Technische Hochschule Brandenburg
University of Applied Sciences

An Analysis Framework for Steganographic Network Data in Industrial Control Systems

Tom Neubert, Bjarne Peuker, **Eric Schueler**, **Henning Ullrich**, Laura Buxhoidt, Claus Vielhauer

cyberagentur

Gefördert durch:

Bundesministerium
für Umwelt, Naturschutz, nukleare Sicherheit
und Verbraucherschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

The research in this work has been performed in context of the project ATTRIBUT (https://omen.cs.uni-magdeburg.de/itiamsl/deutsch/projekte/attribut.html) jointly by a teaching project at “Brandenburg University” in term 2023/2024. This comprises in particular the conceptional design of the experimental analysis framework and embedding method EM_3 , software realization in Python of all algorithms for embedding and feature extraction in Section 3. It was further supported by the evaluation dataset (Section IV-B) generously contributed by the project “SYNTHESIS”, funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV, project no. 1501666B) in the framework of the German reactor safety research program.

Agenda

Gefördert durch:
Bundesministerium
für Umwelt, Naturschutz, nukleare Sicherheit
und Verbraucherschutz
aufgrund eines Beschlusses
des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

Agenda

Gefördert durch:

Bundesministerium
für Umwelt, Naturschutz, nukleare Sicherheit
und Verbraucherschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Introduction

- Stealthy malware is increasingly used by attackers [1]
- It uses unobstrusive data to create hidden channels → utilized to embed malicious code or hidden information
- Since the Stuxnet-Attack in 2010, is has been clear that also ICS are under attack with stealthy malware
- Currently, several attack vectors with steganographic embedding methods and potential defense mechanisms are introduced [5],[6],[7]

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Introduction

- To analyze and compare steganographic embedding methods to identify potential similarities, differences and effects on the cover data and to derive defense mechanisms an analysis framework is needed

- A comprehensive analysis could for example enable the possibility to distinguish between analyzed embedding methods after a detection which can lead to the opportunity to identify potential attackers → Attribution

Gefördert durch:
Bundesministerium
für Umwelt, Naturschutz, nukleare Sicherheit
und Verbraucherschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Contribution

- Thus, this work contributes:
  - a novel **analysis framework** for network steganography in ICS and it offers the possibility to:
    - **compare** and **analyze multiple** network steganographic **embedding methods**
    - with only a single uncompromized network traffic capture from an exemplary ICS
  - validation of novel framework and an extensive evaluation of three exemplary selected embedding methods (2 State of the Art, 1 Novel Method) to find out if we can differntiate between embedding methods and embedded types of message (invariant and heterogenous) with machine learning based approach

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Agenda

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Basics in Network Steganography in ICS

- *"Steganography is the art and science of concealing the existence of information transfer and storage"* [8]
- network steganography targets the transfer & storage of hidden information in network communication traffic
- stealthy malware should be inconspicuous in a sense that a warden would not be able to differentiate between genuine communication and communication with hidden information embedding [5]
- In ICS its special, due to lower amount of available data for potential embedding than in traditional IT-networks
- Additionally, transmitted network packets are usually smaller in ICS since only meta-data or few values (e.g., from sensors) are transferred per packet.
- ICS specific protocols like OPC-UA [10] or Modbus-TCP [11] are often encapsulated in TCP/IP
- often transmitted unencrypted, because ICS are considered as closed networks and not subject to attacks.

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

State-of-the-Art

- Synthetic Steganographic Data Generation Concept used to generate steganographic network data from [13]:
    - Offers opportunity for fast and easy generation of data for comparison and analysis with the framework
    - The concept synthesizes only the type and characteristics of a steganographic channel while the rest is taken directly from an uncompromised ICS-setup

- Embedding Method EM_1 [5] & EM_2 [6] *are recent and relevant attack vectors in ICS with timestamp modulations which are analyzed and compared in this work with framework*

