1. **Principal Investigator**

Dr. Scott Graham, Assistant Professor, Air Force Institute of Technology Department of Electrical and Computer Engineering (AFIT/ENG), (937) 255-3636 ext. 4581, SGraham@afit.edu.

2. **Associate Investigators**

Captain (CPT) Brent Nolan, PhD Student, Air Force Institute of Technology Department of Electrical and Computer Engineering (AFIT/ENG), (937) 255-3636 ext. 6621, brent.nolan@afit.edu.

Dr. Brett Borghetti, Assistant Professor, Air Force Institute of Technology Department of Electrical and Computer Engineering (AFIT/ENG), (937) 255-3636 ext. 4612, brett.borghetti@afit.edu.

Dr. Christine Schubert Kabban, Assistant Professor, Air Force Institute of Technology Department of Mathematics and Statistics (AFIT/ENC), (937) 255-3636 ext. 4549, christine.schubert@afit.edu.

3. **Research Monitor**

Dr. Heidi Ries, Dean for Research and Professor, Air Force Institute of Technology Office of Research & Sponsored Programs (AFIT/ENR), (937) 255-3636 ext. 4544, heidi.ries@afit.edu.

4. **Facility/Contractor**

4.1. Sponsor:

Lt. Col Patrick Sweeney, Air Force Institute of Technology Faculty Research Council

4.2. Funding Source and Funding Amount:

Unfunded. Equipment needed for this experiment has already been obtained.

4.3. Contract #/CRADA #/Cooperative Agreement #:

N/A

4.4. Activity location(s) (where activity will be conducted):

Air Force Institute of Technology, Wright-Patterson AFB

5. **Conflicts of Interest**

None.

6. **Background Information and Scientific Rationale**

6.1. Investigative Question 1: Uniqueness of Vehicle Network Behavior

Hypothesis: For a specific make and model, a vehicle network exhibits statistically consistent behavior when repeatedly operated in a scripted manner. This baseline behavior appears to be statistically unique between different makes and models.

Anecdotal observations (by the author) reveal that vehicle networks exhibit unique baseline behavior between makes and models. This behavior is characterized by the subset of possible message IDs observed in the network, the frequency each unique ID is observed over time, the range of potential data values associated with an ID, and the frequency each byte of the data takes on a particular value. The terms make and model are used here to define a specific vehicle configuration including vehicle manufacturer (make), vehicle design (model), vehicle year, vehicle trim line (e.g. 'sport', 'luxury', 'premium', 'hybrid', etc.), and major manufacturer installed options such as an integrated GPS navigation system, sun roof, or driving assistant.

The study focuses on determining whether baseline network behavior is differentiable between makes and models of passenger vehicles. This will be tested by producing statistical models for various makes and models. These network behavior models will be tested for statistical uniqueness. If the models are unique for all vehicle makes and models, these network behavior models will be aggregated into a database to support follow on research objectives.

6.2. Investigative Question 2: Uniqueness of Vehicle Network States

Hypothesis: A vehicle's network behavior is statistically consistent when parked or moving and these two states are consistently differentiable.

Anecdotal observations (by the author) reveal that network data recorded from the same vehicle exhibits statistically significant differences between non-moving and moving observations. In addition to each vehicle network being unique, these 'states' of network behavior are predicted to be unique. The two overarching states of moving and not moving are expected to be consistently identifiable despite variance between individual observations. Additionally, the moving state will be divided into sub-states such as city and highway driving. If vehicles and their states are indeed unique and quantifiable, the states will be incorporated into the previously mentioned database of baseline network behavior.

6.3. Investigative Question 3: Uniqueness of Human Driving Behavior

Hypothesis: Variance between observations of a vehicle network state can be used to precisely quantify how the vehicle is being operated by the occupants. This precise understanding can be used to estimate how many passengers are in the vehicle and profile individual drivers based on their idiosyncratic driving behavior.

Every essential function of modern passenger vehicles is controlled by digital devices sampling occupant behavior thousands of times a second. The acceleration, braking, and clutch pedals generate hundreds or thousands of Controller Area Network (CAN) messages every second specifically quantifying how the driver is pressing them. The time it takes a driver to release the

accelerator, traverse their foot, and begin pressing the brake is assumed to be idiosyncratic to each driver. The speed which they release or apply the accelerator, brakes, and the clutch as well as the average pressure they apply are other idiosyncrasies.

