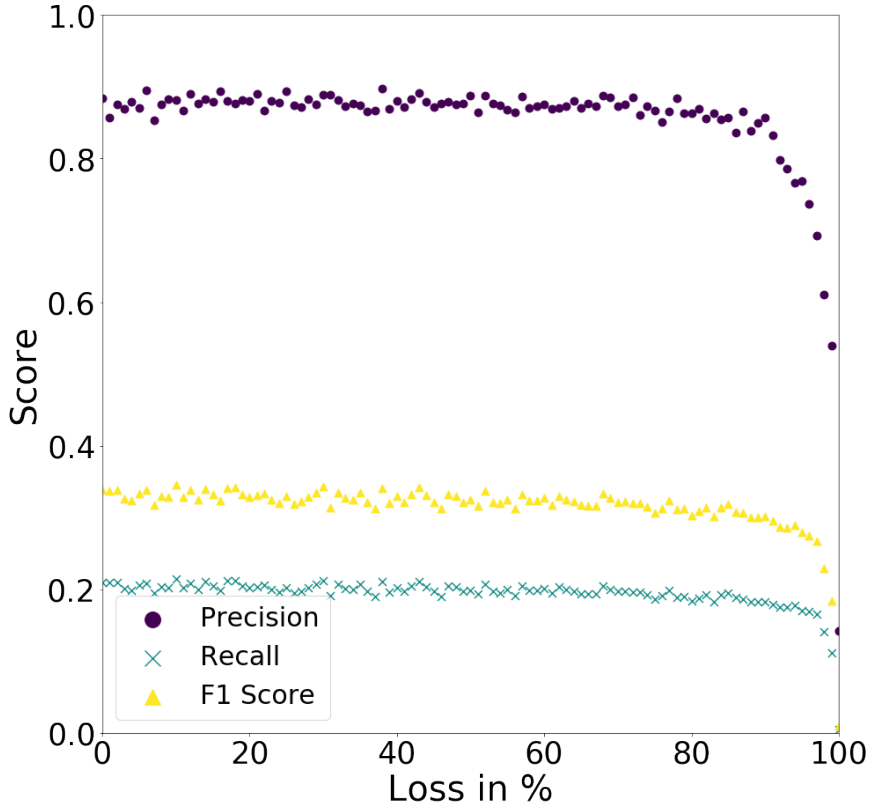
$$\text{Recall} = \frac{tp}{tp + fn}$$

RBED Robustness

- Random Botnet Edge Deletion

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

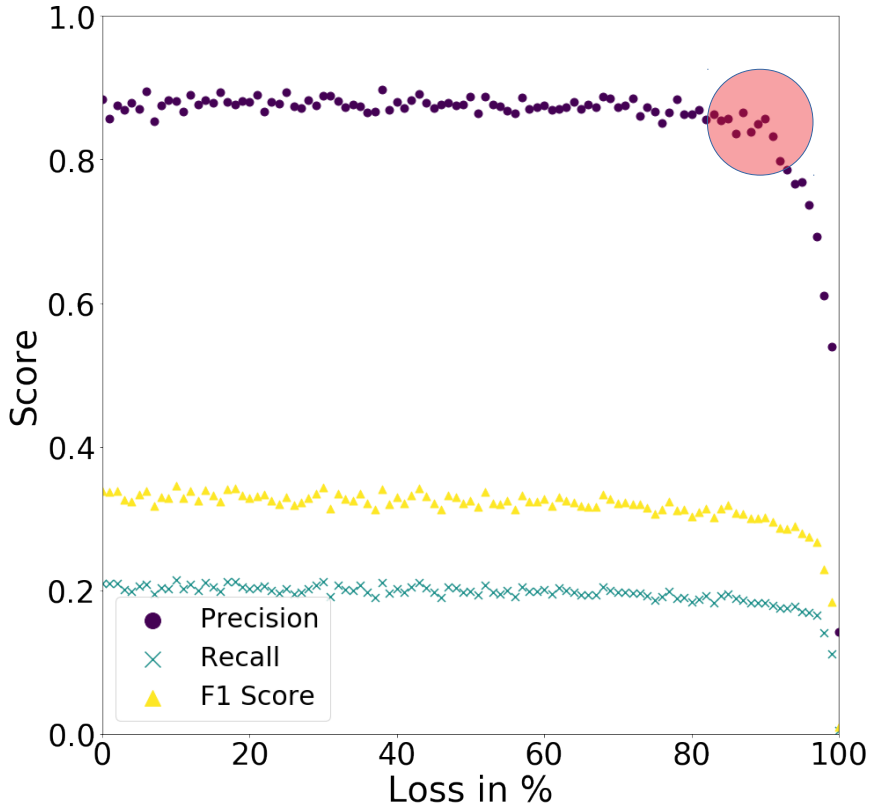


RBED Robustness

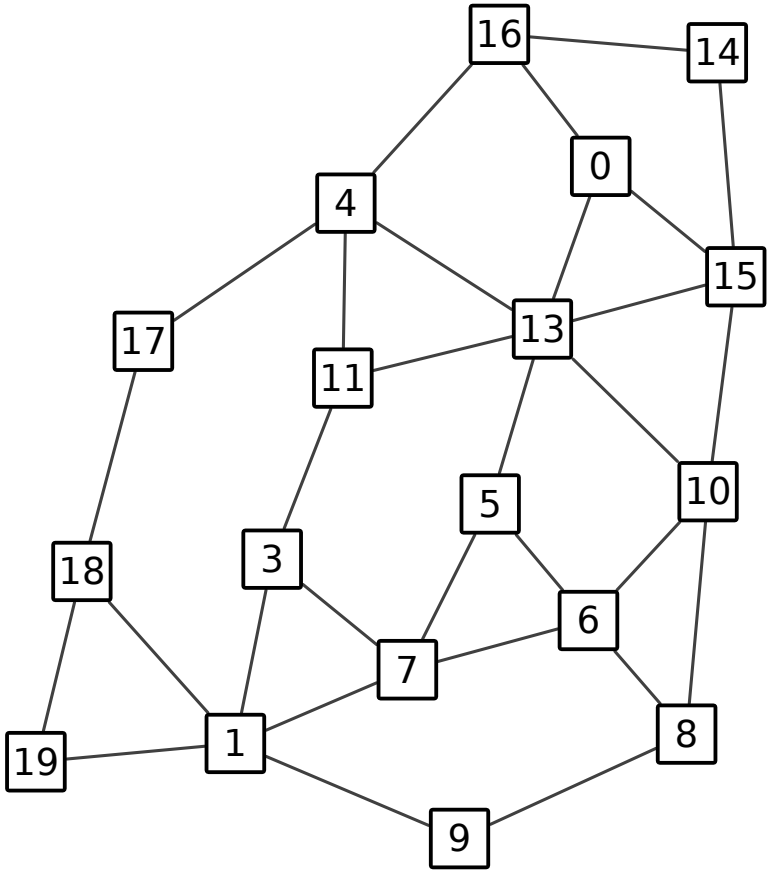
- Random Botnet Edge Deletion
- 90% loss – 83% precision

