smartphone Side-Channel Attacks and Defenses

**Course Project Manual**

**Course Project Development Institution:** Colorado School of Mines

**Course Project Manual Contributor**: Chuan Yue

# A Course Project for Advanced Smartphone Side-Channel Attacks and Defenses

**Course Project Description:** Smartphone motion sensors such as accelerometers and gyroscopes are ubiquitously equipped in modern smartphones. They have enabled mobile apps to have richer functionality and better interactivity, but have also created many new opportunities especially powerful side-channels for attackers to compromise users' security and privacy. This semester-long course project with five modules and eight STEPs aims to help undergraduate and graduate students to deeply learn advanced motion sensor based side-channel attacks and defenses on smartphones.  
  
The high-level **learning outcomes** and the corresponding **assessment** of this course project are summarized as follows. In other words, upon completion of this course project, students should be able to:

* **Build** tools to collect motion sensor data from smartphone users without any restriction.
  + Assessed by the tasks and outputs specified in STEP 3.
* **Construct** tools to preprocess smartphone motion sensor data.
  + Assessed by the tasks and outputs specified in STEP 4.
* **Design** features for machine learning algorithms based on the preprocessed motion sensor data.
  + Assessed by the tasks and outputs specified in STEP 5.
* **Create** powerful side-channel attacks for compromising users’ security and privacy by leveraging the collected data and machine learning techniques.
  + Assessed by the tasks and outputs specified in STEP 6.
* **Design** defense techniques to protect against motion sensor based side-channel attacks.
  + Assessed by the tasks and outputs specified in STEP 7.
* **Compose** a research-paper style course project report.
  + Assessed by the tasks and outputs specified in STEP 8.

The recommendation is that three to four undergraduate and/or graduate students with some basic cybersecurity and machine learning knowledge and skills should form a team to work on this course project. The team is expected to complete the course project based this manual with minimum support or help from an instructor. Note that this manual can also serve as a high-level introduction or guideline for students to work on any of the individual modules as a lab exercise.

**Course Project Environment:** Linux, Mac, or Windows.

**Course Project Files that are Needed:** TheCourse Project Manual file and the CSM\_MotionSensor\_Dataset.zip file.

**Learning Setting:** This course project is for student teams to complete outside the classroom, so it can be used in either face to face or online courses.

**Prerequisites:** Java or Python Programming, Basic Cybersecurity and Machine Learning knowledge and skills, Linux or Windows Systems, Computer Networks.

**Length of Completion:** 3600 to 6000 minutes in total for a team of 3 to 4 students.

**Level of Instruction:** Senior undergraduate students or graduate students in CS or related STEM programs. The course project should be further simplified if it will be used for freshmen, sophomores, or none-CS major students.

**Interconnection with Other Labs:** This course project is standalone by itself; however, if needed, an instructor can use the details in the five lab manuals to provide additional hints to students.

**Assessment Guideline:** Students should follow the steps to perform all the tasks. The grading is based on the final course project report described in STEP 8 and the outputs described from STEP 3 to STEP 7, in terms of their correctness (60%), organization (15%), clarity (15%), and concision (10%). Detailed grading rubrics are provided in each lab manual for each individual question, and they can be used as the references for both instructors and students.

### **Course Project/step 1 (Read A paper to learn about Sensor-based smartphone side-channel Attacks)**

Read the following research paper: Sensor-based Mobile Web Fingerprinting and Cross-site Input Inference Attacks  
([Dr. Chuan Yue's homepage at mines.edu for the paper MoST16.pdf](http://inside.mines.edu/~chuanyue/papers/MoST16.pdf)), by Chuan Yue. In proceedings of the IEEE Workshop on Mobile Security Technologies (MoST), 2016.

### **Course Project/STEP 2 (READ THE FIGURE 1 FOR a general attack framework AND ITS DESCRIPTION)**

The mobile Web fingerprinting and cross-site input inference attacks that you learned in STEP 1 can be formulated as a typical multi-class classification problem. Attackers can train machine learning classifiers in the training phase for identifying users or inferring sensitive inputs in the attacking phase simply based on the motion sensor data. Figure 1 illustrates a reference design of a general framework for performing side-channel attacks and defenses with five modules.