Once a 'baseline' network behavior for a vehicle's moving and non-moving states are established, the remaining variance observed on the network is expected to be the result of individual differences in driving behavior. This network response to a given driver's input will be used to model that driver's behavior. The goal of this phase of the study is to determine whether vehicle network traffic can accurately profile and differentiate between a set of drivers.

1. Study Objective(s) and Purpose

1.1. Purpose:

To statistically evaluate Controller Area Network (CAN) network traffic across various makes and models of passenger vehicles and establish patterns between vehicles and their operational states. To determine if a driver operates a vehicle in an idiosyncratic manner which is significantly consistent and unique compared to other drivers.

1.2. Primary Objective:

To establish a database of baseline network behavior across many makes and models of passenger vehicles when being driven and when parked.

1.3. Secondary Objective:

To identify features of vehicle network traffic which effectively model driving behavior.

2. Study Design

2.1. Description of Study Design:

Since there is no precedent of a detailed statistical analysis of passenger vehicle network behavior, the study will be focused on establishing a baseline for statistical inference. Data will be collected directly from vehicles while parked and moving. The parked state will serve as the experimental control group for analysis of variance and other statistical machine learning methods.

Factors and Levels

The study is a between-subjects evaluation. This is a 3-factor experiment with 3 performance tasks. The factors are vehicle make, specific vehicle model (including year, trim line, and options), and driver. All three factors contain a finite but undefined number of levels. At least two samples of each driving state for a given vehicle and driver combination are required to test the research hypotheses. At least three models per manufacturer from three major manufacturers are needed to test the research hypotheses. At least nine drivers are needed to test the third investigative hypothesis. The parked state of each vehicle will serve as the experimental control.

Task Environment

The task environment for the study will be participant personally owned motor vehicles located in and around Wright-Patterson Air Force Base (WPAFB). Participants will be operating their vehicles during the experiment. Controller Area Network (CAN) messages will be recorded from the vehicle along with GPS and accelerometer data using a vehicle network monitoring device.

Human Participant Task Sequence

This study is a between-subjects experiment conducted for approximately 30 minutes on one day. The sequence of events will cover to the following schedule. Steps 1-5 will be completed before all other steps. Steps 6 through 10 may be performed out of order or partially omitted based on participant time constraints and feedback.

1. Informed Consent Briefing & Signature. This briefing will include safe driving behavior, experiment driving routes, and experiment procedures.
2. Verify participant has a valid driver's license, vehicle insurance, and registration.
3. Verify the vehicle has standard safety equipment including serviceable seatbelts and brakes.
4. Ask the participant if they are comfortable connecting the vehicle monitoring device to their vehicle. Remind them that they may cease participation at any time.
5. Connect the CAN network monitoring device to the vehicle's On-Board Diagnostics (OBD-II) port and secure wires away from the driver's legs.
6. Turn on the vehicle completely (engine start), wait 10 seconds, then turn off the vehicle.
7. Repeat step 5 once the vehicle network has stopped transmitting signals.
8. Turn on the vehicle, roll down the driver's side window, roll up the driver's side window, then turn off the vehicle.
9. Turn on the vehicle, pull out of the parking space, drive for approximately 20 seconds without exceeding the speed limit, then re-park and turn off the vehicle.
10. Turn on the vehicle, pull out of the parking space, and drive normally along a semi-standardized low-traffic route. The route will modified based on starting location, weather, traffic, driver feedback, and to avoid excessive idling time. The route will be approximately two miles of 'city' driving conditions which may include traffic control devices, turns, and moderate speed limits. The start and end of the course will be the same location. The intended driving time is approximately 10 minutes. After completing the route, the driver will park and turn off the vehicle.

Experimental Measures

We are interested in the behavior of vehicles when parked and moving. Four scenarios will be used to measure the variance and nature of vehicle networks. The scenarios are intended to establish data for the vehicle's baseline behavior as well as the driver's dynamic behavior.

