IEEE CSR

Cyber Security and Resilience

Virtual Conference // 26-28 July, 2021

TRUSTY: A Solution for Threat Hunting Using Data Analysis in Critical Infrastructures

P. Radoglou-Grammatikis, P. Sarigiannidis et al.

ITHACA | University of Western Macedonia

pradoglou@uowm.gr

Under SDN-microSENSE Project

Authors

ITHACA | University of Western Macedonia

https://ithaca.ece.uowm.gr/

Panagiotis Radoglou Grammatikis

Athanasios Liatifis

Panagiotis Sarigiannidis

International Hellenic University

https://www.cs.ihu.gr/

Thomas Lagkas

SIDROCO HOLDINGS LIMITED

https://sidroco.com/

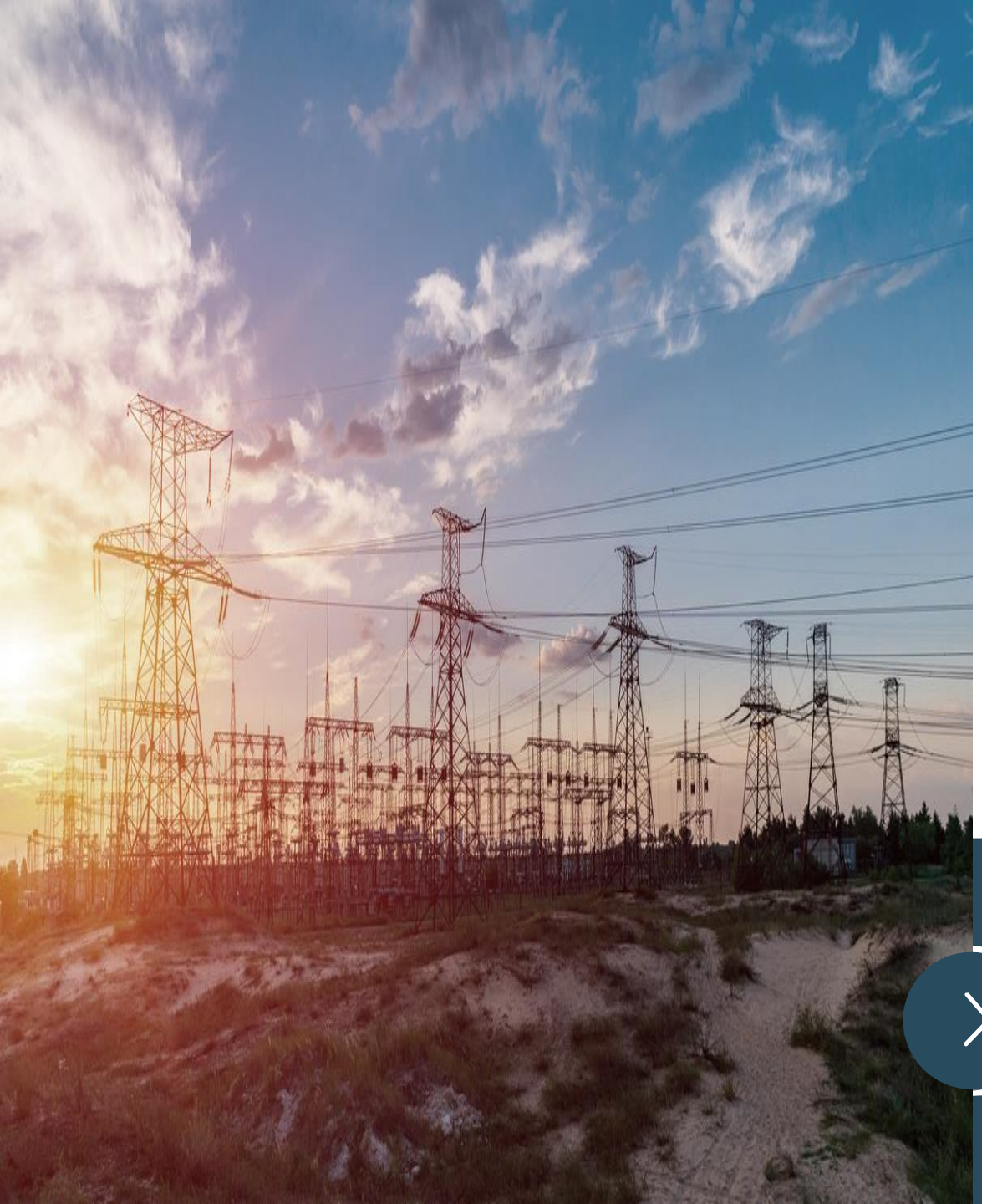
Elisavet Grigoriou

Theocharis Saoulidis

Antonios Sarigiannidis

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833955.

