A Machine-Learning Based Instruction Authentication Mechanism in Smart Home Networks

--- Report for ICSML 2016

Team Info.

Team Number: 11

Team member 1: Weiming Bao (850259672)

Team member 2: Kai Zhang (850259300)

Abstract

In this report, we surveyed the main architecture of Smart Home systems, the security challenges and threats and existing machine-learning based implementations on the specific area. We focus on the authentication in Smart Home networks and propose machine-learning based featured password verification and instruction legitimacy verification based on anomaly detection. Preliminary experiments have been taken regarding the featured password verification part of our proposal.

Contents

1. Introduction

Smart home is a concept that has existed for many years but gained attention among researchers in recent years due to the growth in the domain of Internet of Things (IoT).

1.1 Introduction of Internet of Things (IoT)

Internet of Things (IoT) is a large number of communication and information systems used to support and simplify everyday life by means of technology.

It is a promising area, which has drawn remarkable attention both in industrial and academic areas. According to a white paper released by the McKinsey consulting group in June 2015, the total value of IoT devices and services is estimated to range between 4 and 11 trillion dollars within ten years, which represents up to 11% of the world's economy.[1]

Existing IoT devices mainly contains Smart Home devices, Smart office Devices, Smart City devices (like Smart Grids, etc.), devices used in medical fields and so on.

1.2 Introduction of Smart Homes

One emerging area of interest of IoT devices which is focused on in this paper is the application of IoT devices to Smart Home systems.

Smart Home devices and systems are interconnected devices in people's houses originally designed to automatically and intelligently facilitate people's everyday life and potentially provides additional comfort and even security, as well as enhanced ecological sustainability.

For instance, a smart lightening system can adjust the light based on the sensors and automatically give light at a comfortable level when the house owner leaves his/her bed at night. A smart air conditioning system can use various kinds of household sensors and web-based data sources to operate intelligently, rather than simply run under manual instruction or fixed-schedule control schemes. The smart air conditioning system can predict whether the house is empty or not by tracking and analyzing location data to ensure that the desired comfort level is achieved when the house is occupied, and that energy is saved when it is empty.

Smart Home devices can also assist with independent living for elderly residents. Smart Home devices can assist with daily housework such as cleaning, cooking, shopping and laundry. Those with low level cognitive decline can be supported by specific smart home systems with timely reminders for medication. Home health monitoring systems can signal caregivers to respond before disruptive hospitalization is needed.[2]

Smart Homes can bring much convenience and benefit people a lot in our everyday life.

1.3 Main Architectures of Smart Homes

In order to achieve the functionality of Smart Homes, the devices in the systems need to collect and analyze user data, which will inevitably contain private information that will be harmful to users if leaked.

Collaboration between devices is an important aspect of IoT systems. Many Smart Home systems also require Internet access to obtain supplementary information such as weather forecasts, location data, synchronized time and realize remote control and status push.

Furthermore, Smart Home devices are limited in hardware resources. Constraints exist in computing power, storage capability, so that sophisticated mechanisms like advanced security solution is hard to implement locally. Cloud services are usually needed to support the functionality of the system.

These requires specific design in architecture level of Smart Homes.

1.3.1 Direct Internet Access (DIA) Architecture

In this implementation, Smart Home devices can access the Internet via direct wired or wireless connection with existing LAN based on protocols like IETF's CoAP [3] (The Constrained Application Protocol), as is illustrated in *Fig.1*.

The Smart Home devices can directly upload and download data from Internet servers and sophisticated mechanism and function can be realized with the support of Cloud services.

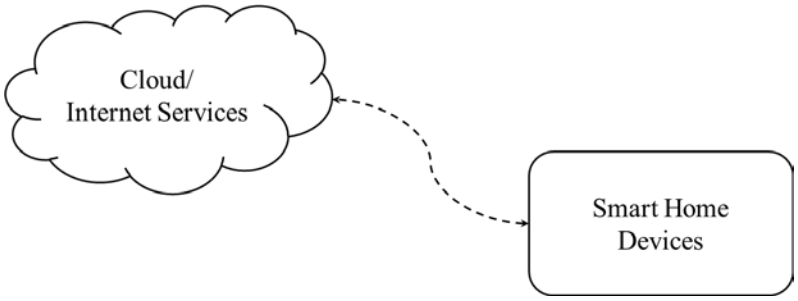

Fig.1 Smart Home system under Direct Internet Access (DIA) architecture

This kind of solution replaces the need for local computation with a need for substantial Internet connection. Large amounts of raw data generated by IoT devices have to be transferred to the cloud, without pre-processing due to limited local resources.

Therefore, the data flow from smart home devices may occupy remarkable network bandwidth. A high-speed, low-latency, always-on Internet connection is required, which is not always available especially in rural or remote areas and will increase the cost to some extent. The functionality of the whole system is highly dependent to the Internet connection. Problems like Internet access failure and control latency may arise.