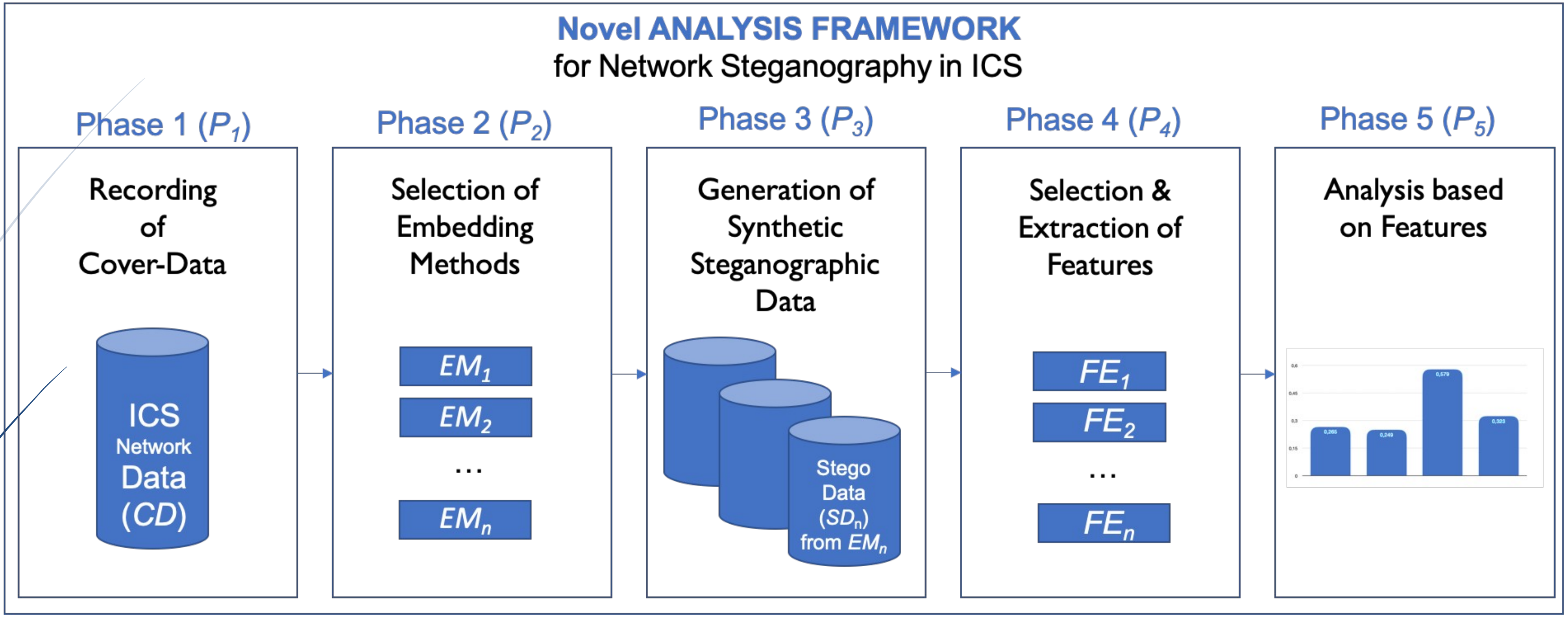
Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Agenda

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

The Analysis Framework

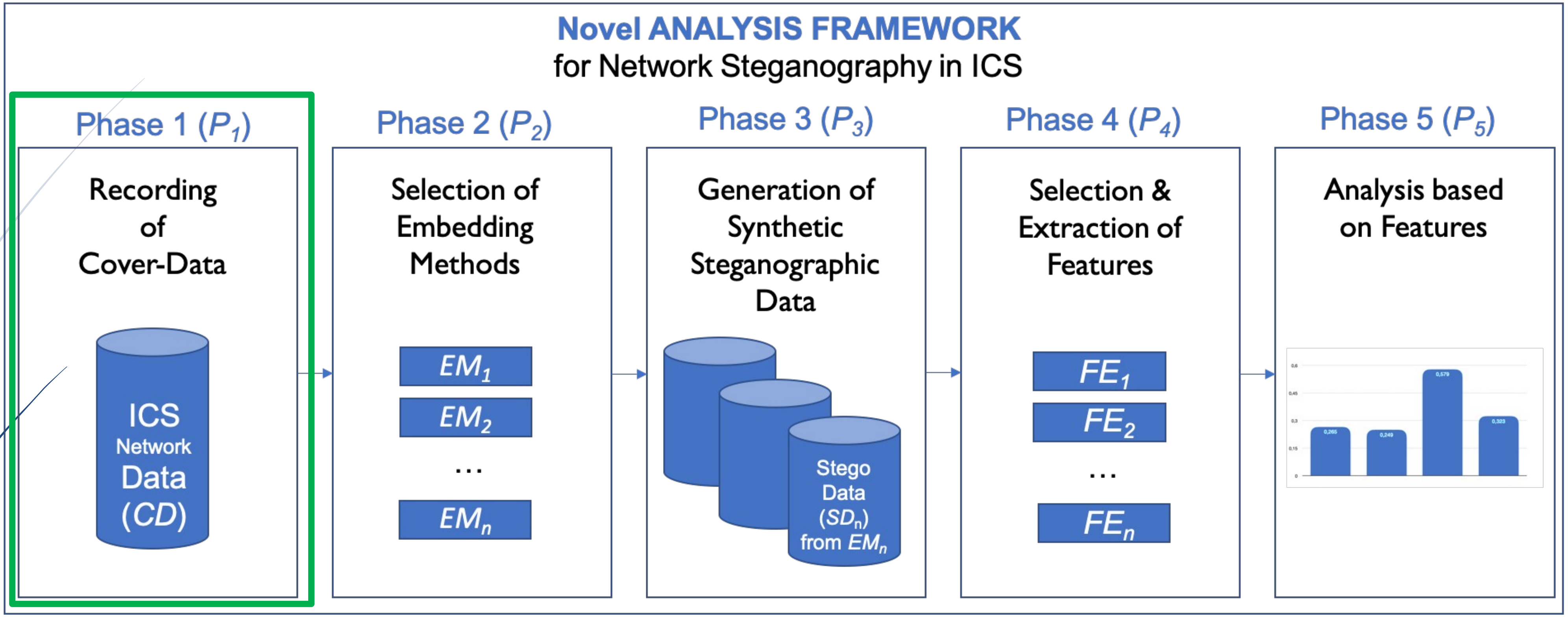


Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

- for **comparison** and **evaluation** of steganographic embedding methods
- to enable the possibility to **distinguish** between methods and to **classify** attackers (Attribution)

