IoT Privacy Preserving in Modern Smart Homes

Keyang Yu, Qi Li, Dong Chen
Knight Foundation School of
Computing and Information Sciences
Florida International University
dochen@cs.fiu.edu

ABSTRACT

The trend of applying Internet of Things (IoT) devices in smart homes is emerging in the recent decades, which raises the concern about user privacy leakage to multiple third parties and Internet Service Providers (ISPs). Extensive prior work has revealed the threat that Internet traffic volume data generated by IoT devices may be analyzed to infer user's in home behaviors, even with encryption and anonymity. And such high-granularity traffic usage data, is being collected by ISPs, IoT device manufacturers, content providers, and being continuously shared with multiple third parties.

To address these issues in smart home users' side, we introduce PrivacyGuard to protect user's sensitive information against data analyzing. By applying user tunable traffic reshaping and injection, we can significantly reduce private information leakage from IoT network traffic data, while still permits sophisticated data analytics or necessary smart home control and management.

However, masking private information requires additional device and traffic overhead, which brings higher cost for smart home users. Based on our research, the performance of user activity information inferring could be reduced by coarser time granularity of traffic data. So, we are invoking efforts from ISPs and device manufacturers to change the way of traffic volume data collection. With minimum influence on normal traffic usage or cost analysis, a coarser traffic data collection could better protect smart home user privacy.

1. INTRODUCTION

The total installed base of IoT devices is projected to increase to 75.44 billion worldwide by 2025, a fivefold rise in 10 years. Such worldwide application of smart devices brings both convenience and privacy threat to various of smart facilities, especially smart homes equipped with multiple IoT devices. Various reports have shown that ISPs, IoT device manufacturers and content providers may collect Internet traffic usage data for analyzing, to provide customized services including promotion, pricing or advertisement, based on statistical analysis on traffic volume data [1]. Unlike traditional devices like PCs, TVs, or smartphones, the traffic generated by IoT devices are relatively less changeable, and easier to be identified. Significant recent research [2, 3, 4, 5, 6, 7, 8, 9] has shown that, even with limited knowledge, on-path adversaries are capable of identify IoT devices by carrying on data analytics to traffic volume data, which shows severe privacy vulnerability. Since most IoT devices are highly user-interactive, such exposure of device information will easily lead to user behavior inferring.

To address these issues, we proposed a new low-cost, open-source, user "tunable" defense system—PrivacyGuard, that enables smart home users to significantly reduce the privacy leakage against on-path adversaries, while still permits sophisticated traffic analytics which are necessary to smart home control and management. The design principles of PrivacyGuard are introduced as follows:

Building Adversarial Machine Learning/Deep Learning Attack Models: As introduced before, the traffic volume data of IoT devices could be analyzed through various static metrics, we picked 8 representative metrics including duration, standard deviation, skewness, etc. to train the attack model. Multiple state-of-art ML and DL models were chosen to better mimic a "smart" adversary, including Decision Tree, Support Vector Machine (SVM), Convolutional Neural Networks (CNNs), etc.

Intelligent Traffic Rate Signature Learning: PrivacyGuard employs intelligent deep convolutional generative adversarial networks (DCGANs)-based traffic signature learning, to store various traffic patterns generated by different devices. These signatures will be artificially injected to the original traffic to mask real user activity based on pre-trained user activity model.

Artificial Traffic Signature Injection: Prior research has proposed various approaches regarding to IoT traffic reshaping. However, randomly inject traffic patterns could easily be detected, and reveal the original user activity information. To address this issue, we compared different user behavior model including Markov Chain and Hidden Markov Model, we picked Long short-term memory (LSTM)-based user activity modeling to best mimic a smart home's user behavior routine and inject proper traffic signature learned from the previous step.

