Interaction matters: A comprehensive analysis and a dataset of hybrid IoT/OT honeypots

ACSAC 2022

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

AALBORG UNIVERSITY

Honeypots

- Traditional, 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 application protocol/service (e.g., SSH, Telnet)
  - Medium – extended simulation, may include a device/profile/vulnerability (e.g., Log4j, Windows XP, Siemens S7)
  - High – actual systems with services configured to work as a honeypot

Value

As non-production systems, there is no real reason for any interaction with honeypots

Any interaction with a “honeypot” system can be suspicious

Limitations in Honeypot studies

- Narrowed scope to a specific vulnerability/protocol/device
- Operation in a single interaction-level (mostly low or medium)
- Limited geographical perspective
- Limited deployment perspective
- Provide the attack landscape to a specific ecosystem (IoT/OT/IT)
- **May contain noise, low-fidelity alerts (Internet scanning services)**

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

Motivation

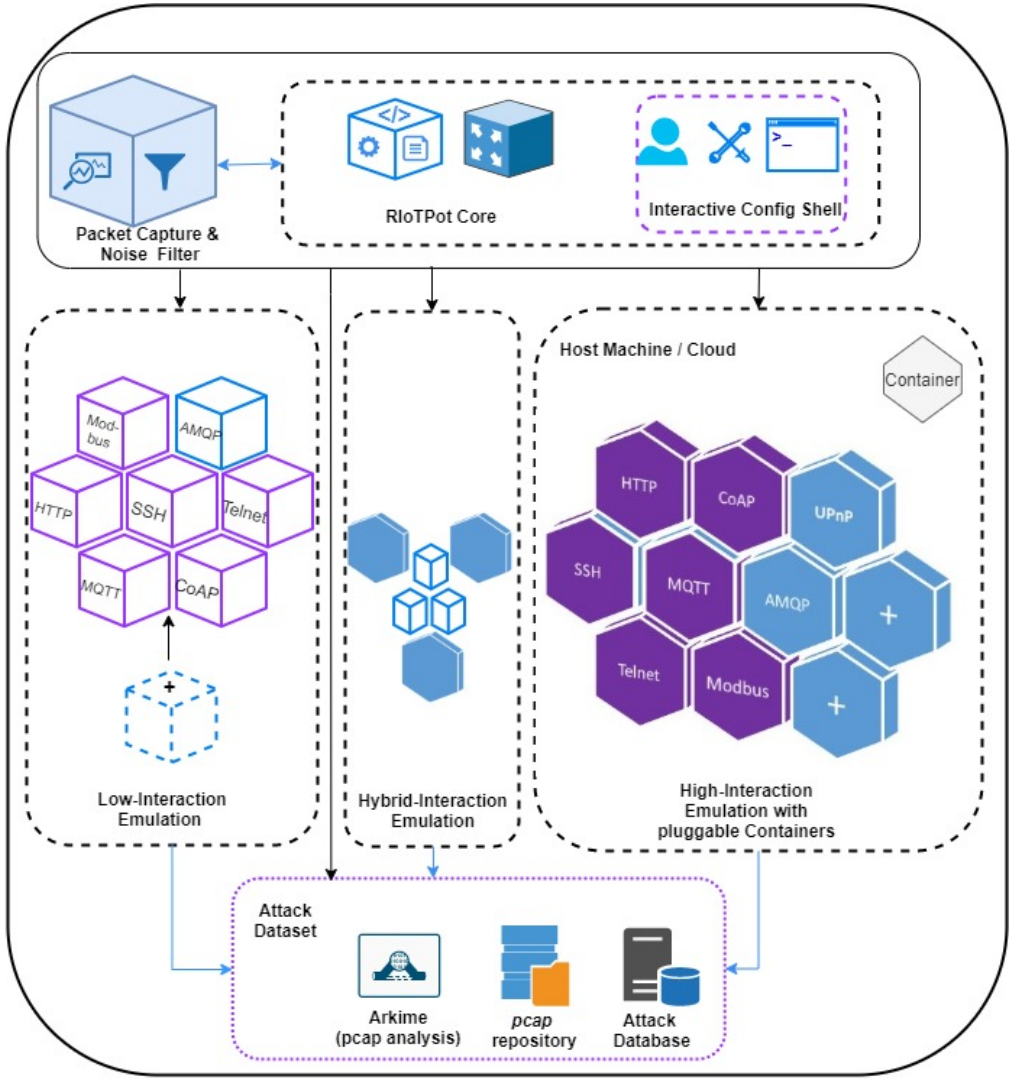
- Do any operational parameters **influence** the type of attacks received on a honeypot?
- What is the influence of known operational parameters on honeypot studies
  - Interaction-levels
  - Simulation environments
  - Deployment infrastructure
  - Geo-location
- Can we capture specific attacks on different parameters?
- Producing a dataset that the research community can use with more freedom, flexibility and **less noise**

Study scope

- Conduct a honeypot study to **evaluate the influence of operational parameters**
- A study of **3 months**
- **6 application protocols** (Telnet, SSH, HTTP, MQTT, Modbus, CoAP)
- **4 Geo-locations**
- **2 deployment environments** (Lab, Cloud)
- **3 interaction levels** (low, high, hybrid)
- **16 hosts** in total

