Design and methodology of a longitudinal honeypot study

Shreyas Srinivasa, Jens M. Pedersen, *Emmanouil Vasilomanolakis
Aalborg University, Denmark

*Technical University of Denmark

AALBORG UNIVERSITY

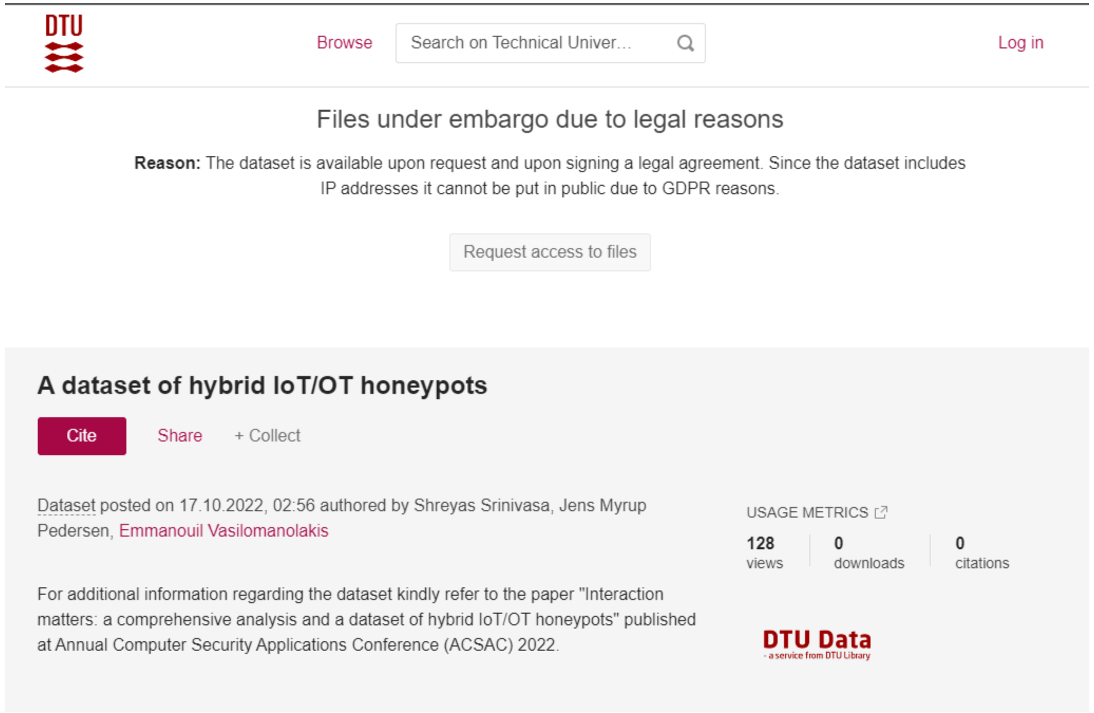
UNIVERSITY OF CAMBRIDGE

$>:whoami()

- Shreyas Srinivasa
- Ph.D. Fellow – Cybersecurity Research Group, Aalborg University, Copenhagen, Denmark
- Research Interests – Threat Intelligence, Cyber Deception, Internet Security Measurements
- Visiting Scholar at University of Cambridge (Cambridge Cybercrime Centre)
- Prior to Ph.D. – worked in SOC team of a bank in Germany
- Masters from TU Darmstadt, Germany

Regarding the dataset/artifact ☹

- Interaction matters: a comprehensive analysis and a dataset of hybrid IoT/OT honeypots (ACSAC 2022)
- No artifact ☹, thanks to GDPR and legal entanglement around it
- Dataset available as embargo, on request (https://doi.org/10.11583/DTU.21088651)
- Ongoing effort to clear the legal hurdles,
- Pseudo-anonymization?
- ~5 TB (comp.)



- **Background**
- Problem
- Design
- Methodology
- Analysis
- Limitations

Honeypots

- deception-based entities that simulate services, gather attack information
- decoys, with a "Know your enemy" concept
- used in defensive security as a trap mechanism
- act as sensors that can be used for malware collection
- study attacker behavior
- insider attacks
- classified based on interaction-levels offered to attackers
  - Low – limited simulation of a protocol (application level)
  - Medium – extended simulation, may include a service/device/profile
  - High – actual systems with services configured to work as a honeypot

Value

Any interaction with a "honeypot" system is suspicious

As they are non-production systems, there is no real reason for any interaction with them

