Paper Review

6thSense: A Context-aware Sensor-based Attack Detector for Smart Devices

Chandrika Mukherjee
email: cmukherj@purdue.edu
Date: 10th October 2021

Summary

The paper attacks the existing permission-based Android Sensor Management System that allows apps to use non-permission-based sensors by directly utilizing sensor API. The attackers can exploit this behavior to steal sensitive information from users. To address the sensor-based threats, the authors proposed "6thSense" - context-aware intrusion detection system to protect all the sensors in a smart device. The study used the android phone as a smart device, considered nine sensors (accelerometer, gyroscope, light, proximity sensors, camera, microphone, GPS, speaker, and headset) which impact daily typical activities of users (e.g.- texting, making calls, playing games, browsing, driving, etc). 6thSense learns the correlation of sensors' data with different activities. The study used efficient machine learning algorithms to build the model that detects if the sensors' current state is benign or malicious. The study evaluated the efficacy of 6thSense using the data collected with their lab-made one benign and three malicious apps. Although The test shows promising results, the data collection was done in a controlled lab environment, therefore, in my view, collected data misses the real-time anomaly. The paper correctly identified the importance of data sampling in particular time interval for the data obtained from different sensors. Overall, the approach taken by the authors to mitigate the risk of sensors-based attacks in smart devices opens relevant future research direction in this topic. Specific comments and suggestion are listed in the next section.

Detailed Comments

- **The Relevance:** Almost all the smart devices use multiple number of sensors that capture data from physical world and provide meaningful information to cyber world. Users allow the apps to use some particular sensors while installation, but these permissions allow the apps to use other non-permission-based sensors as well. Adversaries can thus exploit these sensors to grab sensitive information of the users. On the other hand, the users are unaware of the threats and possible consequences of their action on the apps. Along with this, no comprehensive security mechanism was available at the point of publication of this paper. The authors mentioned higher prevalence of different sensor-based threats in mobile devices, for example - data from motion sensors, recording tap noises on touchpad can be used to infer user input by keystroke analysis, light sensors can be used to trigger malware etc. Therefore, in my view, the research problem is highly relevant.
- **Difference From Previous Work:** Existing tasks are related to anomaly detection that do not take care of the sensors-based attacks. Some intrusion detection study was proposed in Wireless sensor networks(WSN) which are not compatible with smart devices. The paper specifically compare 6thSense against Semadroid, AuDroid and DARKLY. Semadroid does not provide extensive performance evaluation and testing was done against similar attacks. Audroid only considered audio sensor based attacks and Darkly is not tested against sensor based attacks. Contrarily, 6thSense - context-aware intrusion detection system (IDS) provides coverage to all the sensors in smart device and ensures security against three types of sensor based attacks.
- **Threat Model :** The paper considers three types of threats -
  - Triggering Malicious App via sensors.
  - Information leakage via sensors
  - Stealing of information via sensors.

  The authors used three different malicious apps to perform sensor-based attacks on user's smart phone. Android apps triggered by light and motion sensors address Threat model 1. They created a malware that records conversation as audio clip and play back the information (data leakage) that addresses Threat model 2. Another android app that opens camera and records video surreptitiously addresses the Threat model 3. These three lab-made apps were uploaded in VirusTotal to check if the available malware scanners could detect them. The paper shows that only 2 out of 60 scanners could detect, though the type of malicious behaviour could not be detected. This provides reliability on the lab-made apps and the evaluation of 6thSense. Overall, the aim was to secure all the sensors in a smart device as the sensors are co-dependant.
- **Design Assumption:**
  - The study clearly mentions that even if the sensors are independent of each other, but task wise they are dependant - which portrays co-dependency. It substantiates the goal of the authors to secure all the sensors in device, not only one.
  - Different manufacturer builds sensors differently, therefore, the sampling rate of different sensors varies. The study here, didn't consider individual sampling, but captured data in specific time interval to avoid any processing error.

- The study has used some efficient and fast Machine Learning algorithms to handle large volume of data.
- 6thSense monitors all the sensors in the device.