RIoTPot*

- Modular
- Hybrid-interaction
  - Choice of operation (low, medium or high)
  - Choice of operation of specific protocols in either interaction
- extensible
- Extended to adapt to this study



*Srinivasa, S., Pedersen, J. M. & Vasilomanolakis, E., "RIoTPot: a modular hybrid-interaction IoT/OT honeypot", In 26th European Symposium on Research in Computer Security (ESORICS 2021), Darmstadt, Germany, October 4–8, 2021, Proceedings, Part II, Springer, Vol. 2. p. 745-751 7 p. (Lecture Notes in Computer Science, Vol. 12973)

Overview

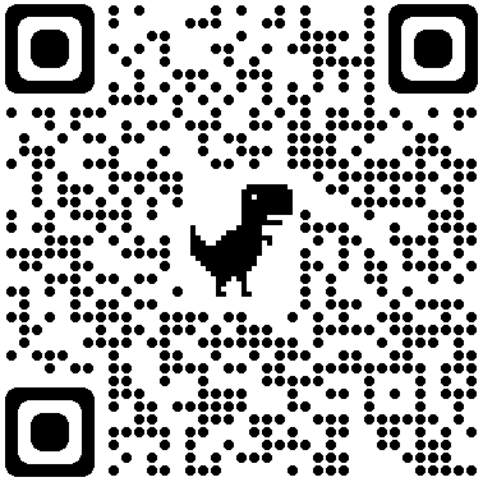
Host	Environment	Geo-Location	Interaction-level	Protocols Emulated
R1	Lab	Denmark	High	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R2	Lab	Denmark	Low	Telnet, SSH, HTTP, MQTT, Modbus, CoAP
R3	Lab	Denmark	Hybrid	High - SSH, MQTT, Modbus, CoAP Low - Telnet, HTTP
C1	Lab	Denmark	Medium	Telnet, SSH, HTTP, Modbus, 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

Dataset

Dataset

- A comprehensive dataset of traffic as *pcaps* and *database dumps*
- 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



https://doi.org/10.11583/DTU.21088651

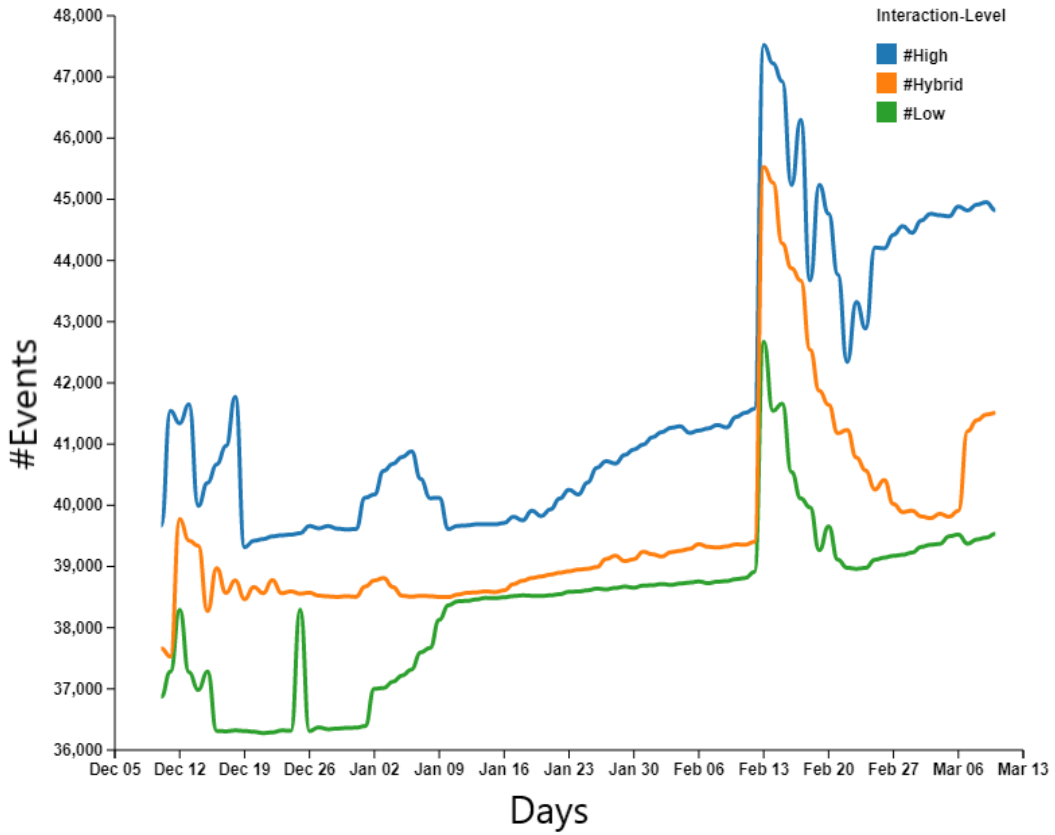
Evaluation and Analysis

Parameter: Interaction-level (Total Events, type)