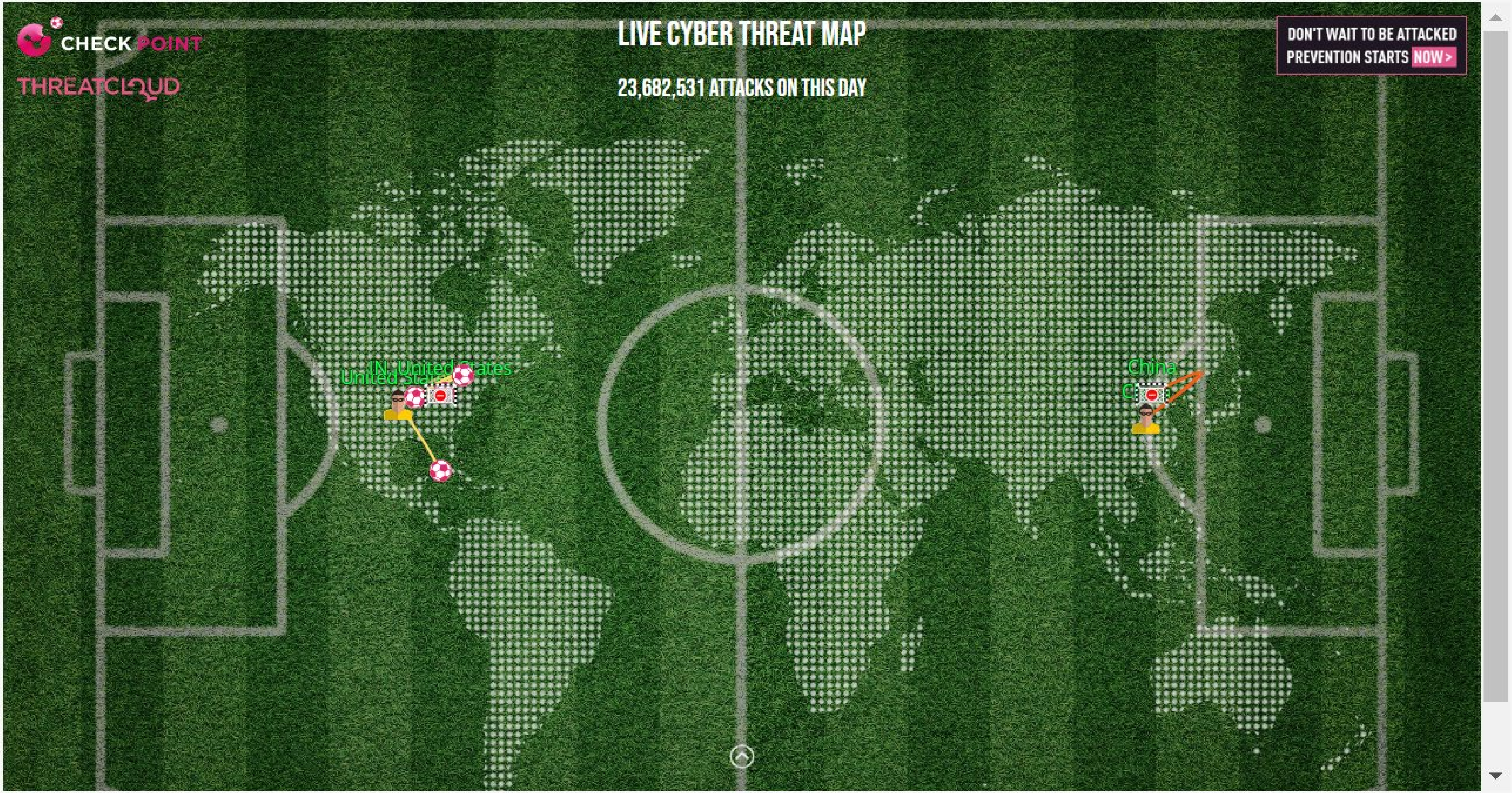
Traditional honeypots

Honeypots	Ports & Services
Kippo	Ports:22/2222 Services: SSH
Cowrie	Ports: 22/2222 23/2323 Services: SSH, Telnet
Glastopf	Ports: 80, 8080 Services: HTTP
Dionaea	Ports: 80, 443, 21 Services: HTTP, FTP
Nepenthes	Ports: 21 Services: FTP
Amun	Ports: 23,21,80,36,143 Services: Telnet, FTP, HTTP, SMTP, IMAP
Conpot	Ports: 80, 502, 102 Services: HTTP, Modbus, S7
Gaspot	Ports: 100001 Services: ATG
MTPot	Ports: 23 Services: Telnet

THE HONEYNET PROJECT

Honeynets / Honeyfarms

- Instead of deploying large number of honeypots or honeypots on every network, you simply deploy your honeypots in a single, consolidated location
- Attackers are redirected to the farm, regardless of of what network they are on / probing
- act as sensors and offer telemetry/feed of events
- Source of Threat Intelligence data
- Can be a one consolidated honeypot host or multiple honeypots deployed in diverse locations



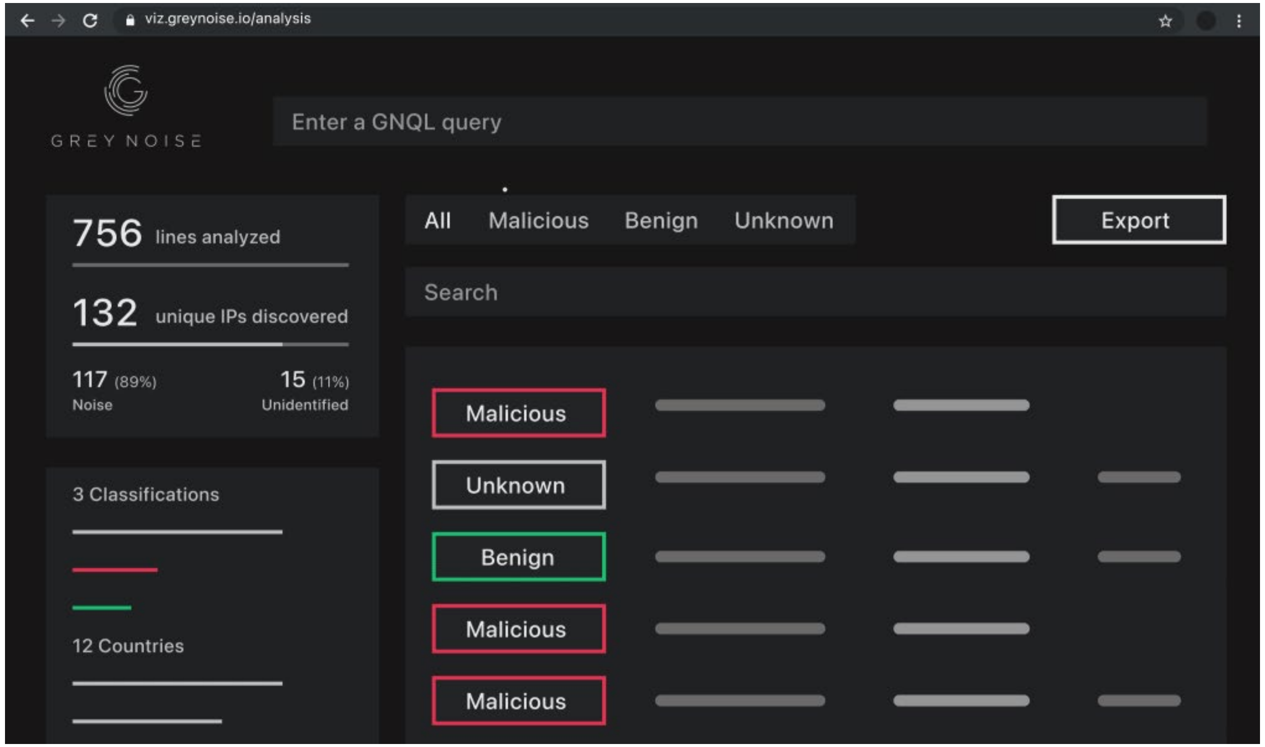
Source: https://threatmap.checkpoint.com/

Web Viewer Terms | Privacy & Cookies

Edit

GREYNOISE
INTELLIGENCE

- Turning Internet scanning noise into intelligence
- Removing false positives from Internet scanners like Shodan, Censys …
- Trending vulnerabilities



- Background
- **Problem**
- Design
- Methodology
- Analysis
- Limitations

RQ

- Do any operational parameters **influence** the type of attacks received on a honeypot?
- What is the influence of known operational parameters
  - Interaction-levels
  - Simulation environments
  - Deployment infrastructure
  - Geo-location

Limitations of current Datasets

- Honeypot datasets are not public (curated)
- Anonymized
- GDPR
- Most honeypots deployed by companies are either in low or medium interaction
- Security corporations have some limitations in what they share, less freedom, low flexibility

Related work – Honeypot Studies

Study	Interaction level	Study period	Geographically distributed	Deployment
Honeycloud [7] (2019)	Medium	12 months	Yes	hardware, cloud
IoTPOT [27](2015)	Low	39 days	No	physical
Open for hire [40] (2021)	Low, Medium	1 month	No	physical
Muti-faceted Honeypot [52](2020)	Low	2 years	No	physical
Honware [48] (2019)	High	14 days	No	physical
Siphon [13](2017)	High	2 months	Yes	physical, cloud
Hornet 40 [44](2021)	Passive	40 days	Yes	cloud
Picky Attackers [3] (2017)	Medium	4 months	Yes	physical, cloud

Designing a longitudinal honeypot study -Challenges

- None of the studies had an empirical focus towards all the parameters in the study
- Traditional honeypots are limited in interaction levels (i.e., offer binary interaction, either low or medium or high)
- Some honeypots known to be vulnerable to fingerprinting attacks (* Vetterl et al.)
- Structured attack data collection
- Staleness

* Vetterl, A., & Clayton, R. (2018). Bitter harvest: Systematically fingerprinting low-and medium-interaction honeypots at internet scale. In *12th USENIX Workshop on Offensive Technologies (WOOT 18)*.

