Distinguishing Tor From Other Encrypted Network Traffic Through Character Analysis

Pitpimon Choorod   Tobias J. Bauer   Andreas Aßmuth

15th April 2024

Special Track Along With Cloud Computing 2024 – Venice, Italy

KMUTNB
Fraunhofer AISEC
Ostbayerische Technische Hochschule Amberg-Weiden
IARIA

Authors

Pitpimon Choorod

King Mongkut's University of Technology North Bangkok, Prachinburi, Thailand

pitpimon.c@itm.kmutnb.ac.th

Tobias J. Bauer

Fraunhofer Institute for Applied and Integrated Security, Weiden, Germany

tobias.bauer@aisec.fraunhofer.de

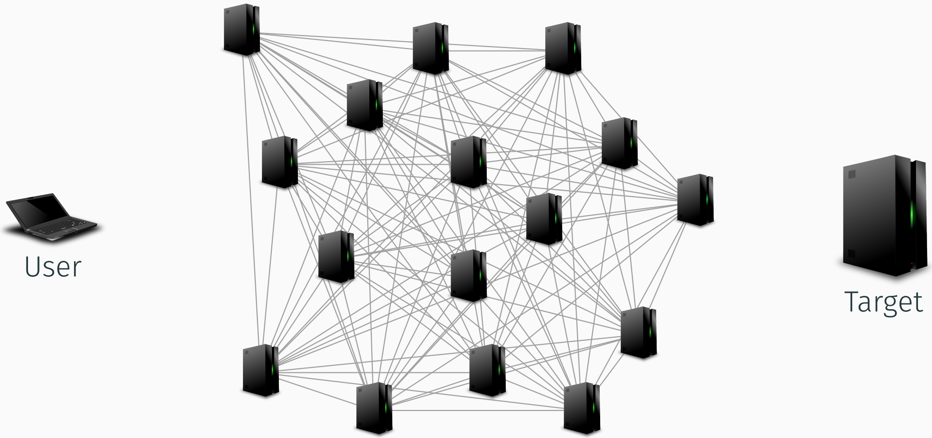
Andreas Aßmuth

Ostbayerische Technische Hochschule Amberg-Weiden, Amberg, Germany

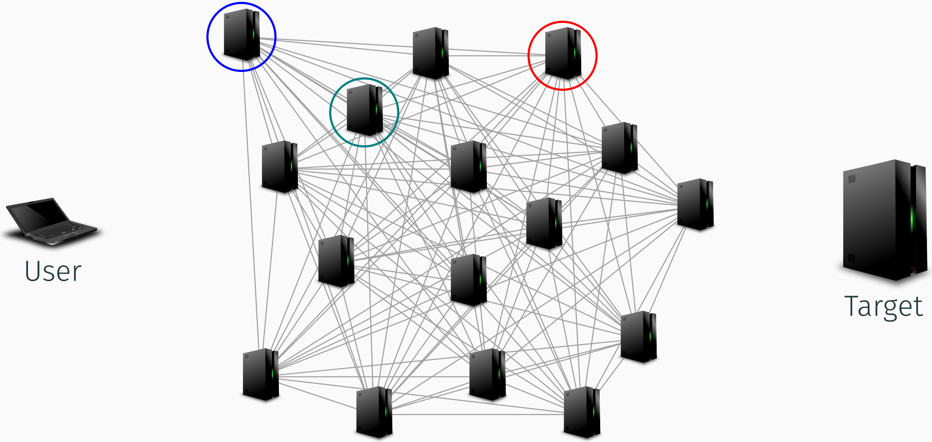
a.assmuth@oth-aw.de

Outline

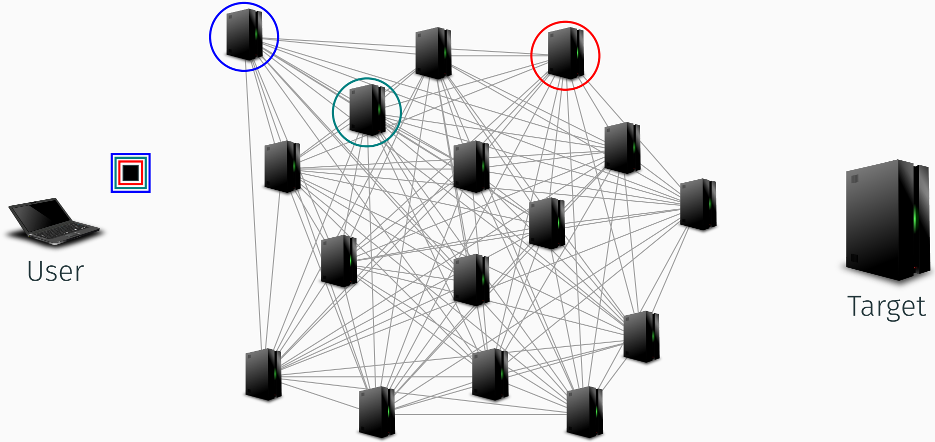