Interaction-level	Even-type	Count
Low-interaction	Scanning-service	2.02 M
High-interaction	Scanning-service	2.02 M
Hybrid-interaction	Scanning-service	2.02 M
Low-interaction	Malicious	1.46 M
High-interaction	Malicious	1.76 M
Hybrid-interaction	Malicious	1.57 M
Total scanning-services events		**6.07 M**
Total malicious events		**4.8 M**
Total events		**10.87 M**

Table 3: Total events by type and interaction level

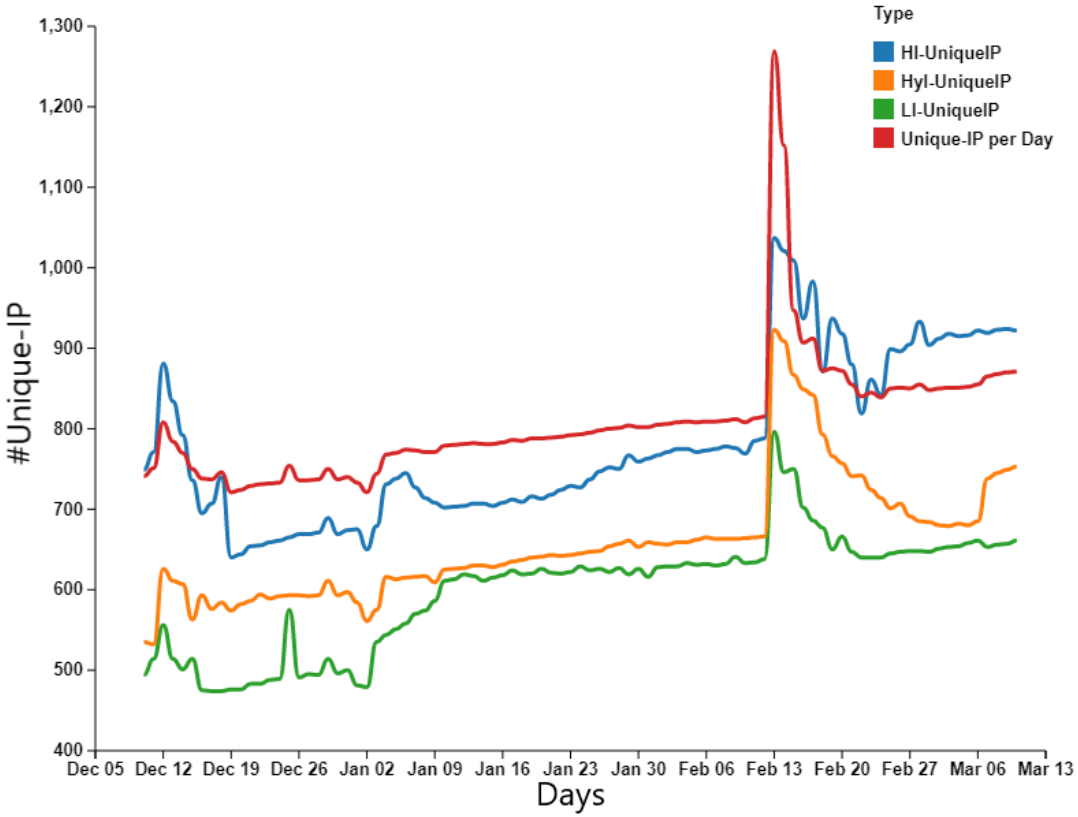
~55%



Parameter: Interaction-level (unique IPs)