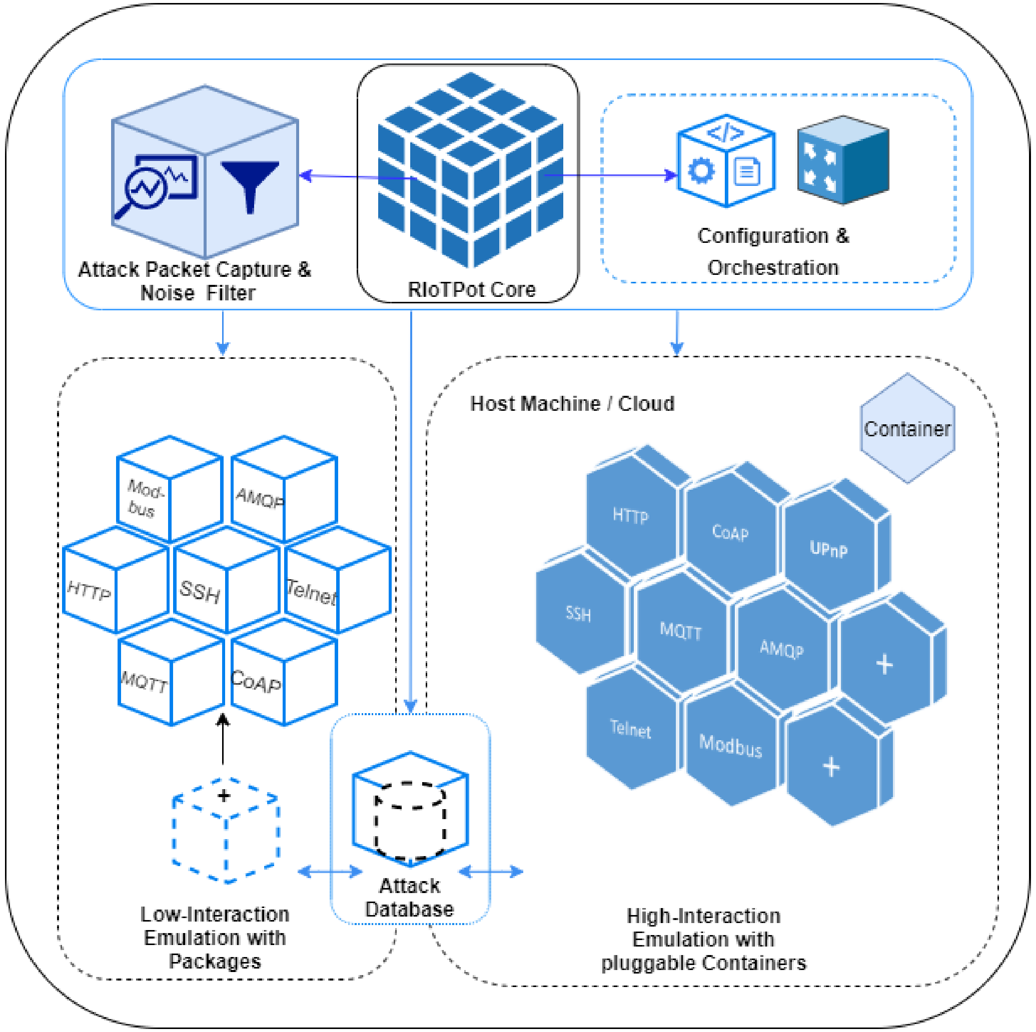
To study the influence

- What is the influence of known operational parameters
  - Interaction-levels → Must have multiple interaction levels
  - Simulation environments → Must simulate multiple protocols (application level)
  - Deployment infrastructure → Deployed on physical (lab env.) and cloud
  - Geo-location → Operational in multiple geo-locations

- Background
- Problem
- **Design**
- Methodology
- Analysis
- Limitations

RIoTPot

- A hybrid-interaction honeypot
- Modular
- Containerized
- Extensibility
- Active noise filter
- Flexible event storage and logging



https://github.com/aau-network-security/riotpot

Related work – Honeypot Studies

Study	Interaction level	Study period	Geographically distributed	Deployment
Honeycloud [7] (2019)	Medium	12 months	Yes	hardware, cloud
IoTPOT [27](2015)	Low	39 days	No	physical
Open for hire [40] (2021)	Low, Medium	1 month	No	physical
Muti-faceted Honeypot [52](2020)	Low	2 years	No	physical
Honware [48] (2019)	High	14 days	No	physical
Siphon [13](2017)	High	2 months	Yes	physical, cloud
Hornet 40 [44](2021)	Passive	40 days	Yes	cloud
Picky Attackers [3] (2017)	Medium	4 months	Yes	physical, cloud
RIoTPot (2022)	**Low, High, Hybrid**	3 **months**	**Yes**	**physical, cloud**

Design - Longitudinal Study

- 3 Interaction levels - Low, High, Hybrid
- 2 Deployment environments - lab, cloud
- 12 independent honeypot hosts per interaction level
- 4 geographical locations - Denmark(Lab), Germany, New York City, Singapore
- 6 application protocols – Telnet, SSH, HTTP, MQTT, Modbus, CoAP
- Comparison with 1 medium interaction honeypot – Conpot
- 3 months of evaluation

