Covert Communication Channels based on Hardware Trojans:

Open-Source dataset and AI-based detection

SemSecuElec

Alán Díaz Rizo, MCF

February 28, 2025

SCIENCES SORBONNE UNIVERSITÉ

cnrs

LIP6

Outline

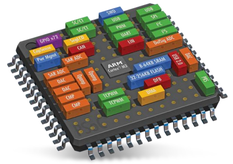
1. Context: Globalized Integrated Circuit (IC) supply chain
2. Problem: Hardware security threats
3. Hardware Trojans (HT)
    a) HT-enabled Covert Communication Channels (HT-CC)
4. Contributions: dataset of HT-CC attacks and AI-based defense
    a) Dataset generation
    b) AI-based detection and classification
5) Conclusion

Outline

1. Context: Globalized Integrated Circuit (IC) supply chain
2. Problem: Hardware security threats
3. Hardware Trojans (HT)
    a) HT-enabled Covert Communication Channels (HT-CC)
4. Contributions: dataset of HT-CC attacks and AI-based defense
    a) Dataset generation
    b) AI-based detection and classification
5) Conclusion

IC supply chain before 1980s

IC Design and Integration



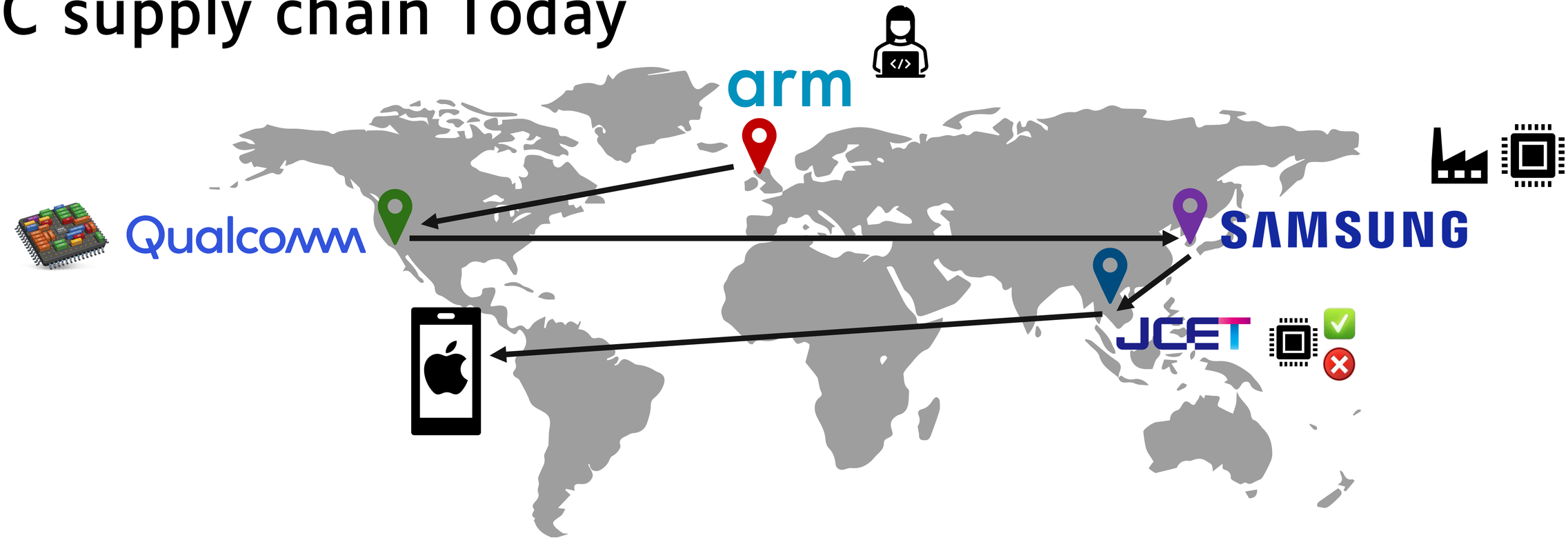
Fabrication

Testing, Packaging, Assembly

End-user

End of IC life

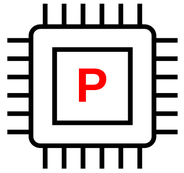
IC supply chain Today



IP/IC Design → SoC Integration → Fabrication → Testing, Packaging, Assembly → End-user → End of IC life

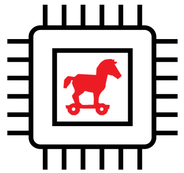
Outline

IC life cycle attacks and hardware security threats

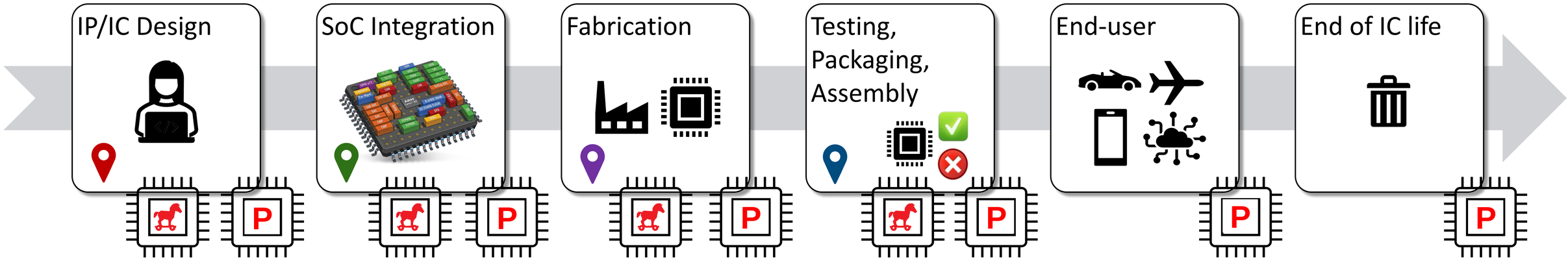


Piracy:

1. Non-authorized use or reutilization of IPs/ICs or SoCs
2. Overproduction and remarking of ICs
3. Netlist extraction via Reverse Engineering
4. Non-authorized recycling



Hardware Trojan (HT) insertion: Malicious modification of a circuit



Outline

Hardware Trojan (HT) threat

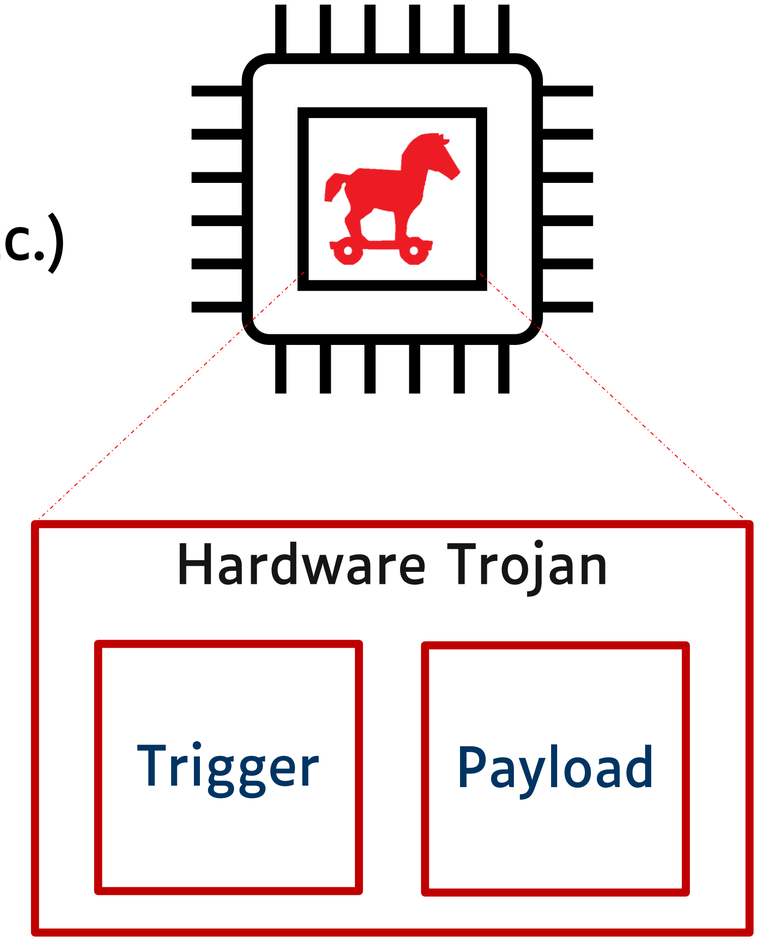
Malicious modification of a circuit

HT design:

a) Triggering mechanism (always on, condition, etc.)

b) Payload mechanism (effect):

+ Changing the function
+ Degrading performances
+ Leaking information from the chip
+ Denial-of-service



Attacker's goal: stealthy, small footprint

Defender's goal: prevention, detection

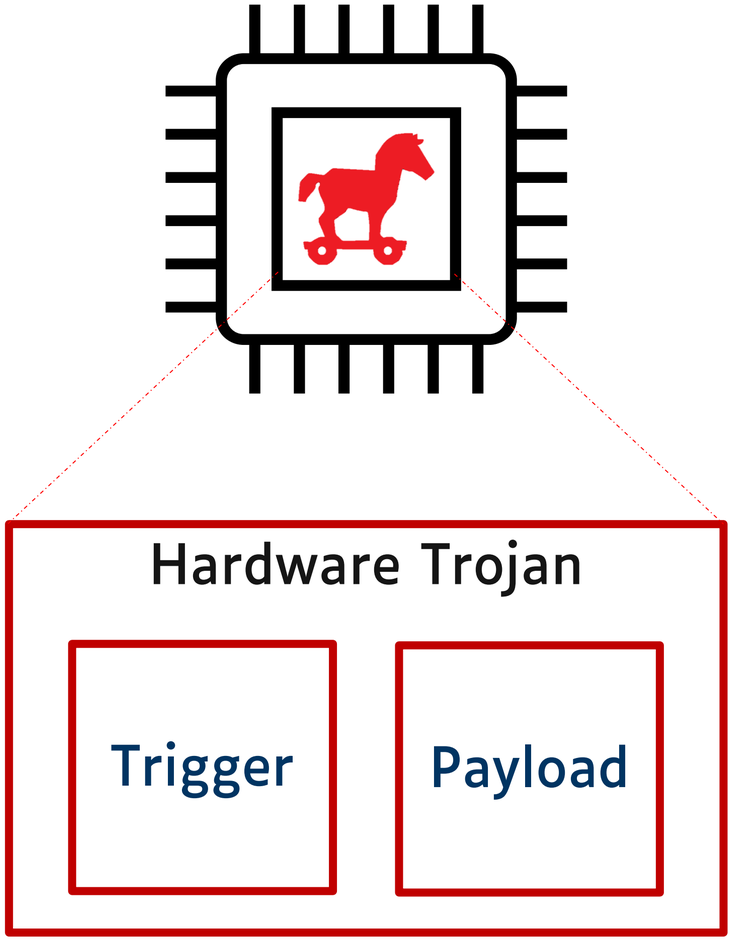
Hardware Trojan (HT) threat

Malicious modification of a circuit

On September 17, 2024, at 15h30, a message was sent to 5000 Gold Apollo branded (Taiwan) pagers of the Hezbollah group. Seconds later, approx. 4000 of the devices exploded, killing several people and injuring thousands others

AR-64 Pager
460 MHz band

Remains of pager after explosion

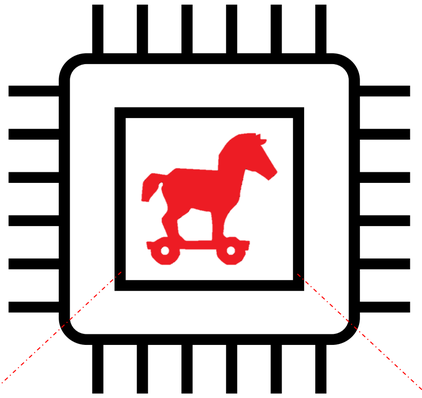


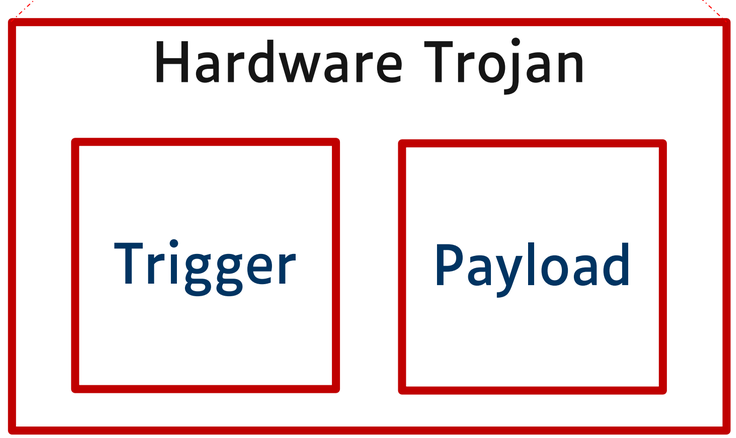
https://www.cryptomuseum.com/covert/radio/apollo/ar924/

Hardware Trojan (HT) threat

Malicious modification of a circuit

On September 18, 2024, at 17h00, about 24 hours after the initial attack, a second wave of explosions occurred, targeting ICOM branded (Japan) handheld radios