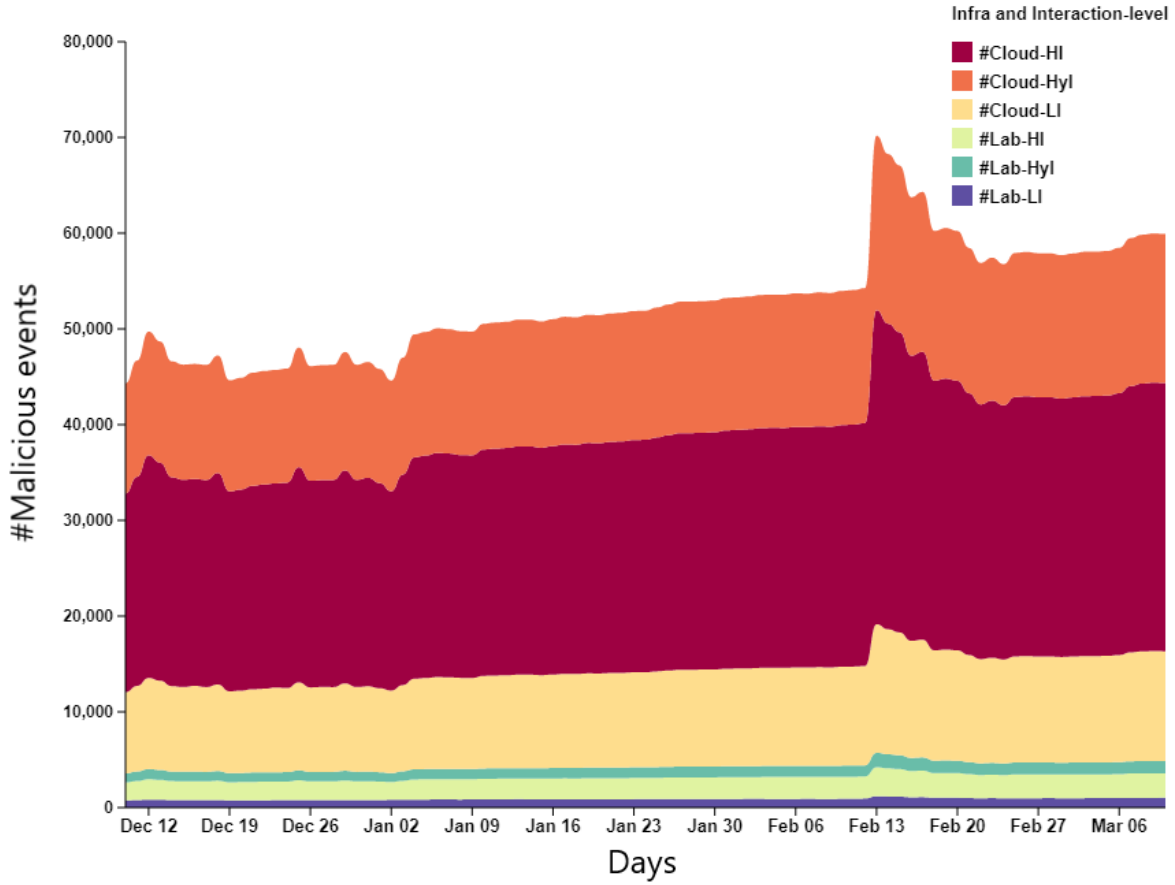
Interaction Level	#Malicious Events	#Unique IPs
High-Interaction	1, 763, 395	18, 431
Hybrid-interaction	1, 575, 807	12, 618
Low-interaction	1, 463, 883	8, 635
Distinct IPs from all interaction levels	22, 518	