Anonymisation Using The Tor Network



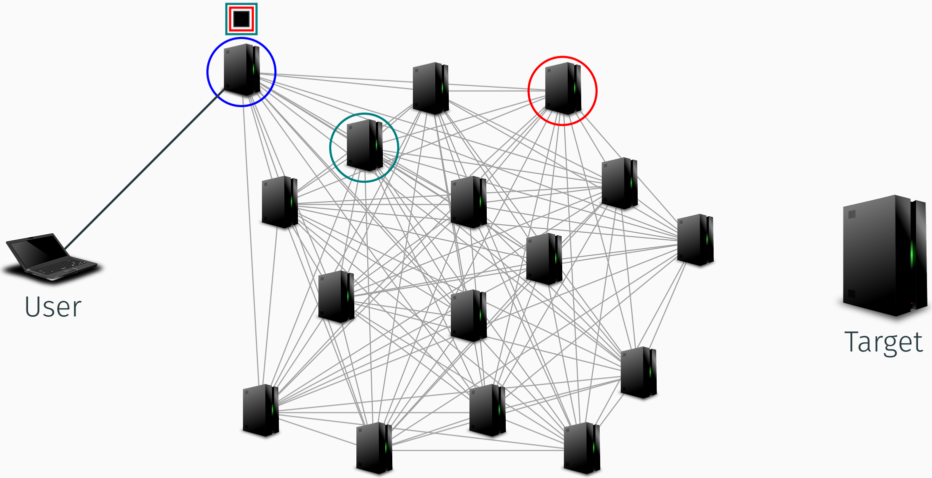
Anonymisation Using The Tor Network



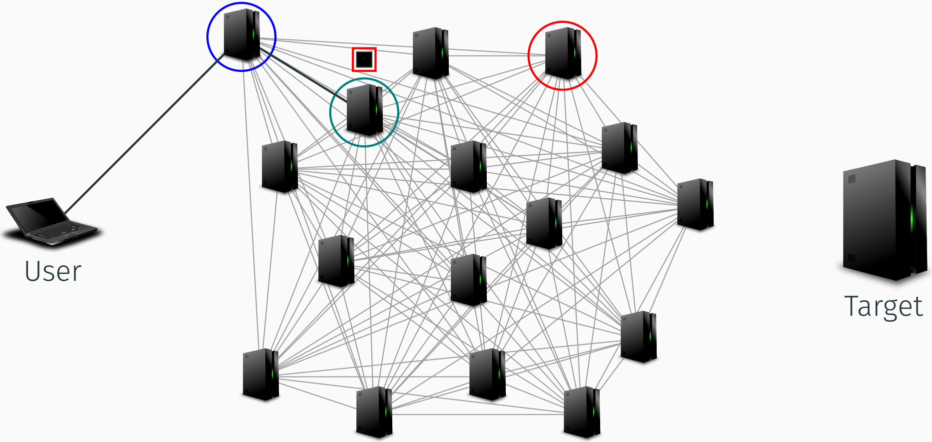
Anonymisation Using The Tor Network



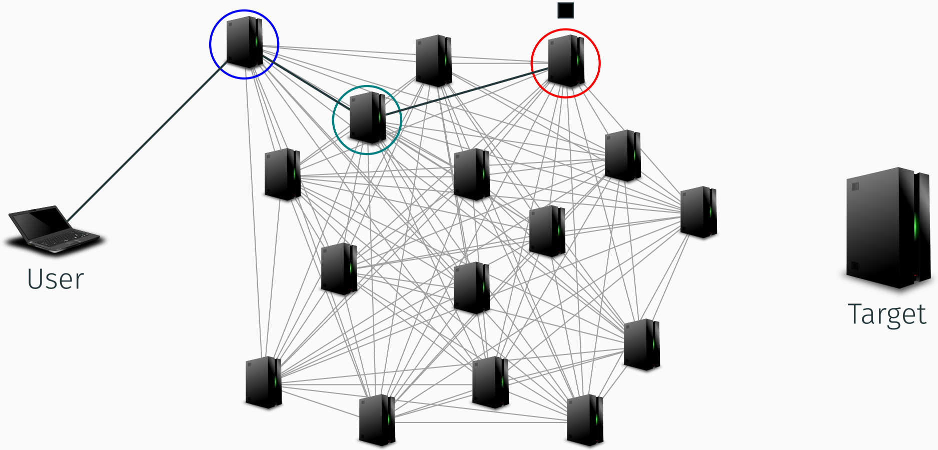
Anonymisation Using The Tor Network



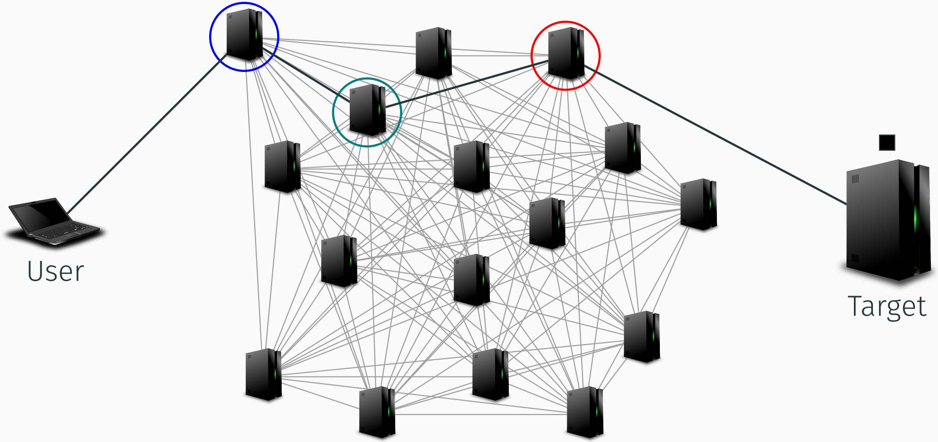
Anonymisation Using The Tor Network



Anonymisation Using The Tor Network

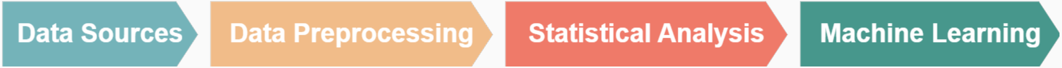


Anonymisation Using The Tor Network



Results From Pitpimon's PhD Thesis

Methodology



1. **Data Sources:** Utilized two distinct datasets:
   a) A **public Tor dataset** from UNB-CIC, encompassing eight application types: audio, browsing, chat, email, FTP, P2P, video, and VoIP.
   b) A **private dataset** focusing on browsing applications.
2. **Data Preprocessing:** Clean and prepare the raw data for analysis.
3. **Statistical Analysis:** Conducted using the Mann-Whitney U Test to identify significant differences and patterns.
4. **Machine Learning Models:** Implemented and evaluated three models—J48, Random Forest, and KNN.

Results From Pitpimon’s PhD Thesis

Analysis of Tor vs. Non-Tor Traffic

Table 1: Number of balanced Tor and non-Tor instances for nine applications

Audio	26,082	**Email**	12,300	**Video**	32,154
Browsing	71,950	**FTP**	514,952	**VoIP**	737,382
Chat	6,504	**P2P**	433,770	**Private**	29,600

Statistical Analysis

- The **Mann-Whitney U test** showed significant differences in traffic, with differentiation rates of ***95.42%*** for the public dataset and ***100%*** for the private dataset.