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

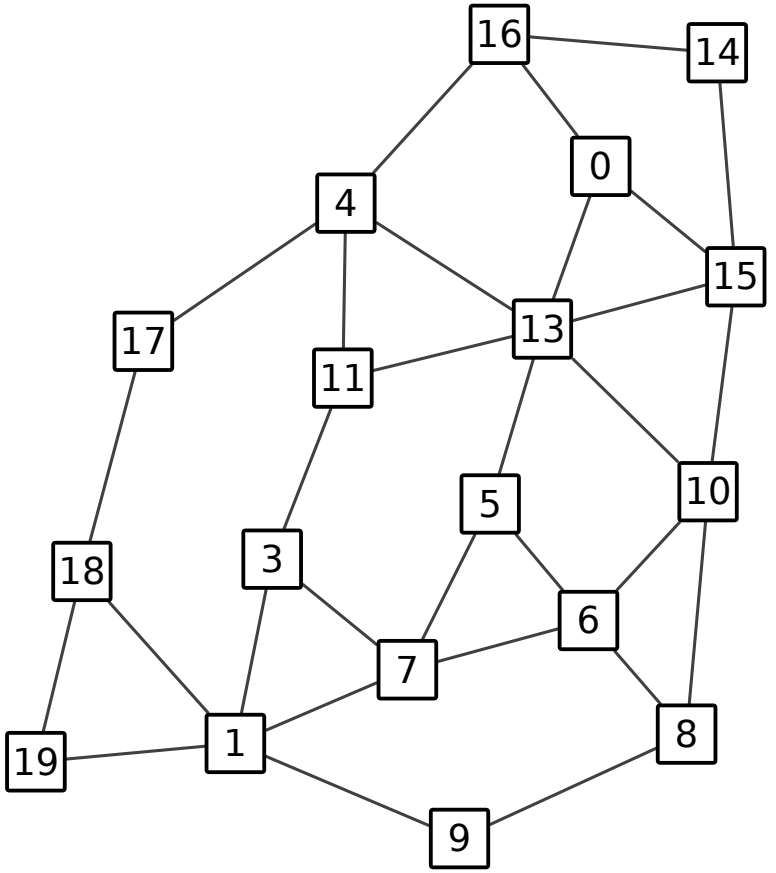


Host-based Visibility



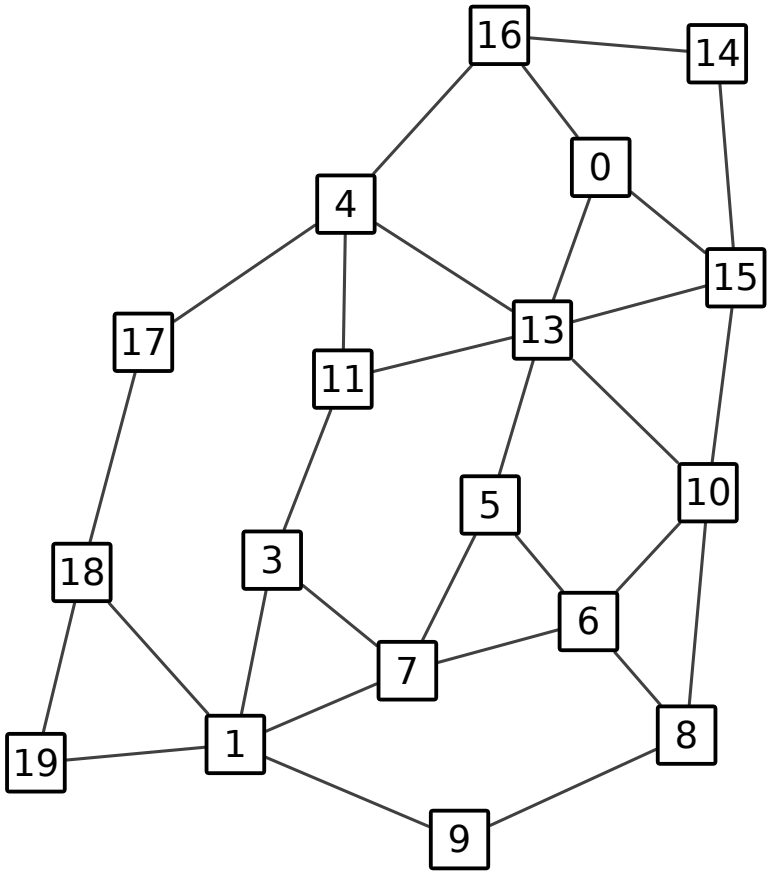
Host-based Visibility

- Sensor deployment



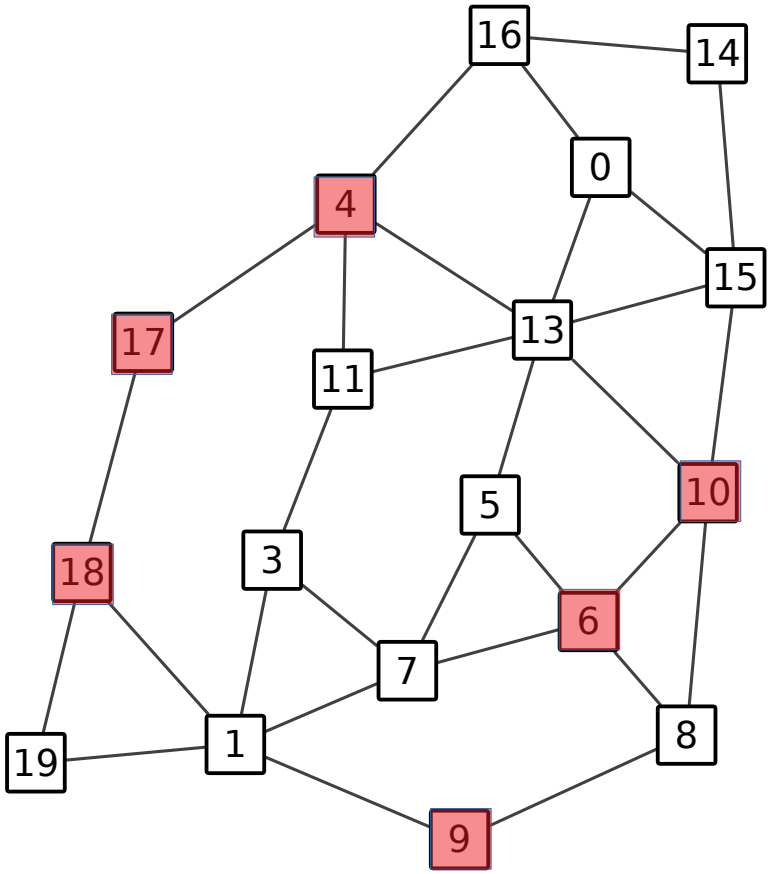
