SensorLoader: Automating the Generation of Software Knowledge Bases for Reverse Engineering Embedded Systems

AnMei Dasbach-Prisk (Cabrillo College), Cory DeWitt (USC), Tristan Scharfenstein-Montgomery (ISI), Luis Garcia (ISI)

USC Viterbi School of Engineering
Information Sciences Institute

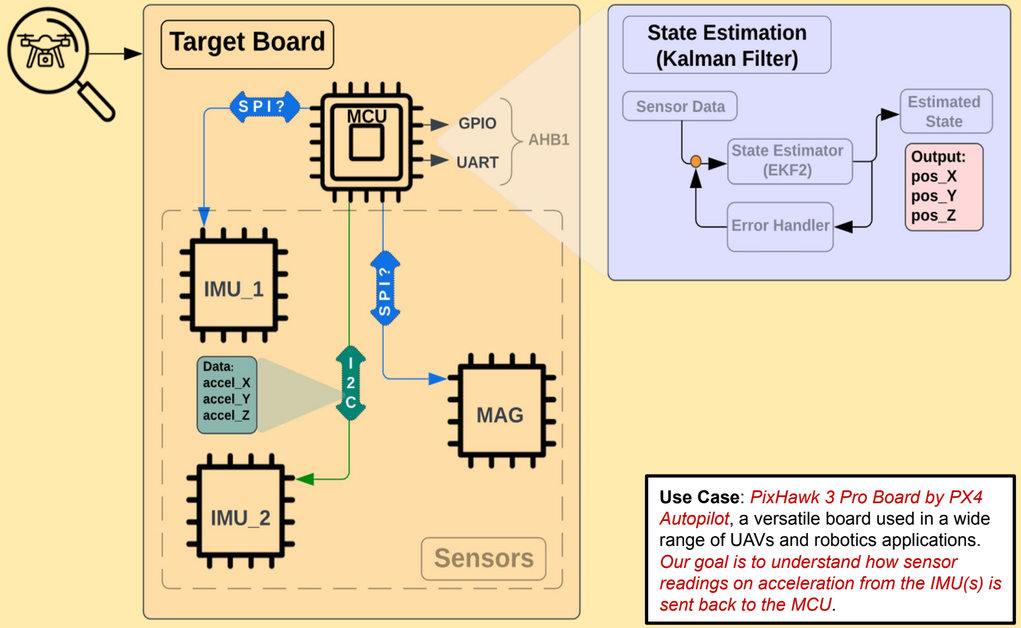
Work performed under REU Site program supported by NSF grant #2051101

Motivation

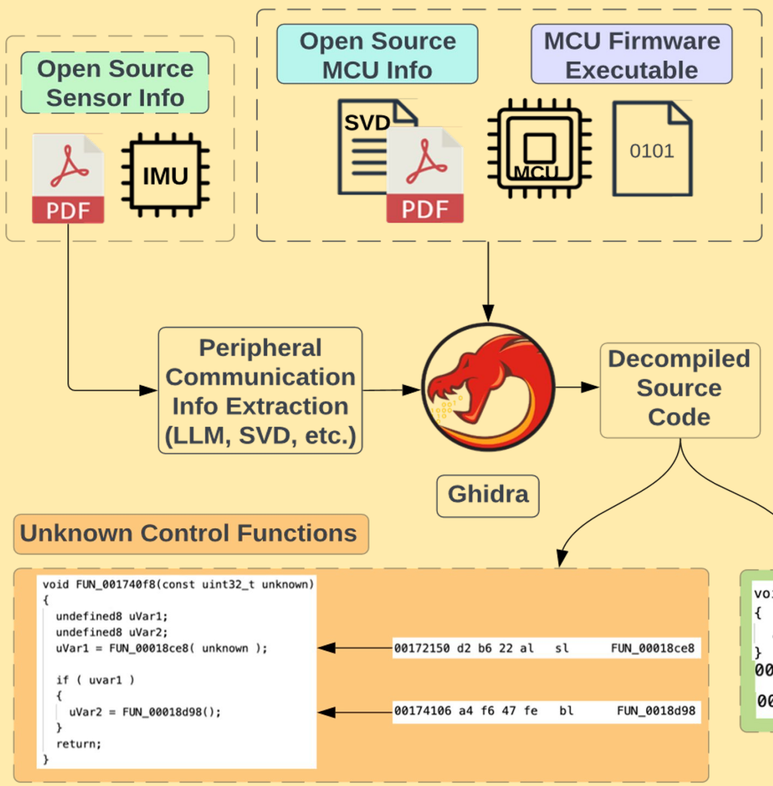
- **Problem**: Cyber-physical systems, particularly Industrial Control Systems (ICS) and Internet of Things (IoT) devices such as drones, form the backbone of various applications. *Their complex communication methods make them susceptible to physical threats.*
- *Static analysis is a part of this reverse engineering process* and involves decompiling board-specific firmware, thereby generating high-level control algorithms. From the binary executable, many of the function names and values initially appear as placeholders.
- *Chips have datasheets containing essential information on their specifications, characteristics, electrical properties, and usage guidelines.* This extracted peripheral information from these datasheets can help label placeholder function names and values. Traditionally, reverse engineers manually cross-reference these datasheets to obtain this information which is inefficient.
- Our research is *inspired by MISMO [1]*, in which they instantiated a framework for reverse engineering ICS and IoT. MISMO assumes that knowledge about how peripherals communicate is already known, and *we attempt to implement their assumption*.

