Parameter Estimation for Decoding Sensor Signals

Matthew Walter Nice
Matt Bunting
Gergely Zachar
Rahul Bhadani
matthew.nice@vanderbilt.edu
Vanderbilt University
Nashville, TN, USA

Paul Ngo
Jonathan W. Lee
Alexandre Bayen
University of California, Berkeley
Berkeley, CA, USA

Dan Work
Jonathan Sprinkle
Vanderbilt University
Nashville, TN, USA

Keywords: cyber-physical systems, in-vehicle network, signal decoding

ACM Reference Format:
Matthew Walter Nice, Matt Bunting, Gergely Zachar, Rahul Bhadani, Paul Ngo, Jonathan W. Lee, Alexandre Bayen, Dan Work, and Jonathan Sprinkle. 2023. Parameter Estimation for Decoding Sensor Signals. In *ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023) (ICCPS '23), May 9–12, 2023, San Antonio, TX, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3576841.3589622

1 Introduction and Related Work

This paper introduces a parameter estimation approach for decoding digital sensor signals in a cyber-physical system. For unknown or not fully characterized digital sensor data, it can be difficult to decipher a desired signal from background or noise. In a cyber-physical system with networked sensors, we can leverage knowledge of the physical system to inform the decoding of the digital signals. This work in progress is a case study on deciphering commercial vehicle on-board sensor networks that communicate through the Controller Area Network (CAN). By understanding the stock vehicle sensor network, a vehicle can be extended into a scalable research platform with minimal instrumentation. Our challenge was to localize desired sensor signals encoded in network traffic that included other sensor data, control messages, as well as encoding and security overhead. Due to the vehicle's unknown sensor network, our approach developed methods to efficiently analyze and identify key signals despite the large state-space for potential signal embeddings. The contribution of this work-in-progress is a formal approach to deciphering pertinent signals uncharacterized cyber-physical system, with a case study in using this approach in vehicle on-board sensor networks. We share a code repository with analysis tools for analyzing digital signals.

Existing works [1, 3] attempt to classify the category of the signal (such as counter or CRC vs. 'physical'). These works do not solve identifying the specific physical signals found, which is a critical function. Data-driven techniques [2] can identify a set of physical signals (e.g. speed) signals in test data. All of these works are brittle to the differences in signal packing between protocols and platforms because their structure assumes the use of the CAN protocol. For example, CAN FD will encode the same signal in different locations based on the dynamically assigned length of the data payload so the bit boundary is dynamic. The protocols and information encodings will continue to change, and there are no standards enforcing the implementation decisions by OEMs.

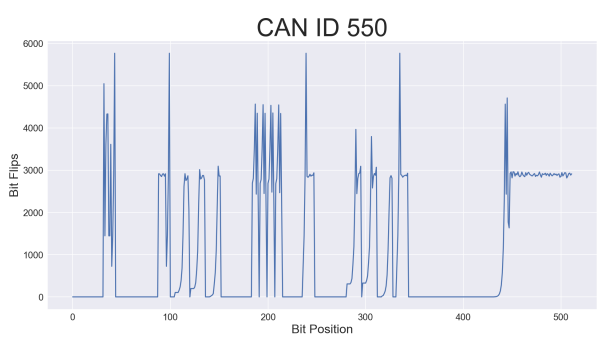
2 Background

Digital signals are packed in several common structures and sub-structures. This information can be employed to inform picking their components apart. Networks must communicate using a mutually understood protocol. Vehicle networks CAN with flexible data rates (CAN FD), Automotive Ethernet (1000BASE-T1), FlexRay, and others.

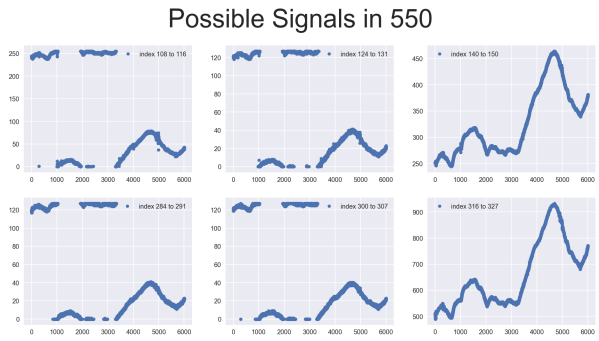
Within the super-structure of the network protocols which include acknowledgements, security handshakes, and metadata, is the main data payload. This is the portion of the message that contains signal data. Though these protocols are public standards, there are little to no conventions for the data payload; implementation decisions here are a guarded corporate secret by OEMs. For this work, we take a look at CAN FD. Our motivation for understanding and mastering the signal data on the vehicle is to empower its use in transportation and mobility research, including as a component of installing experimental control systems at-scale on commodity vehicles.