Technische Hochschule Brandenburg
University of Applied Sciences

The Analysis Framework

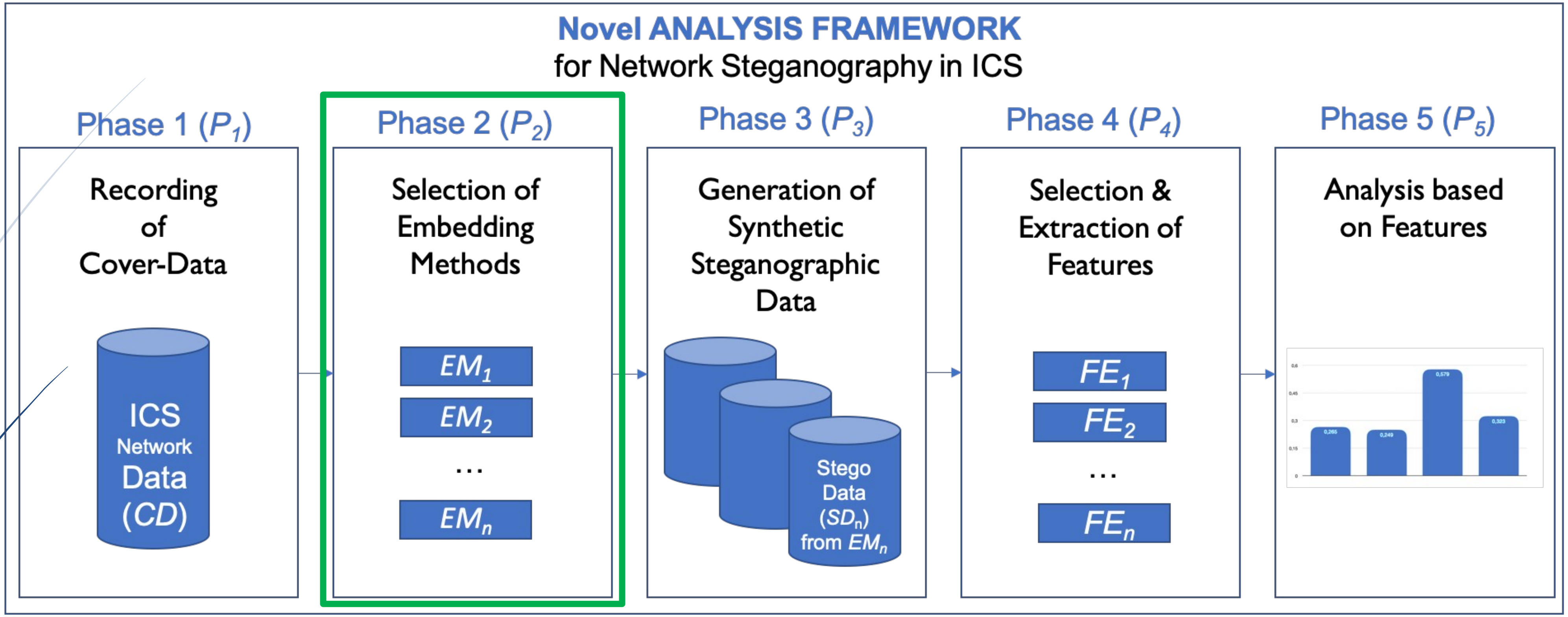


Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

P_1: Recording of Cover-Data:

- **C**over **D**ata (**CD**) has to be recorded from an uncompromized laboratory ICS setup
- *Wireshark* is used, .pcap(ng) file is provided