Host-based Visibility

- Sensor deployment
- Randomly chosen



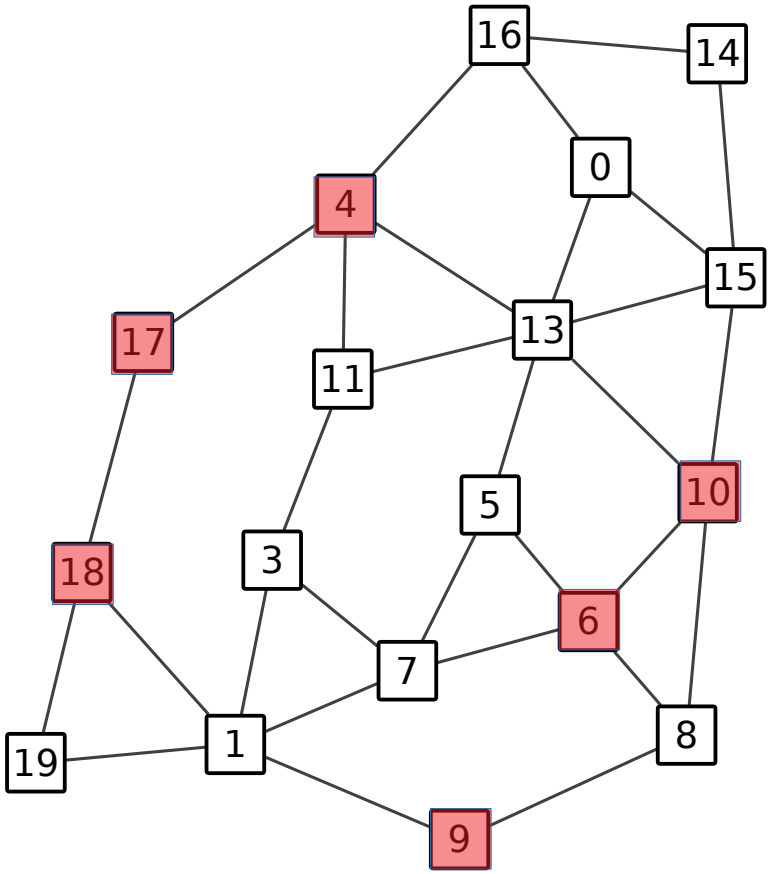
Host-based Visibility

- Sensor deployment
- Randomly chosen



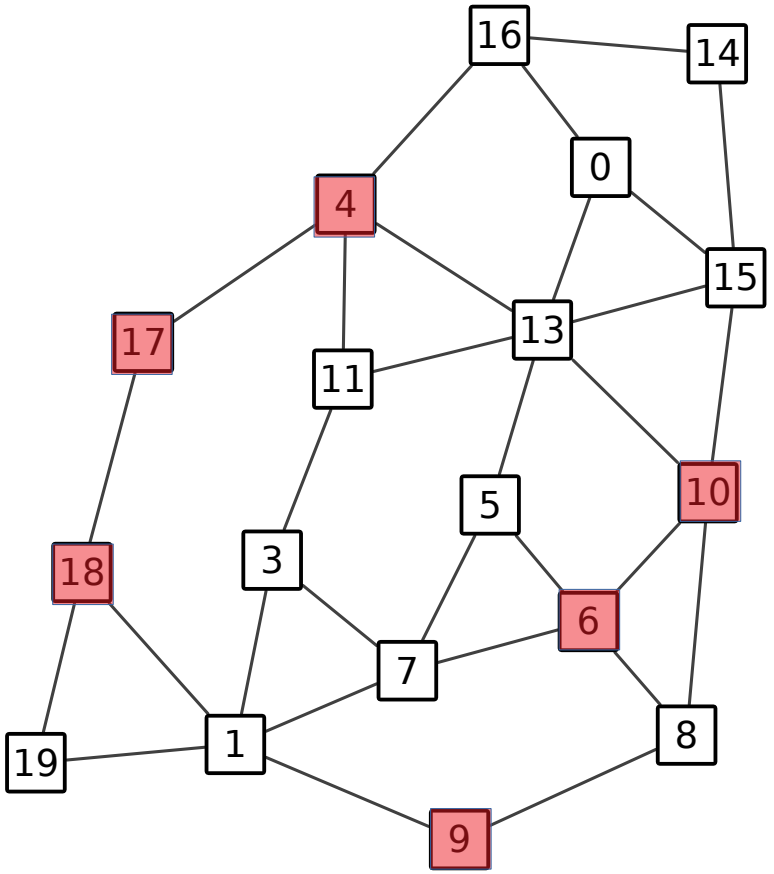
Host-based Visibility

- Sensor deployment
- Randomly chosen
- No communication between unmonitored hosts



Host-based Visibility

- Sensor deployment
- Randomly chosen
- No communication between unmonitored hosts
- Honeypot scenario