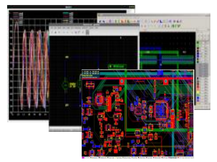
ICOM

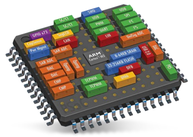
IC-V82
Walkie-talkies
VHF band

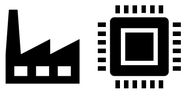
Exploding pagers and spy chips: the rising risk of hardware tampering

Financial Times, Getty Images

HT attacks taxonomy







- **Attacker**
  - EDA tool provider
  - **3PIP provider**
  - SoC integrator
  - Test infrastructure IP provider
  - Foundry
  - Assembly and package
- **Abstraction level**
  - System level
  - **RT level**
  - Gate level
  - Transistor level
  - Layout level
  - Physical level
- **Location**
  - Processor
  - Memory
  - **Analog**
  - Clock
  - Power supply
  - I/O
- **Triggering**
  - **Always on**
  - Internally
  - Time-based
  - Physical condition
  - Externally
- **Payload**
  - Change function
  - Degrade performance
  - Denial of service
  - **Leak information**

R. Karri *et al.*, Computer'10

S. Bhunia *et al.*, Proc. IEEE'14

M. Tehranipoor and F. Koushanfar , D&T '10

HT defenses taxonomy

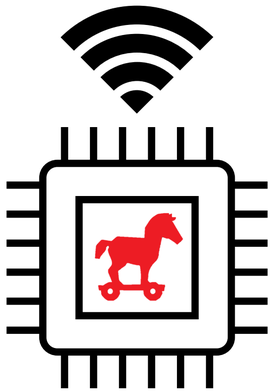
- Pre-silicon prevention
  - Trusted Production
    - 
  - Secure Design
    - Split-manufacturing
    - **Locking**
    - Filling unused layout spaces

    → Make it difficult for an attacker to insert a HT
    - Functional verification
    - Structural analysis
    - Information Flow Tracking
    - Testing
    - **On-chip test infrastructure**

    → Expose the HT
- Post-silicon detection
  - Destructive
    - Reverse engineering
    - Optical circuit analysis
  - Non-destructive
    - Test-time
      - **Standard Testing**
      - Specialized Testing
    - Run-time monitors
      - Using OS
      - Logic assertions
      - **Power monitors**
      - Temp. monitors
      - Invariant properties

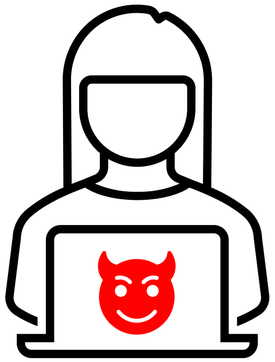
HT-enabled Covert Communication Channels (HT-CC)

Alice



Leaks covert data hidden in a legitimate message

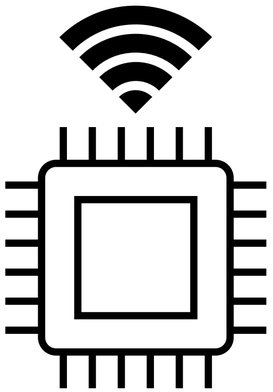
Eve



Retrieves *covert data*

e.g.
Encryption key

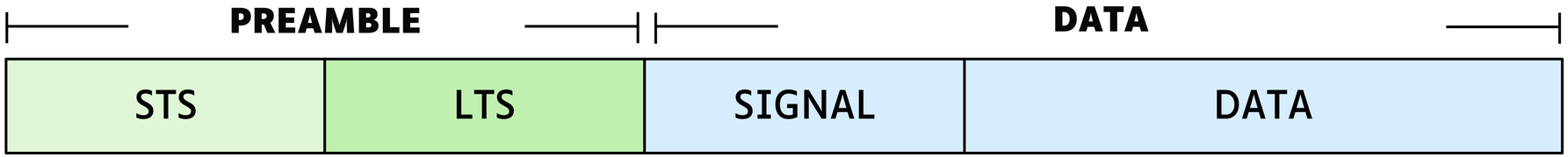
Bob



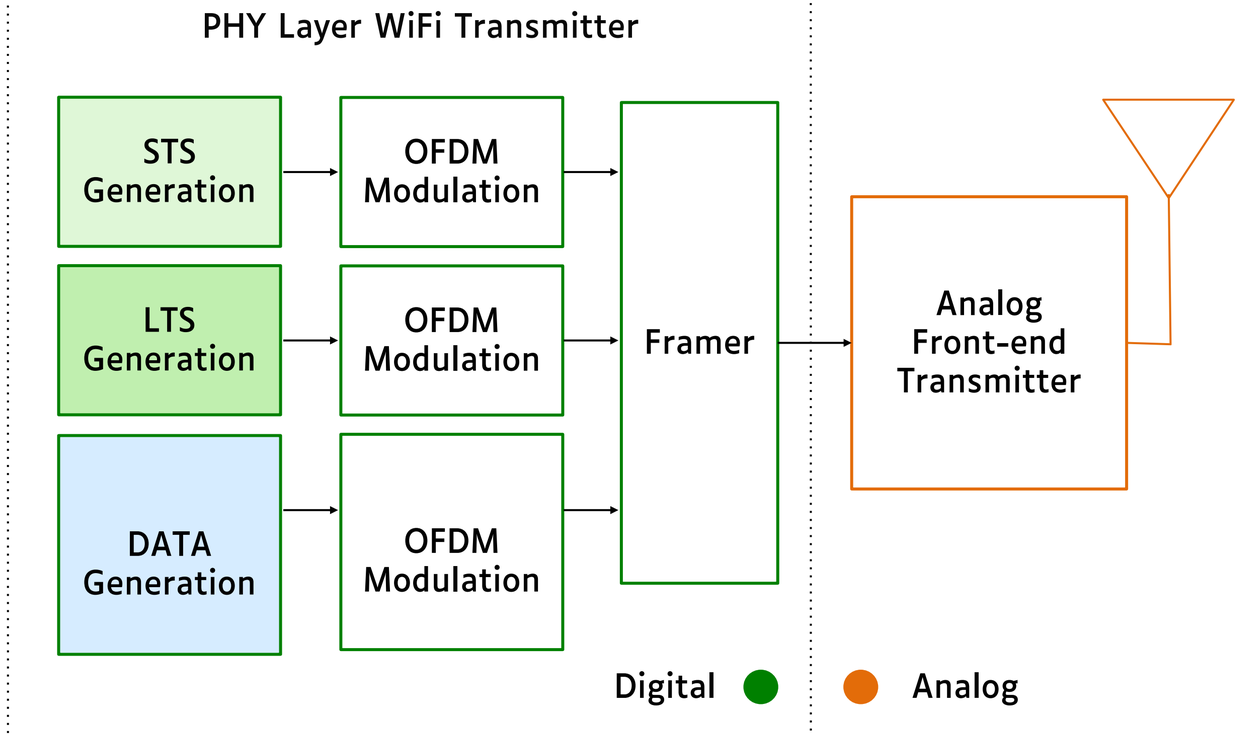
Recovers legitimate message

Y. Jin and Y. Makris, D&T'10, Dutta *et al.*, Information Hiding'13, J. Classen *et al.*, CNS'15, Y. Liu *et al.*, TVLSI'17, K. S. Subramani *et al.*, TIFS'19, K. S. Subramani *et al.*, TIFS'20, S. Chang *et al.*, TODAES'20, A. R. Díaz Rizo *et al.*, IEEE TDSC'22

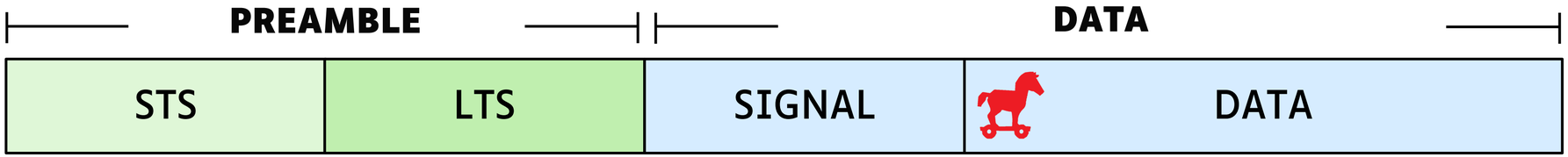
HT-enabled Covert Communication Channels (HT-CC)

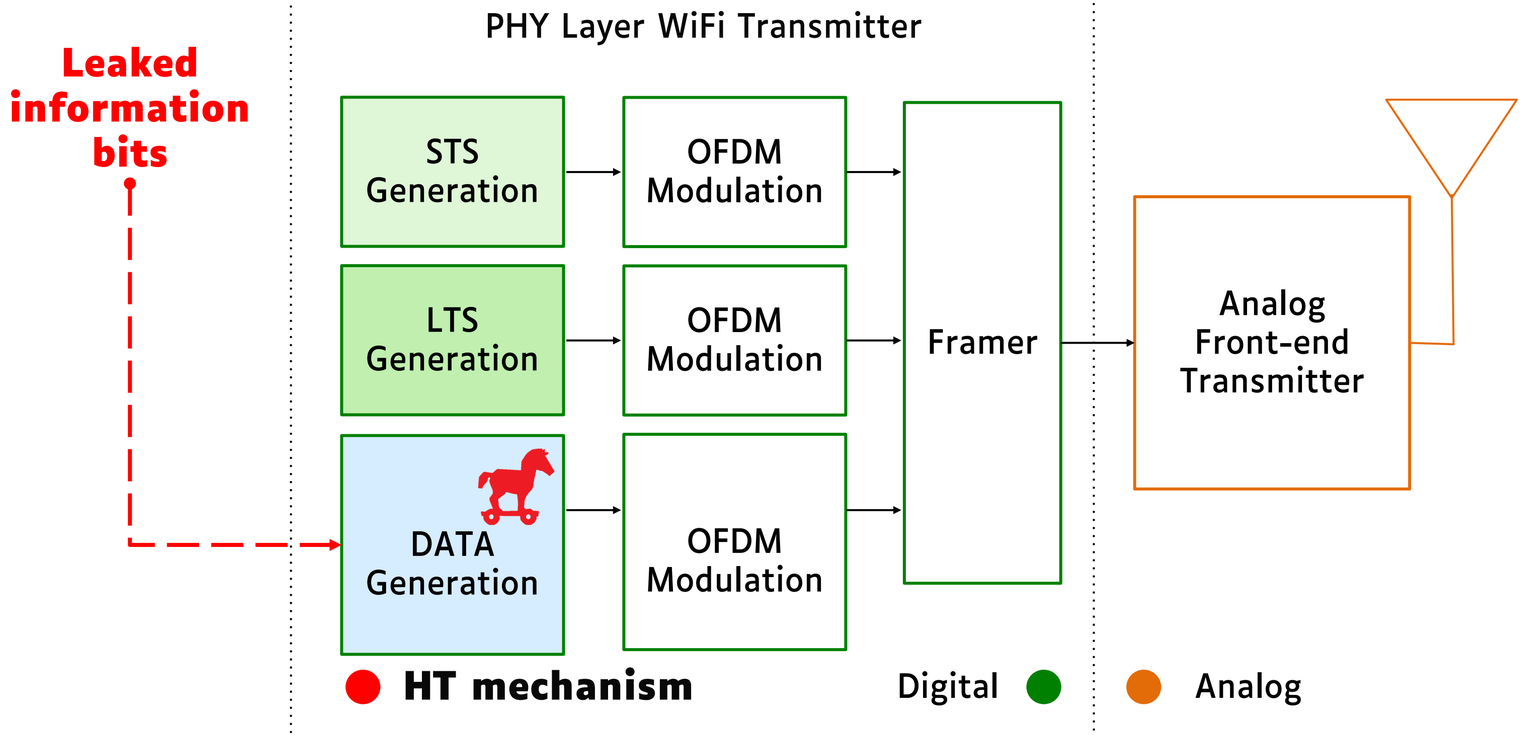


The leaked information bits come from a tampered memory, register, crypto core or any other IP core

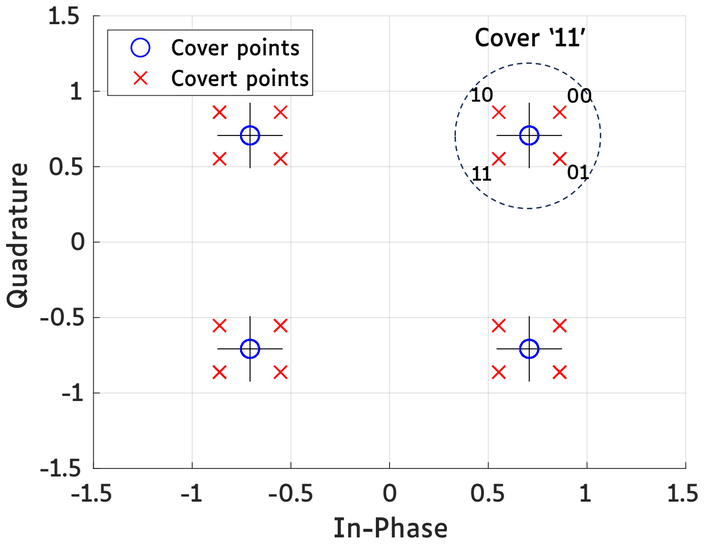


HT-enabled Covert Communication Channels (HT-CC)