The Analysis Framework

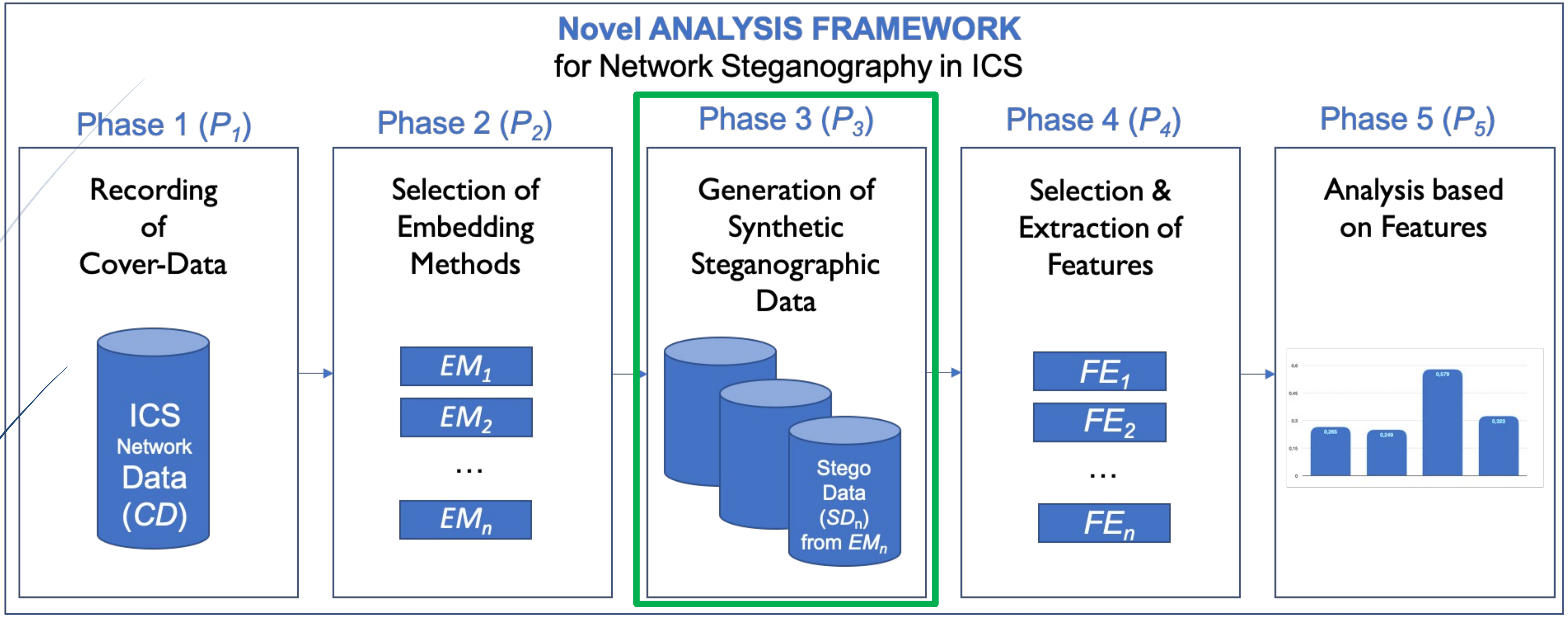


Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

P_2: Selection of Embedding Methods (for Analysis)

- Selection and Formalization of Embedding Methods EM_n (in this work for validation)
  - EM_1 from [5], EM_2 from [6] and novel EM_3 → *all EM are* **Timestamp Modulations**
- Formalization of EM_n in pseudo code representation for better comparison and comprehensibility of methods