https://www.sdnmicrosense.eu/



Introduction

SDN-µSense

Industrial Internet of Things and Smart EPES

In the era of hyper-connected digital economies, the smart technologies play a vital role in the operation of the electrical grid, transforming it into a new paradigm.

Legacy Systems

The presence of legacy systems, such as ICS/SCADA remains a crucial issue, raising multiple threats and vulnerabilities.

Insecure Communication Protocols

Both smart and legacy EPES assets use insecure communication protocols like Modbus, EtherCAT, IEC 60870-5-104, etc. that do not comprise essential authentication and authorization mechanisms.

Existing Countermeasures

Despite the effectiveness of existing cybersecurity solutions they cannot mitigate coordinated EPES cyberattacks, such as Advanced Persistent Threats (APTs)

Lack of Standardization & Certification Activities

The existing countermeasures are not certified dynamically, ensuring their sufficiency.

TRUSTY

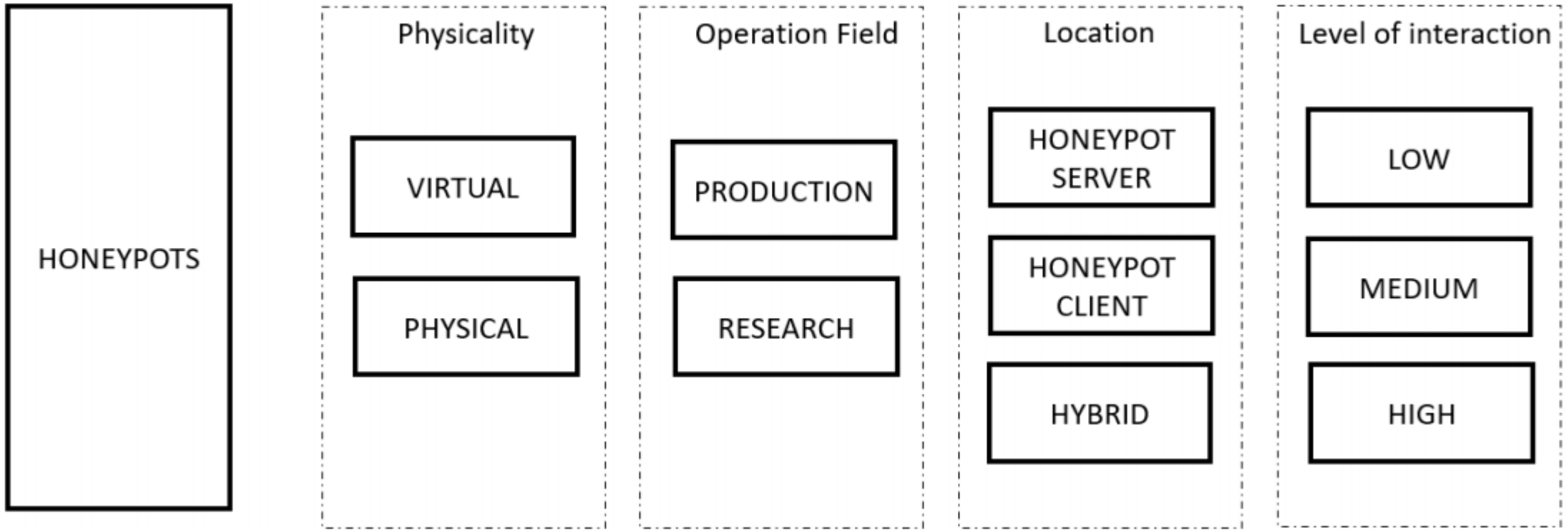
A Solution for Threat Hunting Using Data Analysis in Critical Infrastructures

- Strategic honeypot development
- Honeypot data analysis
- A web-based honeypot analyser
- UOWM Honeypot Dataset

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833955.