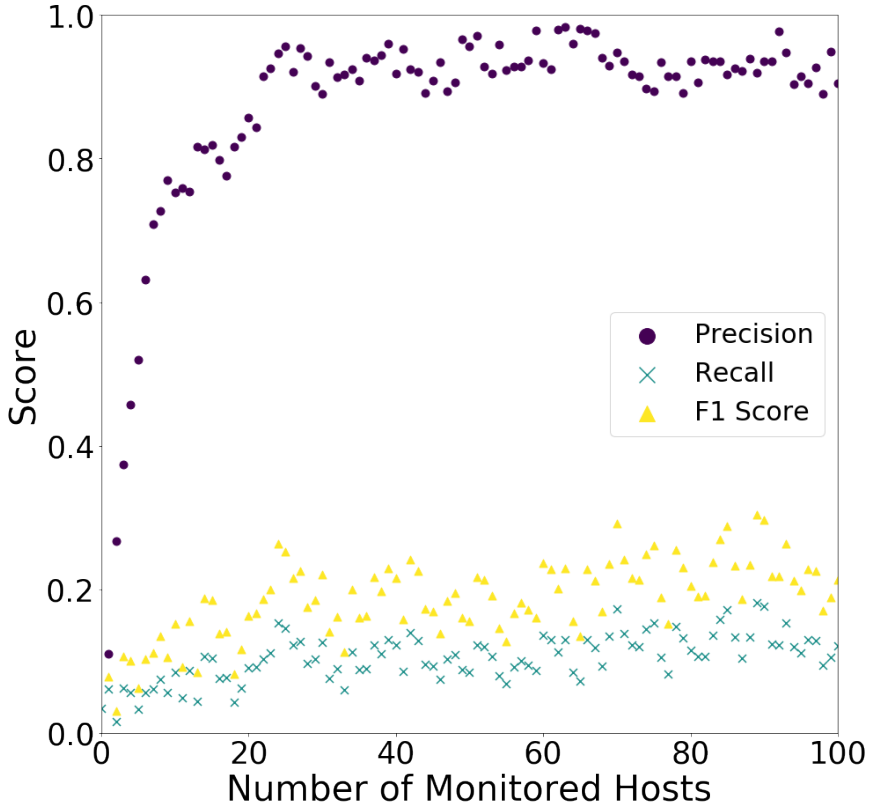


Sensor-Network Robustness

- Sensor deployment

Sensor-Network Robustness

- Sensor deployment

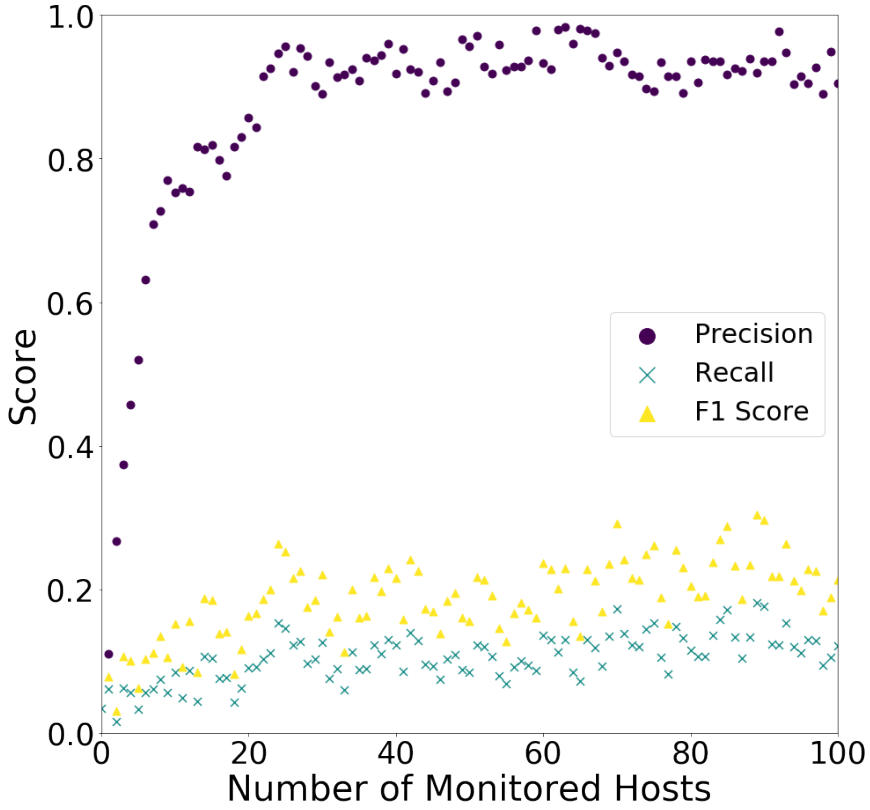