Table 4: Summary of malicious events and unique IPs



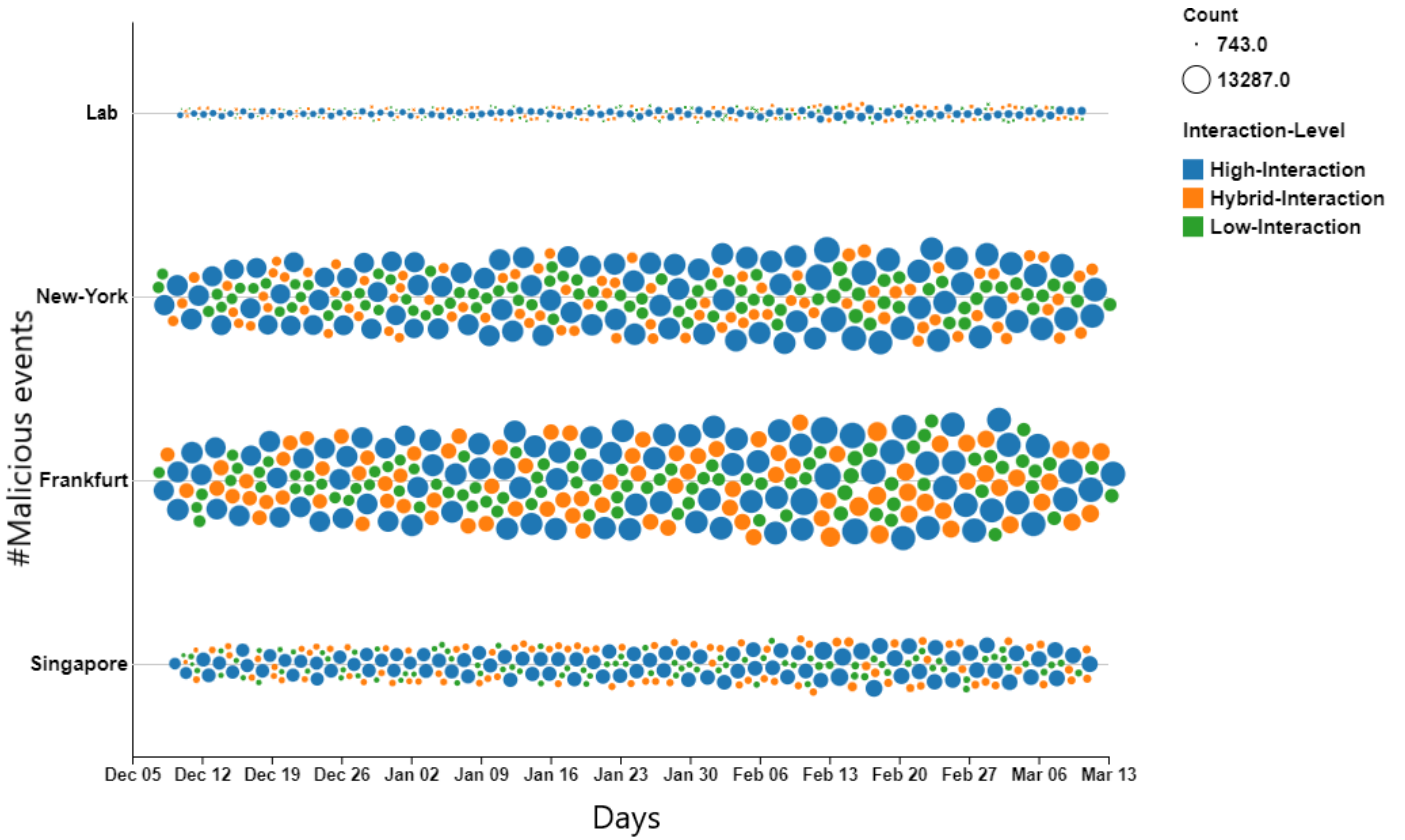
Parameter: Deployment infrastructure

- High-interaction received more attacks than low and hybrid
- Malicious events are seen more in the cloud (more deployments in comparison to the lab)
- Observed minor variations in trend of malicious events in both operating environments



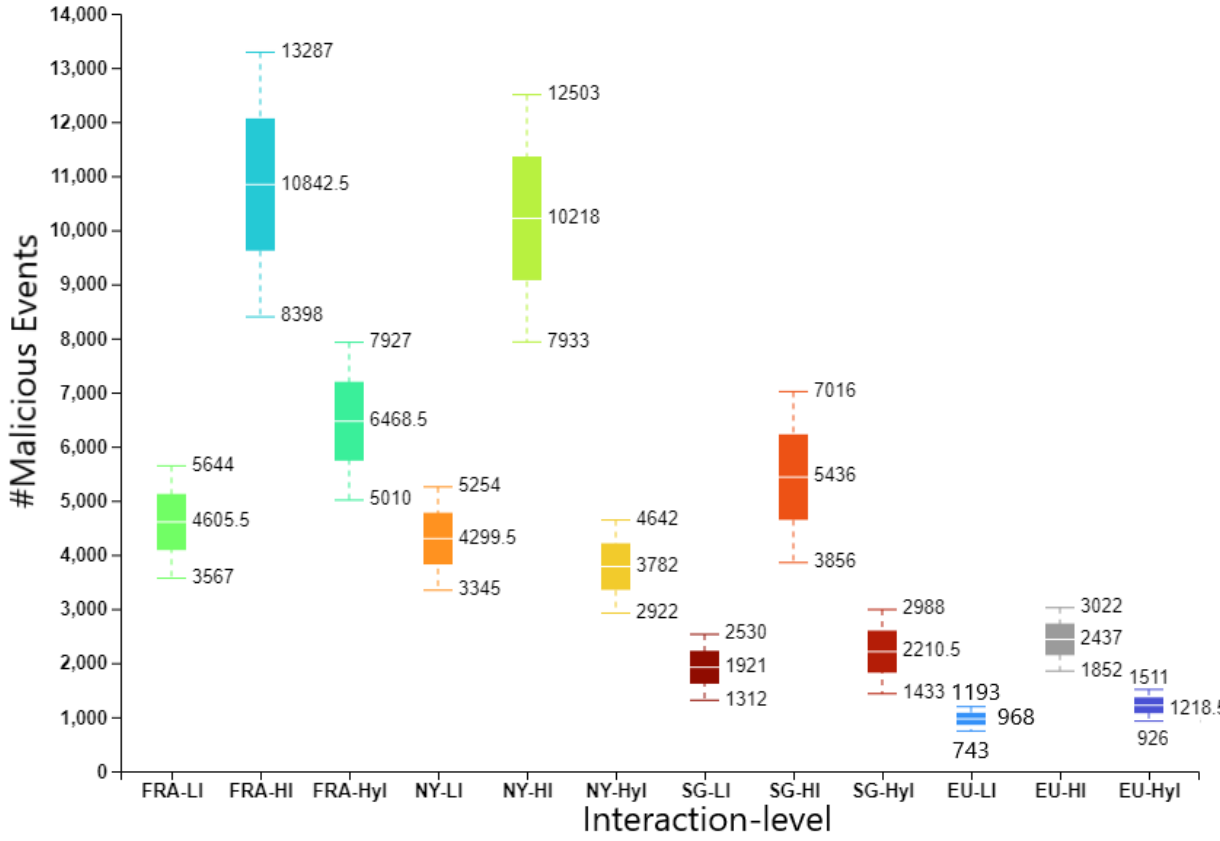
Parameter: Geo-location, city, interaction level, events