Measure 1: Minimal Interaction While Parked [2 minutes]: This observation involved the driver turning on the vehicle, waiting 10 seconds, then turning off the vehicle without adjusting the

steering wheel, brake pressure, or any other input to the vehicle. This measure will be taken twice (steps 5 and 6 from the participation sequence) in order to establish a baseline for further measures. The repeated measurements are intended to validate the hypothesis that the given make and model's CAN network behaves consistently when operated consistently.

Measure 2: Minor Interaction While Parked [1 minute]: This observation will involve turning on the vehicle, lowering the driver's side window, then raising the driver's side window. Using the driver's side control, lock and unlock all of the vehicle's doors. The vehicle will be turned off once the window and locks return to their initial position.

Measure 3: Driving at Low Speed [5 minutes]: This observation will involve turning on the vehicle, safely pulling out of the parking position, driving no faster than the speed limit for approximately 20 seconds, then parking in a legal parking spot. Once parked, the vehicle will be turned off.

Measure 4: City Driving [10 minutes]: This observation will involve turning on the vehicle, safely pulling out of the parking position, driving along a roughly two mile course, then parking near the starting position. Once parked, the vehicle will be turned off. If the start location is within the perimeter of WPAFB Area A, this course will include driving to, along, and return from Spruce Way or Skeel Avenue. If the start location is within the perimeter of WPAFB Area B, this course will include driving to, along, and return from Loop Road W. If the start location is outside of WPAFB, a route will be chosen to closely approximate the low-traffic and changing driving conditions offered by the on post routes.

Participant Removal Criteria / Participation Ending Prematurely

If a participant is unable to operate the vehicle through at least one of the four scenarios or adhere to safe driving practices, they will be asked if they would prefer to return and try participating again at a later time.

If the participant decides to terminate the experiment or depart early, they may do so at any time during the experiment. Data will be included for analysis if the participant agreed to take part in the study but did not complete all of the driving scenarios. If a participant later asks for the collected data to be removed from the study, data collected for that person shall be deleted.

3. Subject Selection

3.1. Inclusion Criteria:

A subject who has met all of the following criteria is eligible for participation in the study:

- Own or have authorization to operate the motor vehicle
- The motor vehicle is manufactured after 2008 (due to OBD-II availability).
- Possess a valid driver's license.
- Possess current vehicle insurance for the vehicle.
- Possess current state registration.

- Appears to be mentally and physically able, and prepared for the driving task (e.g. not under the influence of substances; not experiencing undue emotional stress or physical limitations)

3.2. Exclusion Criteria:

A subject who meets any of the following criteria is disqualified from participation in the study:

- Does not have authorization to operate the motor vehicle
- Possesses a vehicle manufactured prior to the 2008 legislative mandate to use CAN technology or provide an OBD-II connection.
- Is pregnant or believed to be pregnant.
- Under the age of 18 years old.
- Is not a U.S. Citizen
- Possesses a vehicle without standard safety equipment including seatbelts and serviceable brakes.
- Does not possess a valid driver's license, insurance, or state registration.
- Appears to be intoxicated or under the influence of mind altering drugs (including medications).
- Is unable of maintaining clear verbal communication with the researcher.
- Does not have room in the vehicle to seat the researcher in a passenger seat.
- Does not clear rubbish and other obstructions from the area around the driving pedals.
- Is unwilling to connect the network monitoring device to the vehicle's OBD-II port.
- Otherwise appears to be unable to safely operate a motor vehicle.
- Only has access to a specific make and model previously included in the study three times.

3.3. Recruitment Plan

Recruitment Method:

For the purpose of this study, recruitment will be accomplished within the Air Force Institute of Technology, the Air Force Research Laboratory, and other major Air Force organizations located at Wright-Patterson Air Force Base (WPAFB). All participants will be service members, employees of the Department of Defense, or contractors working on WPAFB. An email will be sent to the group distribution mailing address by a person not directly in the chain of command (e.g. secretary). Additionally, the recruiting advertisement will be posted on the electronic AFIT announcement board on the intranet.