Results From Pitpimon's PhD Thesis

Machine Learning Analysis

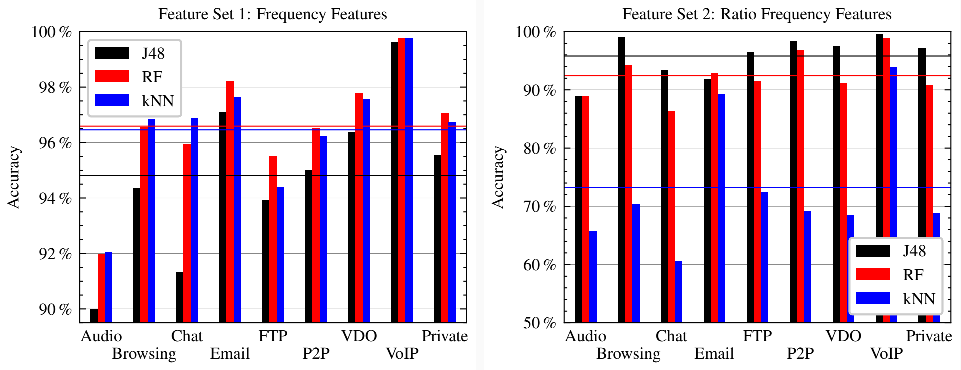


Figure 1: Results of the approach proposed.

Why Are These Results That Remarkable?

Adversarial Indistinguishability Experiment

1. An attacker $\mathcal{A}$ chooses two messages m_0 and m_1 of the same length for a given encryption scheme with security parameter N. The security parameter may be viewed as corresponding to the length of the key.
2. A random key k is generated (depending on N) and a bit $b \in \{0, 1\}$ is chosen at random. $\mathcal{A}$ receives the so-called challenge ciphertext $c \leftarrow \mathsf{Enc}_k(m_b)$.
3. $\mathcal{A}$ outputs a bit $b' \in \{0, 1\}$.
4. The result of the experiment is 1 if $b = b'$, otherwise 0.

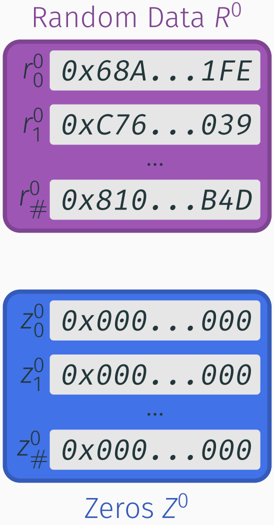
Why Are These Results That Remarkable?

Adversarial Indistinguishability Experiment

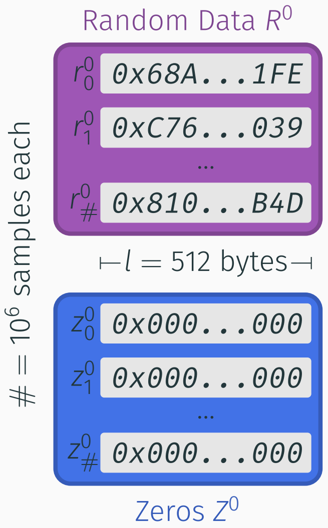
1. An attacker $\mathcal{A}$ chooses two messages m_0 and m_1 of the same length for a given encryption scheme with security parameter N. The security parameter may be viewed as corresponding to the length of the key.
2. A random key k is generated (depending on N) and a bit $b \in \{0, 1\}$ is chosen at random. $\mathcal{A}$ receives the so-called challenge ciphertext $c \leftarrow \mathsf{Enc}_k(m_b)$.
3. $\mathcal{A}$ outputs a bit $b' \in \{0, 1\}$.
4. The result of the experiment is 1 if $b = b'$, otherwise 0.

In a perfect crypto world, distinguishing Tor-encrypted traffic from other encrypted traffic should not be possible!

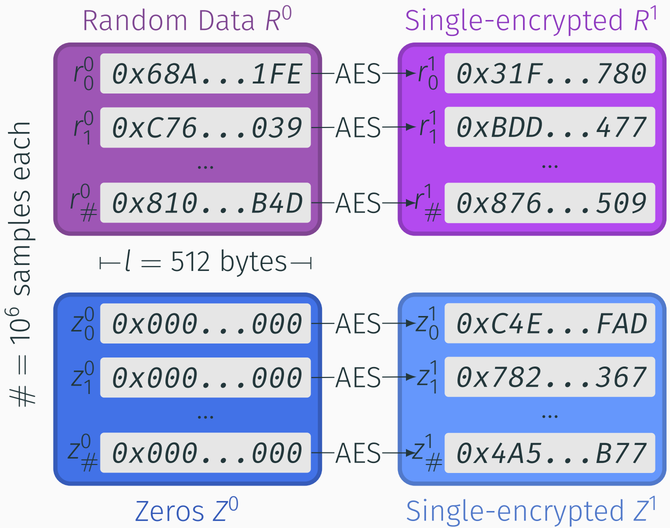
New Experiments: Data Generation and Feature Engineering