- 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



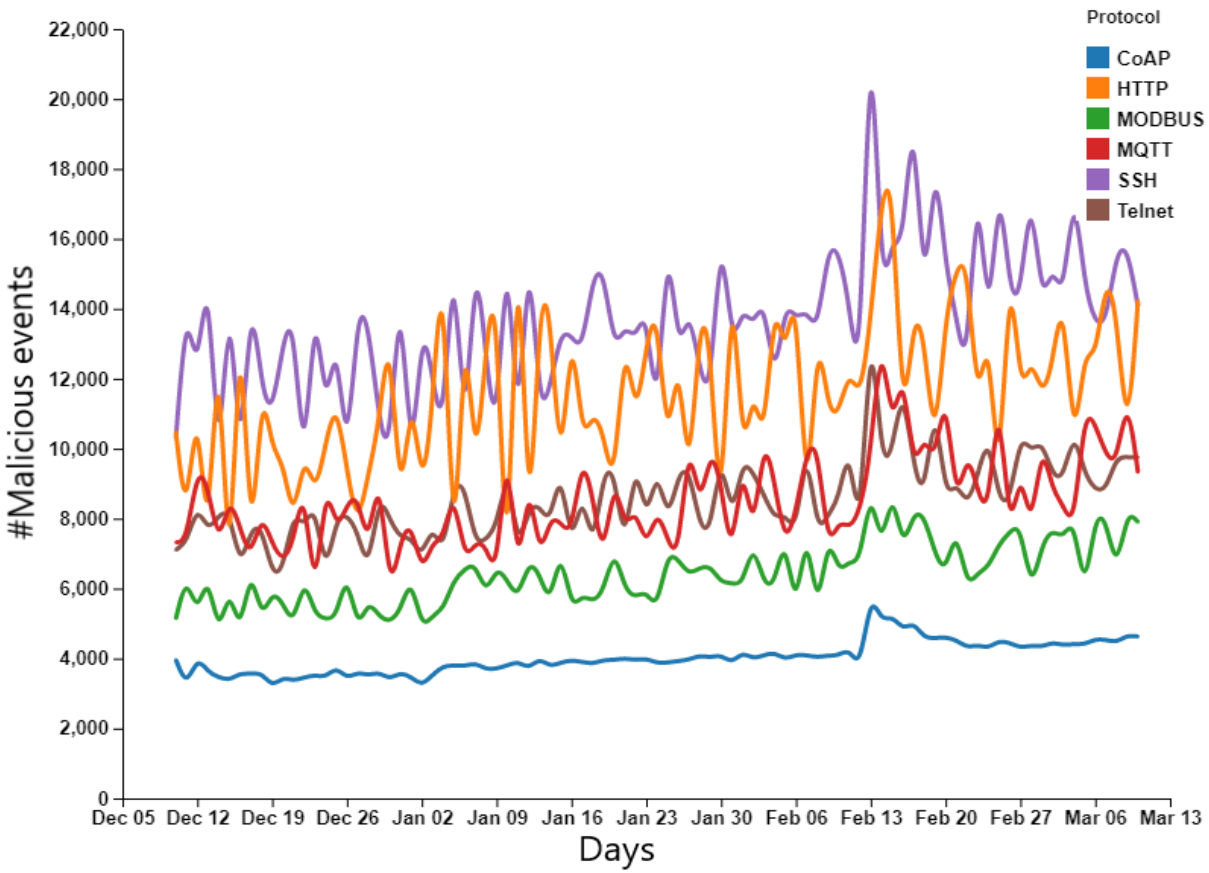
Parameter: Geo-location, lowest-highest, interaction-level

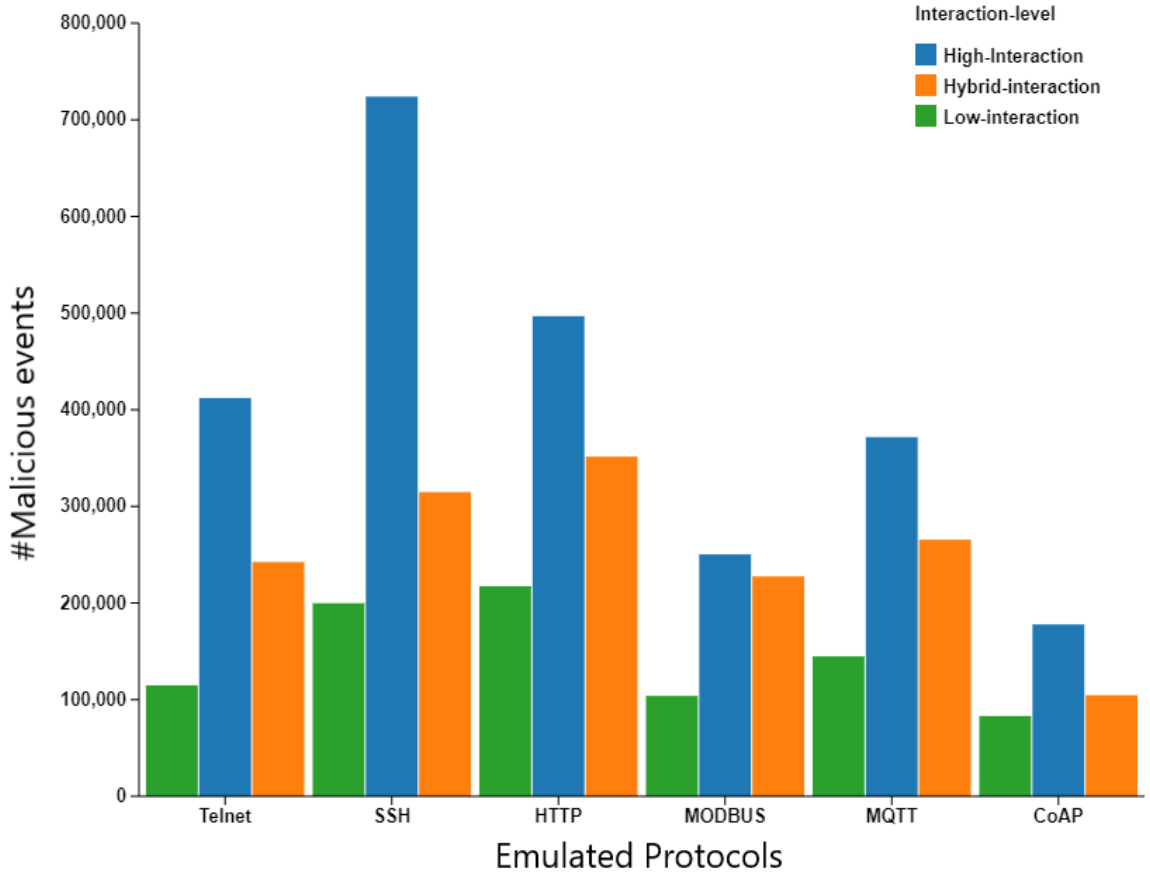
- Highest events recorded in Frankfurt, with High Interaction
- Lowest events recorded in lab deployment, with Low-interaction
- **Regardless, the High-interaction deployments received the highest events**



