AutoHack: A Physically Verified Multi-Bus CAN Dataset for IDS Evaluation

Yuchan Song[1], Sejun Ahn[2], Seungjin Baek[1], Hyeonseong Kim[2], Huy Kang Kim[1], Sanghoon Jeon[2]

[1] HCRL, Korea University· [2] MOSE Lab, Kookmin University

Presenter: Sejun Ahn

VehicleSec'26

KOOKMIN UNIVERSITY

KOREA UNIVERSITY

AutoHack Dataset



AutoHack 2025 — on-vehicle hacking competition
https://mose.kookmin.ac.kr/mose/AutoHack2025.do

USENIX Artifact Evaluation



All three artifact badges awarded

ACADEMIC RESEARCH & EDUCATIONAL USE

Dataset available on Zenodo
https://zenodo.org/records/20321379
Released under CC BY 4.0 — reuse with attribution.

Introduction: Research Background

01 Inherent Vulnerabilities of CAN

• Designed **without baseline security — lacks authentication, integrity, confidentiality.**

- Highly susceptible to **spoofing, replay, arbitrary injection attacks**.

02 Rise of Vehicular IDS & Dataset Dependency

• **Intrusion Detection Systems (IDS)** have emerged as a **core defensive technology.**

- Detection accuracy **depends entirely on the quality of training data**.

03 Imperative for Realistic, High-Quality Datasets

• Must reflect complex real-world traffic — **dynamic ECU interactions, event-driven signals**.

- Essential to ensure the **practical effectiveness and reproducibility of IDS research**.

Introduction: Limitations of Current CAN Datasets

01

Over-reliance on Periodicity

• Legacy benchmarks strictly assume **fixed periodic intervals**.

• Misidentifies legitimate **event-driven aperiodic traffic** as attacks, causing high false-positive rates.

02

Single-Bus Line Limitations

• Public vehicle datasets are restricted to a **single CAN bus interface**.

• Unable to analyze cross-domain interactions or integrated attack propagation across modern gateway architectures.

03

Unrealistic Injection Patterns

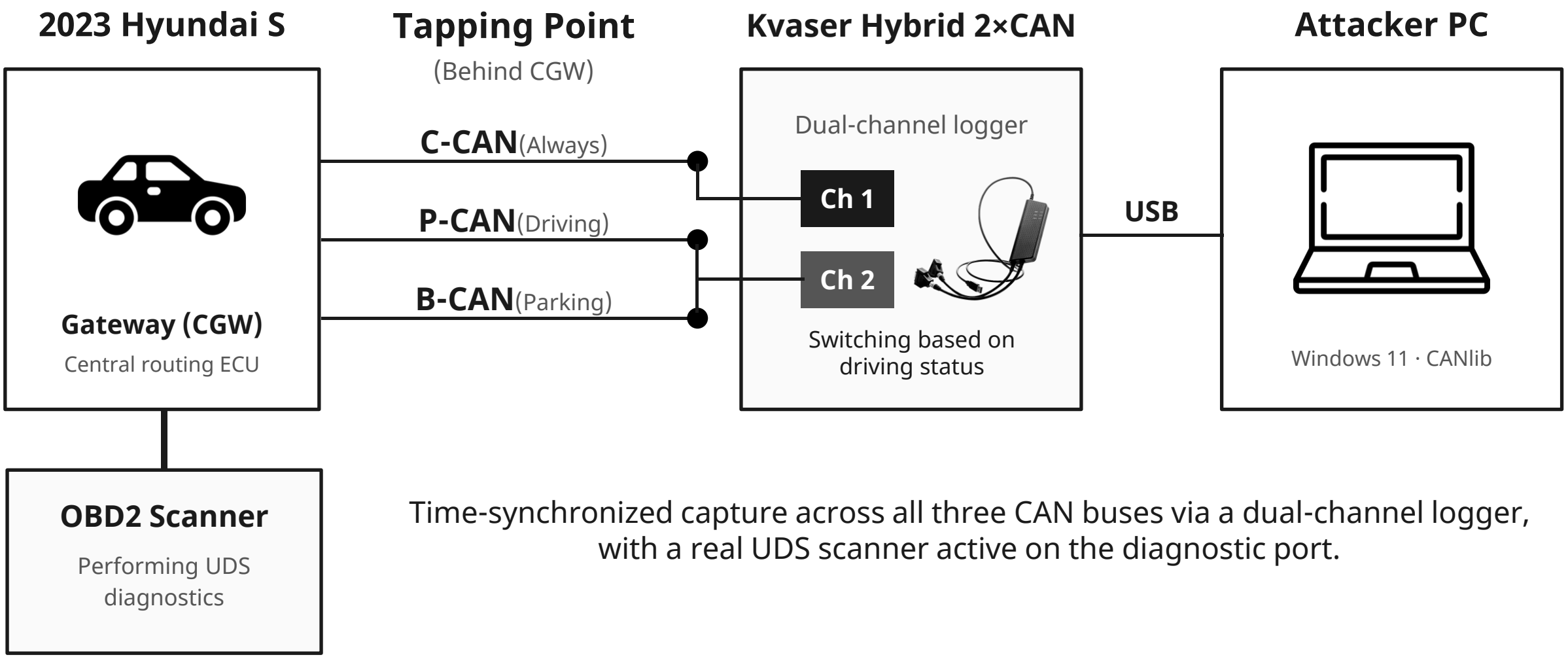
• **Post-editing and simulation** methods fail to capture real-time logical responses.