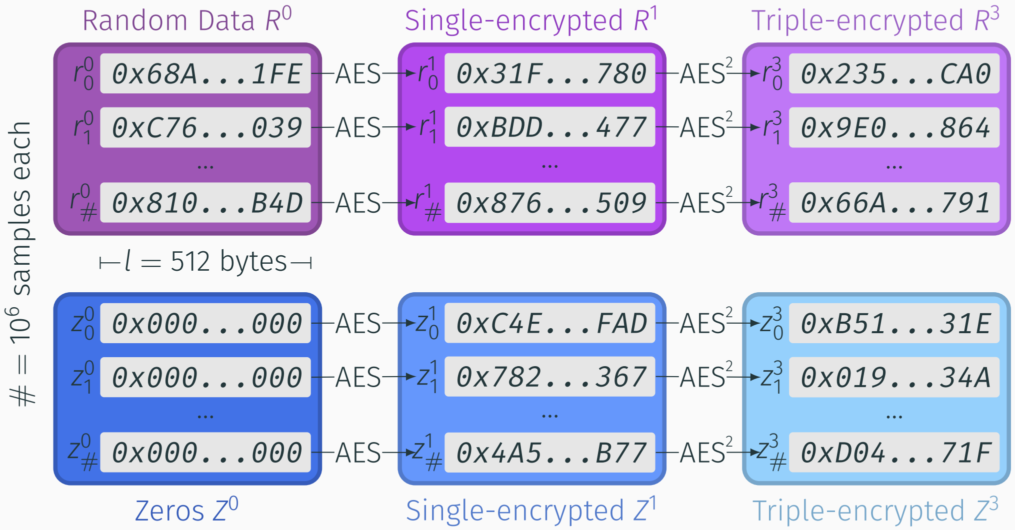
New Experiments: Data Generation and Feature Engineering



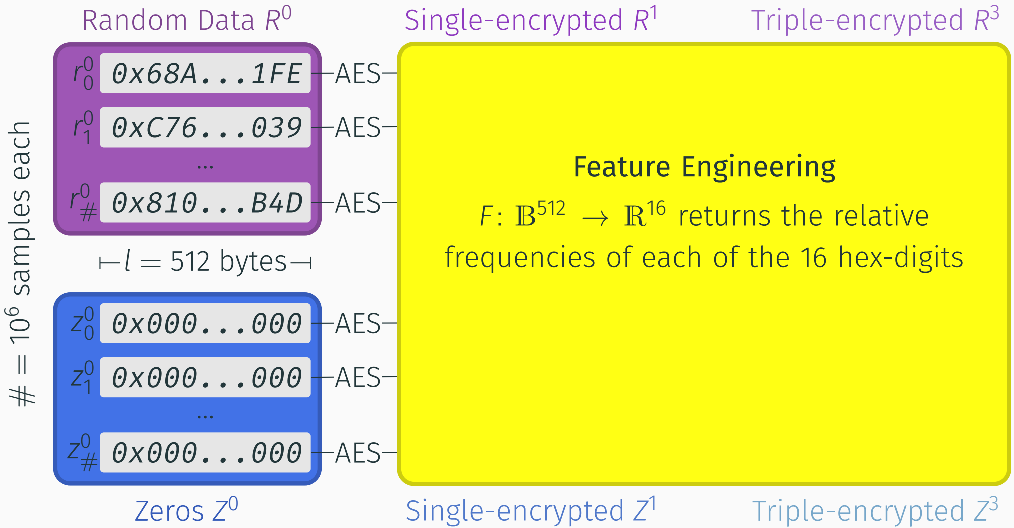
New Experiments: Data Generation and Feature Engineering