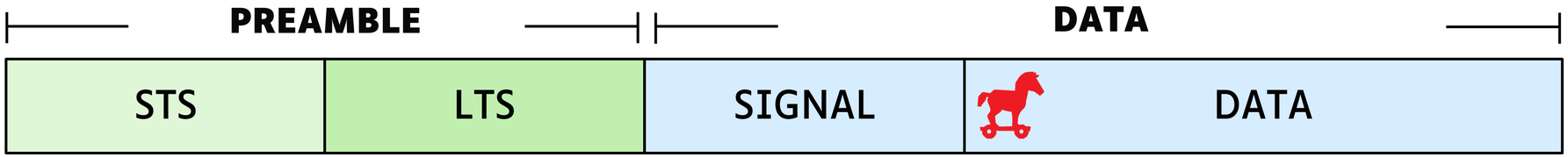


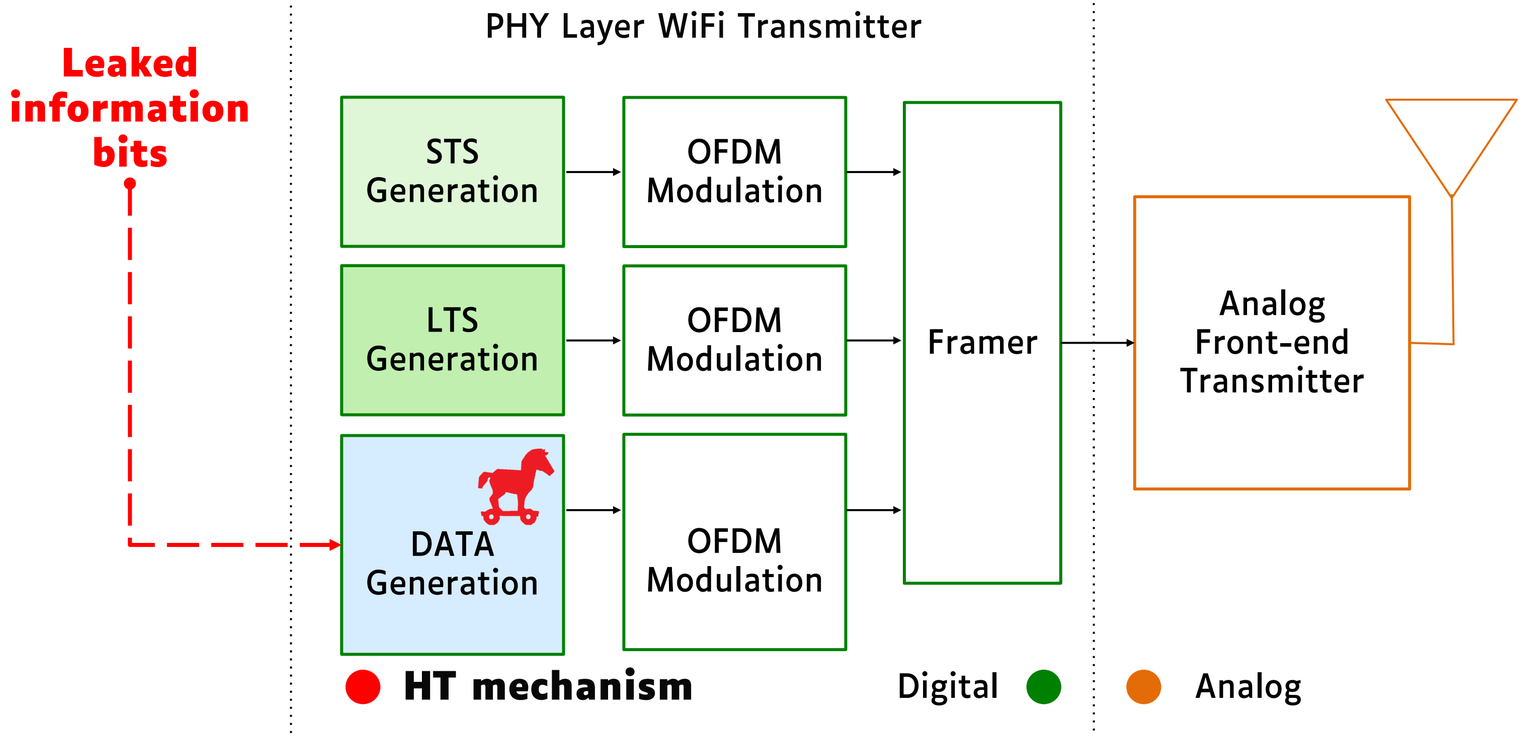
- Encodes leaked data on the I/Q mapping



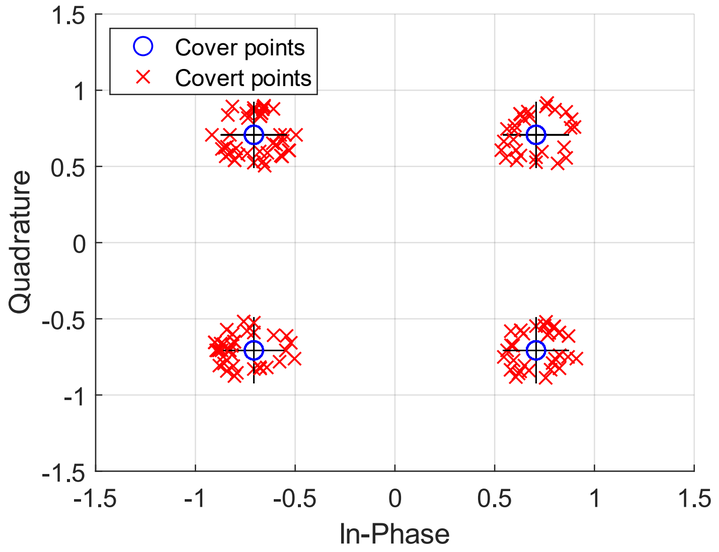
Dutta *et al.*, Information Hiding'13

HT-enabled Covert Communication Channels (HT-CC)



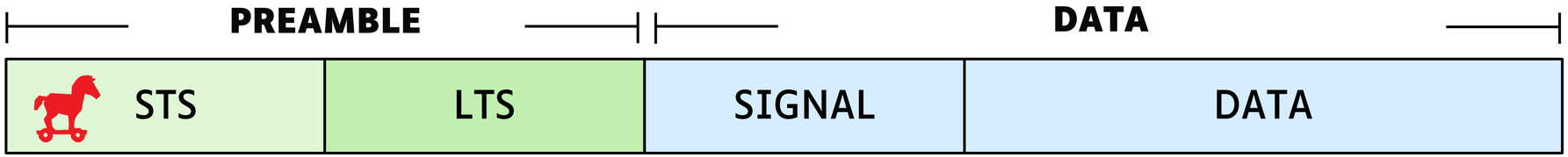


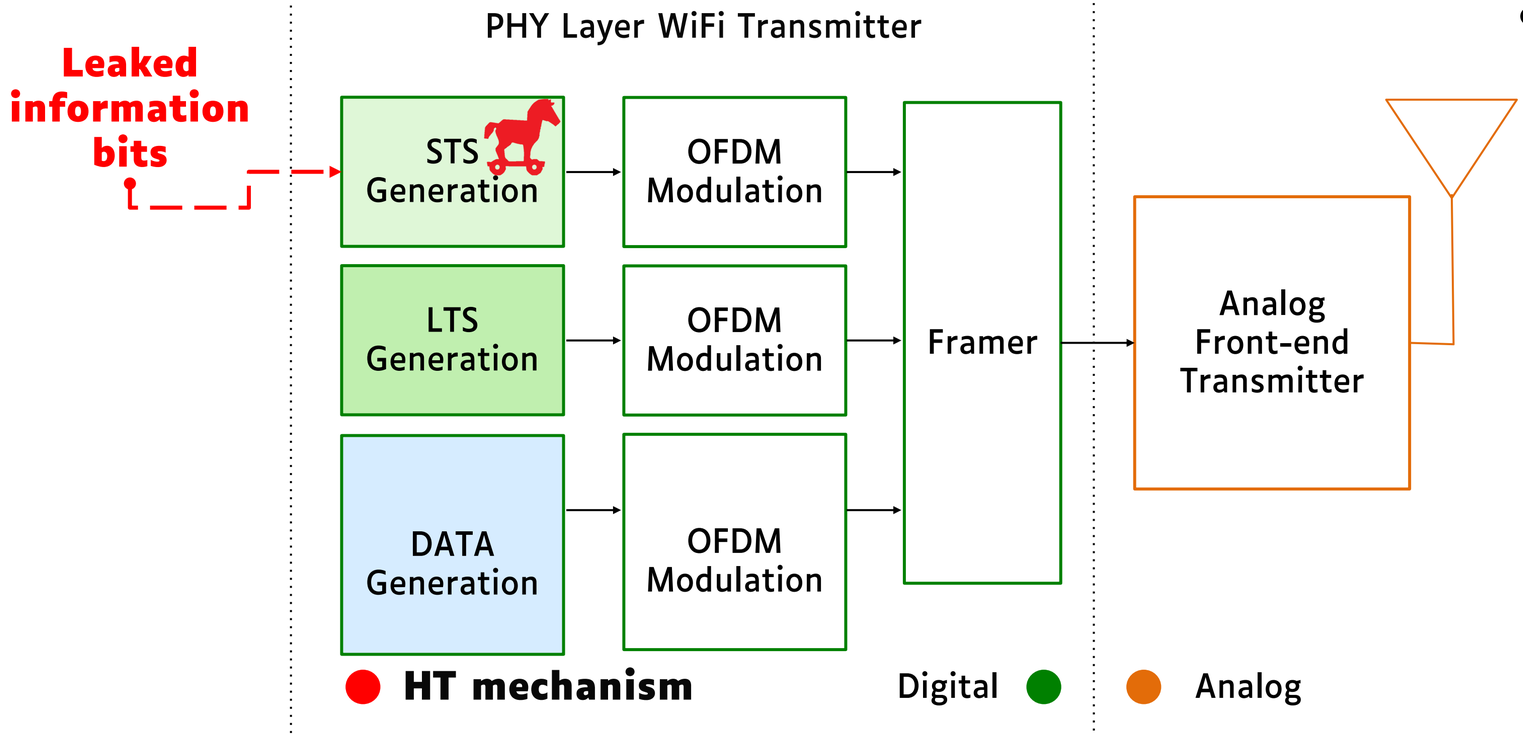
- Encodes leaked data on the I/Q mapping
- Hides encoding with imperfections



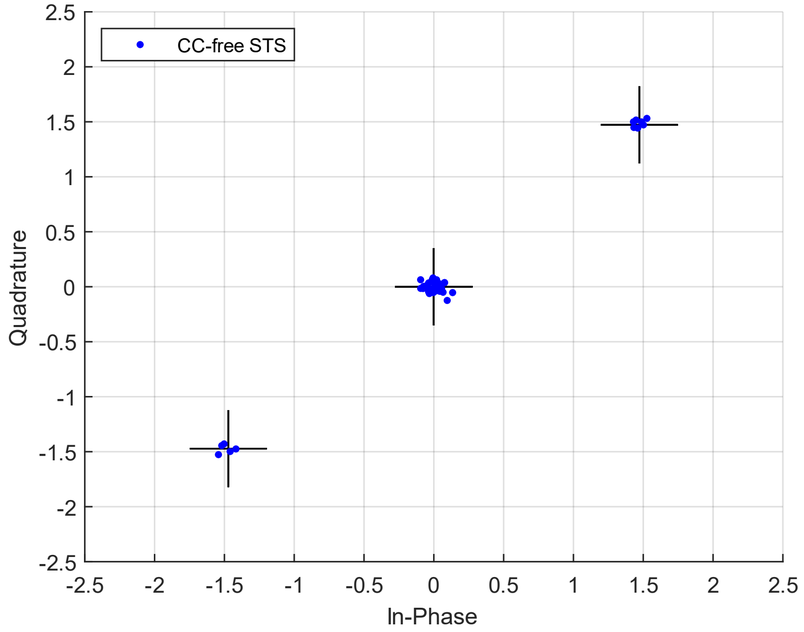
Dutta *et al.*, Information Hiding'13

HT-enabled Covert Communication Channels (HT-CC)