• Miss the dynamic feedback of physical ECUs when attacks are actually injected on the bus.

AutoHack 6 Characteristics

#	Characteristic	Conventional	AutoHack (Ours)
01	**Data Collection Environment**	Limited to simulations / testbeds; fails to capture real-world delays.	Real vehicle driving environments **under true physical conditions.**
02	**Attack Injection Method**	Offline post-editing; missing interactive ECU response and dynamic feedback.	Real-time attack injection, **models learn unpredictable threats.**
03	**Temporal Properties**	Heavily periodic-reliant; ignores event-driven automotive traffic complexity.	Integrated periodic + aperiodic **networks — precise evaluation baselines.**
04	**Network Topology**	Confined to single-bus; no cross-bus routing or attack propagation.	Time-synchronized multi-bus — **modern gateway architecture.**
05	**Protocol Coverage**	Restricted exclusively to standard CAN networks.	Standard CAN + **real-time UDS diagnostic sequences.**
06	**Label Reliability**	Post-collection heuristics based on ID/time patterns; precision errors under load.	Real-time labeling + **physical hazard verification.**

AutoHack Dataset: Data Collection Methodology



Time-synchronized capture across all three CAN buses via a dual-channel logger, with a real UDS scanner active on the diagnostic port.

AutoHack Dataset: Attack Injection Scenarios

DoS

Method: Flood the **highest-priority ID (0x000)** at a strict **0.2ms interval**.
Objective: Seize entire network bandwidth and block legitimate ECU communications.
Physical Response: Data continuity breaks → **warning lights & error alarms in major ECUs.**

Fuzzing

Method: Inject **random combinations** of IDs, DLCs, and payloads at **1ms intervals**.
Objective: Maximize probability of triggering unintended system-level exceptions.
Physical Response: Vehicle **suddenly stops in Drive (D) mode**, unintended eCall activation.

Replay

Method: Retransmit pre-collected normal data, **synchronized with original hardware timestamps**.
Objective: Replicate legitimate traffic intervals to deceive timing/frequency-based IDS.
Physical Response: **Control interference** — arbitrary reproduction / neutralization of specific functions.

AutoHack Dataset: Attack Injection Scenarios

Spoofing

Method: Inject consecutive messages (1-10 frames) using valid IDs (e.g., EPB, PAS) at **1ms intervals**.

Objective: Mimic legitimate burst signals to remain **indistinguishable from normal traffic**.

Physical Response: RPM gauge fluctuation, **EPB warning indicators**, and **PAS logic errors**.