Honeypots

Previous works related to detecting intrusions against healthcare ecosystems



Related Work

Previous Works related to Honeypots

R. K. Shrivastava et al.

Attack detection and forensics using honeypot in iot environment - **2019**

S. Gokhale et al.

Industrial control systems honeypot: A formal analysis of conpot. - **2020**

A. Tiwari et al.

Comparitive study of various honeypot tools on the basis of their classification & features - **2020**

H2020 SPEAR

NeuralPot, RTU Honeypot

H2020 SDN-microSENSE

Modbus Honeypot, IEC 60870-5-104 Honeypot and IEC 61850 Honeypot

H2020 ELECTRON

Honeypot as a Service (HaaS)

P. Diamantoulakis et al.

Game theoretic honeypot deployment in smart grid - **2020**

A. Tiwari and D. Kumar

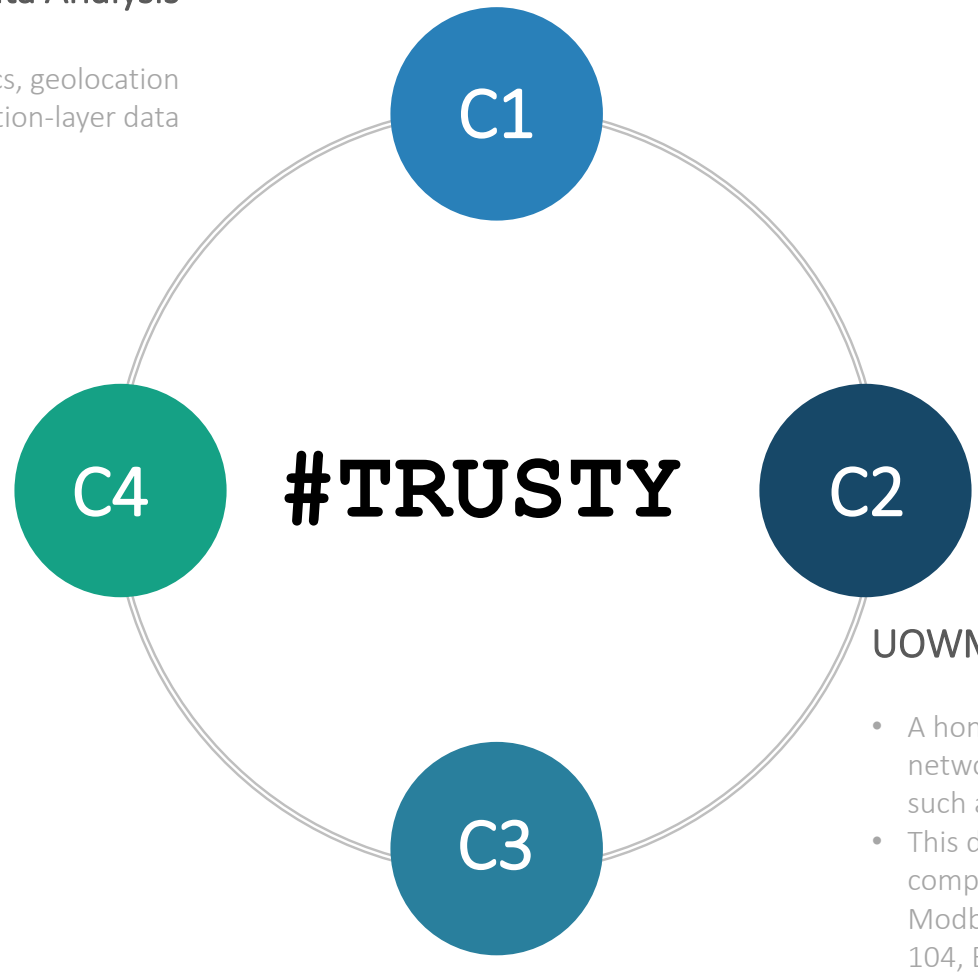
Comparitive study of various honeypot tools on the basis of their classification & features - **2020**

S. Manchekar et al.

Application of honeypot in cloud security: A review - **2018**

Contributions

Four Main Contributions



Honeypot Data Analysis

Network flow statistics, geolocation information, application-layer data

Web-based Honeypot Analyser

We provide a web-based honeypot data analysis platform called TRUSTY, capable of analysing the honeypots' detection results. It provides geolocation information, network layer information (e.g., network flows), application-layer information (e.g., function codes, unit identifiers) and risk estimation per network flow.