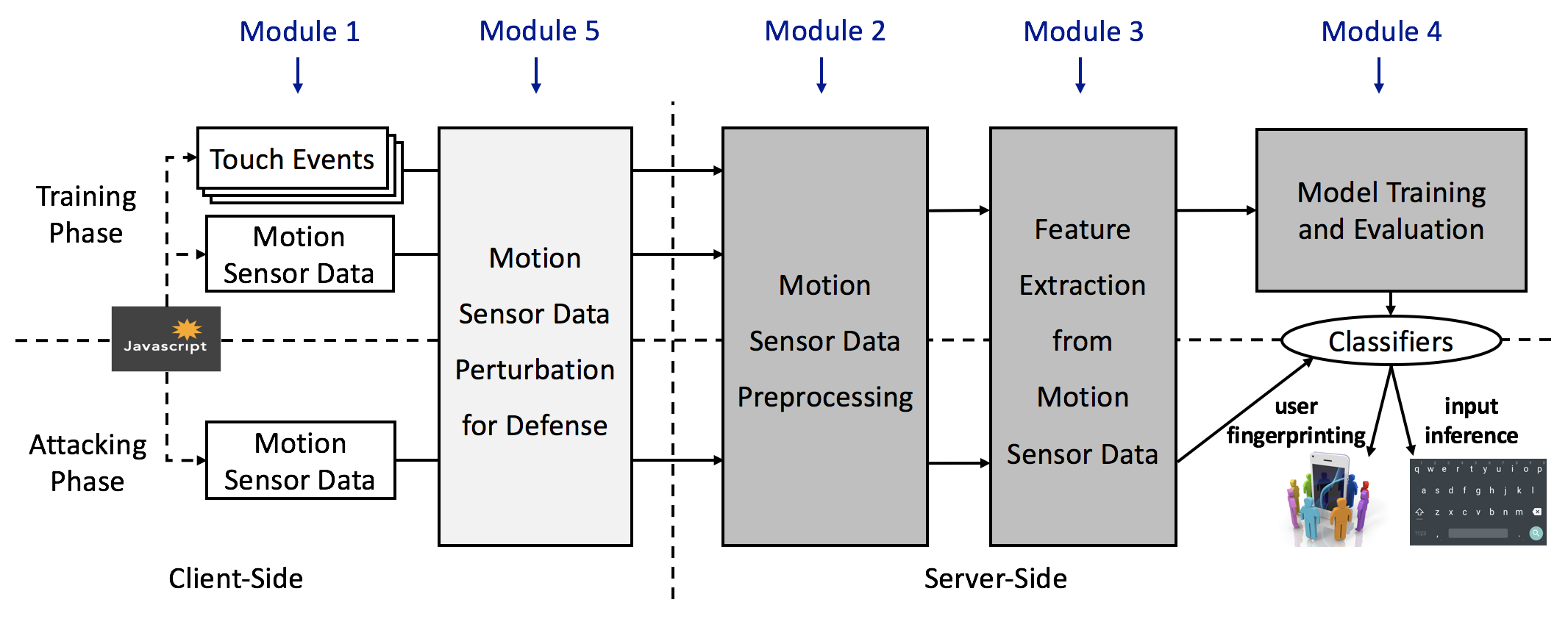


Figure 1 A general framework for performing side-channel attacks and defenses with five modules.

The attack part of the framework consists of four modules from Module 1 to Module 4. Module 1 is “Motion Sensor Data Collection”, which uses JavaScript code on webpages to collect motion sensor data from a user’s smartphone and sends the data to an attacker’s server. Module 2 is “Motion Sensor Data Preprocessing”, which segments motion sensor data that are pertinent to users’ individual keystrokes or other touch activities and reduces the noise in the collected motion sensor data. Module 3 is “Feature Extraction”, which derives time-domain and frequency-domain features from the segmented motion sensor data. Module 4 is “Model Training and Evaluation”, which trains machine learning classifiers that will eventually allow attackers to effectively perform user fingerprinting and input inference attacks, respectively.

Module 5 is “Motion Sensor Data Perturbation for Defense”, which is for students to apply the very representative data perturbation techniques (such as reducing data sampling frequency and adding noises to the data) that can help reduce the quality of the motion sensor data for defending against the mobile Web fingerprinting and cross-site input inference side-channel attacks.

### **Course Project/STEP 3 (Design, implement, and evaluate Module 1 on data collection)**

You will build tools to collect motion sensor data from smartphone users without any restriction. You will write JavaScript code, create “malicious” webpages, collect motion sensor data from the client-side, send the collected data back to a website, and evaluate the correctness of the data.

This ([W3C github.io device orientation specification website](https://w3c.github.io/deviceorientation/spec-source-orientation.html)) provides the information about how to write the client-side JavaScript code to collect motion sensor data from a user’s smartphone. You also need to write the code to send the collected motion sensor data back to your website, for example, using [the Ajax (i.e., XMLHttpRequest) technique illustrated on this w3schools.com website](https://www.w3schools.com/xml/xml_http.asp).

You should define a format for you to save your collected motion sensor data at the server-side, so that the collected data can be easily processed and analyzed for performing the attacks.

The outputs of this step include (1) the code for a website that can collect the motion sensor data from a user who visits the website, and (2) a description of the format of your collected data.

If you need more help about this step of the course project. Please ask your instructor for the detailed Module 1 lab manual for performing this step.

### **Course Project/STEP 4 (Design, implement, and evaluate Module 2 on data preprocessing)**

You will construct tools to preprocess smartphone motion sensor data. You will design and build tools to segment motion sensor data (of multiple consecutive keystrokes collected in a certain time window) into individual data pieces each corresponding to an individual keystroke (e.g., typing letter ‘p’ on the soft-keyboard of the smartphone), and evaluate the segmentation correctness and accuracy.