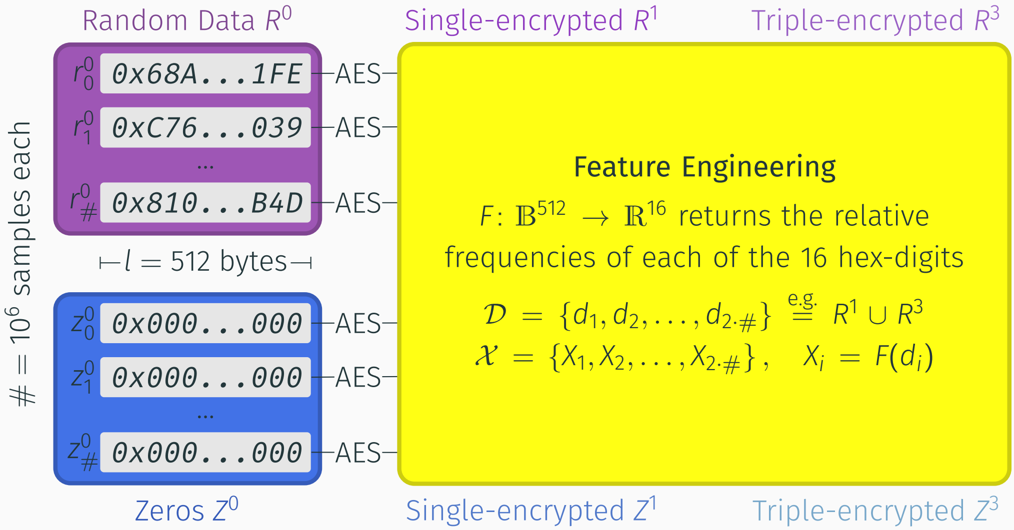
New Experiments: Data Generation and Feature Engineering



New Experiments: Data Generation and Feature Engineering



New Experiments: Data Generation and Feature Engineering



Feature Engineering

$F\colon \mathbb{B}^{512} \to \mathbb{R}^{16}$ returns the relative frequencies of each of the 16 hex-digits

$$\mathcal{D} = \{d_1, d_2, \ldots, d_{2\cdot\#}\} \overset{\text{e.g.}}{=} R^1 \cup R^3$$

$$\mathcal{X} = \{X_1, X_2, \ldots, X_{2\cdot\#}\}, \quad X_i = F(d_i)$$

New Experiments: Algorithms and Experiments

AES Modes of Operation

- Cipher Block Chaining (CBC)
- Counter (CTR)
- Electronic Codebook (ECB)

New Experiments: Algorithms and Experiments

AES Modes of Operation

- Cipher Block Chaining (CBC)
- Counter (CTR)
- Electronic Codebook (ECB)

Machine Learning Algorithms

- Random Forest (RF)
- Decision Tree (DT)
- k-Nearest Neighbor (kNN)

New Experiments: Algorithms and Experiments

AES Modes of Operation

- Cipher Block Chaining (CBC)
- Counter (CTR)
- Electronic Codebook (ECB)

Machine Learning Algorithms

- Random Forest (RF)
- Decision Tree (DT)
- k-Nearest Neighbor (kNN)

Datasets

- $\mathcal{D}_R = R^1 \cup R^3$
- $\mathcal{D}_Z = Z^1 \cup Z^3$

New Experiments: Algorithms and Experiments

AES Modes of Operation

- Cipher Block Chaining (CBC)
- Counter (CTR)
- Electronic Codebook (ECB)

Machine Learning Algorithms

- Random Forest (RF)
- Decision Tree (DT)
- k-Nearest Neighbor (kNN)

Datasets

- $\mathcal{D}_R = R^1 \cup R^3$
- $\mathcal{D}_Z = Z^1 \cup Z^3$

Train-Test-Split

- Split $(\mathcal{X}, \mathcal{Y})$ into $(\mathcal{X}_{\text{tr}}, \mathcal{Y}_{\text{tr}})\,;(\mathcal{X}_{\text{te}}, \mathcal{Y}_{\text{te}})$
- Ratio: 75 % training, 25 % test

A total of 18 experiments were conducted

New Experiments: Results (CBC)

RF (49.90 %)

True \ Prediction	Z^1	Z^3
Z^1	120,550 (24.11 %)	129,450 (25.89 %)
Z^3	121,055 (24.21 %)	128,945 (25.79 %)

DT (49.92 %)

True \ Prediction	Z^1	Z^3
Z^1	131,716 (26.34 %)	118,284 (23.66 %)
Z^3	132,124 (26.42 %)	117,876 (23.58 %)

kNN (50.10 %)