Parameter: Protocol, events

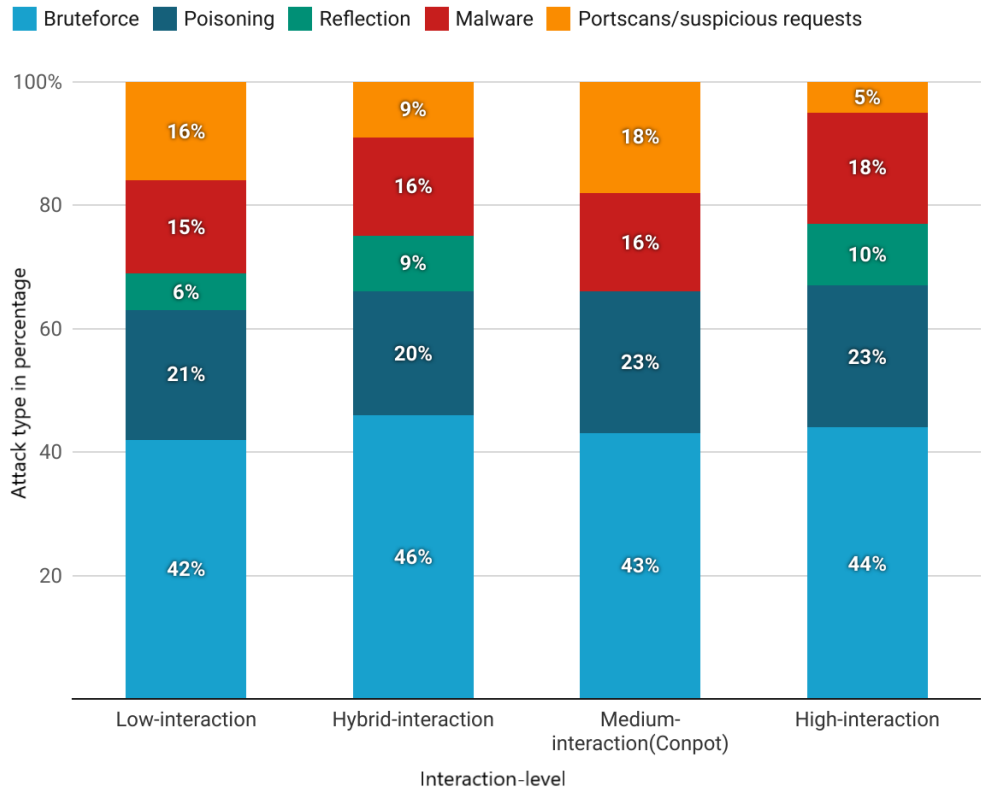
- Highest events on SSH, followed by HTTP, Telnet, MQTT, Modbus and CoAP