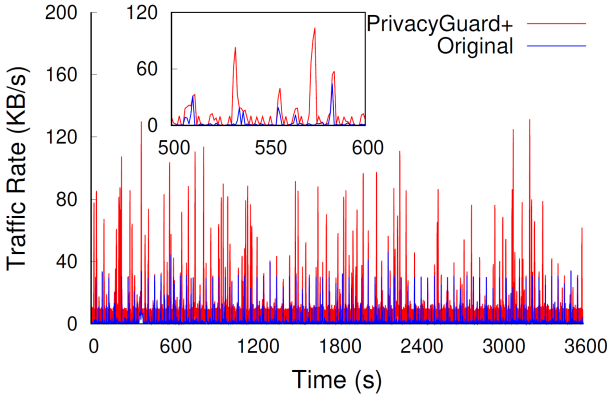

Figure 1. Online Prototype of PrivacyGuard.

User Tunable Partial Traffic Reshaping: Unlike prior approaches which may reveal device activate length, or traffic peak, PrivacyGuard applies a "smarter" reshaping method including randomized extension, and dynamic thresholding. Furthermore, PrivacyGuard provides user "tunable" options for

smart home users to balance the traffic overhead and privacy preserving performance.

We evaluated PrivacyGuard on occupancy detection prevention and user activity detection prevention on 5 different datasets, 5 different time granularities, different user “tunable” preferences, and different adversary confidence. PrivacyGuard is capable of preventing user activity detection in reasonable traffic overhead and cloud service cost.

As shown in Figure 1, the current PrivacyGuard prototype can be deployed on a Raspberry Pi 4, which is a $70 device. The prototype of the PrivacyGuard relies on a remote server running on Amazon EC2, which costs an additional $6.6 per month for a smart home with 30 IoT devices. Our future plan focusses on collecting more IoT traffic traces for a better trade-off point between privacy preserving and traffic overhead. We are also considering deploying the PrivacyGuard prototype to IoT devices or smart routers directly and remove any additional devices from the smart home to further reduce the cost.

REFERENCES

[1] T. Brewster. 2017. Now Those, Privacy Rules Are Gone, This Is How ISPs Will Actually Sell Your Personal Data. https://www.forbes.com/sites/thomasbrewster/2017/03/30/fcc-privacy-rules-how-isps-will-actually-sell-your-data/?sh=360465fc21d1

[2] Noah Apthorpe, Danny Yuxing Huang, Dillon Reisman, Arvind Narayanan, and Nick Feamster. 2019. Keeping the smart home private with smart (er) iot traffic shaping. Proceedings on Privacy Enhancing Technologies2019, 3 (2019), 128–148.

[3] Phuthipong Bovornkeeratiroj, Srinivasan Iyengar, Stephen Lee, David Irwin, and Prashant Shenoy. 2020. RepEL: A Utility-preserving Privacy System for IoT-based Energy Meters. In2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 79–91.

[4] Xiang Cai, Rishab Nithyanand, Tao Wang, Rob Johnson, and Ian Goldberg. 2014. A systematic approach to developing and evaluating website fingerprinting defenses. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM,227 ‑ 238.

[5] Dong Chen, David Irwin, Prashant Shenoy, and Jeannie Albrecht. 2014. Combined heat and privacy: Preventing occupancy detection from smart meters. In2014 IEEE International Conference on Pervasive Computing and Communications. 208–215.

[6] Wenbo Ding and Hongxin Hu. 2018. On the Safety of IoT Device Physical Interaction Control. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security(Toronto, Canada)(CCS ’18). 832–846.

[7] Kevin P Dyer, Scott E Coull, Thomas Ristenpart, and Thomas Shrimpton. 2012. Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail. In2012 IEEE symposium on security and privacy. IEEE, 332–346.

[8] Marc Juarez, Mohsen Imani, Mike Perry, Claudia Diaz, and Matthew Wright. 2016. Toward an efficient website fingerprinting defense. In European Symposium on Research in Computer Security. Springer, 27–46.

[9] Homin Park, Can Basaran, Taejoon Park, and Sang Hyuk Son. 2014. Energy-efficient privacy protection for smart home environmentsusing behavioral semantics.Sensors14, 9 (2014), 16235–16257.