UDS Spoofing

Method: Re-implement legitimate diagnostic sequences extracted from authentic OEM tools targeting **safety-critical ECUs**.

Objective: Transition core ECUs into **restricted diagnostic modes** or deactivate specific functions beyond simple bus injection.

Physical Response: **Unauthorized diagnostic session entry** — restriction of specific ECU functions.

ID (Hex)	Function	Physical Response
0x366	Engine RPM	RPM gauge fluctuation
0x390	Warning Alarm	Unexpected alert trigger
0x490	EPB Error	EPB warning indicator
0x4F4	Parking Assist	PAS logic error
0x50C	Fuel Warning	Fuel warning trigger

Detailed specifications of targeted C-CAN IDs for Spoofing attacks.

ID (Hex)	Targeted ECU	Physical Response
0x755	Right Radar	Diagnostic state entry
0x7B7	Left Radar	Sensor session lock
0x7C4	Front Camera	Camera mode deactivation
0x7D4	Power Steering	Steering assist limitation

Detailed specifications of targeted UDS-based IDs on C-CAN for Spoofing attacks

AutoHack Dataset: AutoHack Dataset Description

Session-based Data Splitting

Separated Train and Test sets into completely **independent driving sessions at a roughly 6.6:3.4 ratio.**
Prevents evaluation bias caused by temporal continuity, ensuring the model is tested on truly distinct driving sequences.

Mutually Exclusive Spoofing IDs

Assigned spoofing target IDs mutually exclusively between the Train and Test sets to prevent model overfitting to specific identifiers

Metric	Train Set	Test Set
Total Frames	6,922,600	2,711,497
Driving Duration (sec)	1,400.8182	583.1505
Unique CAN IDs	2,048	2,048
Normal Frames	6,593,770	2,557,180
Attack Frames	328,830	154,317

AutoHack Dataset statistics

Bus Interface	C-CAN		P-CAN		B-CAN	
Attack Type	Train Set	Test Set	Train Set	Test Set	Train Set	Test Set
Total Frame	4,026,488	1,678,489	2,655,149	883,323	240,963	149,685
Fuzzing	38,191	13,147	39,952	10,319	57,459	29,698
Replay	24,123	16,876	25,797	16,746	5,589	8,408
DoS	14,419	8,490	34,441	7,647	28,244	19,492
Spoofing	22,088	5,384	23,634	5,188	12,846	11,698
UDS Spoofing	580	420	887	344	580	460

Detailed attack statistics by interface and session

AutoHack Dataset: Distinctive Features

01

Overcoming Periodicity Limitations

• **Aperiodic Traffic:** Integrates **event-driven communications like UDS**.

• **Precise Replay Attacks**: Accurately restores **inter-arrival time of original traffic**.

• **Impact:** Neutralizes legacy IDS relying on simple timing analysis — **demands advanced detection**.

02

High Complexity & Entropy from Real Driving

• **Real-World Variations:** Actual roads (acceleration, deceleration, turning), **high message variability**.

• **Independent Data Splits:** Train and Test in **fully independent driving sequences** — not random split.

• **Impact:** Rigorously evaluates the model's **true generalization capabilities**.

03

Synchronized Multi-Bus Data

• **Simultaneous 3-bus capture: C-CAN, P-CAN, B-CAN** via a dual-channel Kvaser Hybrid with switching.