Strategic Honeypot Deployment

We model the honeypots' deployment in a CI as a MAB problem, which is solved via TS, considering the costs and the benefits of the defender and the potential cyberattackers.

UOWM Honeypot Dataset

- A honeypot dataset is provided publicly, including the network traffic and logs from multiple industrial honeypots, such as Conpot and Dionaea.
- This dataset can be utilised for intrusion detection processes, comprising network flows statistics related to Modbus/Transmission Control Protocol (TCP), IEC 60870-5-104, BACnet, Message Queuing Telemetry Transport (MQTT) and EtherNet/IP.

IEEE CSR
Cyber Security and Resilience

Strategic Honeypot Development

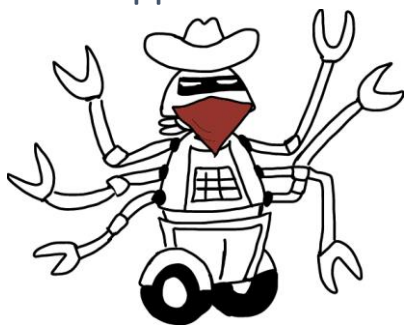
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 833955

Strategic Honeypot Development

TABLE I: Notation

Notation	Explanation
N_{max}	The maximum number of the real EPES assets and honeypots that can be simultaneously connected.
N	The number of the real EPES assets and honeypots that are connected.
a_1	The benefit of the attacker for each attack against a real EPES asset.
a_2	The cost of the attacker for each attack against a honeypot.
a_3	The cost of the attacker for each attack against any machine (honeypot or not).
d_1	The benefit of the defender for each attack against a honeypot.
d_2	The cost of the defender for each attack against a real EPES asset.
d_3	The cost of the defender for each real EPES asset which is replaced by a honeypot.
d_4	The cost of the defender as N increases.
$U_A[t]$	The utility of the $Attacker$ at the time interval t.
$U_D[t]$	The utility of the $Defender$ at the time interval t.
θ	The ratio of N utilised by honeypots.

Multi-Armed Bandit Approach



Attacker & Defender Modeling

Attacker

The goal of attacker is to attack real assets

Defender

Defender aims to deploy the appropriate number of honeypots that will provide the maximum protection, taking into account the available computing resources and the behaviour of the Attacker

Utility of the Attacker

$$U_A[t] = f(a_{i\in\{1,2,3\}}, \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i}, \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i}, \sum_{i=1}^{N} s_{A,i}) \quad (1)$$

$$U_A[t] = a_1 \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i} - a_2 \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i} - a_3 \sum_{i=1}^{N} s_{A,i}) \quad (2)$$

Utility of the Defender

$$U_D[t] = g(d_{i\in\{1,2,34\}}, \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i}, \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i}, \sum_{i=1}^{N} \frac{1+s_{D,i}}{2}, N) \quad (3)$$

$$U_D[t] = d_1 \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i} - d_2 \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i} - d_3 \sum_{i=1}^{N} \frac{1+s_{D,i}}{2} - d_4 N \quad (4)$$

Goal: Maximising Defender's Utility

$$max(U_D[t]) = max(d_1 \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i} - d_2 \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i} - d_3 \sum_{i=1}^{N} \frac{1+s_{D,i}}{2} - d_4 N) \quad (5)$$

Strategic Honeypot Development

Multi-Armed Bandit Approach

- Based on the security events received by the Suricata our goal is to set the appropriate ratio θ in order to maximise UD[t] each time (Equation 4).
- To re-define, the appropriate number of θ for each security event in the time interval t can be expressed as a MAB problem, where exploitation intends to maximise UD[t] (Equation 5) and exploration aims to test different values of θ to discover more information for the Attacker in terms of Equation 4.
- T RUST Y plays the role of the gambler and the various values of theta represent the slot machines. To solve the MAB problem, we adopt the **e – Greedy** method, where we commonly select that mean of theta providing the maximum value UD[t] (Equation 5) and there is a small probability e where other values of θ are selected in order to discover how Equation 4 ranges. The Algorithm reflects how T RUST Y decides to deploy θ honeypots, utilising **e – Greedy**.