Subject Recruitment Message:

"The Air Force Institute of Technology (AFIT) is conducting a study in which participants will operate their personally owned vehicle for brief periods of time. Data will be collected from the vehicle's Controller Area Network (CAN) via its On-Board Diagnostics (OBD-II) port located under the steering column. No data will be transmitted to the vehicle and no electronic

modifications of any kind will be made. The main goal of this study is to analyze and model commercial motor vehicle computer networks. Participation in this study is voluntary and there is no compensation. However, participation in this study will allow you to take part in important research about vehicle network defense and help the investigators develop meaningful suggestions for the automotive industry and legislators. Volunteers will be asked to operate their motor vehicle in a parking lot and along an approximately two mile course. Participants will be asked to follow three scripted scenarios for up to 30 minutes. This research project has been approved for the use of human subjects by the Air Force Research Laboratory's Institutional Review Board in according with AFI 40-402 and AFRLI 40-402."

3.4. Consent Plan

Information on the informed consent document will be presented by the PI or associate investigators, and consent will be obtained from the participant before any data collection is carried out.

3.5. Compensation

There are no plans to provide compensation for participation in the research.

4. Experimental Plan

4.1. Equipment:

The following equipment will be used to collect data for analysis from the participant vehicle:

Intrepid Control Systems NeoVI FIRE 2 and NeoVI MIC Accessory

While performing the experiment, vehicle networks will be monitored by the NeoVI FIRE 2 multi-protocol data logger shown in figure 1. The NeoVI FIRE 2 will connect to vehicle Controller Area Networks (CAN) using a physical cable run from the device to the OBD-II port located under the steering column. Each Controller Area Network (CAN) message recorded will be tagged with the current GPS coordinates of the vehicle using the NeoVI MIC accessory shown in figure 2. Additionally, accelerometer data will be correlated to the CAN messages using a cell phone or a standalone small form factor accelerometer connect to the NeoVI. The accelerometer will be connect to the back of the Neo VI FIRE 2 using Velcro, elastic bands, or adhesives. The length of accelerometer and Neo VI FIRE 2 will be oriented parallel to the ground and perpendicular to the front of the vehicle. Depending on the vehicle interior, both will be placed resting on the floor of the vehicle and held in place by the researcher's foot or secured in the center console. If both need to be placed on the floor, a small cloth rag will be used to mitigate damage to the equipment.

FIGURE 1: INTREPIDCS NEOVI FIRE 2 DATA LOGGER

FIGURE 2: INTREPIDCS NEOVI MIC ACCESSORY

All data will be collected on a password-protected laptop authorized to operate on government networks only. The IntrepidCS Vehicle Spy software running on the laptop or the NeoVI FIRE 2's internal memory will be used to capture all signals. The decision to use the laptop or memory inside the Neo VI FIRE 2 to record network data will be based on the serviceability of the Neo VI FIRE 2 and Neo VI MIC. If internal memory is successfully recording traffic and the NeoVI MIC is correctly tagging GPS coordinates without the laptop, then the researcher will not actively use the laptop. It will be stored closed and by the researcher's feet in sleep mode. However, if any technical issues occur then the researcher will use the laptop running Vehicle Spy to record and coordinate data from the Neo VI FIRE 2 and the Neo VI MIC. If the laptop is required, it will be operated in the open position while resting on the researcher's lap. After capturing the network traffic from the NeoVI FIRE 2, analyses will be accomplished using government computers and networks running Vehicle Spy, SAS, JMP, R, Python and its open source libraries, TensorFlow, MATLAB, C, C++, Java, Excel, or similar statistics, programming, and machine learning software.

Incidental Findings:

Immediately after collection, all data collected will be searched using regular expressions for partial and complete instances of Vehicle Identification Numbers (VIN). If VIN numbers are found, they will be obfuscated to a VIN number of all zeros (which masks the ID number of the

vehicle) to prevent the vehicle's information from being used as a source of determining the identity of the owner.

Data will labeled by the make, model, driving state, driver number, and collection date. Driver numbers will start with 1 (for the first participant who participates) and increase sequentially with each additional participant. Make and model will be as detailed as needed to specifically identify a particular vehicle configuration. Thus, the exact directory structure may change to reflect the manufacturer's model identification scheme. An example directory and file name for a vehicle network data file is shown in figure 3. The purpose of recording the capture data is to help account for changes recalls or software updates may have on a specific model relative to the data capture date. For example, an update to a Tesla Model S's autopilot system my significantly impact its CAN network behavior.