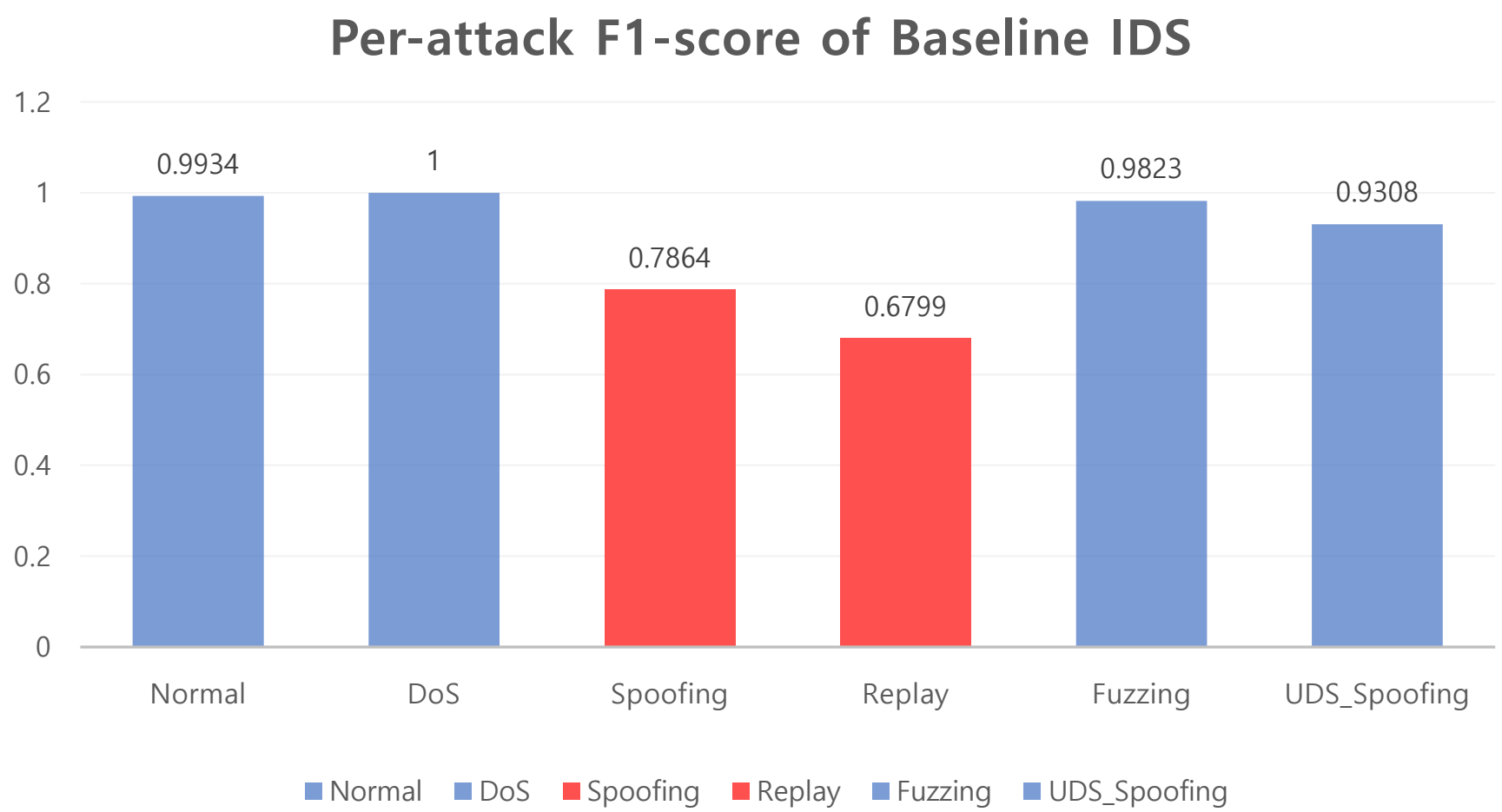
• **Cross-bus label consistency:** Enables analysis of **overlapping IDs and cross-domain attack propagation**.

• **Impact: First public benchmark** that mirrors the actual gateway topology of modern vehicles.