Design - Longitudinal Study

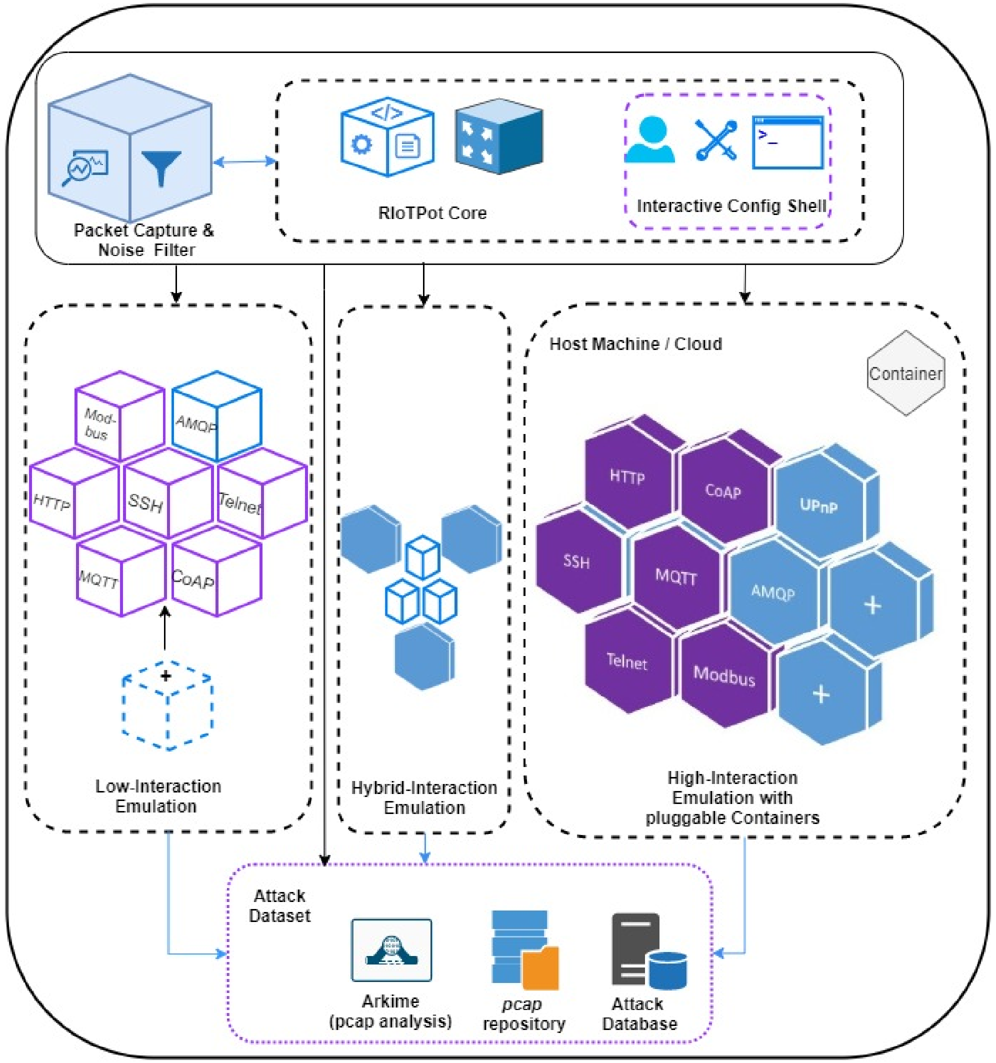
Host	Environment	Geo-Location	Interaction-level	Protocols Emulated
R1	Lab	Denmark	High	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R2	Lab	Denmark	Low	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R3	Lab	Denmark	Hybrid	High - SSH, MQTT, Modbus, CoAP Low - Telnet, HTTP
C1	Lab	Denmark	Medium	Telnet, SSH, HTTP, Modbus, S7
R4	Cloud	New York City	High	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R5	Cloud	New York City	Low	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R6	Cloud	New York City	Hybrid	High - SSH, MQTT, Modbus, CoAP Low - Telnet, HTTP
C2	Cloud	New York City	Medium	Telnet, SSH, HTTP, Modbus, S7
R7	Cloud	Frankfurt	High	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R8	Cloud	Frankfurt	Low	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R9	Cloud	Frankfurt	Hybrid	High - SSH, MQTT, Modbus, CoAP Low - Telnet, HTTP
C3	Cloud	Frankfurt	Medium	Telnet, SSH, HTTP, Modbus, S7
R10	Cloud	Singapore	High	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R11	Cloud	Singapore	Low	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R12	Cloud	Singapore	Hybrid	High - SSH, MQTT, Modbus, CoAP Low - Telnet, HTTP
C4	Cloud	Singapore	Medium	Telnet, SSH, HTTP, Modbus, S7

Table 2: Experimental setup overview

- Background
- Problem
- Design
- **Methodology** ←
- Analysis
- Limitations

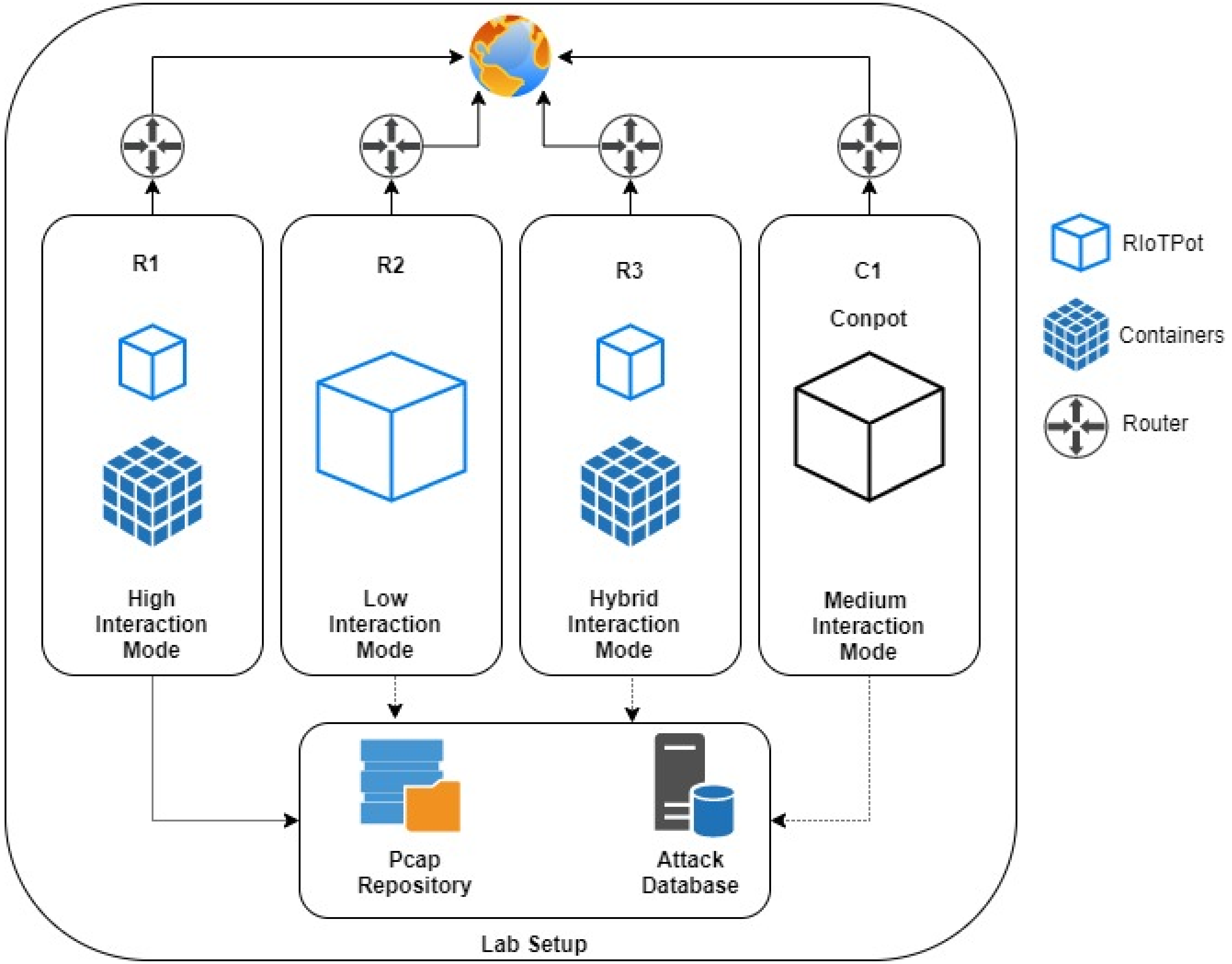
RIoTPot – adapting for the study

- Interactive setup and configuration shell
- Enhancing the emulation of SSH, Modbus, HTTP, MQTT, CoAP protocols
- Inclusion of verified docker images for the high-interaction emulation
- pcap analysis with Arkime and a pcap repository for extended packet-level capture and analysis