Sensor-Network Robustness

- Sensor deployment

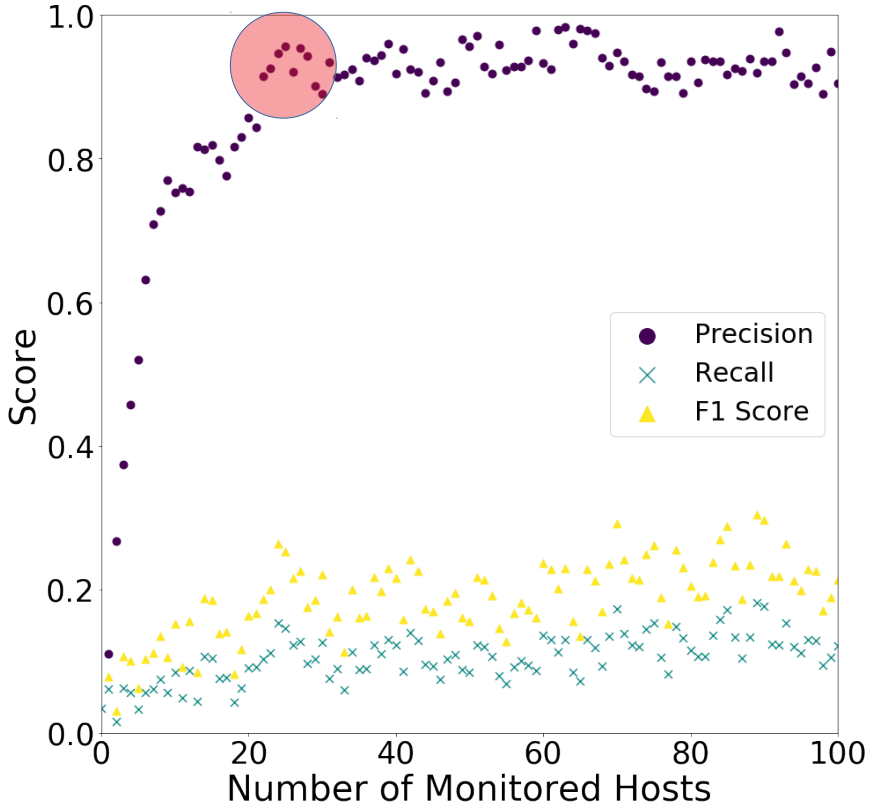
$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$