Evaluation Setup

Task	**6-class classification** — Normal, DoS, Spoofing, Replay, Fuzzing, UDS Spoofing
Models	Random Forest, XGBoost
Feature Sets	**8-feature baseline, 38-feature multi-bus context**
Metrics	Macro F1, Confusion Matrix, Feature Importance
Reproducibility	Benchmark code and experimental scripts publicly released on Zenodo · **Reproduced Artifact Badge** https://zenodo.org/records/20321379

Observation 1: Variations in Detection Difficulty AcrossAttack Types

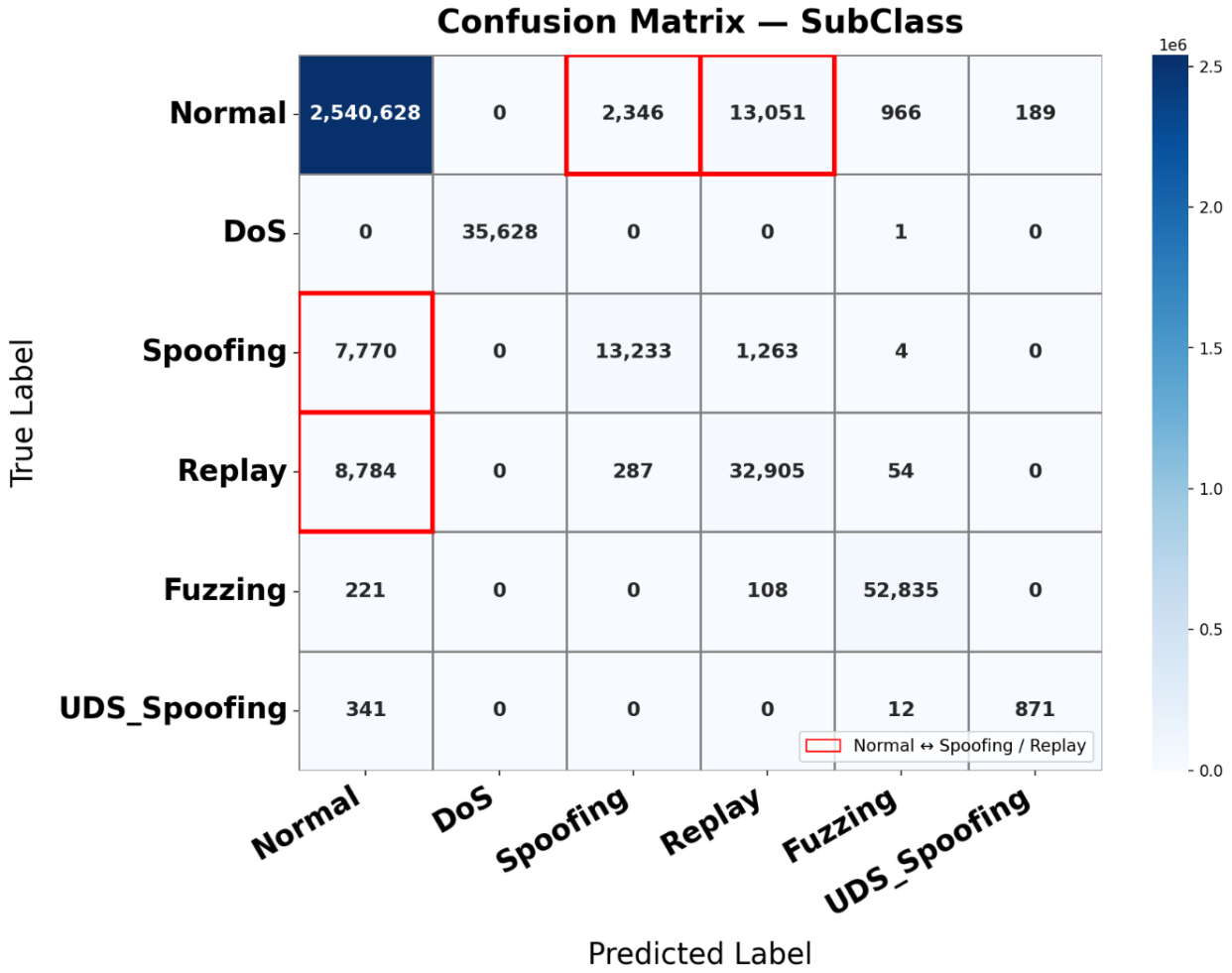


Key Observations

- Detection difficulty **varies by attack type**.
- Most attacks achieve high F1-scores (DoS: **1.0000**, Fuzzing: **0.9823**, UDS Spoofing: **0.9308**).
- Replay and Spoofing show clear drops (Replay: **0.6799**, Spoofing: **0.7864**).