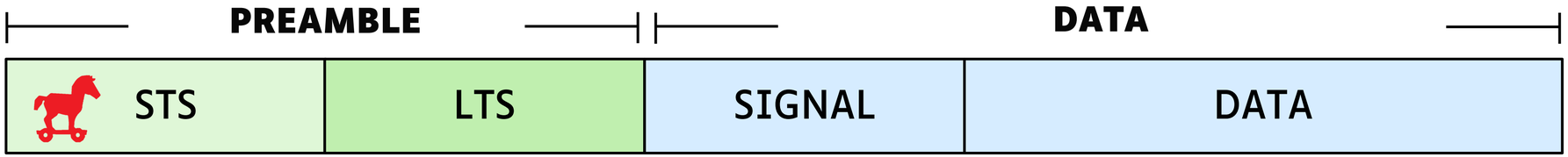


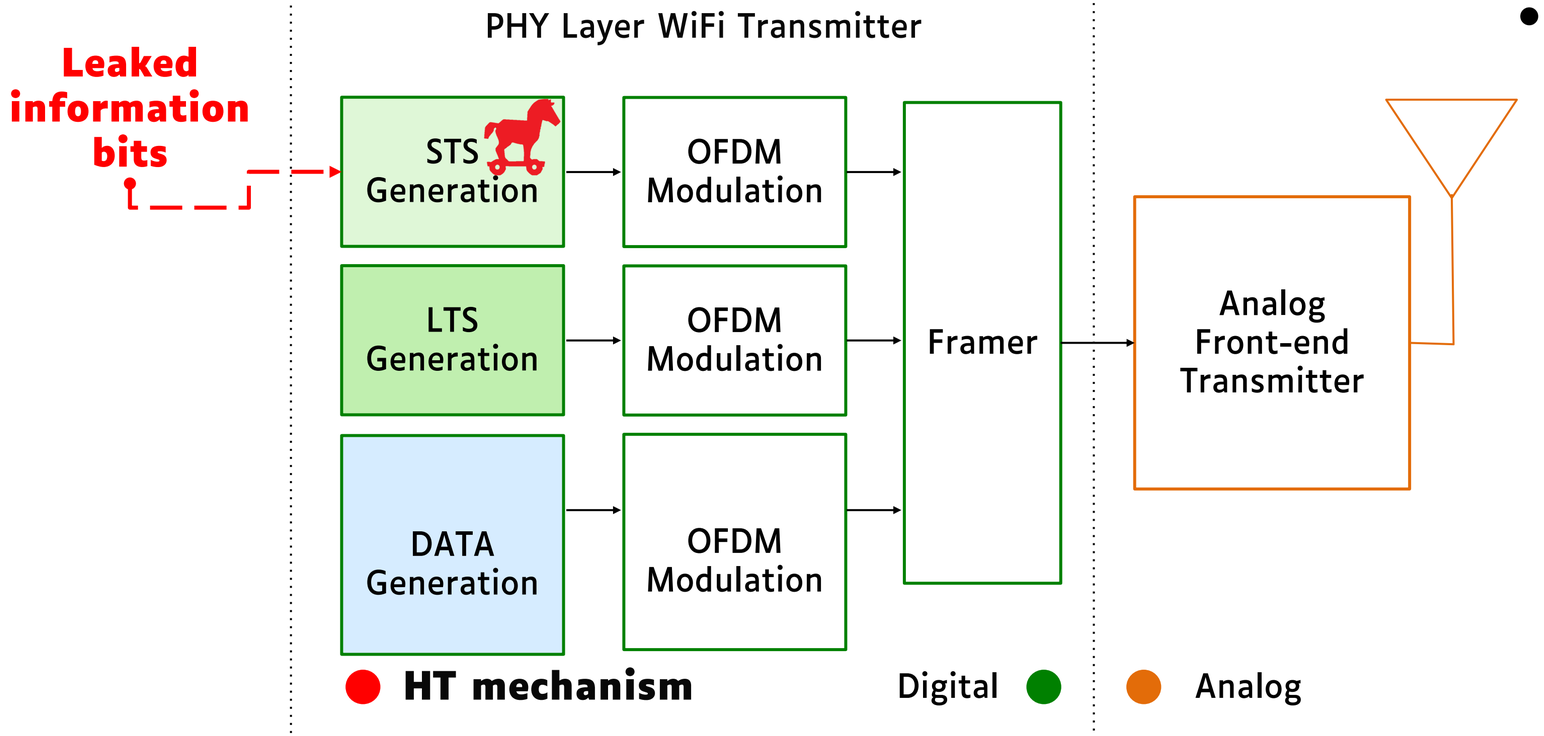
- STS for detecting the start of the frame (synchronization)



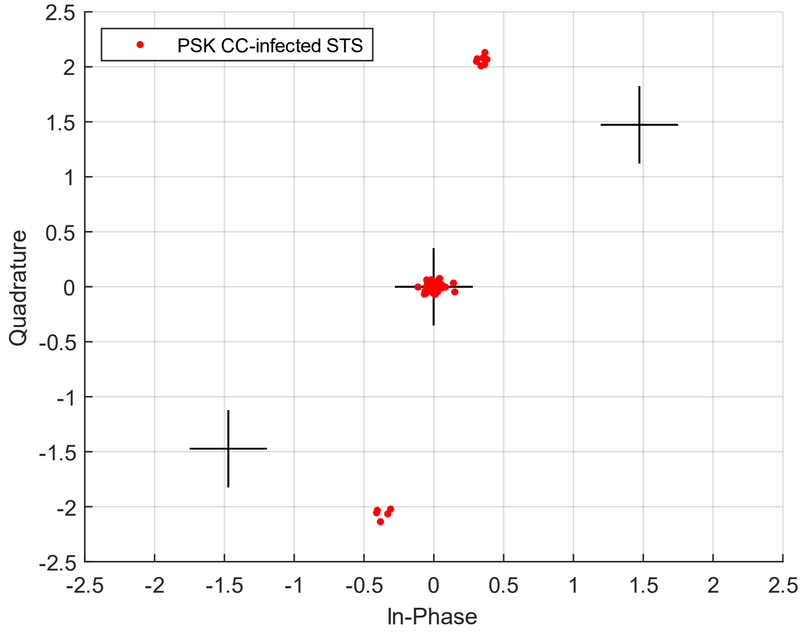
J. Classen *et al.*, CNS'15

HT-enabled Covert Communication Channels (HT-CC)



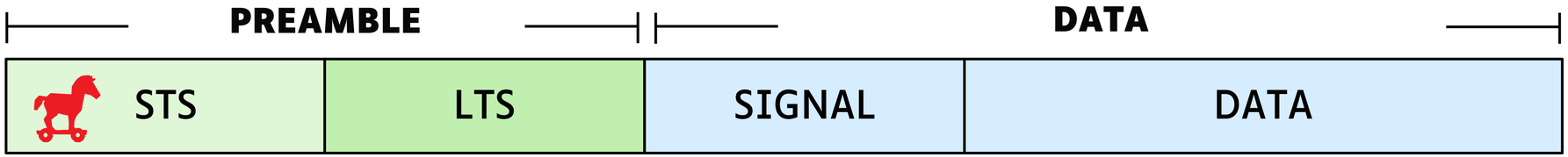


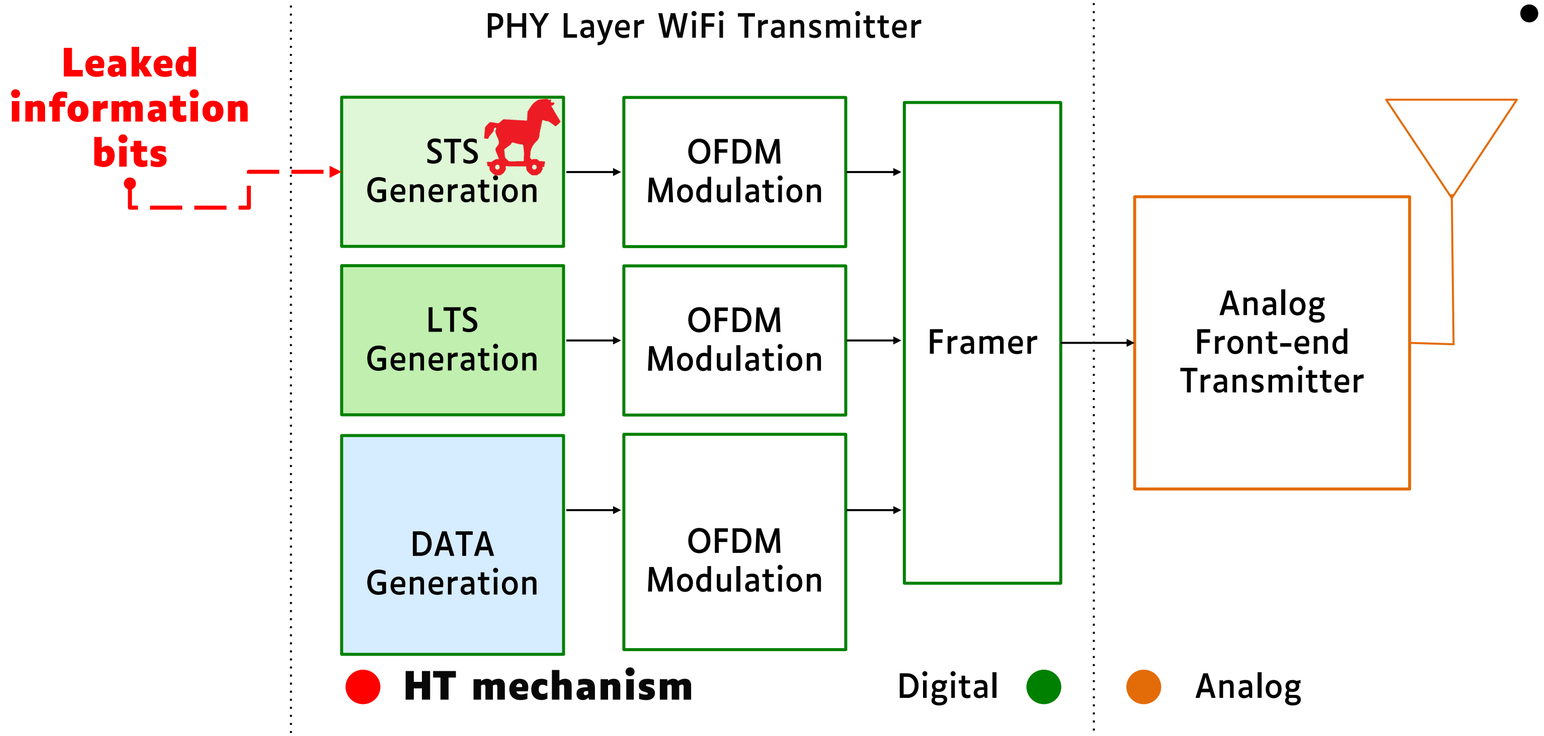
- Additional **phase shift** in the STS symbols



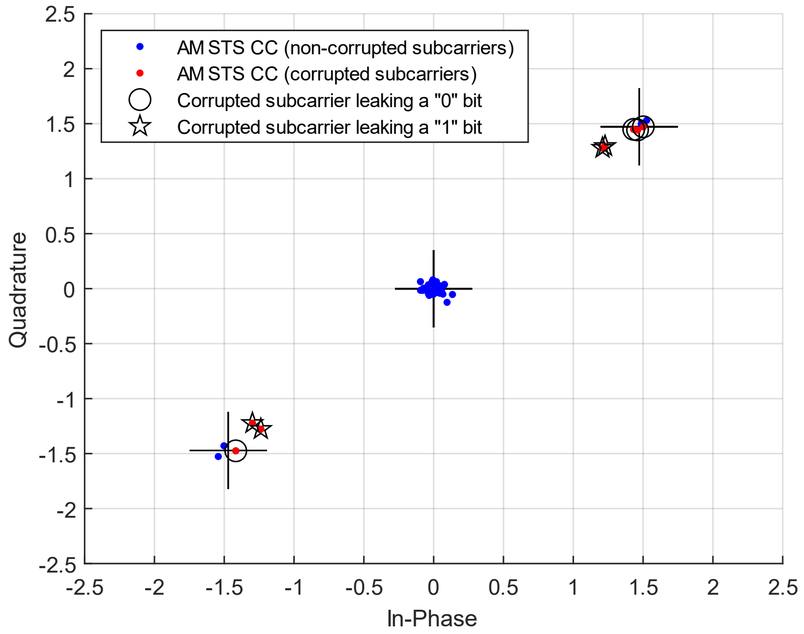
J. Classen *et al.*, CNS'15

HT-enabled Covert Communication Channels (HT-CC)