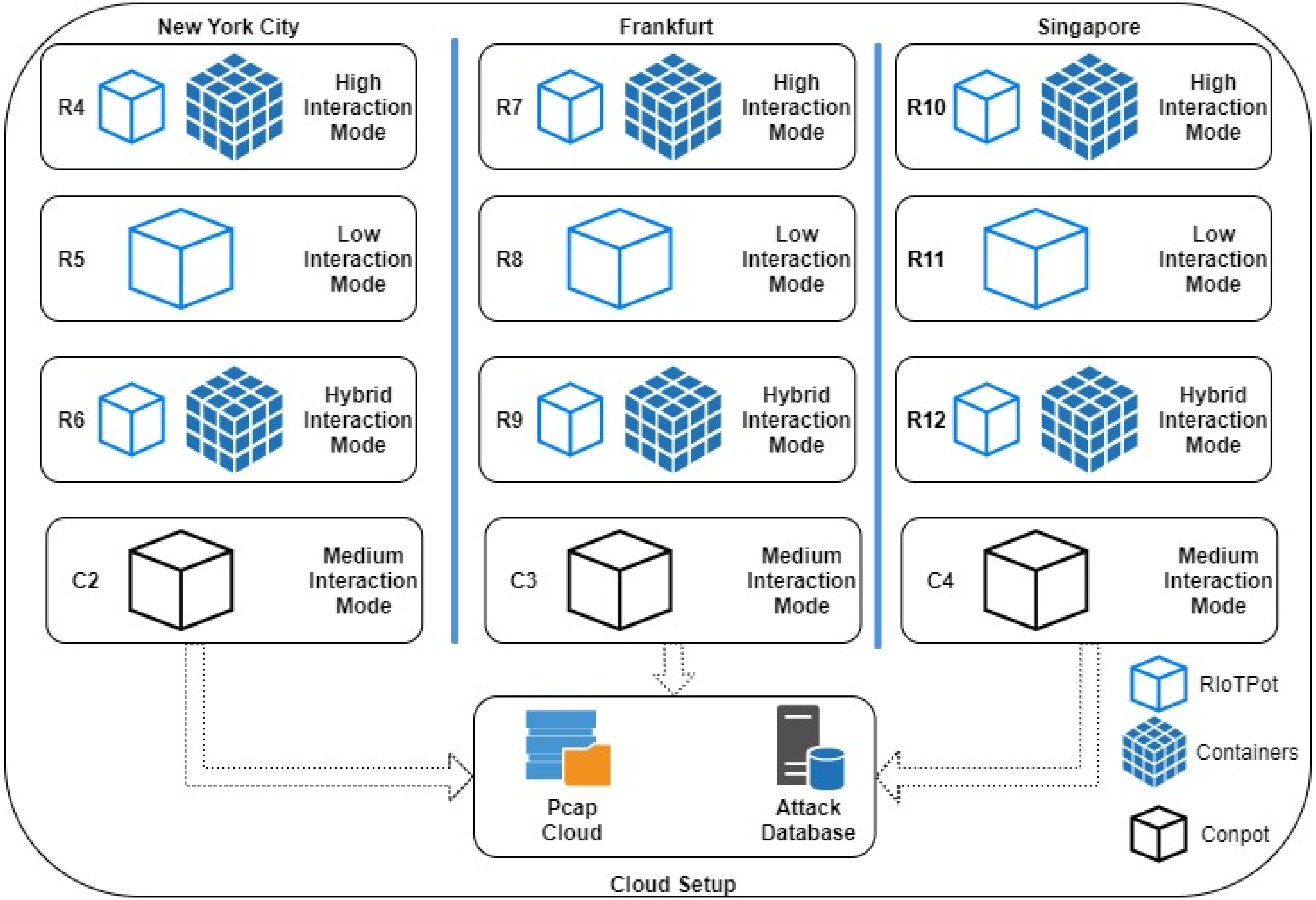


https://github.com/aau-network-security/riotpot

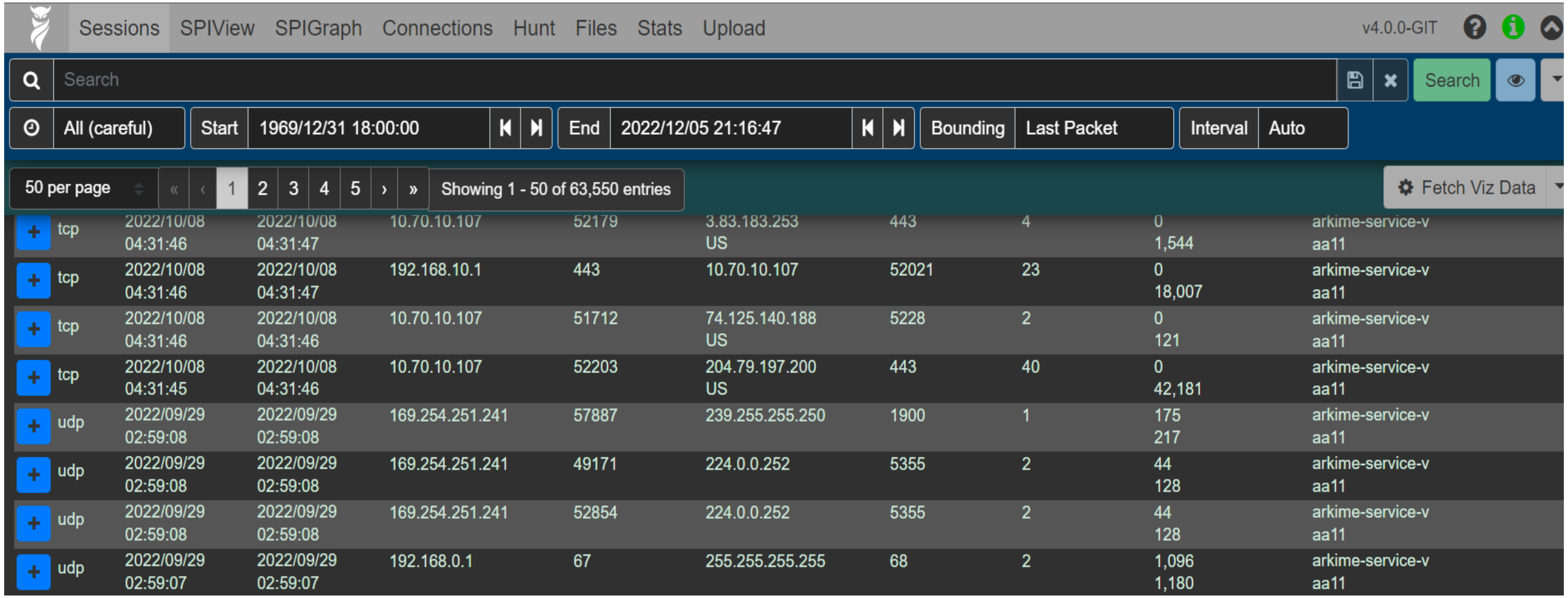
Lab Setup (Denmark)