The Analysis Framework

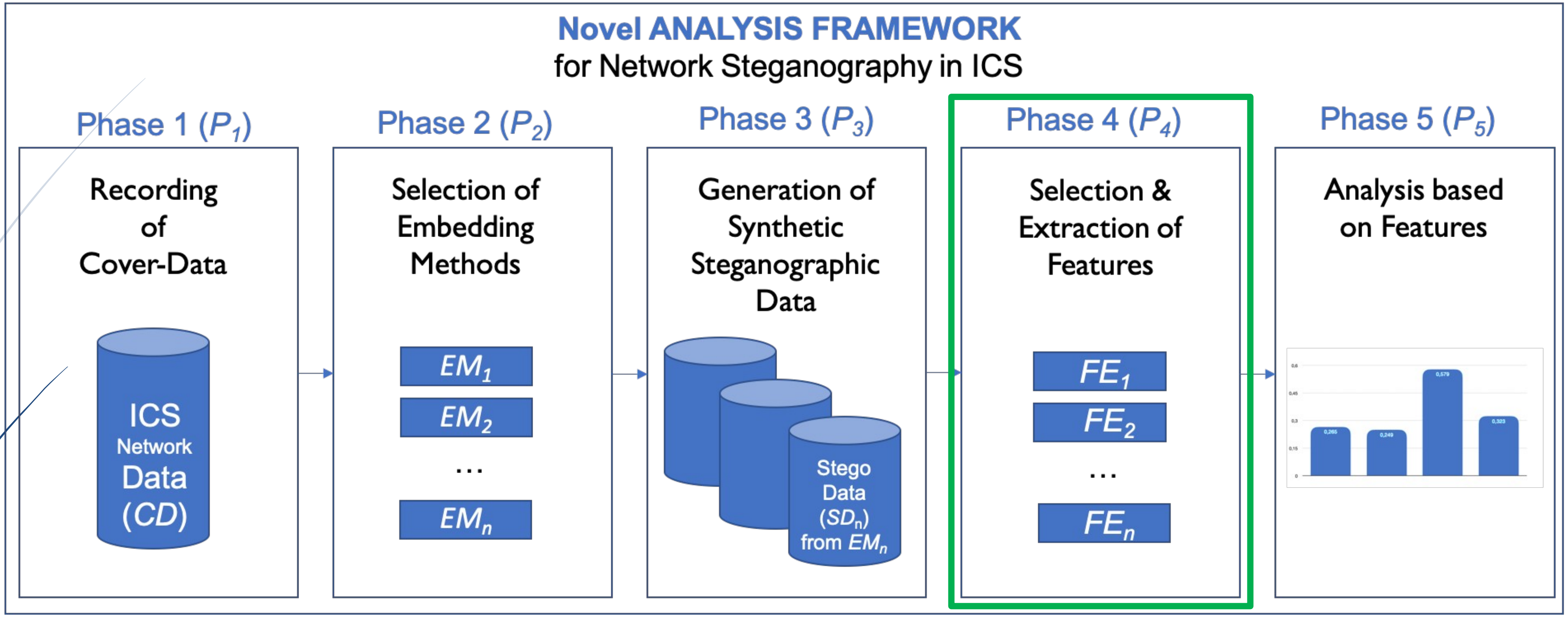


Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

P_3: Generation of Synthetic Steganographic Data (with all EM_n):

- SSE-Concept from [13] is used for easy and fast generation of steganographic data
- No need of physical ICS setup for all embedding methods → very time consuming and complex to elaborate corrupted ICS setup

The Analysis Framework



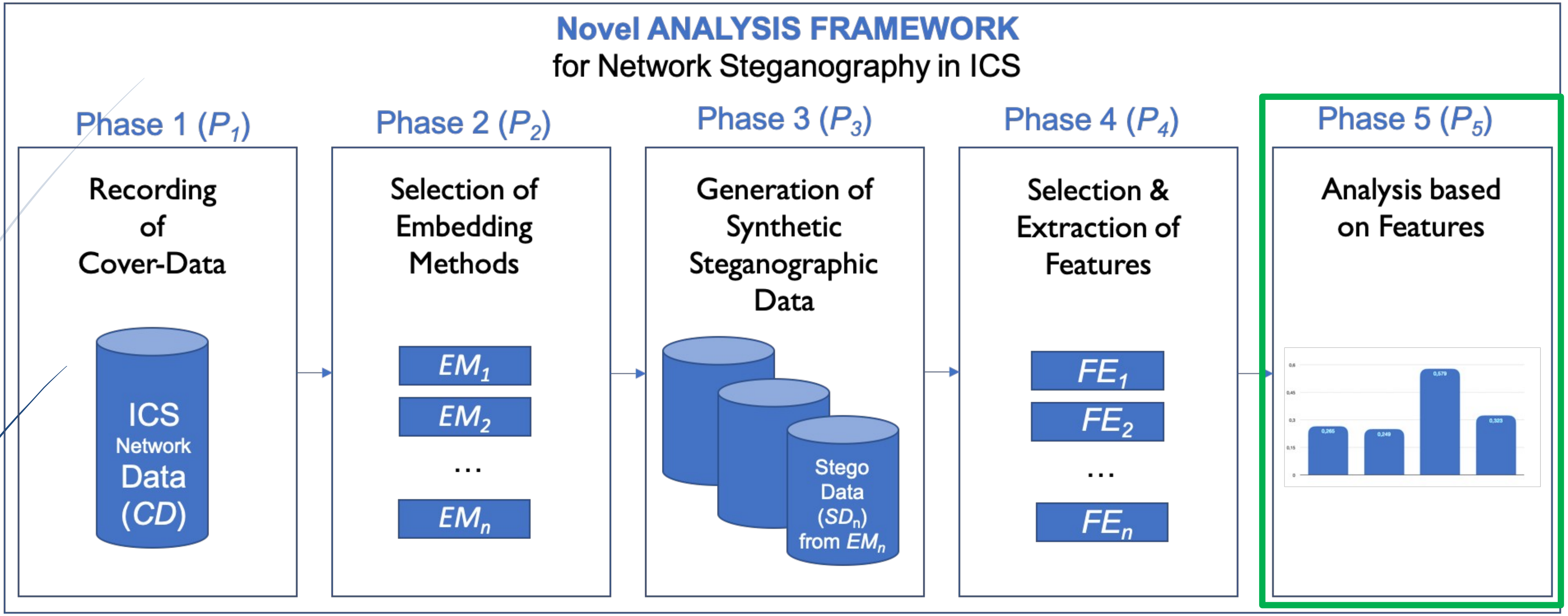
Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

P_4: Selection & Extraction of Features:

- for feature extraction from .pcap recordings, relevant structural elements of network packets should be converted to .csv or .txt (more details in paper)
- handcrafted feature space with discriminatory power should be used for successfull analysis
- we use handcrafted feature space from state-of-the-art [15]

Technische Hochschule
Brandenburg
University of
Applied Sciences

The Analysis Framework



Gefördert durch:
Bundesministerium
für Umwelt, Naturschutz, nukleare Sicherheit
und Verbraucherschutz
aufgrund eines Beschlusses
des Deutschen Bundestages

P_5: Analysis:

- Based on extracted features a statistical analysis can be carried out
- Generally, the analysis can focus different use case specific aspects, for example: detectability, attributability, embedding scheme and more depending on goals and objectives of a study