True \ Prediction	Z^1	Z^3
Z^1	156,050 (31.21 %)	93,950 (18.79 %)
Z^3	155,546 (31.11 %)	94,454 (18.89 %)

RF (49.98 %)

True \ Prediction	R^1	R^3
R^1	121,993 (24.40 %)	128,007 (25.60 %)
R^3	122,116 (24.42 %)	127,884 (25.58 %)

DT (50.04 %)

True \ Prediction	R^1	R^3
R^1	113,512 (22.70 %)	136,488 (27.30 %)
R^3	113,320 (22.66 %)	136,680 (27.34 %)

kNN (50.01 %)

True \ Prediction	R^1	R^3
R^1	155,438 (31.09 %)	94,562 (18.91 %)
R^3	155,387 (31.08 %)	94,613 (18.92 %)

New Experiments: Results (CTR)

RF (50.07 %)

True \ Prediction	Z^1	Z^3
Z^1	110,909 (22.18 %)	139,091 (27.82 %)
Z^3	110,566 (22.11 %)	139,434 (27.89 %)

DT (49.88 %)

True \ Prediction	Z^1	Z^3
Z^1	139,964 (27.99 %)	110,036 (22.01 %)
Z^3	140,569 (28.11 %)	109,431 (21.89 %)

kNN (50.11 %)

True \ Prediction	Z^1	Z^3
Z^1	155,935 (31.19 %)	94,065 (18.81 %)
Z^3	155,368 (31.07 %)	94,632 (18.93 %)

RF (50.06 %)

True \ Prediction	R^1	R^3
R^1	104,927 (20.99 %)	145,073 (29.01 %)
R^3	104,634 (20.93 %)	145,366 (29.07 %)

DT (50.04 %)

True \ Prediction	R^1	R^3
R^1	126,658 (25.33 %)	123,342 (24.67 %)
R^3	126,479 (25.30 %)	123,521 (24.70 %)

kNN (49.90 %)

True \ Prediction	R^1	R^3
R^1	155,333 (31.07 %)	94,667 (18.93 %)
R^3	155,848 (31.17 %)	94,152 (18.83 %)

New Experiments: Results (ECB)

RF (49.99 %)

True \ Prediction	Z^1	Z^3
Z^1	129,216 (25.84 %)	120,784 (24.16 %)
Z^3	129,249 (25.85 %)	120,751 (24.15 %)

DT (49.97 %)

True \ Prediction	Z^1	Z^3
Z^1	125,043 (25.01 %)	124,957 (24.99 %)
Z^3	125,205 (25.04 %)	124,795 (24.96 %)

kNN (49.95 %)

True \ Prediction	Z^1	Z^3
Z^1	155,541 (31.11 %)	94,459 (18.89 %)
Z^3	155,771 (31.15 %)	94,229 (18.85 %)

RF (49.91 %)

True \ Prediction	R^1	R^3
R^1	127,775 (25.55 %)	122,225 (24.45 %)
R^3	128,250 (25.65 %)	121,750 (24.35 %)

DT (49.83 %)

True \ Prediction	R^1	R^3
R^1	120,618 (24.12 %)	129,382 (25.88 %)
R^3	121,479 (24.30 %)	128,521 (25.70 %)

kNN (50.03 %)

True \ Prediction	R^1	R^3
R^1	155,696 (31.14 %)	94,304 (18.86 %)
R^3	155,553 (31.11 %)	94,447 (18.89 %)

Conclusion and Outlook

Conclusion

Pitpimon's PhD Thesis shows distinguishability of Tor and non-Tor packets

Q: Is this due to the different number of encryption passes?
I.e.: Can *single-encrypted* data be distinguished from *triple-encrypted* data via analysis of hex characters?

A: All three ML models failed to do so regardless of the type of encrypted data.
The accuracy is $\approx 50\,\%$, which is the guess probability.

Conclusion and Outlook

Conclusion

Pitpimon's PhD Thesis shows distinguishability of Tor and non-Tor packets

Q: Is this due to the different number of encryption passes?
I.e.: Can *single-encrypted* data be distinguished from *triple-encrypted* data via analysis of hex characters?

A: All three ML models failed to do so regardless of the type of encrypted data.
The accuracy is $\approx 50\,\%$, which is the guess probability.

Outlook

Further experiments needed to gradually rule out possible explanations for the distinguishability and to identify the actual cause.

Thank you!