Sensor-Network Robustness

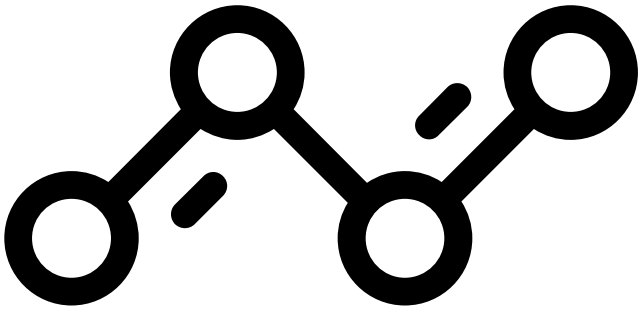
- Sensor deployment
- 25 sensors – 90% precision



$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

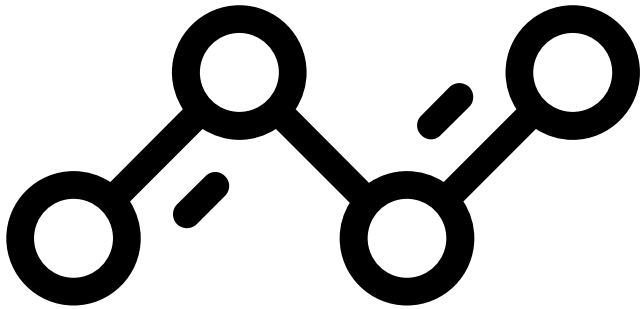
Conclusion



[7]

Conclusion

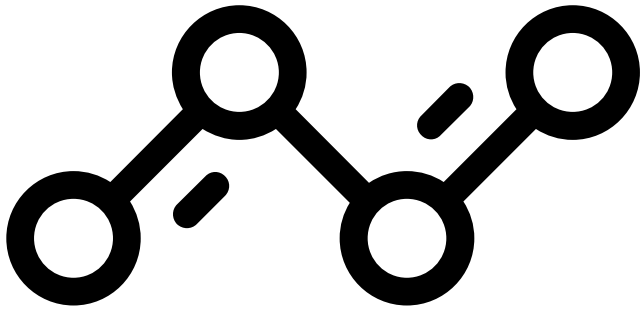
- Structured and unstructured botnets: fast-mixing



[7]

Conclusion

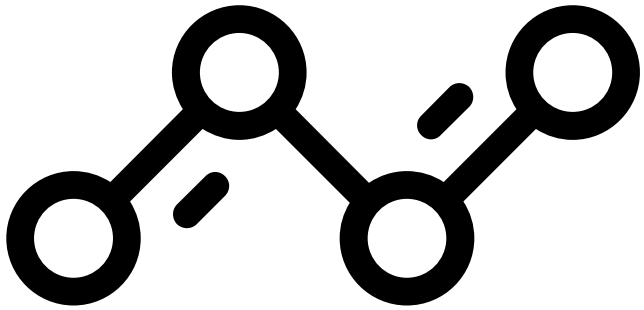
- Structured and unstructured botnets: fast-mixing
- High-precision detection
  - 83% precision



[7]

Conclusion

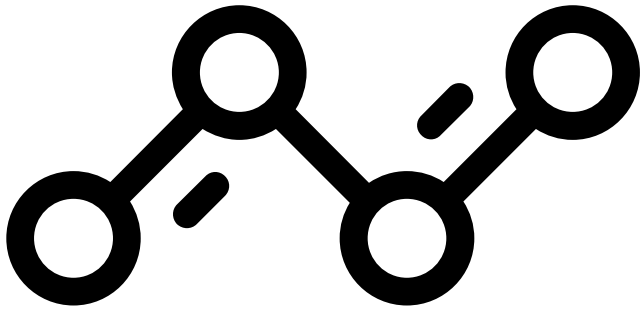
- Structured and unstructured botnets: fast-mixing
- High-precision detection
  - 83% precision
  - With 90% missing edges