Technische Hochschule Brandenburg
University of Applied Sciences

Agenda

Gefördert durch:

Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz

aufgrund eines Beschlusses des Deutschen Bundestages

Evaluation Setup - Goals

- In our evaluation, we presented the following **GOALS** for our framework to analyze the introduced embedding methods EM_1, EM_2 and EM_3:
  - **G_1:** Analysis of the three exemplary embedding methods (EM_1, EM_2 & EM_3) based on the extracted features (see paper for brief description) to determine whether a potential **distinction between the methods** is possible for a potential detection of attackers.
  - **G_2:** Analysis of different **message types** (invariant (*IV*) {'*a*'} and heterogeneous (*HE*) {'*securware2024*'}) embedded with EM_1, EM_2 & EM_3 to determine whether a potential distinction between embedded messages is possible.

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

Evaluation Setup - Data

- Uncomprimized laboratory ICS setup with lean server-client-communication
  - Siemens S7-1500 **P**rogrammable **L**ogical **C**ontroller (Server)
  - **H**uman-**M**achine-**I**nterface (Client)
  - Exemplary automation tasks running on PLC (traffic light control, temperature measuring)
  - Packets requested from HMI every 100 ms
  - → *Cover Data* ***REC_{CD}***: 61 Minutes of recording → 38,189 packets (half requests, half responses)

- Attack Scenario:
  - PLC corrupted via Supply-Chain-Attack and sends corrupted packets via timing delay to embed steganographic message (thus only server responses from PLC are relevant packets)
- Steganographic Embedding with EM_1, EM_2 & EM_3 in **REC_{CD}** with synthetic steganographic embedding concept (SSE-concept)

Name	Type of Recording	Embedding Method	Message Type	Hidden Message	No. of relevant Packets
REC_{CD}	Cover Data Recording	-	-	-	19,094
$REC_{EM1_{IV}}$	Steganographic Data	EM_1	invariant	a (repeated)	19,094
$REC_{EM1_{HE}}$	Steganographic Data	EM_1	heterogenous	$securware2024$ (repeated)	19,094
$REC_{EM2_{IV}}$	Steganographic Data	EM_2	invariant	a (repeated)	19,094
$REC_{EM2_{HE}}$	Steganographic Data	EM_2	heterogenous	$securware2024$ (repeated)	19,094
$REC_{EM3_{IV}}$	Steganographic Data	EM_3	invariant	a (repeated)	19,094
$REC_{EM3_{HE}}$	Steganographic Data	EM_3	heterogenous	$securware2024$ (repeated)	19,094

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

Evaluation Setup

- We iterate through every recorded network data set and extract a feature vector after 100 relevant packets, which results in 190 samples per data set
- Used to train machine learning based approach
- For $\mathbf{G_1}$ a Multi Layer Perceptron (MLP_{4C}) with **4-classes** (CD, EM_1, EM_2, EM_3) is trained to identify **Embedding Method** of sample
- For $\mathbf{G_2}$ a Multi Layer Perceptron (MLP_{7C}) with **7-classes** (CD, EM_{1IV}, EM_{2IV}, EM_{3IV}, EM_{1HE}, EM_{2HE}, EM_{3HE}) is trained to identify **Message Type** of sample

Name	Label of Vectors	In MLP_{4C} included vectors: extracted from:	Number of Vectors	Goal
VEC_{CD}	CD	REC_{CD}	190	G_1
VEC_{EM1}	EM_1	$REC_{EM1_{IV}}$, $REC_{EM1_{HE}}$	380 (2x190)	
VEC_{EM2}	EM_2	$REC_{EM2_{IV}}$, $REC_{EM2_{HE}}$	380 (2x190)	
VEC_{EM3}	EM_3	$REC_{EM3_{IV}}$, $REC_{EM3_{HE}}$	380 (2x190)	
		In MLP_{7C} included vectors:		
VEC_{CD}	CD	REC_{CD}	190	G_2
$VEC_{EM1_{IV}}$	$EM1_{IV}$	$REC_{EM1_{IV}}$	190	
$VEC_{EM1_{HE}}$	$EM1_{HE}$	$REC_{EM1_{HE}}$	190	
$VEC_{EM2_{IV}}$	$EM2_{IV}$	$REC_{EM2_{IV}}$	190	
$VEC_{EM2_{HE}}$	$EM2_{HE}$	$REC_{EM2_{HE}}$	190	
$VEC_{EM3_{IV}}$	$EM3_{IV}$	$REC_{EM3_{IV}}$	190	
$VEC_{EM3_{HE}}$	$EM3_{HE}$	$REC_{EM3_{HE}}$	190	

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

Evaluation Setup

- **5-fold Cross Validation** performed to evaluate MLPs and achieve $\mathbf{G_1}$ and $\mathbf{G_2}$

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Agenda

Gefördert durch:

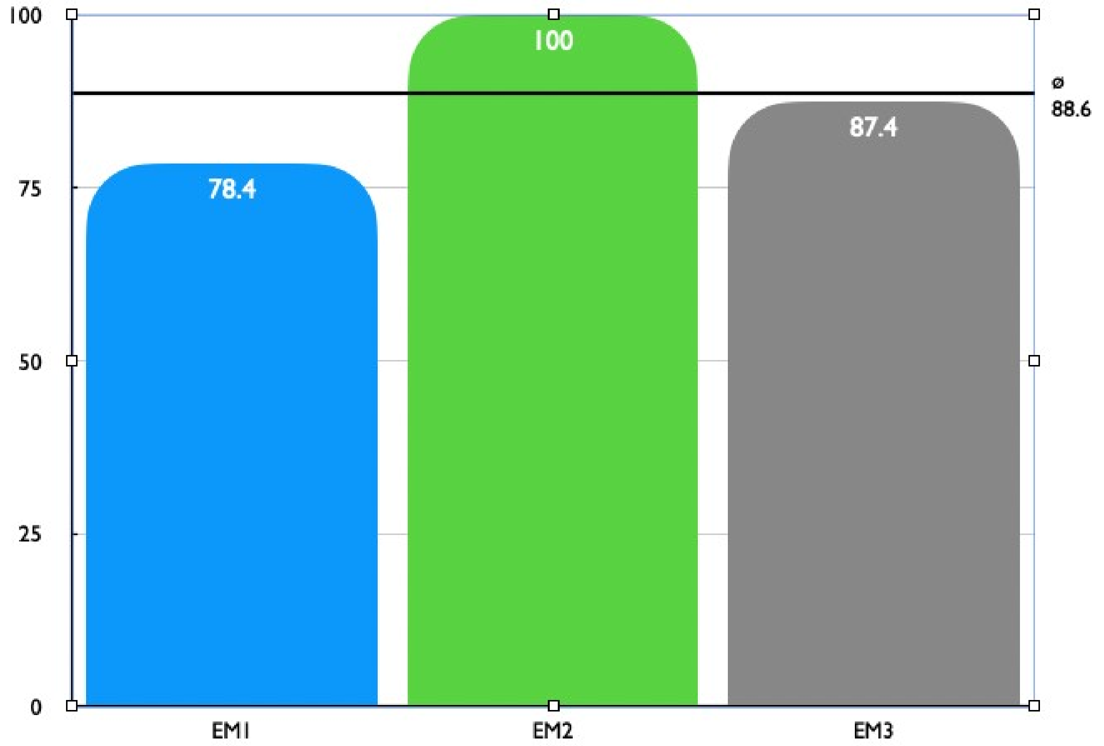
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz

aufgrund eines Beschlusses des Deutschen Bundestages

Evaluation Results - G_1

- MLP_{4C} classifies ~77% of samples correctly
- It can distinguish between Embedding Methods with accuracy of 88.6%
- Challenge: distinction between Cover Data (CD) and EM_1 (due to sophistication of EM_1)

classified as −> Actual	CD	EM_1	EM_2	EM_3
CD (190)	**12**	150	0	28
EM_1 (380)	78	**298**	0	4
EM_2 (380)	0	0	**380**	0
EM_3 (380)	27	21	0	**332**



Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Evaluation Results – G_2

- MLP_{7C} can distinguish between Embedding Methods comparable to MLP_{4C}
- The Message Type can be distinguish for EM_2 with accuracy of **61.3%**
- Challenge: for EM_1 and EM_3 most samples are misclassified due to the embeddings
  - The formalizations of these embeddings show, that the embedded message (type) should not result in statisticially measuable differences with our features

classified as −> / Actual ($\sum$)	CD	$EM1_{IV}$	$EM1_{HE}$	$EM2_{IV}$	$EM2_{HE}$	$EM3_{IV}$	$EM3_{HE}$
CD (190)	**80**	7	8	19	20	39	17
$EM1_{IV}$ (190)	66	**18**	*28*	16	17	31	14
$EM1_{HE}$ (190)	58	*23*	**22**	16	16	38	17
$EM2_{IV}$ (190)	9	0	5	**126**	*35*	15	0
$EM2_{HE}$ (190)	2	0	4	*68*	**107**	9	0
$EM3_{IV}$ (190)	36	2	7	23	26	**62**	*34*
$EM3_{HE}$ (190)	38	1	7	29	22	*69*	**24**

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Agenda

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Summary and Future Work

- **Summary**:
  - Novel Analysis Framework to compare and analyze network stego embedding methods in ICS
  - Exemplary Analysis of 3 *EM*
  - With a MLP as classification engine based on a state-of-the-art feature space we are able to distinguish between 3 embedding methods with an accuracy of 88.3%
  - The classification of embedded message types is challenging for $EM_{1,3}$, but decent for EM_2
- **Future Work:**
  - Analysis of various embedding methods from state-of-the-art with framework
  - Additionally, we would like to analyze the opportunity to differentiate between message types more accurately with for a example a novel handcrafted feature space
  - Can improved features spaces lead to a attribution of attackers with different types of embeddings and message types that are not involved into training

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Thank you for your Attention!

Do you have any Questions?