Algorithm 1: TRUSTY Honeypot Deployment

Data: N_{max}, N, UD_Matrix, sum_θ_Matrix, $mean\_\theta\_Matrix$, max_mean, securityEventCounter, a_1, a_2, a_3, d_1, d_2, d_3, d_4
Result: $\theta_{selected}$
$size\_\theta\_Matrix = []$, $UD\_Matrix = []$, $sum\_\theta\_Matrix = []$, $mean\_\theta\_Matrix = []$, securityEventCounter = 0, max_mean = 0, $\theta_{selected}$ = 0, a_1, a_2, a_3, d_1, d_2, d_3, d_4 = init();
while *True* **do**
  Receive a security event;
  securityEventCounter = securityEventCounter +1;
  max_mean = 0;
  p = random number in [0,1];
  if $p < e$ **then**
    $\theta_{selected}$ = random integer number in [1, N];
    $UD\_Matrix[\theta] = d_1 \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i} - d_2 \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i} - d_3 \sum_{i=1}^{N} \frac{1+s_{D,i}}{2} - d_4 N$;
    $sum\_\theta\_Matrix[\theta] = sum\_\theta\_Matrix[\theta] + UD\_Matrix[\theta]$;
    $mean\_\theta\_Matrix = sum\_\theta\_Matrix[\theta]$ / securityEventCounter;
  end
  else
    for $\theta \leftarrow 1$ **to** N **by** 1 **do**
      $UD\_Matrix[\theta] = d_1 \sum_{i=1}^{N} \frac{1-S_{D,i}}{2} s_{A,i} - d_2 \sum_{i=1}^{N} \frac{1+S_{D,i}}{2} s_{A,i} - d_3 \sum_{i=1}^{N} \frac{1+s_{D,i}}{2} - d_4 N$;
      $sum\_\theta\_Matrix[\theta] = sum\_\theta\_Matrix[\theta] + UD\_Matrix[\theta]$;
      $mean\_\theta\_Matrix = sum\_\theta\_Matrix[\theta]$ / securityEventCounter;
      if $mean\_\theta\_Matrix[\theta] > max\_mean$ **then**
        $max\_mean = mean\_\theta\_Matrix[\theta]$;
        $\theta_{selected} = \theta$;
      end
    end
  end
end

IEEE CSR
Cyber Security and Resilience

TRUSTY: Web-based Honeypot Data Analyser

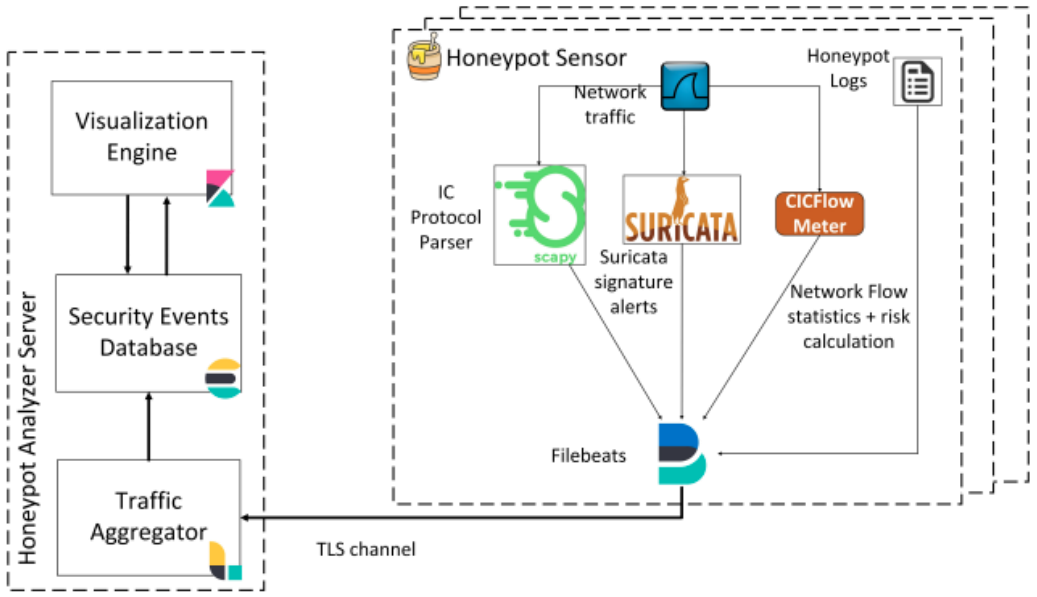
iTHACA

https://ithaca.ece.uowm.gr/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 833955