- **Main Contributions:** 6thSense has three main phases a) Data Collection, b) Data Processing c) Data Analysis.
  - for Data Collection, authors have used nine sensors as they only considered typical user activities to build the contextual model. To gather the data, they built a custom android app - sensoreventlistener API captures numerical values from data-oriented sensors and the App also determines the state of the sensors (0/1).
  - Collected data is sampled every second - for data-oriented sensors, average value is taken per second. While generating the transition matrix, 6thSense only considers condition change. if the value changed from previous state, then it is considered 1, otherwise, 0. Sampling frequency for logic based sensor is 0.2Hz. So, the study considered sensor condition of logic-based sensors to be same in this 5 seconds time period to correctly assemble data from both type of sensors.
  - In the final Data Analysis phase, several efficient Machine Learning algorithms are used - Markov Chain is used to predict the probability of transition occurring between two states. Number of consecutive malicious states were observed, if it crossed a threshold, then the activity was considered malicious.
    Naive Bayes algorithm was used on sensor data set. 6thSense uses nine typical user activities, and ground truth are used to define these activities. Training data set is used to determine frequency of sensor condition changes for a particular activity. If the computed probability for all known benign activities is not over a predefined threshold, then it is detected as malicious activity.
    The paper also demonstrates the use of different supervised machine learning algorithms like - PART, Logistic Function, LMT, Hoeffding Tree, J48, Multilayer perceptron using WEKA to detect threats on sensors.

- **Evaluation :** If number of consecutive malicious states is considered 3, then Markov Chain achieves accuracy 98% without introducing False Positives(detecting malicious states as benign which is harmful for the system). Naive Bayes approach performs best with 60% as threshold, but performance is worse than Markov Chain, it achieves accuracy of 95%, but with False Positive values. Using WEKA, LMT gives better accuracy among other algorithms. Overall, LMT provides better accuracy than other two approaches. Performance overhead wise, Markov Chain outperforms other two algorithms. Main overhead concern is power consumption. As all senors are kept on and historical data storage lead to more power consumption. The authors also mention that the real-time monitoring will not store historical data which will reduce overhead by several times.

  Standard algorithms are used to detect malicious state, but the whole experiment is conducted in controlled environment. Within this data set, 75% is used for training and rest 25% is used for testing. However, because of the controlled environment, the training data set contains standard behavior and less anomaly, which makes the calculation easier and distant from real-world calculation. Therefore, I think, with real time data, accuracy would not reach this high.

Recommendations

This paper makes great effort in designing and developing 6thSense - context-aware task-oriented sensor-based attack detector for smart devices. The paper clearly establishes the importance of this research in practical world. The paper mentions that how users are unaware of data leakage or stealing of their sensitive information while they do typical day-to-day activities using their smart devices. The authors also raised the issue in permission based android sensor management system - for example, allowing an app to access camera will also allow the app to access the light sensor. The paper presents a comprehensive study to secure and monitor all the sensors in a smart device which are co-dependant on each other. Evaluation of the system was done on anonymous data from 50 users, although in controlled setting, the high accuracy reached by selected algorithms - Markov Chain, LMT and Naive Bayes adds reliability for future research in this domain.

Observation of sensor usage in application level can be combined with 6thSense's sensor level detection to make the system more robust. To avoid constraints of offline training, online training method can be used which also defines a future scope for 6thSense. Power consumption was the primary overhead for 6thSense, therefore, in future, frequency-accuracy/ battery-accuracy/ battery-frequency trade-off can be studied.

6thSense used data in controlled environment, real-time collection of data could have provided more reliability for real-world use of 6thSense. Also, while gathering training data from users, 6thSense didn't consider if any other malicious app was present in user's mobile phone that was suppressing the real values from sensors. If this was true, then can put the results obtained by 6thSense in uncertainty. Though the study used VirusTotal to check reliability of their lab-made malicious apps, they could have also uploaded their malicious apps on Google Play Store. As Google does its own security check on submitted apps, it could have provided more reliability on 6thSense's ability to detect real malicious behaviours.