For this step and all the following steps in this course project, you will use a dataset consisting of the motion sensor data collected from 30 participants. Note that collecting such user related real data is often very time and effort consuming, and must be approved by a formal IRB (Institutional Review Board) process. Therefore, you do not need to collect the real motion sensor data from many users, and only need to use the provided dataset in the CSM\_MotionSensor\_Dataset.zip file.

Download and unzip the file CSM\_MotionSensor\_Dataset.zip, which also contains a detailed description of the dataset and the data format.

You should define a format for you to save your preprocessed motion sensor data into a file, so that features can be easily derived to train machine learning classifiers in the following steps.

The outputs of this step include (1) the code for preprocessing the data in the file CSM\_MotionSensor\_Dataset.zip, (2) the description of the format of your preprocessed data, and (3) your preprocessed motion sensor data.

If you need more help about this step of the course project. Please ask your instructor for the detailed Module 2 lab manual for performing this step.

### **Course Project/STEP 5 (Design, implement, and evaluate Module 3 on feature extraction)**

You will design features for machine learning algorithms based on your preprocessed motion sensor data. You will analyze and identify a set of features that can be derived from your preprocessed motion sensor data, and select or write tools to derive both time-domain and frequency-domain features that will be later used to train the machine learning classifiers in the training phase and also perform the attacks in the attacking phase.

You should define a format for you to save your derived features into a file, so that you can directly supply the file to some popular machine learning software packages such as [Weka](https://www.cs.waikato.ac.nz/ml/weka/) and [scikit-learn](http://scikit-learn.org/stable) to train and evaluate machine learning classifiers in the next step.

The outputs of this step include (1) a table of the features that you identified and their detailed descriptions, (2) the code for you to derive all those features from your preprocessed motion sensor data, (3) the description of the format of the file for your derived features, and (4) the file(s) for your derived features.

Similarly, if you need more help about this step of the course project. Please ask your instructor for the detailed Module 3 lab manual for performing this step.

### **Course Project/STEP 6 (Design, implement, and evaluate Module 4 on model training and evaluation)**

You will create powerful side-channel attacks for compromising users’ security and privacy by leveraging the collected data and machine learning techniques. You will compare, select, and apply some appropriate machine learning algorithms such as Random Forest and SVM (Support Vector Machine), identify certain popular machine learning software packages such as [Weka](https://www.cs.waikato.ac.nz/ml/weka/) and [scikit-learn](http://scikit-learn.org/stable), create programs to train the machine learning classifiers, and evaluate the accuracy of your trained classifiers using techniques such as cross-validation.

You should describe and justify your choices on machine learning software packages and algorithms. You should describe and justify your choices on the evaluation methods and process. You should also describe and discuss your evaluation results. Note that you may want to train different machine learning classifiers to perform user fingerprinting and input inference attacks, respectively.

The outputs of this step include (1) a description of your machine learning algorithms and a justification of your choices, (2) a description of your evaluation methods and a justification of your choices, (3) the code for you to train and evaluate your machine learning classifiers, and (4) a description and discussion of your evaluation results on user fingerprinting and input inference attacks, respectively.

Similarly, if you need more help about this step of the course project. Please ask your instructor for the detailed Module 4 lab manual for performing this step.

### **Course Project/STEP 7 (Design, implement, and evaluate Module 5 on data perturbation for defense)**

You will design defense techniques to protect against motion sensor based side-channel attacks. You will analyze and apply some representative data perturbation techniques (such as reducing data sampling frequency and adding noises to the data) to reduce the quality of the original motion sensor data for defending against both user fingerprinting and input inference attacks, and will evaluate the effectiveness of such defense techniques.

You should describe and justify the basic idea and intuition of your defense techniques, and compare the effectiveness of the attacks between with and without using your defense techniques.

The outputs of this step include (1) a description of your defense techniques and a justification of your choices, (2) the code for you to implement and deploy your defense techniques, (3) a description and discussion of your evaluation results on reducing the effectiveness of user fingerprinting and input inference attacks, respectively, and (4) a discussion of the potential side-effects of your defense techniques.

Similarly, if you need more help about this step of the course project. Please ask your instructor for the detailed Module 5 lab manual for performing this step.

### **Course Project/STEP 8 (Write a final course project report)**

You will compose a research-paper style course project report. You will write a final course project report based on what you have done and achieved for this project. You are recommended to use the [Manuscript Templates for IEEE Conference Proceedings](https://www.ieee.org/conferences_events/conferences/publishing/templates.html) (two-column and font size 10 or 11) to write an 8-page project report. Your project report can be structured to contain the following components:

1. title of the project and full names of the team members
2. an abstract
3. introduction (including the background, motivation, and goal)
4. key idea and approach including system design, implementation, and algorithms, etc.
5. activities such as literature study and experiments that you performed
6. description and analysis of the key results and observations
7. discussion of the limitations and potential future work
8. conclusion
9. references including the title of a paper or article, author list, conference or journal name, and year of publication, etc.

This structure is similar to that of a typical academic research paper. Please ask your instructor for additional suggestions if you need.

### **Puzzler (N/A)**

This is an advanced activity for students who complete the regular activities early. N/A for this course project.

## What to submit

A final course project report described in STEP 8, and the outputs described from STEP 3 to STEP 7.