Observation 1: Variations in Detection Difficulty AcrossAttack Types

Confusion matrix of the Random Forest model

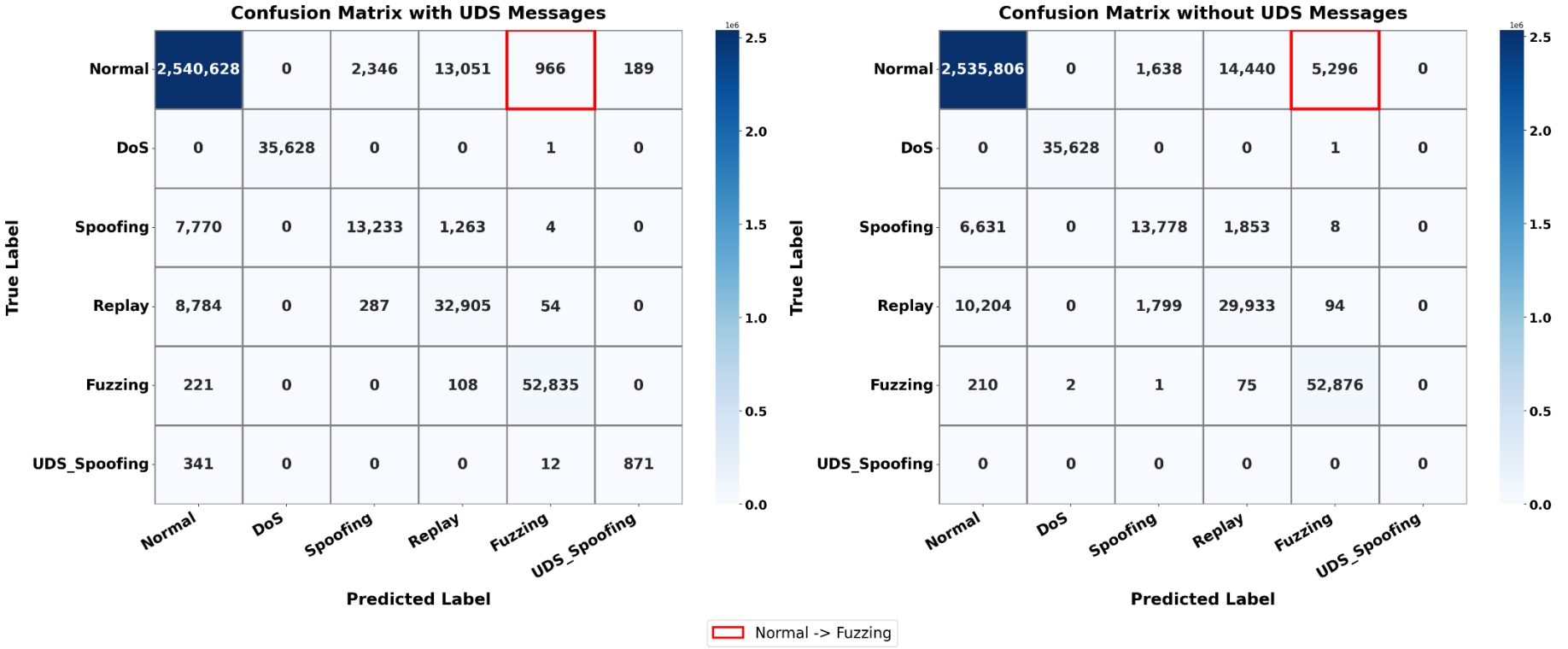


Key Observations

- Replay and Spoofing are often **confused with Normal traffic timing**.
- Replay reproduces **legitimate traffic timing**.
- Spoofing uses **valid control IDs and short burst messages**.
- This confirms AutoHack as a **realistic benchmark for evaluating IDS generalization**.

