**Decision Tree to Find Signal Relations for Rule-Based Intrusion Detection**

## Abstract

## Introduction

1. Background information — CAN

2. Importance of your study

Recent approaches: Signature detection (attacks are known)

Anomaly detection: statistical, knowledge based, machine learning based.

3. Introducing your study

## Background

CAN structure

Attack model

## Current work

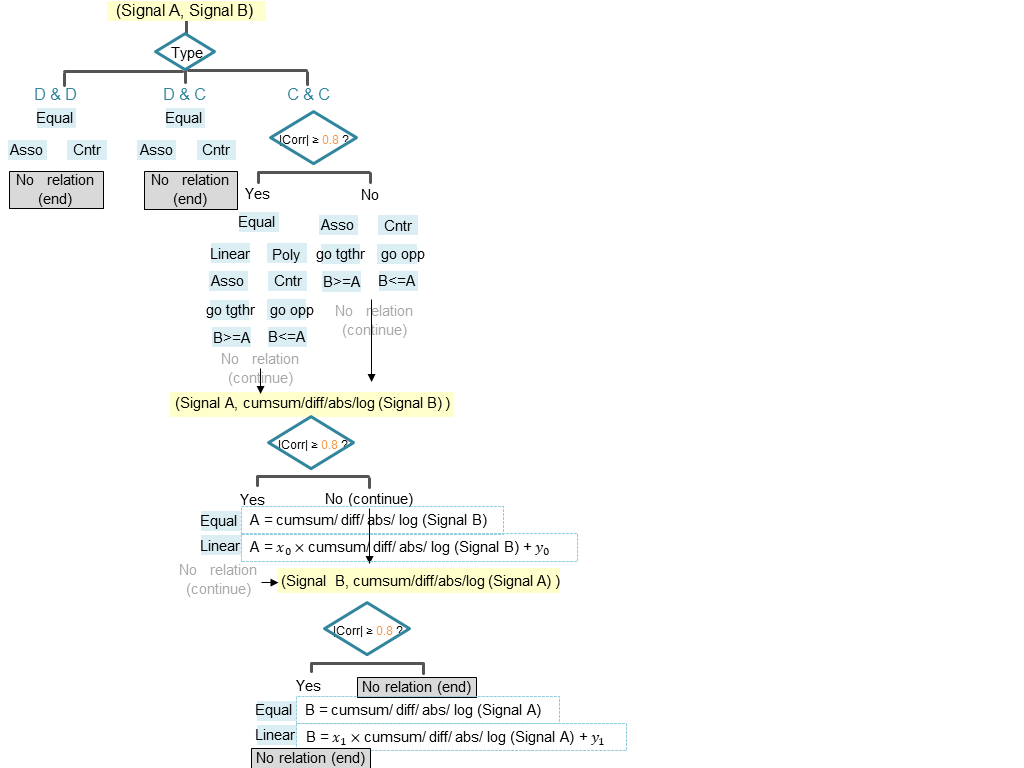
Rule based – weak

ML based – overhead

## Method (A systematic, step by step account)

1. Relation types
2. **Automatically find relation using decision tree**

In order to find the relationship of every signal pairs appeared in a CAN log, we designed a decision-tree algorithm, shown in fig. . After preliminary cleanup, every signal is paired with the others and goes through the decision tree to see whether the two signals have relation or not, making sure no relations are overlooked.



***Signal types***

After analyzing CAN matrixes, we found that there are several signals indicating the status of switches (e.g. 0 for releasing the brake pedal and 1 for braking) or directions (e.g. 0 for go straight, 1 for turn right, 2 for turn left). There are also many signals changing smoothly with time (e.g. speed signal). The two kinds of signals are both important in signal relationship, but the relationship types between them can have different meanings. We classified the first type of signal as “Discrete” and the second type as “Continuous” based on the number of unique values in signals’ data payload. The chosen of “unique value threshold” should be done carefully to prevent misclassifying. When pairing two signals, there should be 3 kinds of signal pairs in total.

D & D: both signals are discrete

D & C: one signal is discrete, the other is continuous

C & C: both signals are continuous

For the three different types of signal pairs, we check different kinds of relationships.

***Relation types***

**Equal:** two signals are equal to each other

**Associate (Asso):** A and B should be zero/non-zero together

**Counteract (Cntr):** A and B cannot be non-zero together

**Linear:** A = B +

**Polynomial (Poly):** A = + + … + +

**Go together (go tgthr):** A and B increase/ decrease together

**Go opposite (go opp):** A and B increase/ decrease oppositely

**B>=A:** B always greater or equal to A

**B<=A:** B always less or equal to A

**Cumsum:** Cumulative Sum of …

**Diff:** Differentiate of …

**Abs:** Absolute value of …

**Log:** log of …

***Correlation Coefficient***

1. Intruder

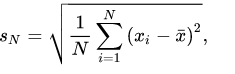
## Experiment

***Cleanup***

In a full CAN log file, there could be hundreds of Arbitration IDs and even much more signals, including some that remained unchanged all the time or some unimportant signals. Pairing each of the signals and find their relations can be time wasting, so we decided to discard signals in the following cases.

1. Low Standard Deviation

In statistics, the standard deviation represents the variation or spreading of values. Lower standard deviation indicates that the values are likely to be close to the mean of dataset, while higher standard deviation indicates that the values are more spread out. The formula for population standard deviation (of a finite population) is:



We discarded the signals with their value having low standard deviation (threshold = standard deviation threshold), which means the signals’ value seldom changed during the car running time.

1. Non-periodic signals

There are some event signals (e.g. door open) which are not sent periodically. We discarded them because …

***Find relations from logs***

***Test***

## Results and discussion

## Conclusion

## Future work