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Technische Hochschule Brandenburg
University of Applied Sciences

References

[1] - MITRE-ATT&CK, “Data obfuscation: Steganography”, *https://attack.mitre.org/versions/v14/techniques/T1001/002/, last access: 19/09/24*, 2020.

[5] - M. Hildebrandt, K. Lamshoeft, J. Dittmann, T. Neubert, and C. Vielhauer, “Information hiding in industrial control systems: An opc ua based supply chain attack and its detection”, IH&MMSec 2020, pp. 115–120, 2020. DOI: 10.1145/ 3369412.3395068.

[6] - T. Neubert, C. Kraetzer, and C. Vielhauer, “Artificial stegano- graphic network data generation concept and evaluation of de- tection approaches to secure industrial control systems against stegano- graphic attacks”, *In The 16th International Conference on Availability, Relia- bility and Security (ARES 2021), August 17–20, 2021, Vienna, Austria. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3465481.3470073*, 2021.

[7] - K. Lamshoeft, T. Neubert, J. Hielscher, C. Vielhauer, and J. Dittmann, “Knock, knock, log: Threat analysis, detection & mitigation of covert channels in syslog using port scans as cover”, *Digital Investigation 2022 (DFRWS EU 2022)*, 2022.

[8] - S. Wendzel *et al.*, “A generic taxonomy for steganography methods”, Jul. 2022. DOI: 10.36227/techrxiv.20215373.v1.

[13] - T. Neubert, B. Peuker, L. Buxhoidt, E. Schueler, and C. Vielhauer, “Synthetic embedding of hidden information in industrial control system network protocols for evaluation of steganographic malware”, *Tech. Report, arXiv, https://doi.org/ 10.48550/arXiv.2406.19338*, 2024.

[15] - T. Neubert, A. J. C. Morcillo, and C. Vielhauer, “Improving performance of machine learning based detection of network steganography in industrial control systems.”, *In the Proceed- ings of 17th International Conference on Availability, Reliability and Security (ARES 2022), Article No.: 51, pp. 1 - 8, August 23– 26, 2022, Vienna, Austria. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3538969.3544427*, 2022.

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Appendix

Algorithm 1 Steganographic Embedding Method EM_1

$AM \leftarrow A$
$i \leftarrow 0$
$K \leftarrow 4\ Digit\ Key$
$I \leftarrow 4\ Digit\ Initialization\ Vector$
while $i < Length(A)$ **do**
  $D \leftarrow Hour\ value\ of\ T_i$
  $E \leftarrow Minute\ value\ of\ T_i$
  $F \leftarrow Second\ value\ of\ T_i$
  $G \leftarrow Value\ of\ digit\ 1\ after\ floating\ point\ of\ T_i$
  $H \leftarrow Value\ of\ digit\ 2{-}6\ after\ floating\ point\ of\ T_i$
  $S \leftarrow G \oplus DigitSum(K)\ mod\ 2$
  $O \leftarrow D \times E \times F\ mod\ 10000$
  $K' \leftarrow \sum_{n=0}^{3}((K_n \oplus (G + I_n))\ mod\ 10) \times 10^n$
  $K'' \leftarrow O \oplus K'\ mod\ 10000$
  $c \leftarrow m \oplus K''\ mod\ 8192$
  if $S == 0$ **then**
    $H_0, H_1, ..., H_3 \leftarrow c$
  else if $S == 1$ **then**
    $H_1, H_2, ..., H_4 \leftarrow c$
  end if
  $AM[i] \leftarrow T_i$
end while

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Appendix

Algorithm 2 Steganographic Embedding Method EM_2

$AM \leftarrow A$
for Bit in $Bitstream$ **do**
  for $i \leftarrow 1$ to 3 **do**
    if Bit_i is 0 **then**
      $T_i[\mu_{i \ mod \ 3}] \leftarrow 4$
    else if Bit_i is 1 **then**
      $T_i[\mu_{i \ mod \ 3}] \leftarrow 9$
    end if
    $AM[i] \leftarrow T_i$
  end for
end for

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages

Appendix

Algorithm 3 Steganographic Embedding Method EM_3

$AM \leftarrow A$
$i \leftarrow 0$
$K \leftarrow$ "$SyntheticStegoKey$"
for Bit in $Bitstream$ **do**
  for $i \leftarrow 1$ to 3 **do**
    $C_0 \leftarrow 0$
    $C_1 \leftarrow 0$
    while $C_1 == C_2$ **do**
      $C_0 \leftarrow Random(K)\ mod\ 9$
      $C_1 \leftarrow Random(K)\ mod\ 9$
    end while
    $j \leftarrow C_0 + C_1\ mod\ 3$
    if Bit_i is 0 **then**
      $T_i[\mu_j] \leftarrow C_0$
    else if Bit_i is 1 **then**
      $T_i[\mu_j] \leftarrow C_1$
    end if
    $AM[i] \leftarrow T_i$
  end for
end for

Gefördert durch:
Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
aufgrund eines Beschlusses des Deutschen Bundestages