ICCPS '23, May 9–12, 2023, San Antonio, TX, USA

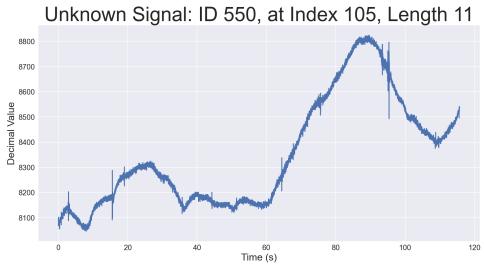
ACM ISBN 979-8-4007-0036-1/23/05…$15.00
https://doi.org/10.1145/3576841.3589622



(a) CAN Data by Bit Flips



(b) Automated Guessed Decodings



(c) Isolated Sensor Signal

Figure 1. (a) Pattern and structure in CAN FD 64 byte bit-flip plot. (b) Automated possible signal identifier guessing signal decodings based on bit flip counts. (c) The first signal from (b) identified as a continuous value by manually adjusting parameters.

3 Methods

Assume that you can examine some output of the sensor signals y', and you are looking for a pertinent physical signal y. The examined output is, $y' = f(\alpha, \beta, \gamma, y_?)$, where $y_? = a * f_1 + b * f_2$ some linear transformation of a measured signal. We use $y_?$ to indicate that *a priori*, we do not know if y' is a function of y, or some irrelevant but real signal. The parameters, $\alpha, \beta, \gamma, a, b$, stand in for a larger set of possible attributes which dynamically are informed by the knowns of a given system. Since we know we were dealing with CAN FD, we can use parameters like bit position, bit length, endianness, message ID, message length, scaling factors, and offsets as part of the signal function f.

Since we are in a cyber-physical system we can use deduce characteristics of y by leveraging the physical attributes of the system. This is a distinction from previous works which explicitly do not rely on *a priori* knowledge of the system. The the physical components of a system should be leveraged in a reverse engineering effort of a cyber-physical system.

We, for example, moved in a repeated structured way to induce the capture of a pattern that we know must be captured in the sensors, thereby creating the form of y to search for. This work uses a set of software tools we created to easily manipulate parameters α, β, γ and so on to find these pertinent signals. This often requires semantic examination by an expert at some point in the process. A collection of these analysis functions are collected in a GitHub repository, RosettaCAN at https://github.com/jmscslgroup/rosetta_can/.

4 Results and Future Work

Figure 1 summarizes the a typical search process. Part 1a showcases a bit flip vs. bit position plot. This figure presents a lot of encoded information on from the signal. First, a bit flip is defined as the changing from 1 to 0 or vice versa; the bit position is the 0-indexed location of each bit of the signal. This plot is counting the number of bit flips at that position throughout the period of the collected data. Suppose we are looking for a signal y which is the speed, and we collected data while increasing and then decreasing the vehicle's speed. Examining 1a we can see that where the bit flips are 0, no signal changes occurred and therefore the signal is not contained any of those bits. Furthermore, we know that when we change our speed, the higher order bits will change less often than low order bits. We also know that the space to encode speed must have a large enough space in binary form to allow for all the signal values of the speed. With whole numbers, y will need more than 64 values; $2^7 = 128$ so at least 7 bits are needed. As more significant figures are captured by the signal, more bits will be needed. Part (b) shows how we can leverage heuristics like these to algorithmically to identify likely candidates for the pertinent signal by guessing their bit boundary and plotting the signal as a function of time and decimal conversion. Part (c) then shows how explorations and refinements can iteratively narrow down on the bit boundary and other parameters.

In previous works [1, 3], and in this one, the algorithmic techniques are heavily reliant on a non-essential characteristic of digital signals: the structures of bit flips. Future work

could create a set of formal theorems to motivate and inform more generalizeable parameter estimation.

Acknowledgments

This material is based upon work supported by NSF (CNS-2135579), the Dwight D. Eisenhower Fellowship program under Grant No. 693JJ32345023 (Nice), and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Vehicle Technologies Office award number CID DE-EE0008872. The views expressed herein do not necessarily represent the views of the U.S. DOE or the U.S. Government.

References

[1] Mirco Marchetti and Dario Stabili. 2018. READ: Reverse engineering of automotive data frames. *IEEE Transactions on Information Forensics and Security* 14, 4 (2018), 1083–1097.

[2] Paul Ngo, Jonathan Sprinkle, and Rahul Bhadani. 2022. CANClassify: Automated Decoding and Labeling of CAN Bus Signals. (2022).

[3] Mert D Pesé, Troy Stacer, C Andrés Campos, Eric Newberry, Dongyao Chen, and Kang G Shin. 2019. LibreCAN: Automated CAN message translator. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 2283–2300.