TRUSTY: Web-based Honeypot Analyser

Two Main Components



Honeypot Sensors

- Virtual machines hostig the honeypot applications (a) Conpot and (b) Dionaea as well as external tools like Tshark, CICFlowMeter, Scapy and Suicata
- Protocols: Modbus, IEC 60870-5-104, BACnet and EtherNet/IP and MQTT
- Risk Estimation per CICFlowMeter Network Flows

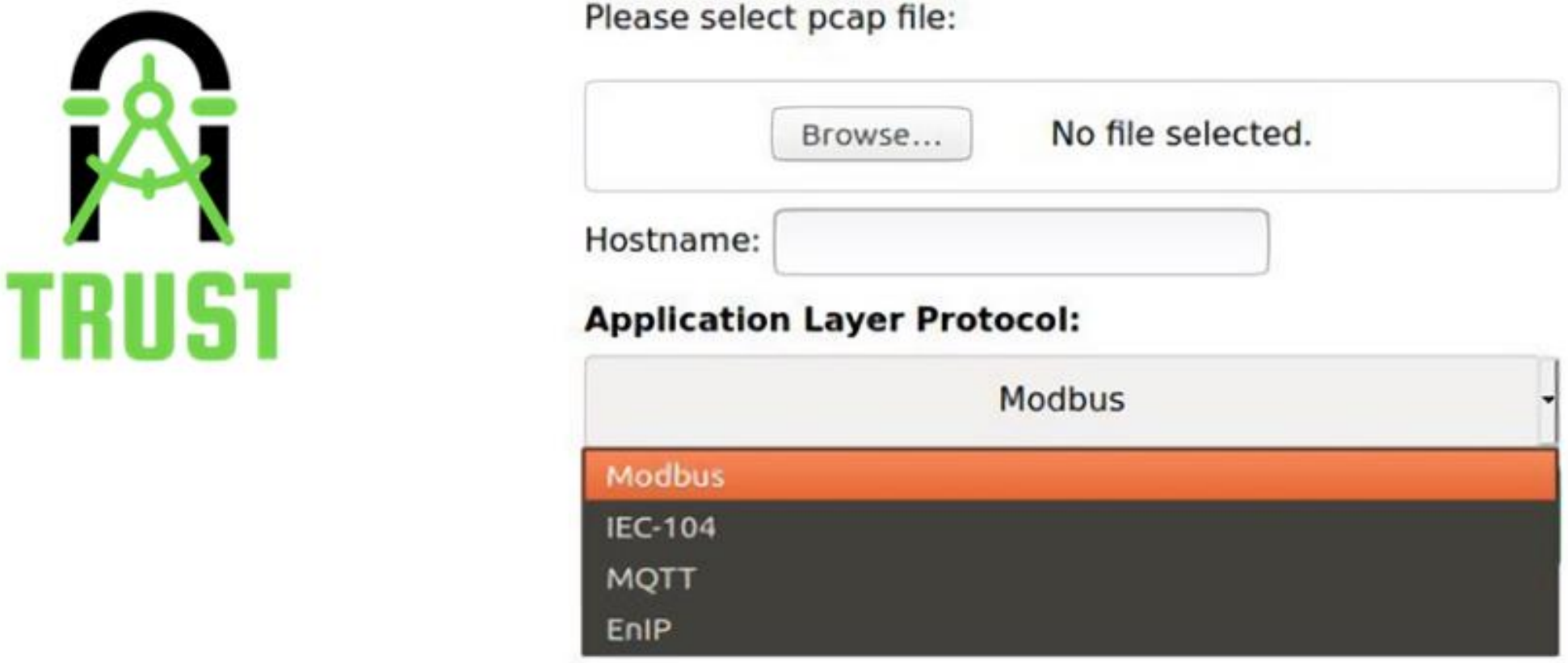
$$Risk = log(nPackets) + log(nBytes) + log(duration + 1) \quad (6)$$

Honeypot Analyser Server

- Three modules: (a) Traffic Aggregator, b) Security Events Database and (c) Visualisation Engine
- The **Traffic Aggregator** relies on **Logstash** and is responsible for receiving data from multiple honeypots applications.
- **Security Events Database** uses **Elasticsearch** in order to store honepot-related security events
- **Visualization Engine** uses **Kibana** and is responsible for presenting the data through interactive visualisations in a web-based environment.