Observation 2: Impact of Non-Periodic Communication on Detection Boundaries

RF-based IDS performance with and without non-periodic UDS messages



Key Observations

- Removing UDS increased Normal → Fuzzing errors by **5.5×** (**966 → 5,296**).
- UDS enriches the Normal class and **prevents periodicity-biased detection**.
- Other classification patterns remained broadly similar.

Observation 3: Empirical Characteristics of the Time-Synchronized Multi-Bus Setting

Performance Comparison: Single-bus vs. Multi-bus

Model	Bus	Single-bus	Multi-bus	Δ
RF	B-CAN	0.8411	0.6840	−0.1572
	C-CAN	0.8368	0.7299	−0.1069
	P-CAN	0.8880	0.7775	−0.1104
XGBoost	B-CAN	0.7879	0.7339	−0.0540
	C-CAN	0.8613	0.7505	−0.1109
	P-CAN	0.8236	0.7791	−0.0445

F1-score per bus: models trained on a single bus vs. jointly on all buses (Δ = Multi − Single)

Key Observations

- Single-bus models were **trained separately** for each CAN bus.
- The multi-bus model was **trained jointly** and evaluated on each bus.
- Both models showed **lower performance** under multi-bus training.
- This raises the question of whether **the degradation is solely due to domain shift**.

Top 4 Feature Importance Across Models

Random Forest (RF)

Regime	Rank 1	Rank 2	Rank 3	Rank 4
Single-B-CAN	Data_Frequency	ID_Prev_Interval	Data_Prev_Interval	ID_Frequency
Single-C-CAN	ID_Prev_Interval	ID_Frequency	IAT_MEAN	Data_Prev_Interval
Single-P-CAN	ID_Prev_Interval	ID_Frequency	IAT_MEAN	Data_Prev_Interval
Multi-bus	ID_Frequency	ID_Prev_Interval	IAT_MEAN	Data_Prev_Interval

XGBoost

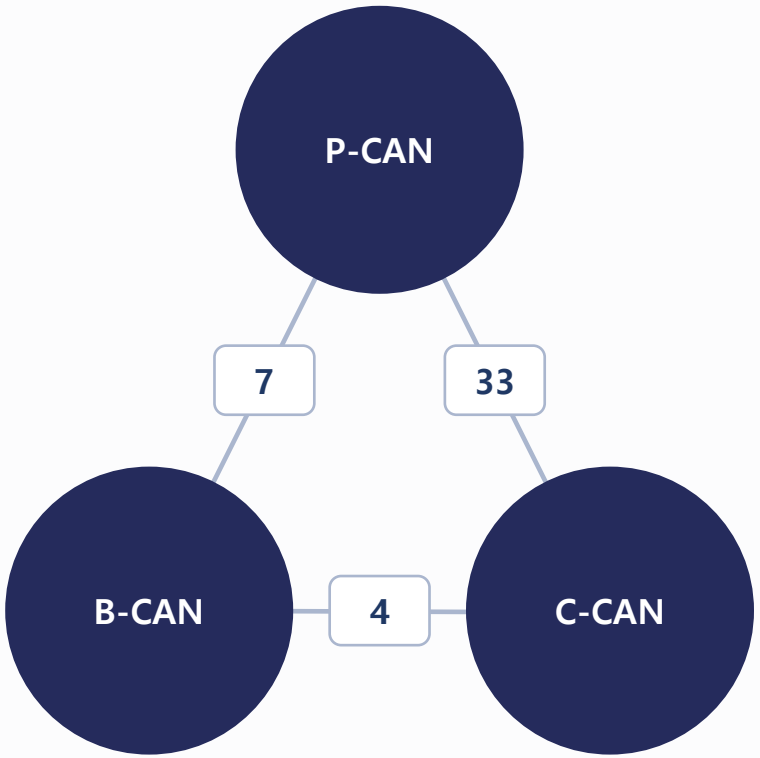
Regime	Rank 1	Rank 2	Rank 3	Rank 4
Single-B-CAN	ID_Frequency	Data_Frequency	Data_Prev_Interval	DLC
Single-C-CAN	IAT_MEAN	ID_Prev_Interval	ID_Frequency	CAN_ID
Single-P-CAN	IAT_MEAN	CAN_ID	ID_Prev_Interval	ID_Frequency
Multi-bus	ID_Frequency	IAT_MEAN	CAN_ID	ID_Prev_Interval