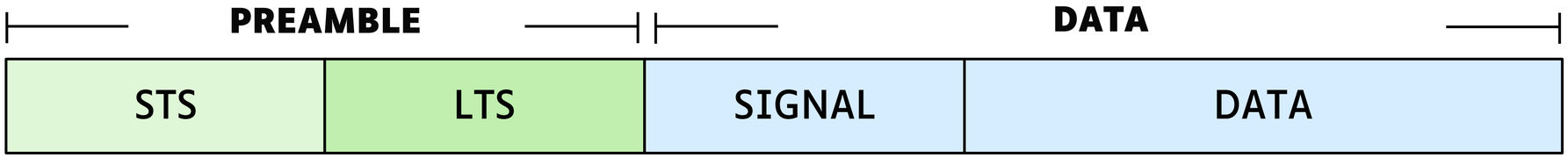


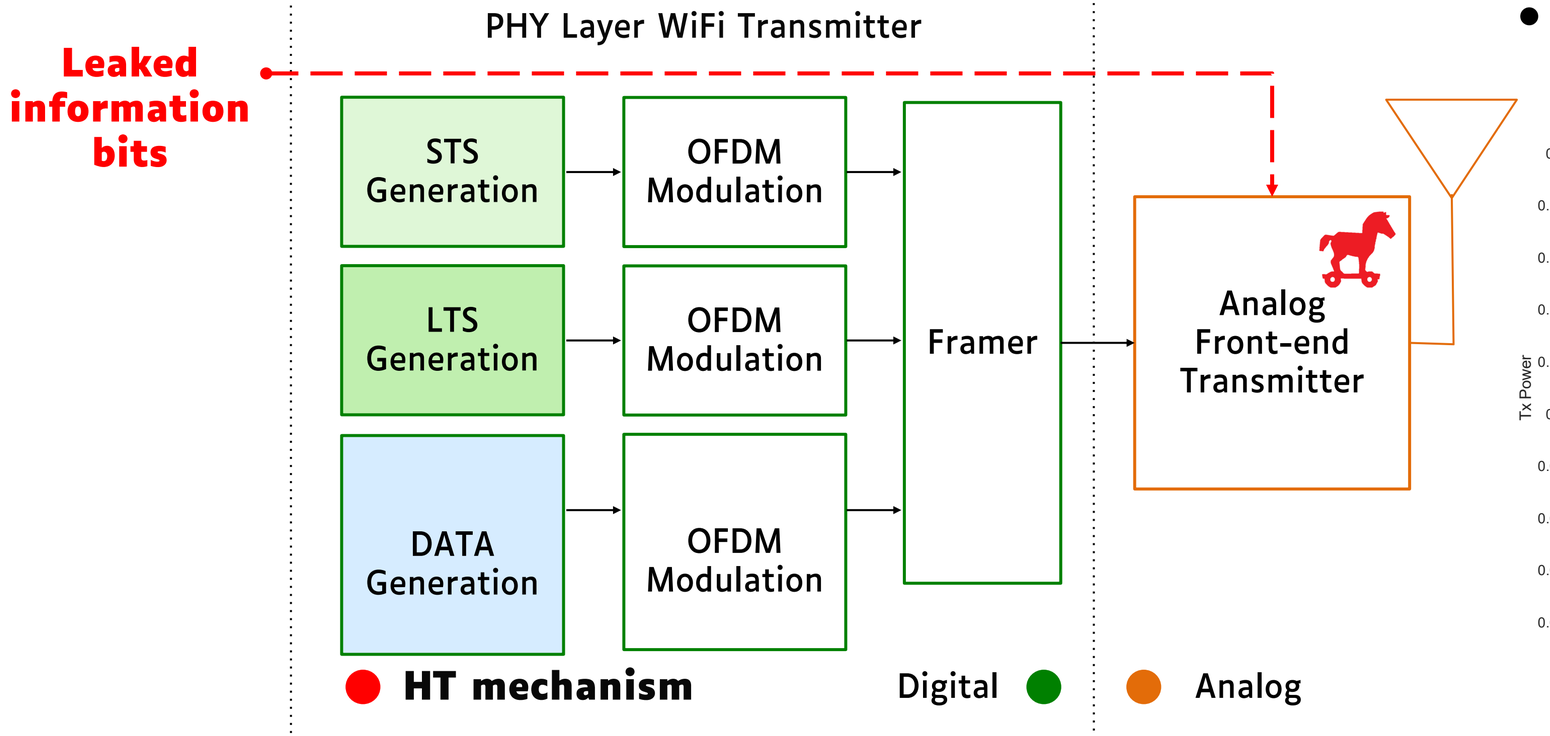
- **Amplitude modulation** of some STS symbols



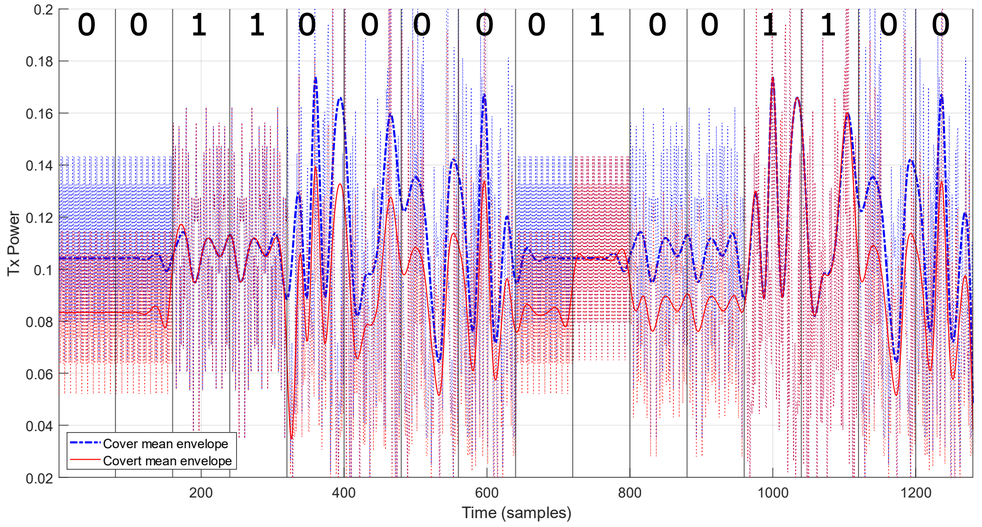
A. R. Díaz Rizo *et al.*, IEEE TDSC'23

HT-enabled Covert Communication Channels (HT-CC)



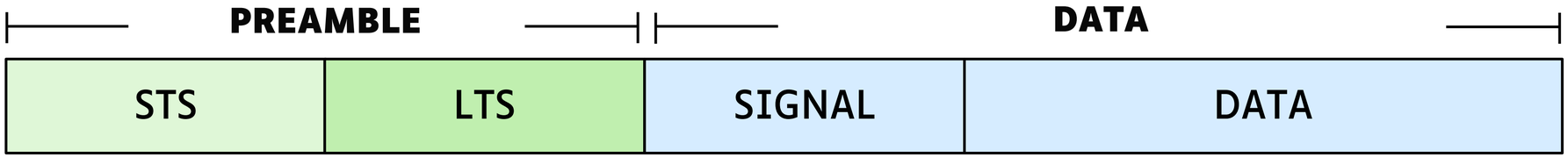


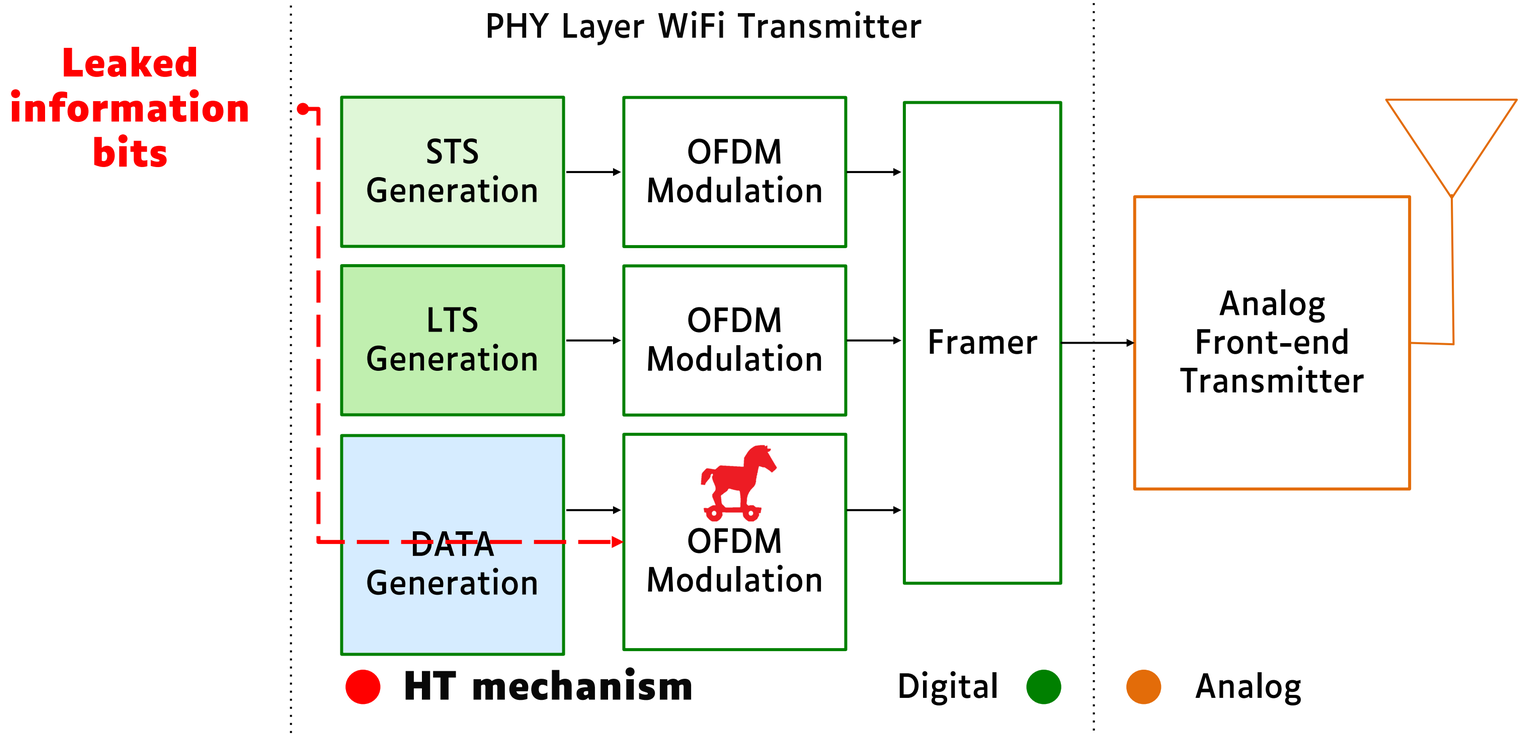
- **Amplitude modulation** of entire Tx Signal



Y. Jin and Y. Makris, D&T'10, Y. Liu *et al.*, TVLSI'17, K. S. Subramani *et al.*, TIFS'20, S.

HT-enabled Covert Communication Channels (HT-CC)





- Extra subcarriers added to the OFDM symbol
- Leaks data into the OFDM's Cyclic Prefix

J. Classen et al., CNS'15

State-of-the-Art on HT-CC detection techniques

How can these attacks be detected?

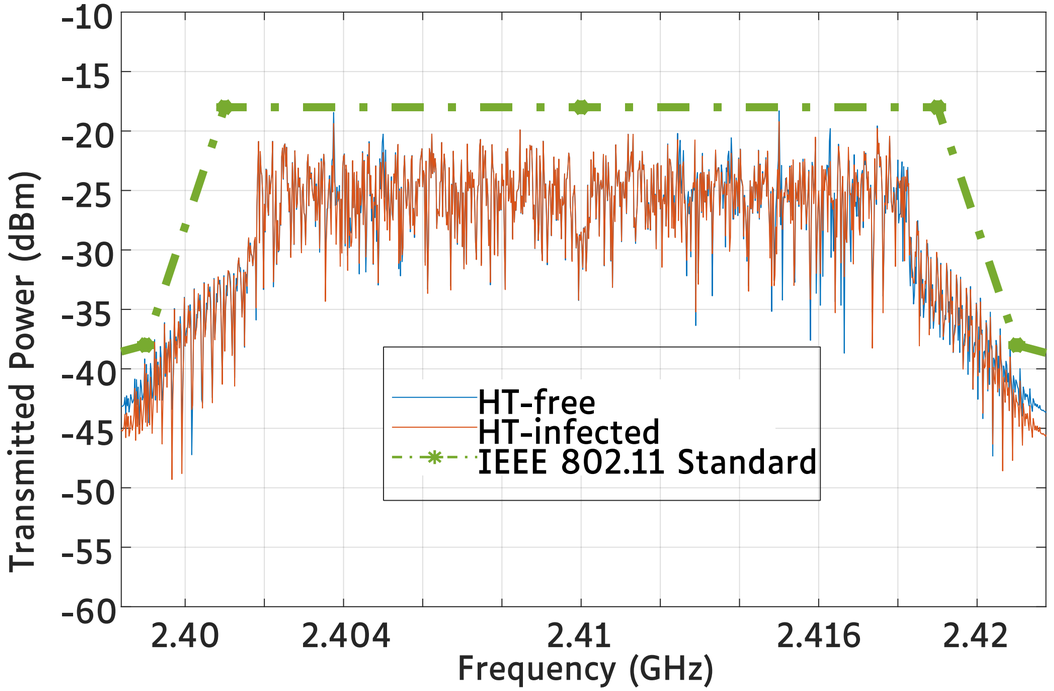
State-of-the-Art on HT-CC detection techniques

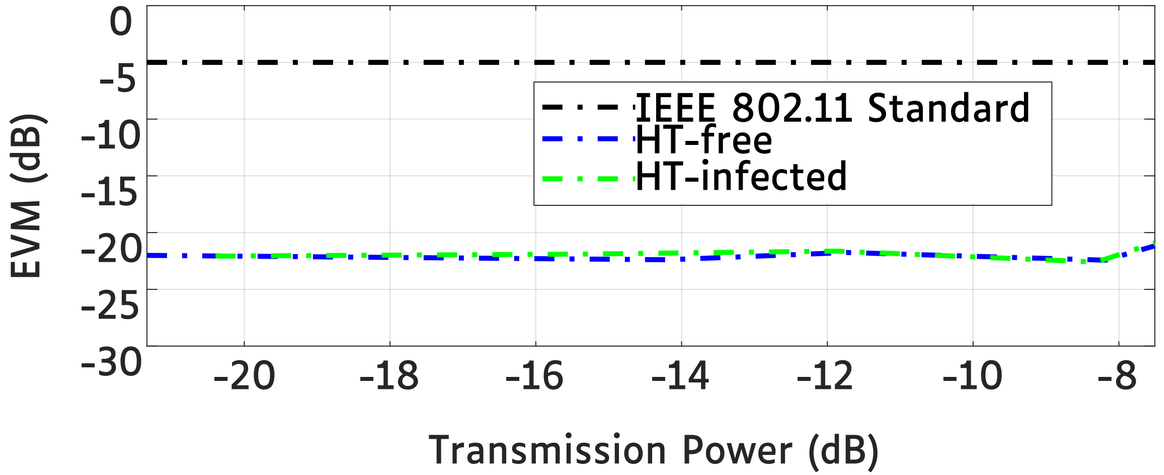
Ref.	Attack Model	Defense Mechanism
[1]	Encodes leaked data on the **I/Q mapping** and hides the encoding by introducing imperfections to the transmitted signal.	Certain tests, such as **EVM**, show a distinguishing behavior **compared to HT-free** operation.
[2]	Leaks data in **extra subcarriers** added to the OFDM signal.	Decode the signal field to determine if the **number of subcarriers is correct; spectrum analysis**
[2]	Leaks data into parts of the OFDM **Cyclic Prefix (CP)**.	Compare the last 16 samples of an OFDM symbol with its CP; **spectrum analysis**.
[3]	Leaks data using **spread spectrum** techniques.	**Spectral analysis**
[4,5,6]	Leaks data by **modulating amplitude and/or frequency of transmitted signal**.	Statistical Side Channel Fingerprinting (**SSCF**); Adaptive Channel Estimation (**ACE**).
[7]	Leaks data through **amplitude modulation** of some subcarriers in the **Synchronization Sequence,** a.k.a. Short Training Sequence (STS), of the Preamble.	**Evades** any known **test-time and run-time defense** for an **amplitude modulation** α <15%.