TRUSTY: Web-based Honeypot Analyser

A Web-based application using Kibana



IEEE CSR
Cyber Security and Resilience

Evaluation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 833955

Evaluation

Evaluation Methodology

Honeypot Sensors Deployment

A

`Conpot` and `Dionaea`

Honeypot Data Collection

B

Modbus/TCP, IEC 60870-5-104, MQTT, BACnet and EtherNet/IP

C

Honeypot Data Analysis

`Suricata` and `Shodan`

D

Strategic Honeypot Development Evaluation

a1, a2, a3, d1, d2, d3, d4, e=0.1

E

Evaluation Results

PDF of UD[t]

$$Accuraccy = \frac{TP + TN}{TP + TN + FP + FN}$$

Honeypot Data Analysis

Geolocation, Network Traffic & Malware Information

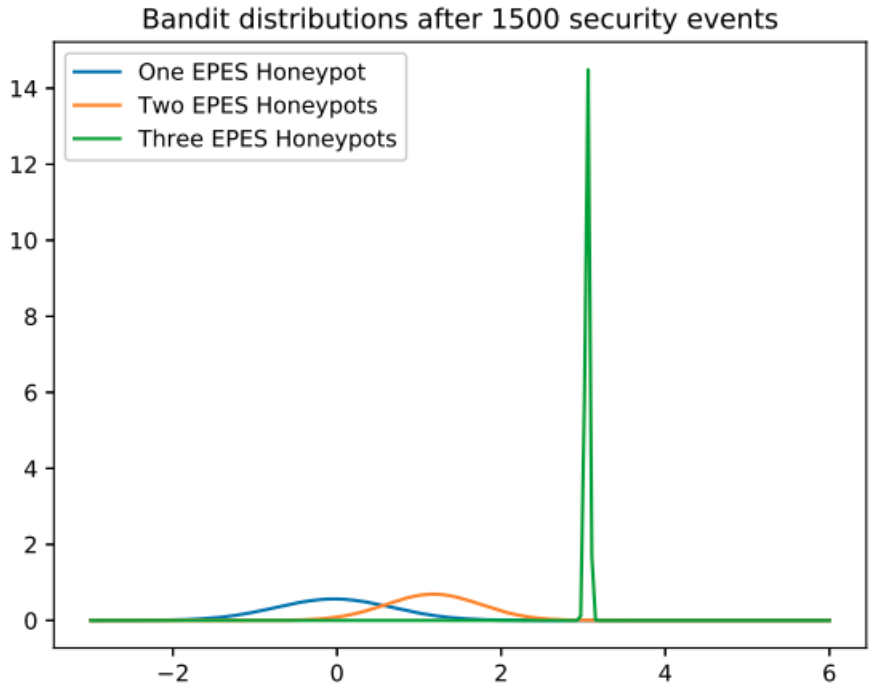
USA and Netherlands are popular choices fordeploying bots. This is also validated by the alerts generated by Suricata on each honeypot sensor. Countries such as Russia and China are ranked third and fourth respectively.