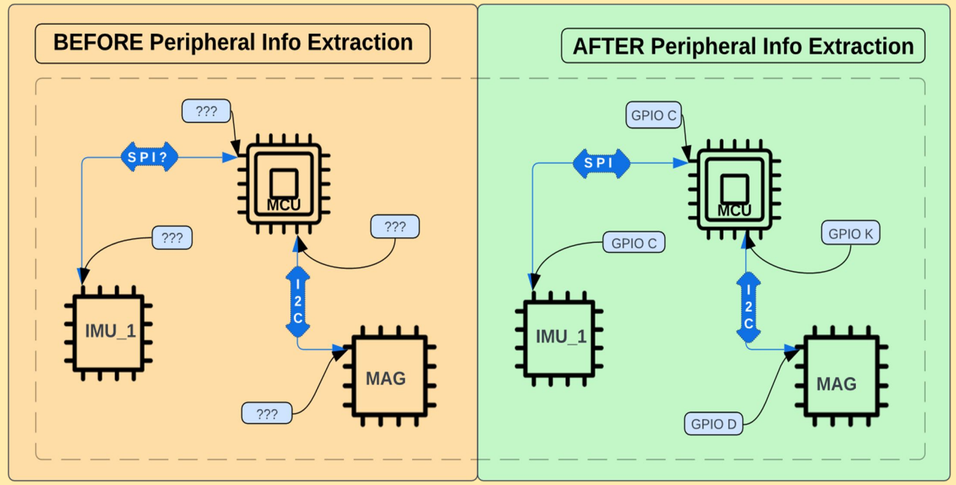
Contributions

- We propose a domain-specific reverse engineering solution, called SensorLoader, *to extract sensor communication information from embedded microcontrollers (MCU)*. This information provides reverse engineers with a better understanding about how a microcontroller is communicating with onboard sensors.
- The framework *automatically maps sensor communication information by leveraging open-source documentation and system description files.* This works with both structured and unstructured data.
- SVD-Loader gathers structured information, while LangChain LLM parses sensor peripheral communication protocols from unstructured documents. *Both tools are integrated into a Ghidra plug-in, mapping sensor semantics to the reverse engineering framework*.



Use Case: *PixHawk 3 Pro Board by PX4 Autopilot*, a versatile board used in a wide range of UAVs and robotics applications. *Our goal is to understand how sensor readings on acceleration from the IMU(s) is sent back to the MCU.*





Relevant Frameworks

- **Ghidra (Reverse Engineering Tool)**
  - An *NSA-developed, open-source reverse engineering tool*, decompiles firmware executables into a high-level source code. We used to decompile target MCU firmware.
- **SVD-Loader.py (Peripheral Info Extraction) [2]**
  - An open-source script that analyzes MCU hardware description files in SVD format. We leveraged this tool to overlay the peripheral communication information into Ghidra.
- **LangChain (Large Language Learning Model)**
  - We used this framework, which utilizes OpenAI's "text-davinci-002" model, and augmented it to collect peripheral communication information from the sensors' datasheets using QA Retrieval.

Challenges

- SVD-Loader only supports ARM based chips.
- Some sensor datasheets are publicly unavailable.

Future Work

- Generalize our project for more use cases.
- Have a repository where all the datasheets of the peripherals of a specific board are stored.

1 - Pengfei Sun, Luis Garcia, and Saman Zonouz. 2019. *Tell me more than just assembly! reversing cyber-physical execution semantics of embedded iot controller software binaries.* In 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE, 349–361.
2 - Roth Thomas, Pavlik Ryan. 2019. *SVD-Loader-Ghidra*. Github Repo, https://github.com/leveldown-security/SVD-Loader-Ghidra

Checkout our repo: https://github.com/cjdewitt/PX4_device_analysis.git

If interested contact, anmei.dasbachprisk@gmail.com & cjdewitt@usc.edu