[1] Dutta *et al.*, Information Hiding'13,
[2] J. Classen *et al.*, CNS'15,
[3] S. Chang *et al.*, TODAES'18,
[4] Y. Jin and Y. Makris, D&T'10,
[5] Y. Liu *et al.*, TVLSI'17,
[6] K. S. Subramani *et al.*, TIFS'20,
[7] A.R. Díaz Rizo *et al.*, TDSC'23.

State-of-the-Art on HT-CC detection techniques

Standard measurements: Spectral Analysis, EVM test

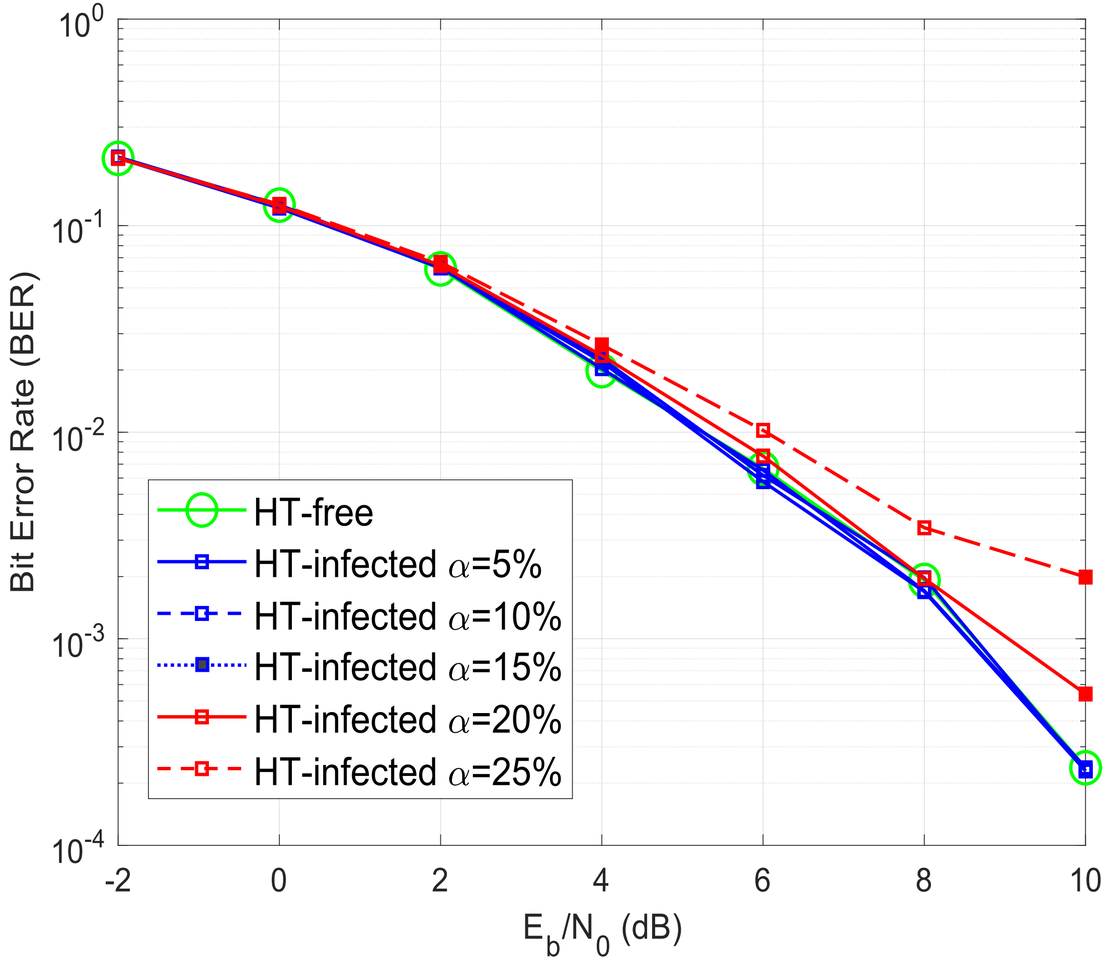




Check if transmissions comply with the wireless standard

State-of-the-Art on HT-CC detection techniques

Standard measurements: Bit Error Rate (BER) performance

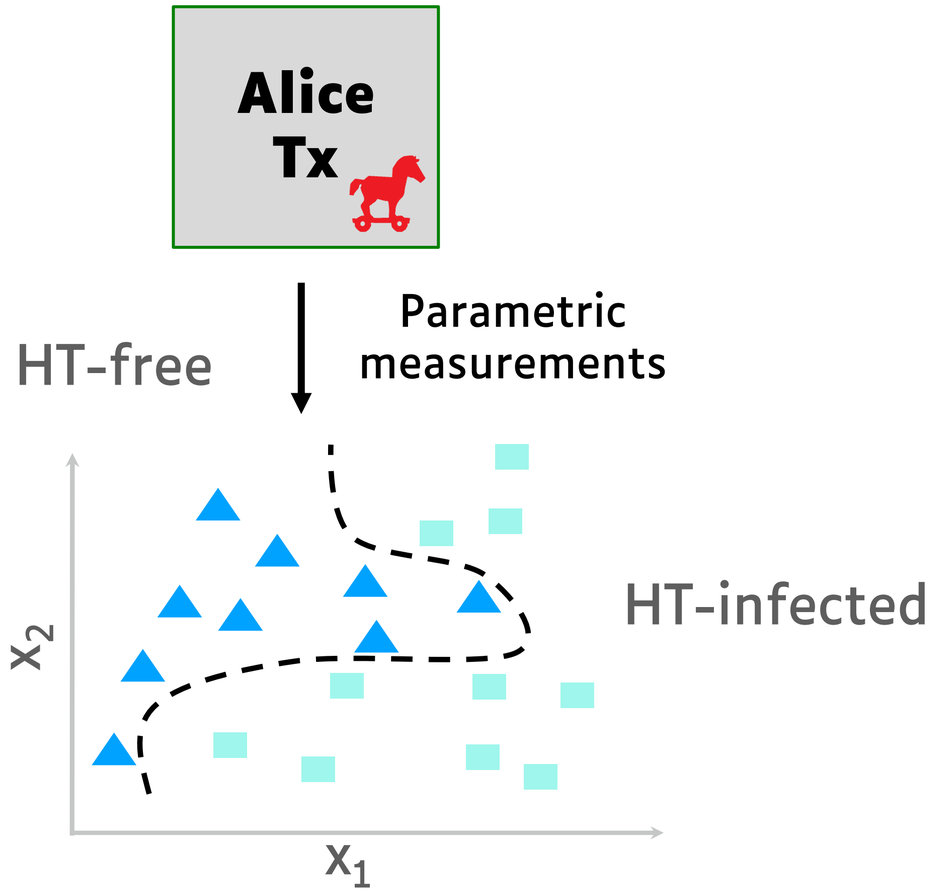


- Since there is no performance penalty for α ≤ 15%, HT is undetectable.
- We may not have a golden model (CC-free chips)

Check if transmissions have performance penalties compared to the golden model

State-of-the-Art on HT-CC detection techniques

Specialized measurements: Statistical Side-Channel Fingerprinting

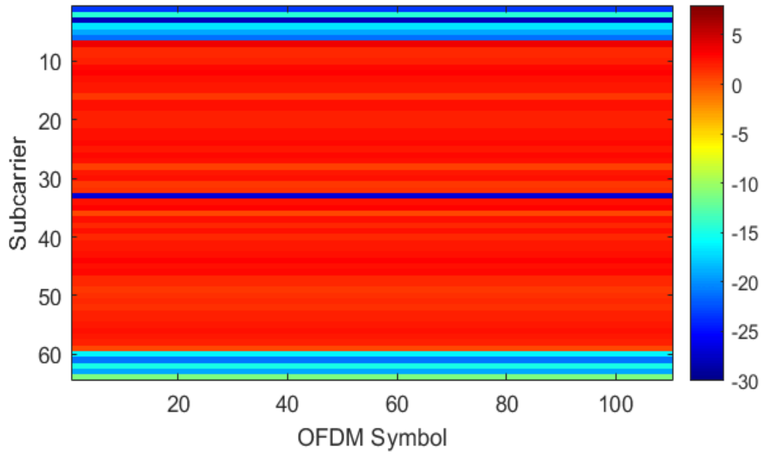


- A machine learning one-class classifier is trained in a space of parametric measurements (e.g. transmission power) to identify HT-free devices
- We may not have a golden model (CC-free chips)
- Only efficient for HT-CCs distorting transmission power

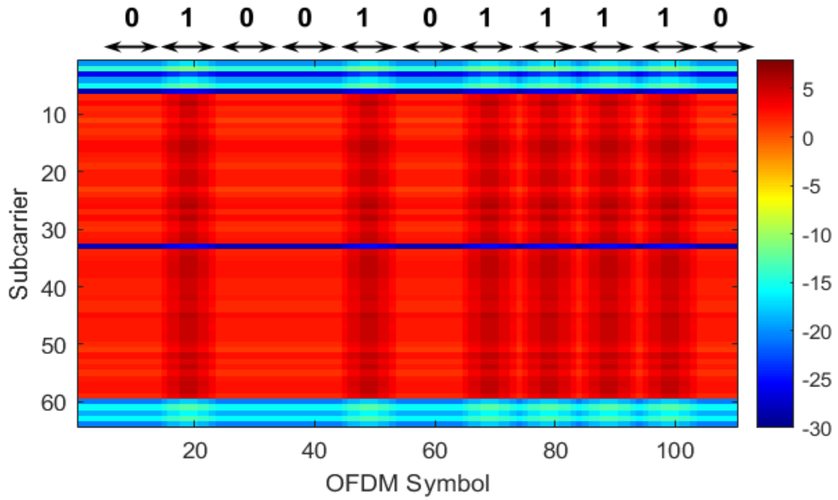
Check if transmissions have a different statistical fingerprint penalties compared to the golden model

State-of-the-Art on HT-CC detection techniques

Specialized measurements: Adaptive Channel Estimation



HT-free



HT-infected

Subramani *et al.*, TIFS'20

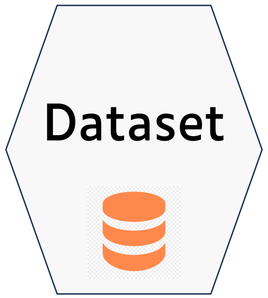
- Differentiates between channel impairments (noise, carrier frequency offset, etc.) and HT activity modulating the amplitude of the transmitted power
- Only effective for HT-CCs located in the DATA

Check for HT activity hidden in channel impairments

State-of-the-Art on HT-CC detection techniques

Defenses are very diverse and attack specific.