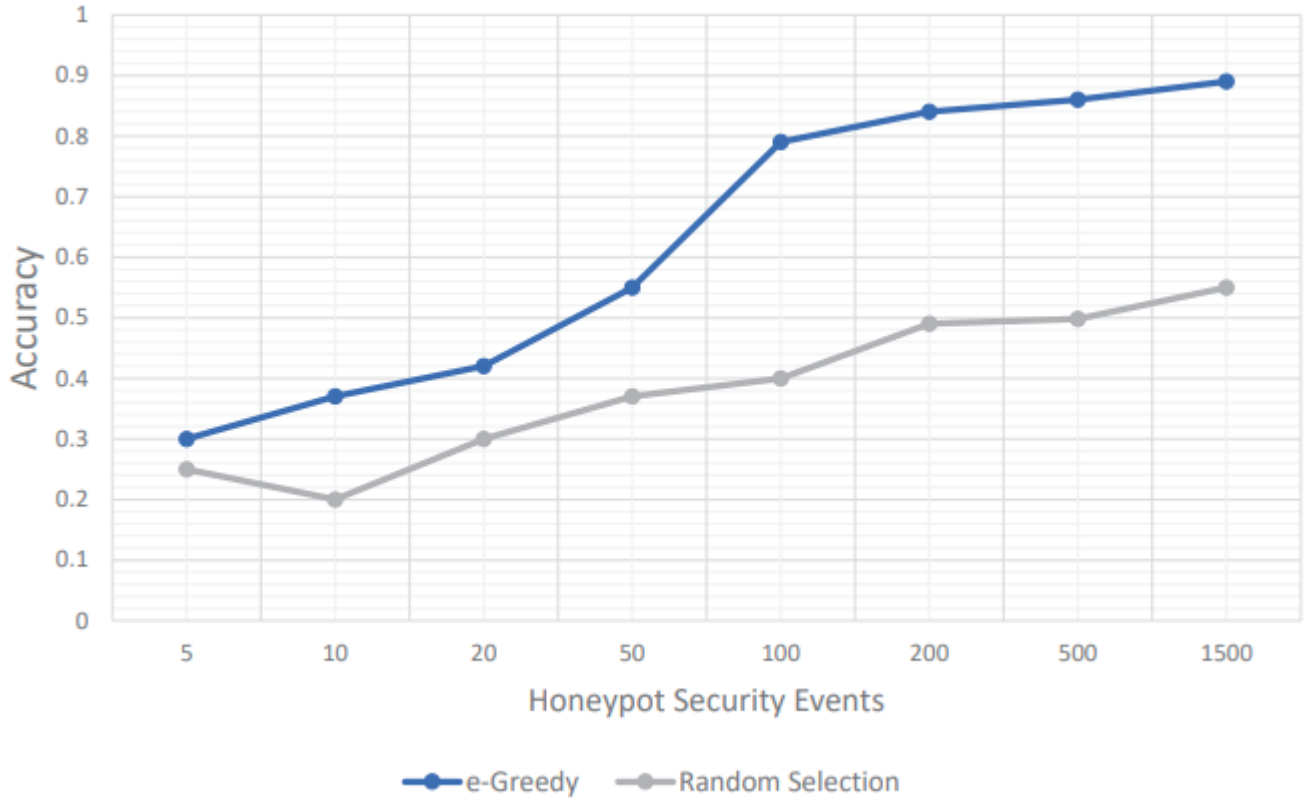
The majority of the network flows was short TCP sessions. This is a strong indication related to reconnaissance cyberattacks

Protocol	Collected Data
MQTT	topic, QOS
Modbus	Function code, Unit ID, length, Type
EnIP	command ID, length, session, status, sender, context, start, ADPU length, testfr_connection, testfr_action,
options IEC104	stopdt_con, stopdt_act, startdt_con, startdt_act, octet_1_1_2 octet_2, octet_3

Honeypot Strategic Deployment

Evaluation Results



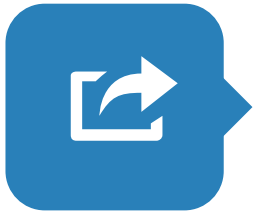


Conclusions

Final Remarks

01

- The rise of IoT offers multiple benefits and raises new cybersecurity risks that require appropriate intrusion detection and prevention mechanisms.
- Honeypots constitute an emerging technology that can trap potential cyberattackers and gather valuable information about their malicious activities.

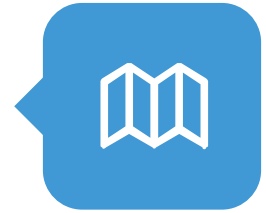


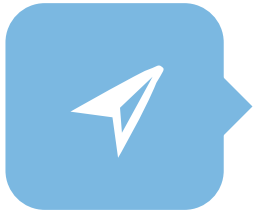
02

Strategic and dynamic way for deploying honeypots in an industrial environment, taking into account the costs and benefits of the defender and the cyberattacker.

03

TRUSTY, a web-based honeypot analyser which comprises industrial honeypot applications and analyses their detection outcomes in terms of network traffic data, network flow statistics and honeypots' logs.





04

- The evaluations results demonstrate the strategic honeypot development
- A dataset with honeypot-related security events was created

Thank You & Q/A

Contact us

pradoglou@uowm.gr

https://pradoglougrammatikis.com/

https://www.linkedin.com/in/panagiotisrg/

https://www.youtube.com/channel/UCw6-d5G01ToBhCmaUnHIcpw

Thank You

Q/A ?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833955.