[7]

Conclusion

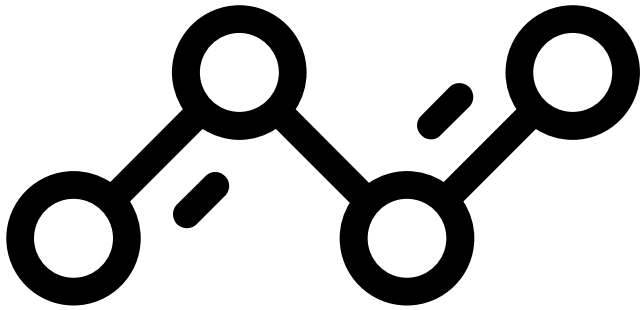
- Structured and unstructured botnets: fast-mixing
- High-precision detection
  - 83% precision
  - With 90% missing edges
- Simple architecture



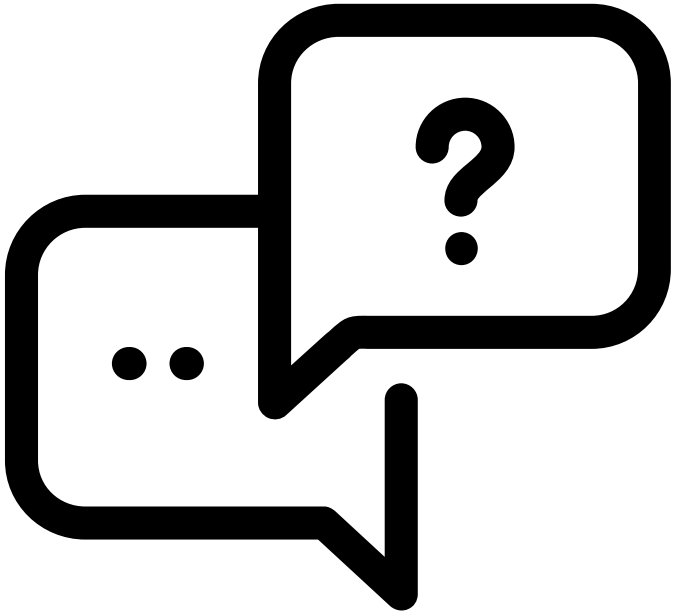
[7]

Conclusion

- Structured and unstructured botnets: fast-mixing
- High-precision detection
  - 83% precision
  - With 90% missing edges
- Simple architecture
- Only open-source algorithms



[7]



[7]

Thanks!

Questions?

References

[1] http://www.theregister.co.uk/2017/04/27/hajime_iot_botnet/

[2] https://www.zdnet.com/article/satori-botnet-successor-targets-ethereum-mining-rigs/

[3] https://arstechnica.com/information-technology/2017/12/100000-strong-botnet-built-on-router-0-day-could-strike-at-any-time/

[4] https://www.scmagazine.com/malicious-bot-traffic-climbs-95-percent-in-2017-says-report/article/754164/

[5] https://www.zdnet.com/article/new-mirai-style-botnet-targets-the-financial-sector/

[6] https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/-hide-n-seek-botnet-uses-peer-to-peer-infrastructure-to-compromise-iot-devices

[7] Icon made by Freepik from https://www.flaticon.com/

[8] Icon made by dDara from https://www.flaticon.com/

[9] Icon made by Kiranshastry from https://flaticon.com/

[10] Shishir Nagaraja et al. “BotGrep: finding P2P bots with structured graph analysis”. In: USENIX Security Symposium. 2010, p. 7.

[11] Pratik Narang et al. “PeerShark: Detecting peer-to-peer botnets by tracking conversations”. In: Proceedings – IEEE Symposium on Security and Privacy. Vol. January 20. 2014, pp. 108–115.

[12] Guofei Gu, Junjie Zhang, and Wenke Lee. “BotSniffer : Detecting Botnet Command and Control Channels in Network Traffic”. In: Proceedings of the 15th Annual Network and Distributed System Security Symposium. 53.1 (2008), pp. 1–13.