Cloud Setup


New York City
Frankfurt
Singapore
R4
High Interaction Mode
R5
Low Interaction Mode
R6
Hybrid Interaction Mode
C2
Medium Interaction Mode
R7
High Interaction Mode
R8
Low Interaction Mode
R9
Hybrid Interaction Mode
C3
Medium Interaction Mode
R10
High Interaction Mode
R11
Low Interaction Mode
R12
Hybrid Interaction Mode
C4
Medium Interaction Mode
Pcap Cloud
Attack Database
RIoTPot
Containers
Conpot
Cloud Setup

Arkime


Sessions SPIView SPIGraph Connections Hunt Files Stats Upload
v4.0.0-GIT
Search
Search
All (careful)
Start 1969/12/31 18:00:00
End 2022/12/05 21:16:47
Bounding Last Packet
Interval Auto
50 per page
1 2 3 4 5
Showing 1 - 50 of 63,550 entries
Fetch Viz Data
tcp 2022/10/08 04:31:46 2022/10/08 04:31:47 10.70.10.107 52179 3.83.183.253 US 443 4 0 1,544 arkime-service-v aa11
tcp 2022/10/08 04:31:46 2022/10/08 04:31:47 192.168.10.1 443 10.70.10.107 52021 23 0 18,007 arkime-service-v aa11
tcp 2022/10/08 04:31:46 2022/10/08 04:31:46 10.70.10.107 51712 74.125.140.188 US 5228 2 0 121 arkime-service-v aa11
tcp 2022/10/08 04:31:45 2022/10/08 04:31:46 10.70.10.107 52203 204.79.197.200 US 443 40 0 42,181 arkime-service-v aa11
udp 2022/09/29 02:59:08 2022/09/29 02:59:08 169.254.251.241 57887 239.255.255.250 1900 1 175 217 arkime-service-v aa11
udp 2022/09/29 02:59:08 2022/09/29 02:59:08 169.254.251.241 49171 224.0.0.252 5355 2 44 128 arkime-service-v aa11
udp 2022/09/29 02:59:08 2022/09/29 02:59:08 169.254.251.241 52854 224.0.0.252 5355 2 44 128 arkime-service-v aa11
udp 2022/09/29 02:59:07 2022/09/29 02:59:07 192.168.0.1 67 255.255.255.255 68 2 1,096 1,180 arkime-service-v aa11

Arkime

Sessions
SPIView
SPIGraph
Connections
Hunt
Files
Stats
History
Settings
Users
Search
All (careful)
Start
1969/12/31 19:00:00
End
2021/09/23 13:06:45
Bounding
Last Packet
Interval
Auto
Query Size
100
Src:
Src IP
Dst:
Dst IP
Src & Dst
Min. Connections
Node/Link Weight
Sessions
Baseline
disabled
190.0.0.1
Type
Source
Links
12
Sessions
12
Bytes
15,720
Data bytes
0
Packets
120
Arkime Node
test
Hide Node

- Background
- Problem
- Design
- Methodology
- **Analysis**
- Limitations

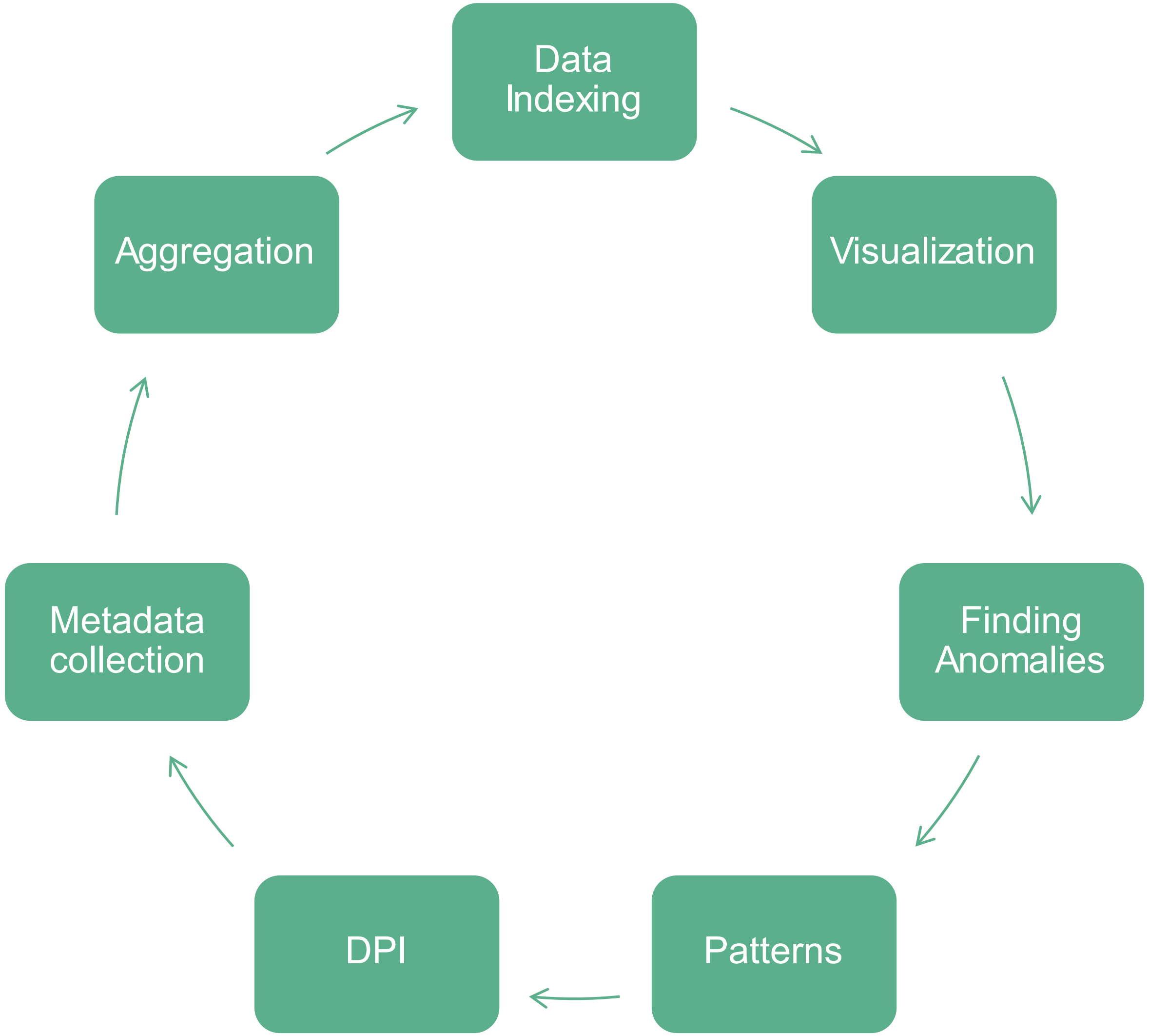
Dataset

- A comprehensive dataset of *pcaps* and events in database
- The database schema contains
  - Source IP address (attacker)
  - Destination IP addresses (honeypots, anonymized)
  - Source IP ports
  - Destination IP ports
  - Timestamps
  - Geolocation of the attacker IPs
  - Interaction level of the honeypots and protocols (where the attack event was observed)
  - Deployment environment information of the honeypots (Cloud/Lab)
  - IP layer traffic and flags
  - Transport layer traffic and flags
  - Application layer data transmitted