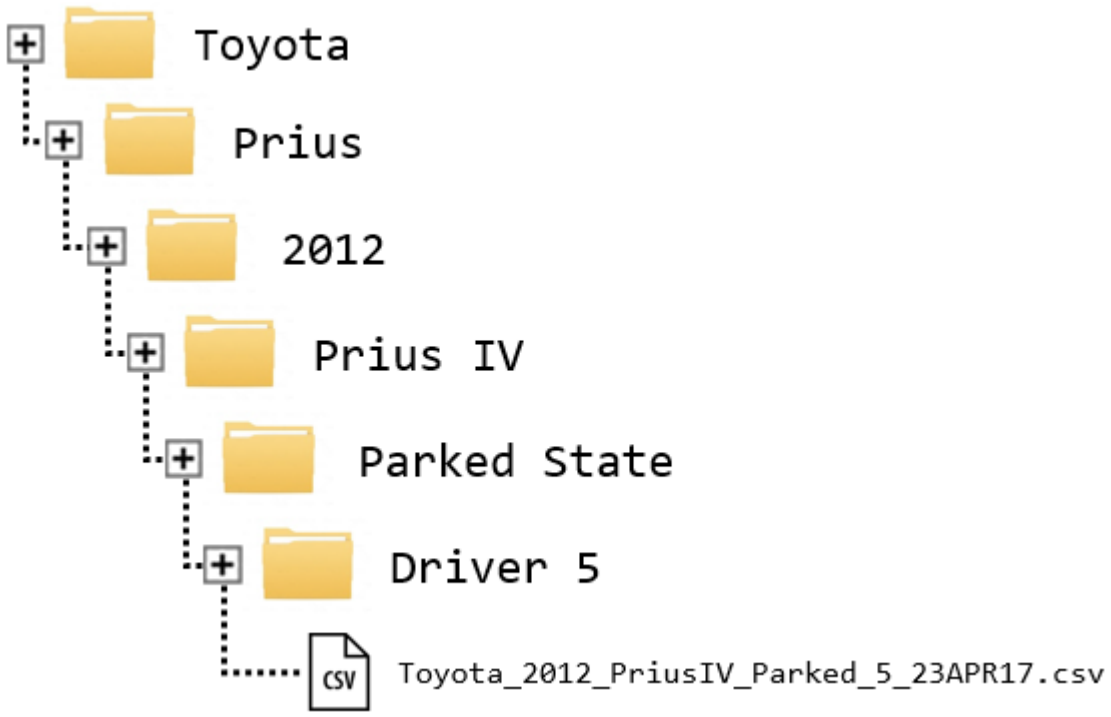

FIGURE 3: DATA LABELING EXAMPLE

5. Risk/Benefit Analysis

5.1. Benefits:

There is no benefit to the subjects.

5.2. Risks:

There are minimal known risks. Physical risk posed by operating a motor vehicle is the most significant risk to participants. This risk is identified in three areas—physical risk posed by collisions, physical risk by obstructed driving pedals, and physical risk posed by the data cable.

All participants and their vehicles will be subjectively evaluated by the researcher to be in good health and safe running order. Any debris in and around the driving pedals will be cleared before operating the vehicle. The data collection cable running near the driver's legs will be secured with Velcro straps wrapped around natural anchor points in the car so it doesn't distract or obstruct the driver. All standard safety measures will be taken including fastening seatbelts and obeying the rules of the road. Furthermore, the researcher will act as a safety officer advising the driver of any foreseeable hazards around the vehicle in order to mitigate the risk of accidents and injury.

Second, no data will be collected from the participants. Only vehicle network data will be collected. The vehicle data will be searched for VIN numbers using regular expressions which will be obfuscated to all zeros. No other information capable of identifying the specific vehicle or the driver is present on the CAN network. Thus, there is no potential risk to participant's confidentiality, employability, or reputation as a result of a data breach.