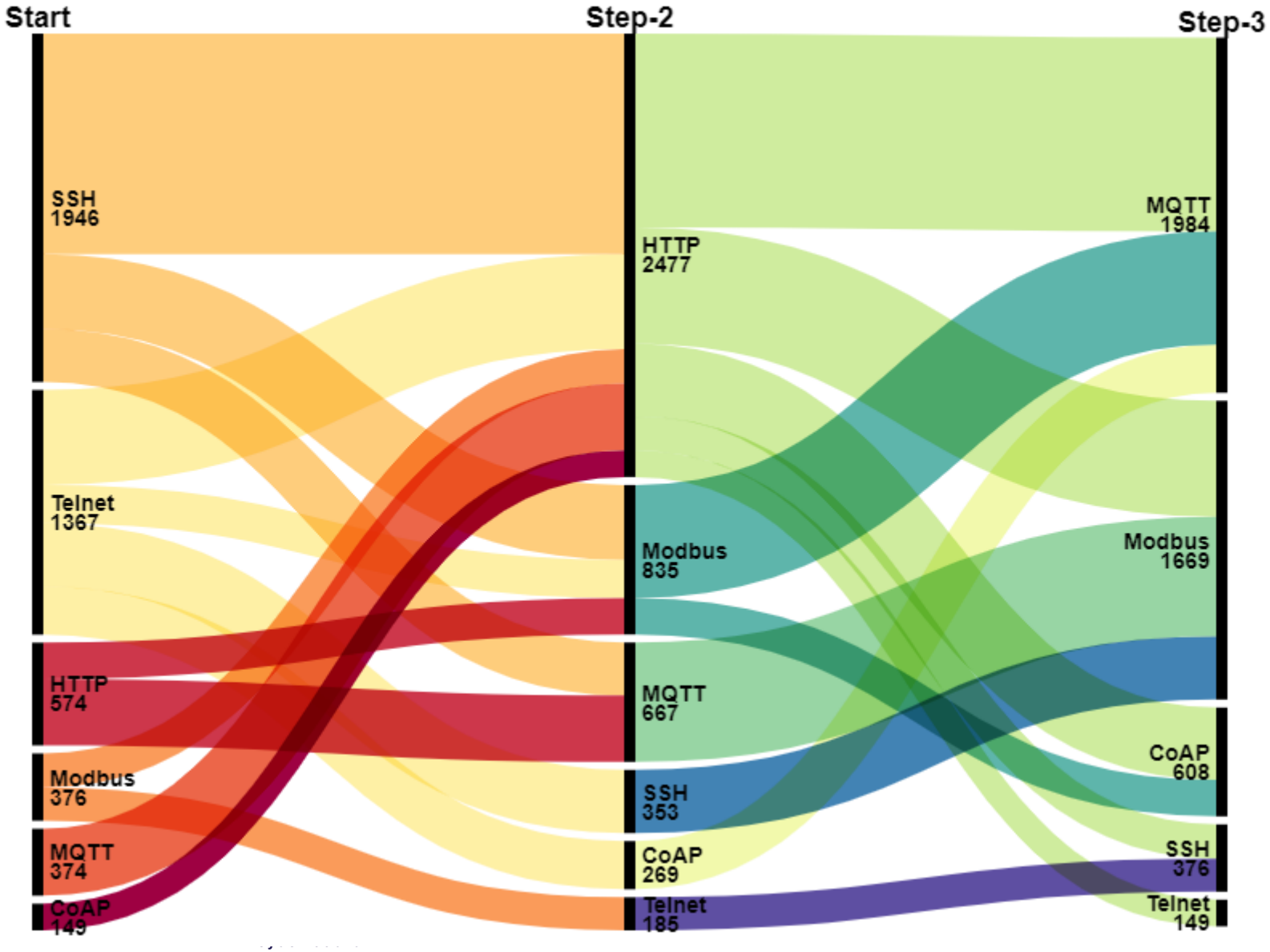
Attack types – by interaction-level

- Diverse attack types observed
- Persistent volume of brute-force attacks observed across all interaction-levels
- *events from known scanning-services are filtered



Multistage attacks

- Total of 4786 attacks across all instances



Attack sources

Device type	Protocol	Count
Router	HTTP	1819
DVR	HTTP	1621
Router	Telnet	721
IP Phone	HTTP	311
Switch	HTTP	287
Switch	Telnet	211
IP Printer	HTTP	176
NAS	HTTP	118
Total		5264

Table 5: Attack-source types

22,518 Unique attack sources

Region-specific attacks

Instance	Region	Attack-type	P…	…que attacker IP	Volume
R1	Denmark(lab)	Brute-force…			7%
R4	New-York	Brute…			11%
R7	Frankfurt	…			14%
R10	Singapore				14%
R5	New-Yor…				17%
R7	Frankfurt				18%
R10	Singapore				12%
R10	Singapore			6	16%

Table 7: Sumr… …gion-specific attack types

More results on the paper…

Limitations

- The honeypot deployed in the lab had an IP address associated with the University Research Network
- Operating honeypots/honeyfarms as a research individual is challenging
- Nation-level CERTS are very efficient in tracking vulnerable systems exposed to the Internet
- Over-counting as a “connection” definition differs on protocols

Summary

- Honeypots are still an effective tool to study attack landscape; if configured carefully
- Carefully configured honeypots (High interaction) can provide with more effective data for Threat Intelligence
- The parameters play an important role in honeypots and honeypot studies
- 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

Future Work

- Statistics of malware identified on specific honeypot types and geo-location
- Longer study
- Study of interesting traffic received during conflict period (beyond scope in this study)

Thank you

Questions?

Contact:
Shreyas Srinivasa
https://sastry17.github.io
Email: shsr@es.aau.dk

Reach out for:
Curated datasets on Internet Scanning, Honeypots, DarkWeb and more…