Data analysis

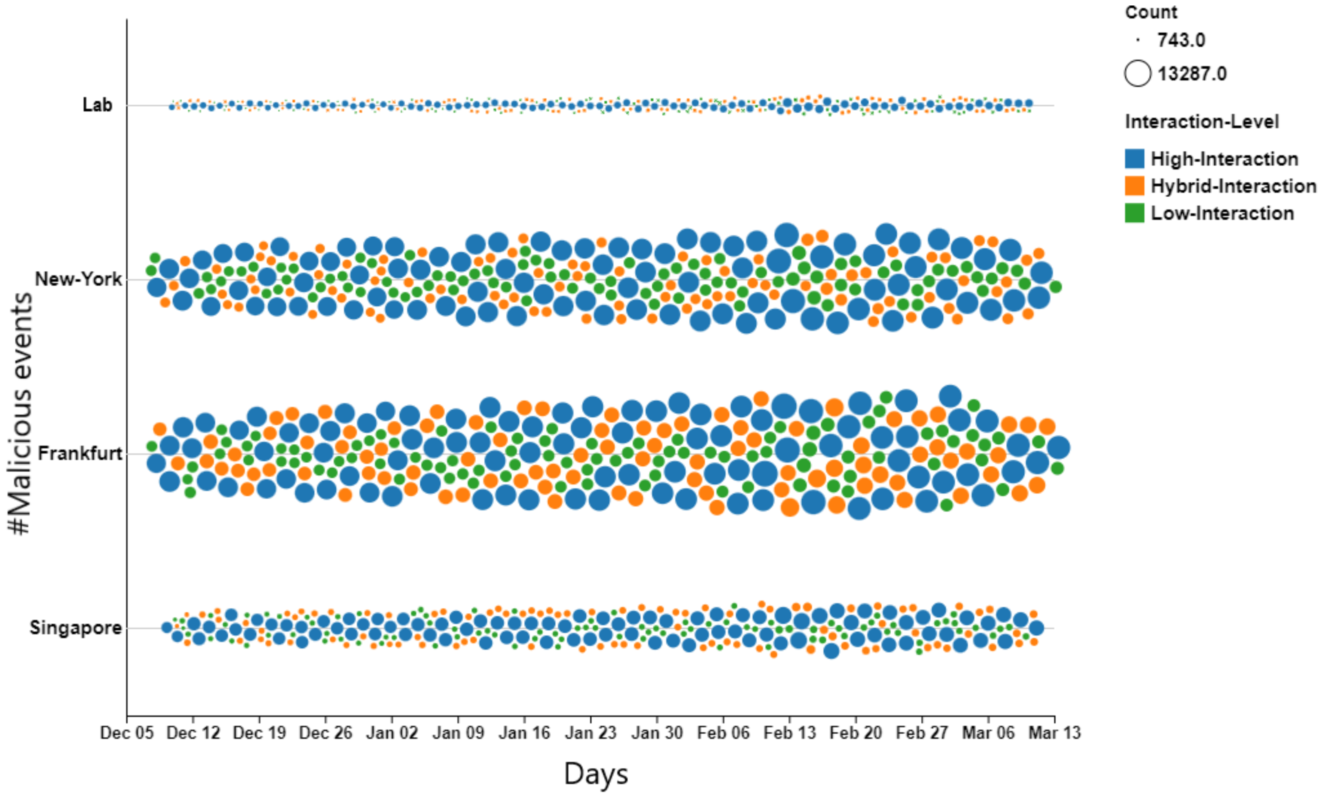
- The analysis was done on events recorded in json format in MongoDB
- The packet level inspection was done with Arkime
- The metadata for further analysis was requested from Greynoise

Combing/breakdown



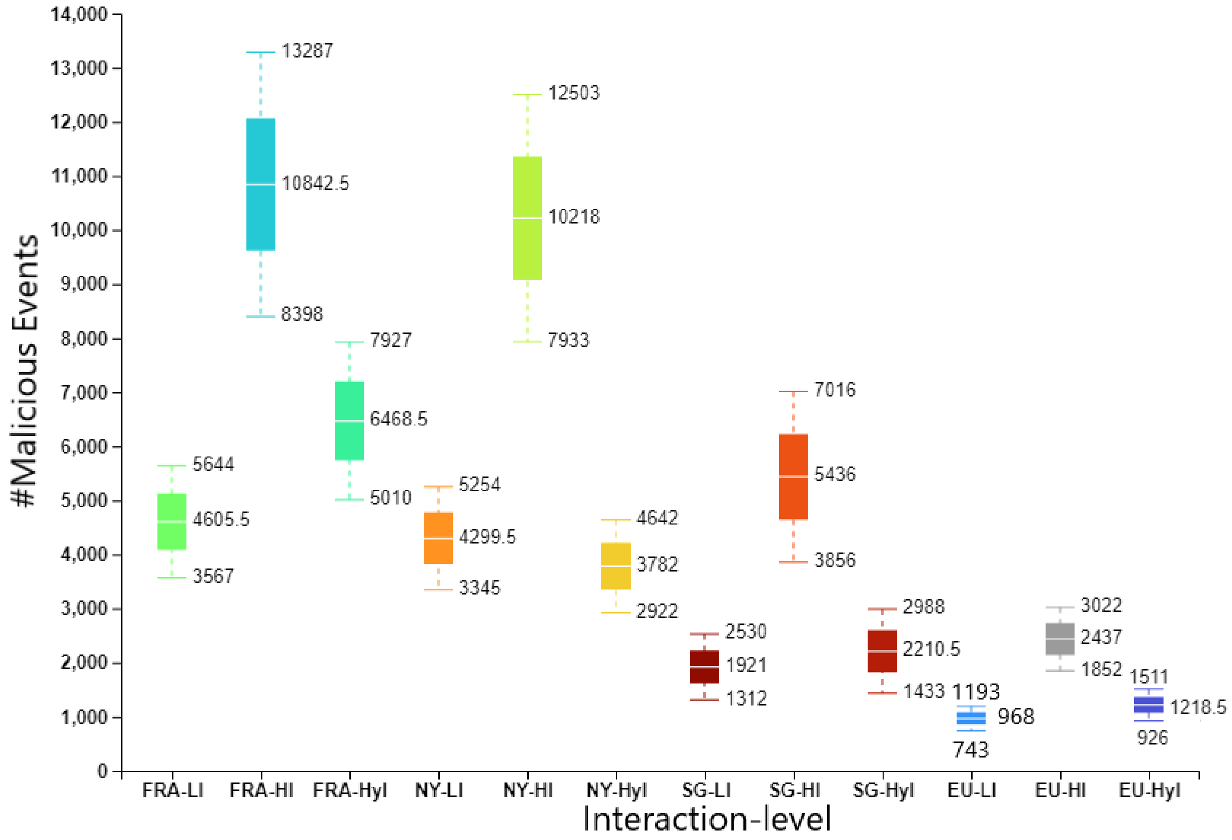
Parameter: Geo-location, city, interaction level, events

- Sphere size denotes the number of daily events per day by interaction-level
- lowest received: 743, highest: 13,287
- The lab instances received lower malicious events
- The Frankfurt instances (cloud) received the highest traffic overall