Third, there is a risk of malfunction by the NeoVI FIRE 2 data capture device resulting in the transmission of digital data to the vehicle. The NeoVI FIRE 2 is an industrial grade CAN network interface routinely used by General Motors and other vehicle manufacturers. There are no recorded instances of the device malfunctioning or transmitting data when operating in a read-only mode. The NeoVI FIRE 2 will be used exclusively in a read-only mode during this study. In the event of a malfunction, the researcher will immediately disconnect the device from the data collection cable in order to disconnect it from the vehicle and restore normal network operation. Disconnecting the Neo VI FIRE 2 from its end of the cable will not affect the portion of the cable near the driver. Malfunctions will be identified using a combination of error messages in Vehicle Spy 3, integrated LED status lights on the Neo VI FIRE 2, and abnormal vehicle behavior. Other than these risks, there are minimal risks to this experiment beyond what occurs in normal daily life.

6. Statistical Consideration and Plan

6.1. Sample Size (Power analysis):

Because of the novel domain for this research and the multitude of makes and models of passenger vehicles, there is no predictable limit on the number of vehicles useful for this study. Vehicles generate hundreds of CAN messages every second so individual data samples are expect to contain thousands of messages and pose no limitations for statistical analysis. A minimum of three models per manufacturer from at least three major vehicle manufacturers as well as nine different drivers are needed to test the research hypotheses. No more than three representatives of a specific make and model (including year and trim line) will be included in the study. New volunteers will be excluded from the study if they can only participate using a

specific make and model which was already studied three times. This measure is to partially limit the number of subjects used to meet the research objectives.

7. Safety Monitoring and Reporting

The participants will be monitored by research personnel throughout the entire test session. Participants will be told (verbally by the experimenter, as well as in the informed consent form) that they are free to terminate their participation at any time. Participants will be informed of steps to be taken, requirement to employ standard vehicle safety features, etc. The PI or associate investigators will act as the on-site monitor and will notify participants in person in the case of emergency. For example, participant notification is necessary if they became unconscious, there are dangerous driving conditions they don't notice, or they are driving dangerously. In the case of an adverse incident, WPAFB emergency services will be notified immediately. The PI will ensure that mishaps or injuries sustained during research will be reported as required pursuant to AFI 91-204.

8. Confidentiality

No data regarding the participants will be collected or maintained. No contact will be made with participants after their participation in the study. Vehicle data is generic with the exception of the potential presence of VIN numbers. All vehicle network data collected will be scanned for VIN numbers. If whole or partial VIN numbers are found, they will be replaced by a generic VIN of all zeros.

9. Data Management/ Data Sharing Plan

Vehicle network data will be sorted by make, model, and observation state. The data will be copied to and stored securely on an access-controlled folder on AFIT's internal network, and access will be managed by the principal investigator and enforced by AFIT/SC via network policy. Dual protection for this folder requires CAC & pin network login plus account-based access to the folder. Only researchers listed on this document will have access to the data.

All data will be maintained beyond the completion of this study in order to support related work in the future. Vehicle data will also be analyzed and the results will be used for publication via conference and journal papers as well as theses and dissertations. No published results will identify participants, specific vehicles, or associate them with any findings.

10. References

Enev, M., Takakuwa, A., Koscher, K., & Kohno, T. (2016). Automobile Driver Fingerprinting. *Proceedings on Privacy Enhancing Technologies*, *2016*(1), 34–51. http://doi.org/10.1515/popets-2015-0029

International Organization for Standardization. (2003). ISO Standard 11898-1:2003(E) Controller Area Network (CAN). *International Organization for Standardization*. http://doi.org/10.1017/CBO9781107415324.004

Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, Savage, S. (n.d.). Experimental Security Analysis of a Modern Automobile. Retrieved from http://www.autosec.org/pubs/cars-oakland2010.pdf

Wang, Q., & Sawhney, S. (2014). VeCure: A practical security framework to protect the CAN bus of vehicles. In *2014 International Conference on the Internet of Things, IOT 2014*. http://doi.org/10.1109/IOT.2014.7030108

11. Attachments

Informed Consent Document
Curriculum Vitae of Investigators (4 Separate Documents)
Intrepid Control Systems NeoVI FIRE 2 Multi-Protocol Vehicle Network Interface Brochure