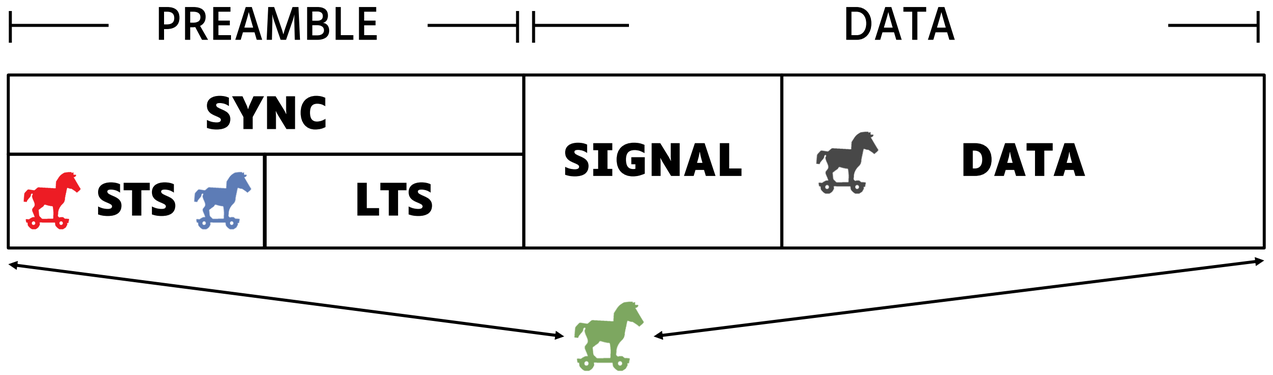
How to evaluate if a new defense is effective against all existing attacks?



Outline

HT-CC dataset planning

Ref.	Attack Model	Defense Mechanism	In Dataset
[1]	Leaks data through **amplitude modulation** of some **STS symbols** of the Preamble	**Evades** any known **test-time and run-time defense** for an **amplitude modulation** α <15%	**HT1-CC**
[2]	Leaks data by introducing an additional **phase shift** into all **STS symbols** of the preamble	Analysis of the preamble **constellations**	**HT2-CC**
[3]	Encodes leaked data on the **I/Q mapping** and hides the encoding by introducing imperfections to the transmitted signal	Certain tests, such as **EVM**, show a distinguishing behavior **compared to HT-free** operation	**HT3-CC**
[4,5,6]	Leaks data by **modulating amplitude and/or frequency of transmitted signal**	Statistical Side Channel Fingerprinting (**SSCF**); Adaptive Channel Estimation (**ACE**)	**HT4-CC**



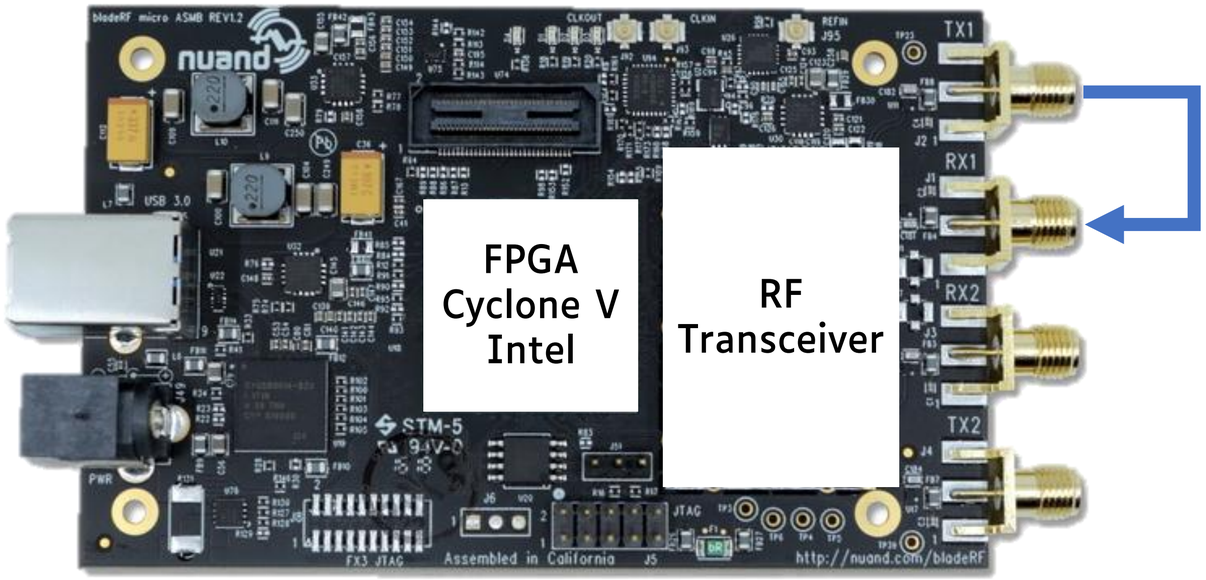
[1] A.R. Díaz Rizo *et al.*, TDSC'22,
[2] J. Classen *et al.*, CNS'15,
[3] Dutta *et al.*, Information Hiding'13,
[4] Y. Jin and Y. Makris, D&T'10,
[5] Y. Liu *et al.*, TVLSI'17,
[6] K. S. Subramani *et al.*, TIFS'20,

HT-CC dataset generation: hardware platform

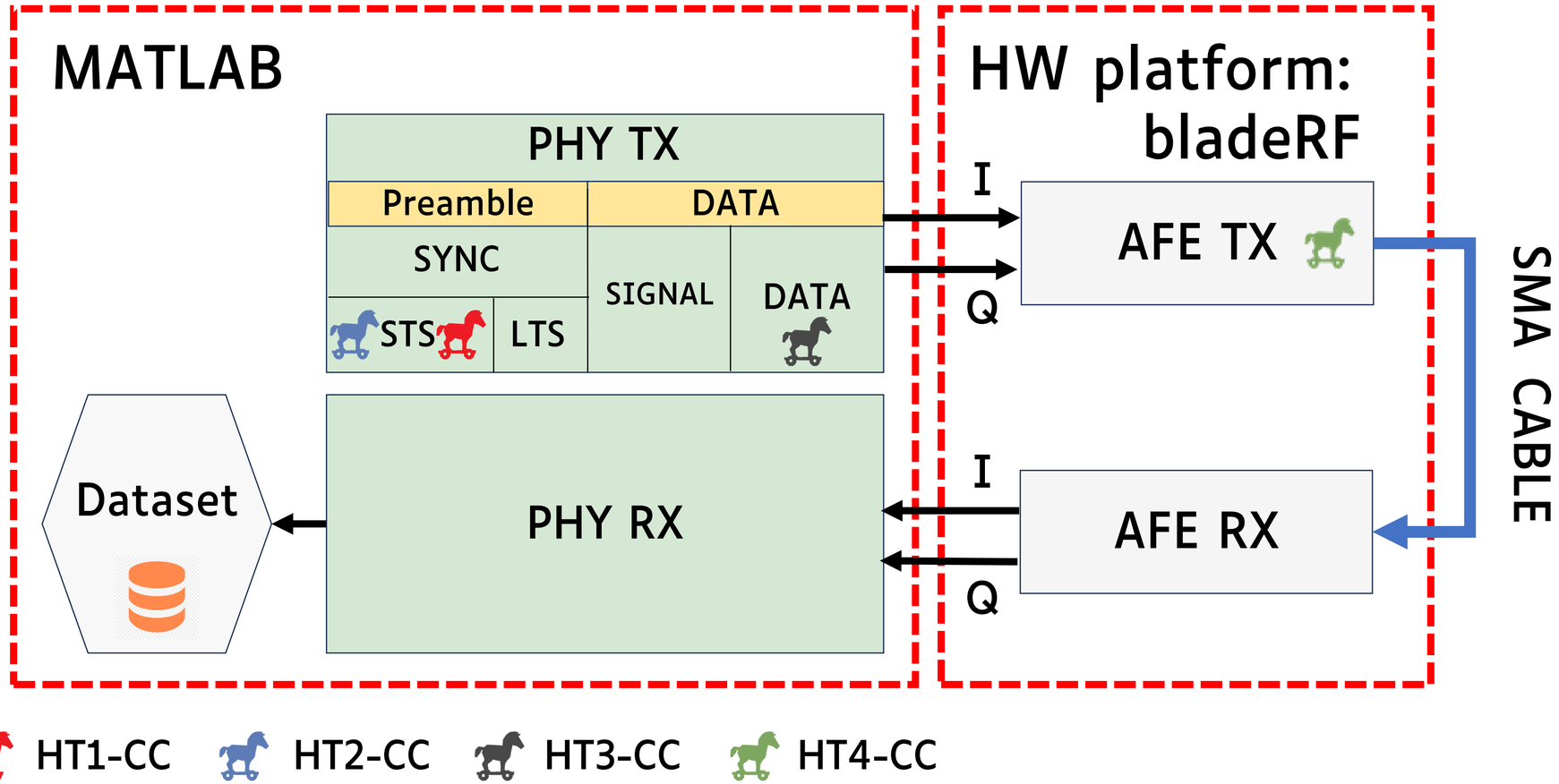
Software Defined Radio (SDR) bladeRF board from Nuand.

Hardware impairments affecting the signal at baseband and RF:

- Flicker Noise
- Quantification error
- DC offset
- I/Q imbalance
- Carrier Frequency Offset (CFO)
- Phase Noise
- Jitter



HT-CC dataset generation: framework



Parts : HT0-CC (CC-free), HT1-CC, HT2-CC, HT3-CC, HT4-CC
Number of acquisitions : 8 per part with SNR values ranging from 1dB to 29dB with a step of 4dB.
Number of frames : 2000 fixed-length OFDM IEEE 802.11 frames received with the hardware platform per acquisition.
Frame size : 640 complex-value I/Q samples.

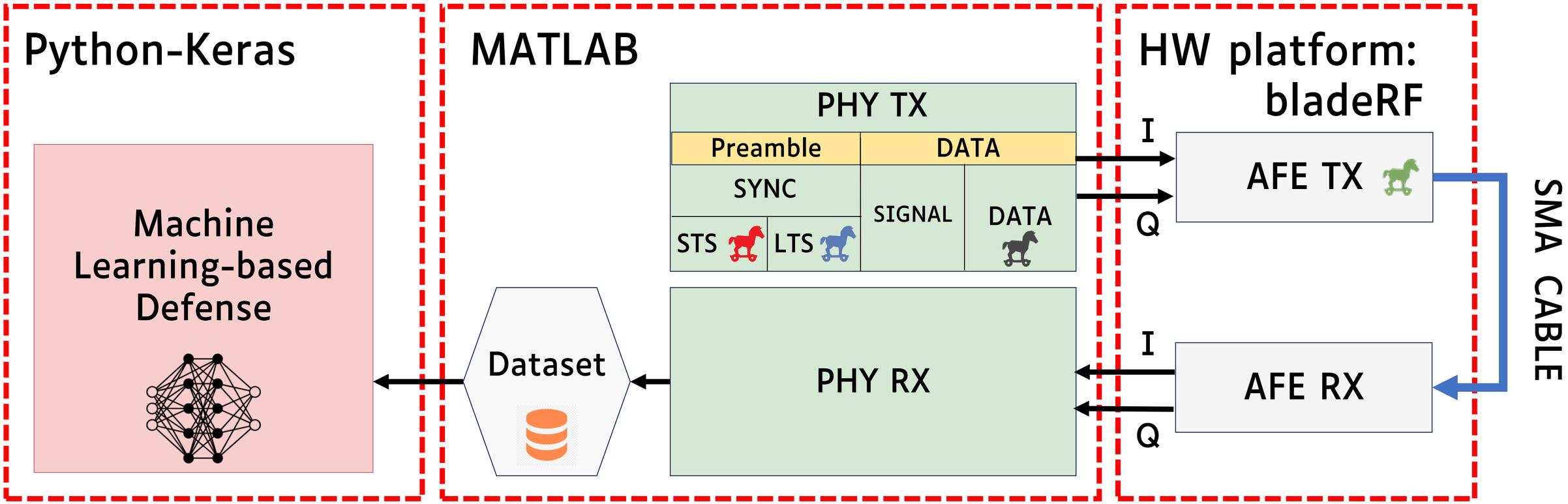
State-of-the-Art on HT-CC detection techniques

To provide maximum security, the defender is forced to combine simultaneously many of these countermeasures.

Defense cost rise exponentially.

Is it possible to have an effective defense against all existing attacks?

HT-CC dataset generation: framework



Parts : HT0-CC (CC-free), HT1-CC, HT2-CC, HT3-CC, HT4-CC
Number of acquisitions : 8 per part with SNR values ranging from 1dB to 29dB with a step of 4dB.
Number of frames : 2000 fixed-length OFDM IEEE 802.11 frames received with the hardware platform per acquisition.
Frame size : 640 complex-value I/Q samples.