Key Observations

- Feature importance was evaluated across **38 features**.
- ID-related features **ranked highly in both models**.

Observation 3: Empirical Characteristics of the Time-Synchronized Multi-Bus Setting

Overlapping CAN IDs in the Multi-bus Setting



40

total overlapping CAN IDs identified

P-CAN ↔ B-CAN	**7**
P-CAN ↔ C-CAN	**33**
B-CAN ↔ C-CAN	**4**

Key Observations

- **40 unique CAN IDs** were shared across multiple buses (P ↔ B: 7, P ↔ C: 33, B ↔ C: 4).
- Some overlapping IDs had **inconsistent labels across buses**, potentially complicating cross-bus analysis.

Observation 3: Empirical Characteristics of the Time-Synchronized Multi-Bus Setting

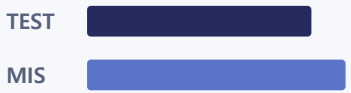
Overlap-ID Ratio Analysis: Full Test Set vs. Misclassified Samples

r_{test} = overlap-ID ratio in the full test set r_{mis} = overlap-ID ratio among misclassified samples (RF, multi-bus)

Normal

46.12% → 53.14%

r_test r_mis



↑

Overlap ratio increases (+7.02%p)

Spoofing

25.47% → 25.75%

r_test r_mis



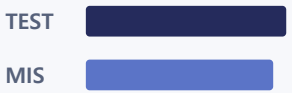
≈

Overlap ratio stable (+0.28%p)

Replay

41.20% → 38.52%

r_test r_mis



↓

Overlap ratio decreases (−2.68%p)

No single uniform tendency across classes → the multi-bus degradation cannot be explained by domain shift alone; overlap-ID structure shapes each class's error composition differently.

Key Observations

- We compared the share of **overlap-ID messages** in the full test set (r_test) with that in the misclassified subset (r_mis).
- The direction differs by class: **Normal increases, Spoofing remains stable, and Replay decreases**.
- This class-wise divergence indicates that the multi-bus setting is not a single uniform effect — **domain shift alone cannot explain the degradation**.

Limitations & Future Work

Limitations

- **Single-vehicle scope**
  Collected from one 2023 Hyundai model under a finite set of driving conditions —
  results are not evidence of vehicle-independent generalization.
- **Vehicle-specific attributes**
  CAN IDs, payload semantics, gateway routing, and diagnostic implementations **vary across manufacturers, models, and E/E architectures.**
- **Representative attack set**
  Five major attack types are covered; the dataset **does not comprehensively span all possible attack vectors.**

Future Work

- **Extend across platforms**
  Apply the same collection & evaluation protocol to multiple vehicle models and manufacturers —
  enabling cross-vehicle, cross-bus, and cross-manufacturer IDS evaluation as a natural next step.
- **Advanced IDS design**
  Develop architectures that explicitly model **inter-bus interactions and aperiodic (UDS) communication patterns.**

Thank You

DATASET & ARTIFACT

AutoHack Dataset & Benchmark Code
zenodo.org/records/20321379
DOI 10.5281/zenodo.20321379 · CC BY 4.0







CONTACT

HCRL, Korea University
Hacking & Countermeasure Research Lab
hksecurity.net · cenda@korea.ac.kr

MoSE Lab, Kookmin University
Mobility Cybersecurity Lab
mose.kookmin.ac.kr · sh.jeon@kookmin.ac.kr

Presenter — Sejun Ahn
s3037444@kookmin.ac.kr