Parameter: Geo-location, lowest-highest, interaction-level

- Highest events recorded in Frankfurt, with High Interaction
- Lowest events recorded in lab deployment, with Low-interaction
- **Regardless, the High-interaction deployments received the highest events**



- Background
- Problem
- Design
- Methodology
- Analysis
- **Limitations**

Limitations

- One Lab deployment environment; uneven comparison with the cloud deployments
- Limited to 4 cities in 3 continents
- 6 protocols
- We consider each connection as an event, entailing limitations in terms of over-counting
- Not in Netflow format (flexible integration)
- Sharing limitations; GDPR issues in Europe (IP is considered sensitive information)

Failures

- Hosting “vulnerable” instances is tricky
- The National CERTS don’t want vulnerable instances around
- Also, in the cloud (ingress, egress rules)
- Cost!
- Monitoring

Summary

- Honeypots are still an effective tool ; if configured carefully
- The parameters play an important role in honeypots and honeypot studies
- Configuring the parameters based on studies provide a broader overview of the attack landscape

- Supplementary findings
  - High-interaction honeypots receive higher attack events
  - Location-specific attacks observed
  - There is an increase in “scanning-service” traffic, many new services observed

Lessons learnt

- Deploying, managing and operating honeypots is challenging
- Attackers could exploit honeypots to launch attacks
- Deception–based systems are a great resource, however you must have a strategy and look for what you need
- Threat Hunting is a tedious task, especially when you have billion events per day
- **The dataset is precious; however, the GDPR issues make the public sharing challenging – Open Question!**

References

- Shreyas Srinivasa, Jens Myrup Pedersen, and Emmanouil Vasilomanolakis. 2021. Open for hire: attack trends and misconfiguration pitfalls of IoT devices. In Proceedings of the ***21st ACM Internet Measurement Conference (IMC '21).*** Association for Computing Machinery, New York, NY, USA, 195–215. https://doi.org/10.1145/3487552.3487833
- Srinivasa, S., Pedersen, J. M., & Vasilomanolakis, E. (2021). RIoTPot: a modular hybrid-interaction IoT/OT honeypot. In *26th* ***European Symposium on Research in Computer Security (ESORICS) 2021***. Springer.
- Vetterl, A., & Clayton, R. (2018). Bitter harvest: Systematically fingerprinting low-and medium-interaction honeypots at internet scale. In ***12th USENIX Workshop on Offensive Technologies (WOOT 18)*.**
- S. Srinivasa, J. M. Pedersen and E. Vasilomanolakis, "Deceptive directories and “vulnerable” logs: a honeypot study of the LDAP and log4j attack landscape," ***2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)***, 2022, pp. 442-447, doi: 10.1109/EuroSPW55150.2022.00052.

Acknowledgement

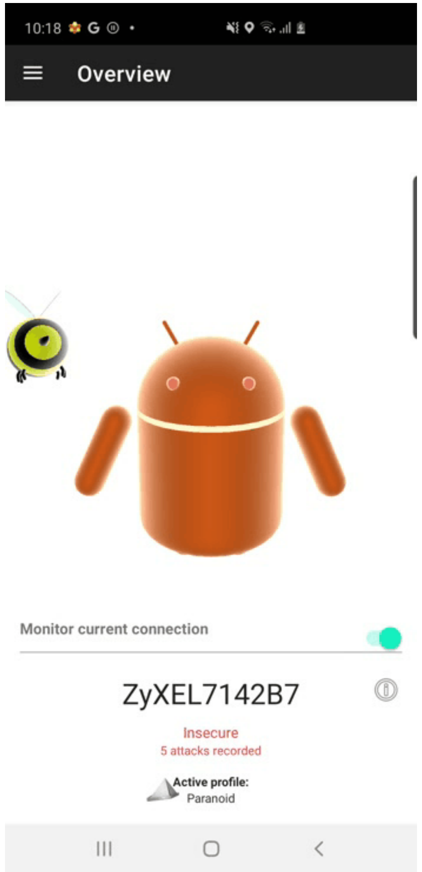
UNIVERSITY OF CAMBRIDGE

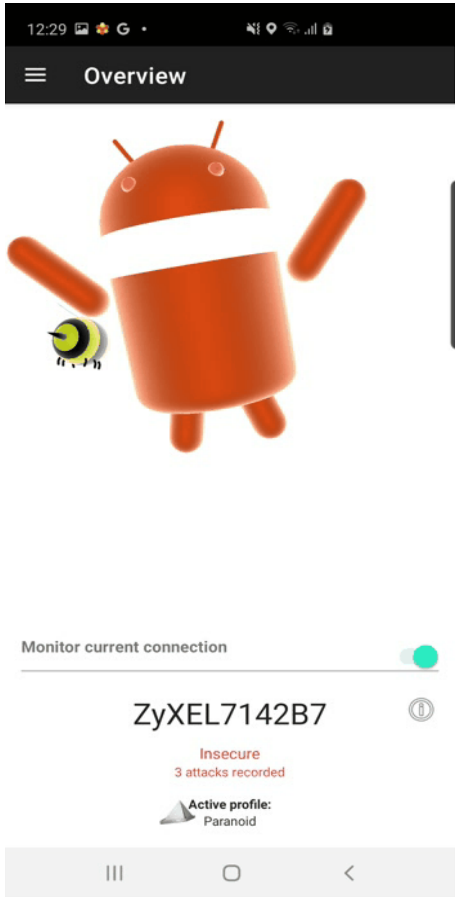
- Dr. Richard Clayton
- Dr. Alice Hutchings

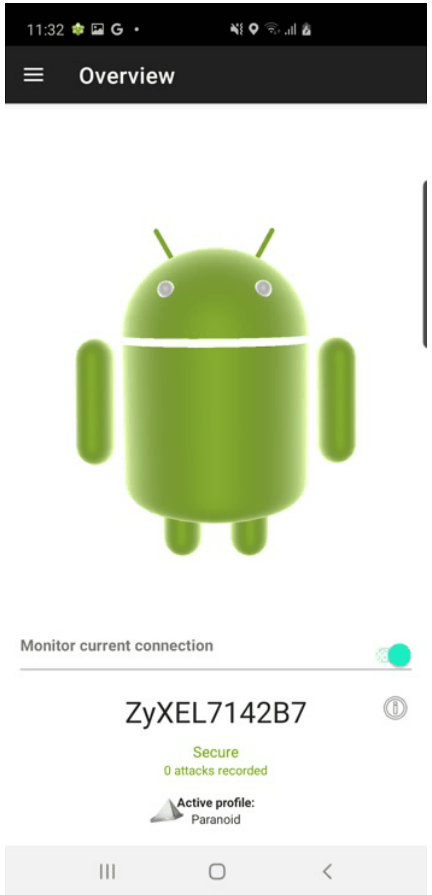
- Cambridge Cybercrime Centre, University of Cambridge
- Rich, curated datasets on Internet scanning, honeypots, DarkWeb, DeepWeb and more..

More from our research group

HosTaGe- an Interactive, mobile-based honeypot







Google Summer of Code

black hat EUROPE 2020

Contact

- Shreyas Srinivasa
- shsr@es.aau.dk
- https://sastry17.github.io
- Datasets on Selective Internet Scanning, Honeypots, Darkweb (marketplaces, forums)