Outline

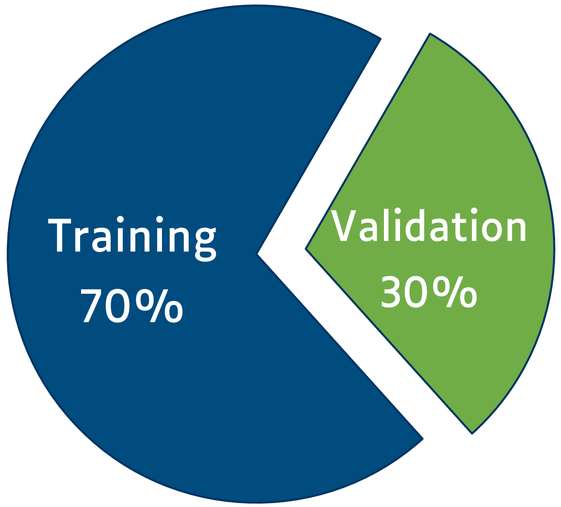
Proposed AI-based defense for HT-CC detection

Task: Binary classification, i.e., distinguishing CC-infected frames

1st Try: One-class Support Vector Machine (SVM) classifier

Feature: Raw received frames

Training set: HT0-CC (CC-free)



Result: Poor prediction accuracy
Max accuracy: 75%
Average accuracy (All SNRs): 66%

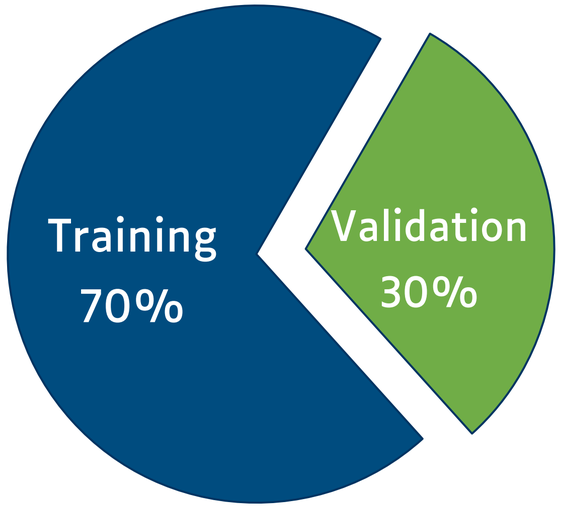
Proposed AI-based defense for HT-CC detection

Task: Binary classification, i.e., distinguishing CC-infected frames

2nd Try: Deep Neural Network (DNN) classifier ➔ Convolutional Neural Network

Feature: Raw received frames encoded as a 2x640 "image"

Training set: HT0-CC (CC-free) and HTX-CC (all CC-infected combined)



Result: Very good prediction accuracy

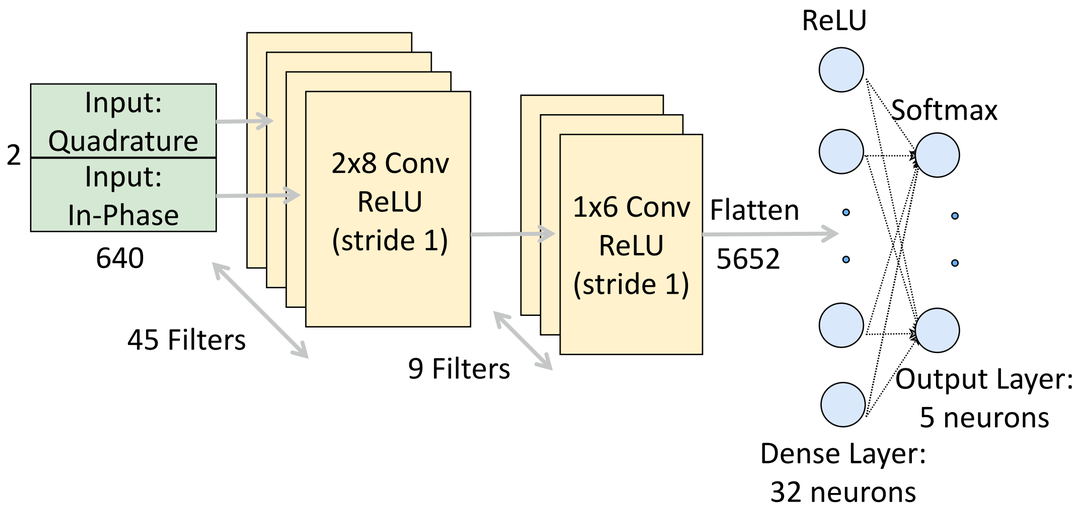
Max accuracy: 99% for SNR > 20dB

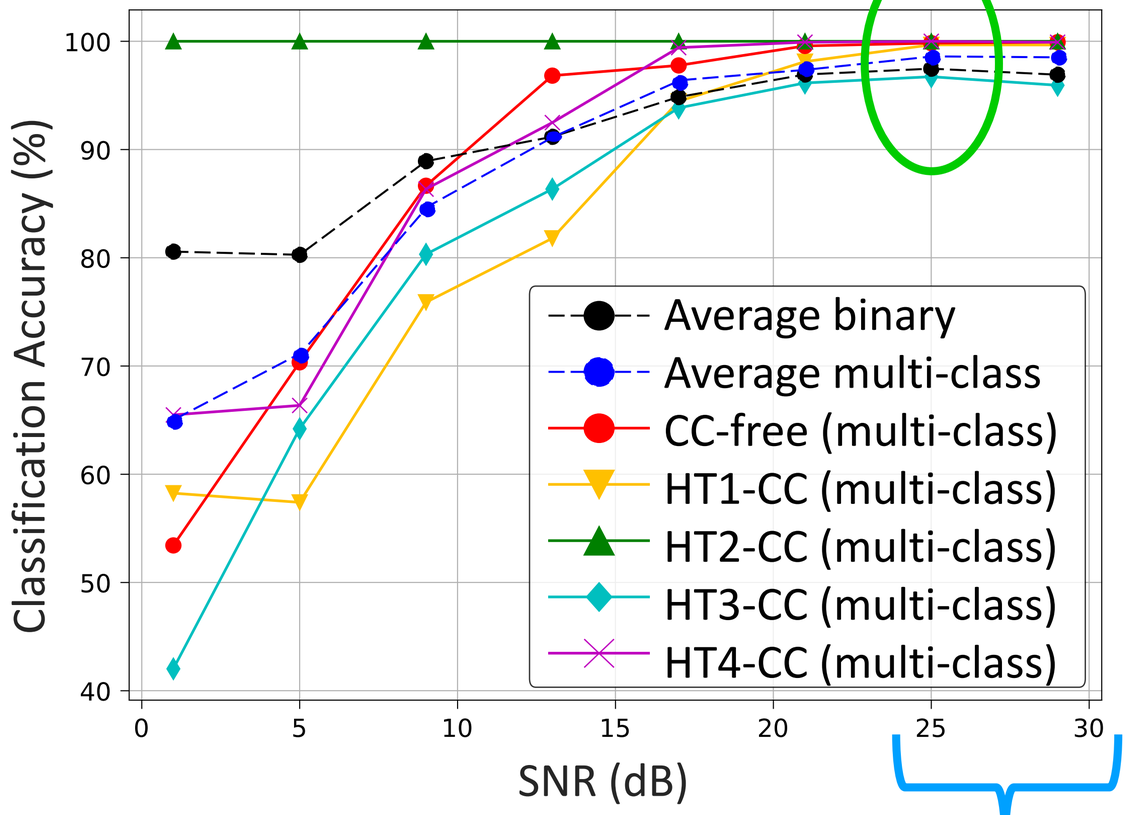
Proposed AI-based defense for HT-CC detection

Task 1: Binary classification, i.e., distinguishing CC-infected frames

Task 2: Multi-class classification, i.e., distinguishing every class

CNN architecture

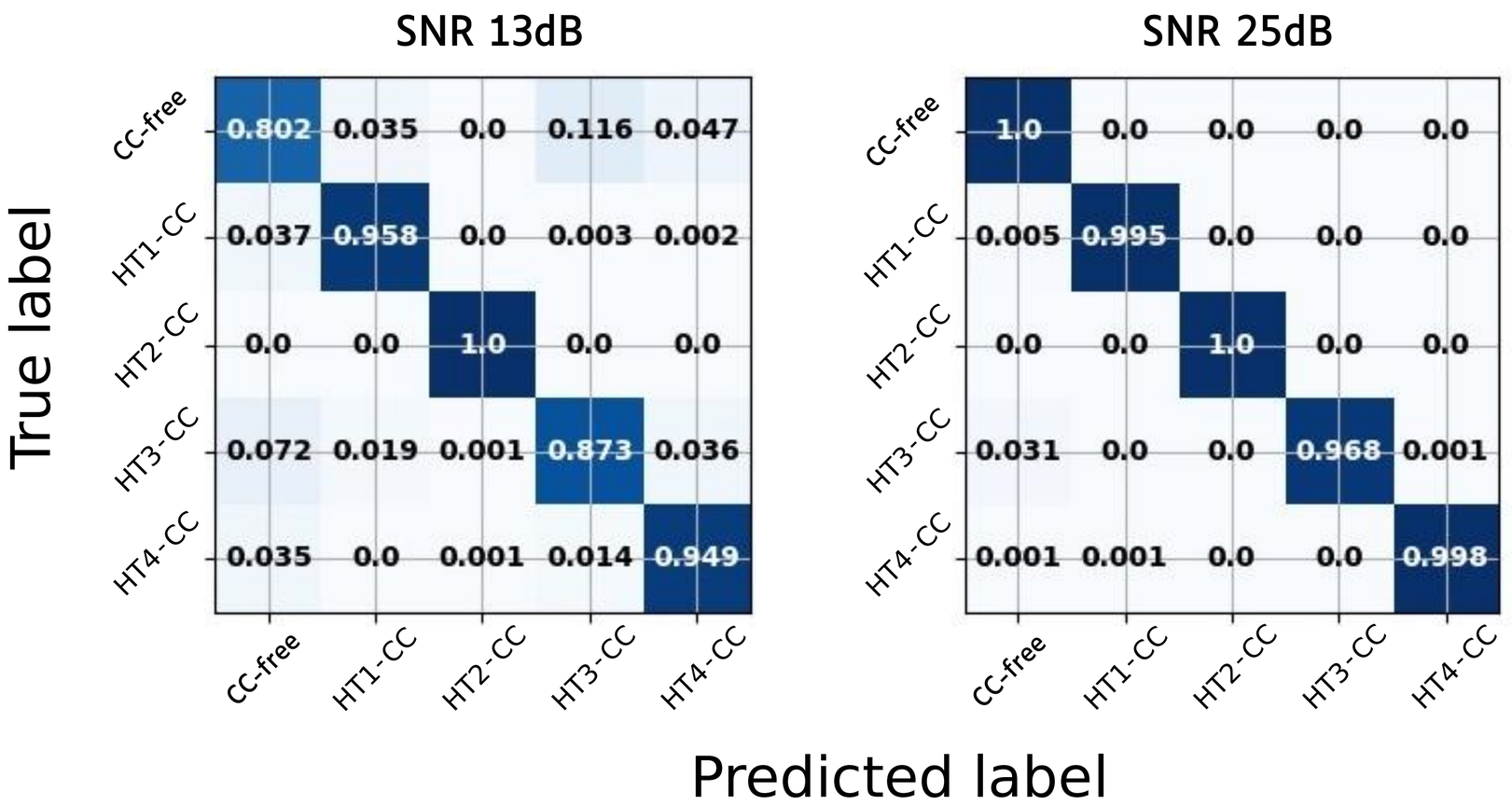




Following a trial-and-error approach, we reduced the architecture maintain the maximum accuracy

For robust WiFi communication, SNR > 25 dB

AI-based defense for HT-CC detection



CC-free as CC-free 100%

Misclassifications:

HT1-CC as CC-free 0.5%

HT3-CC as CC-free 3.1%

HT4-CC as CC-free 0.1%

HT3-CC as HT4-CC 0.1%

HT4-CC as HT1-CC 0.1%

The AI-based defense can detect CC-infected communications and the underlying HT mechanism

Outline

Conclusion

1. We generated on hardware and made publicly available the first HT-CC dataset comprising CC transmission originated from infected transmitters with four different HT mechanisms.

2. We proposed a novel single run-time defense based on deep learning that achieved over 99% detection accuracy on the dataset for the SNR range of interest.

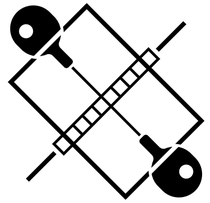
Publications:

- A. R. Díaz-Rizo, H. Aboushady and H. -G. Stratigopoulos, "Leaking Wireless ICs via Hardware Trojan-Infected Synchronization," in IEEE Transactions on Dependable and Secure Computing (**IEEE TDSC**), vol. 20, no. 5, pp. 3845-3859, 1 Sept.-Oct. 2023
- A. R. Díaz-Rizo, A. E. Abdelazim, H. Aboushady and H. -G. Stratigopoulos, "Covert Communication Channels Based On Hardware Trojans: Open-Source Dataset and AI-Based Detection," 2024 IEEE International Symposium on Hardware Oriented Security and Trust (**HOST**), Tysons Corner, VA, USA, 2024

Open-source : https://github.com/alandr918/Hardware-Trojan-Covert-Channel-dataset

Work in Progress (WP) and Future Work (FW)

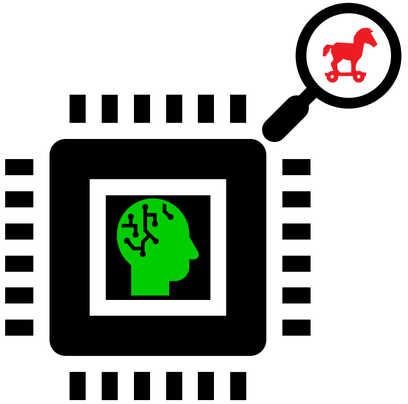
WP-1: Enhance the dataset



+ more HT-CCs

+ more features (e.g., channel conditions)

WP-2: Hardware accelerator embedded inside wireless IC



FW: Dataset of HT-CC in other wireless protocols

Wi Fi · NFC · zigbee · 4G · 5G · 6G · NB-IoT · Z-WAVE · sigfox · LoRa®

Thank you for your attention!

alan-rodrigo.diaz-rizo@lip6.fr