1.3.2 Gateway-based Internet Access (GIA) Architecture

An gateway is a relatively resource-rich node in the Smart Home network collaborating with the other endpoints. It can be a central management point to coordinate, monitor dataflow and devices. Also, it can act as a bridge to connect the local system to the Cloud as is illustrated in *Fig.2*.

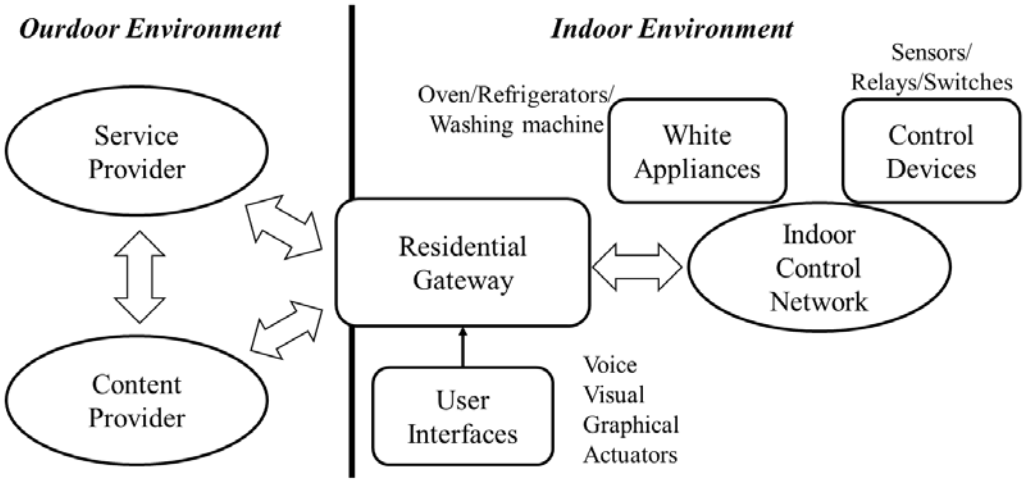

Fig.2 Smart Home system under Gateway-based *Internet Access (GIA) architecture*

In local Smart Home network, there're many standards that can be implemented like Zigbee, Zwave, Insteon, Bluetooth, Ethernet, WIFI, RS232, RS485, C-bus, UPB, KNX, EnOcean, Thread, each among which has its strengths and weaknesses. [4] Usually, the

implementation is different from the widely-used WIFI (IEEE 802.11) protocol, which meets the requirement of less energy consumption and more flexibility.

Since the gateway has more computing power and resources, tasks that require relatively high computing power and rich storage resources can be conducted locally.

In terms of security, the gateway can centralize user authentications and apply access control to guard against unauthorized access or modification of restricted data. It also acts as a firewall to protect the smart devices and privacy from attackers, and to reduce the attackable surface.[5]

2. Security Challenges and Threats

None of the foreseeable benefits from the Smart Home systems is possible and realistic if the systems are not secure and trusted. Privacy leakage, functionality failure of devices and even fatal threats can happen to people using Smart Homes.

In most cases, Smart Home IoT devices are likely to be retrofitted to an existing smart home system piece by piece as needs arise. Deploying security mechanism in smart home environments is a challenging issue due to their heterogeneous and complex nature and constrained resources.

2.1 Main Threats and vulnerabilities

Although the Smart Home system is an environment different from networks in other domains, the main aspects that can be compromised by potential threats are confidentiality, integrity, authentication and availability.

Confidentiality threats are those threats that result in the unexpected release of sensitive information. Leakage of keys and passwords will lead to unauthorized access. Some Smart Home devices and the dataflow between devices contain private information, like medical information. Besides, even seemingly inessential data, such as the internal domestic temperature and the status of air conditioning system could be used to derive whether a house is empty or not.

Integrity threats can infect the normal operation of the whole system, compromise the performance of the system, leading to troublesome experience or even financial losses on both the manufacturer and the customer or fatal threats to users specifically using household medical assistance system.

Authentication threats can lead to harmful intrusion in sense of either sensing or control information. For instance, unauthenticated system status alerts might confuse a house controller into an emergency situation and opening doors and windows, while in fact allowing illicit entry. Malware or malicious injection can happen if the authentication process of software or firmware update is compromised.

Availability threats are those when unauthorized devices connected to the network take over the control or steal network bandwidth, resulting in a Denial of Service (DoS) to legitimate users. Since many Smart Home devices may operate only depending on embedded batteries, flooding a network with requests can lead to a specific form of denial of service.

The vulnerabilities of Smart Home networks may include both remote and physical accessibility, constrained system resources, system heterogeneity, fixed firmware, lack of dedicated security professionals and so on.

For instance, a non-export may ignore the policy where a web camera is accessible by users who have been given its host name and port number. With the help of Internet device-

scanning search engines such as Shodan[6] (https://www.shodan.io), we can find vulnerable devices, using keywords “has_screenshot:true port:554”, as is illustrated in *Fig. 3*.

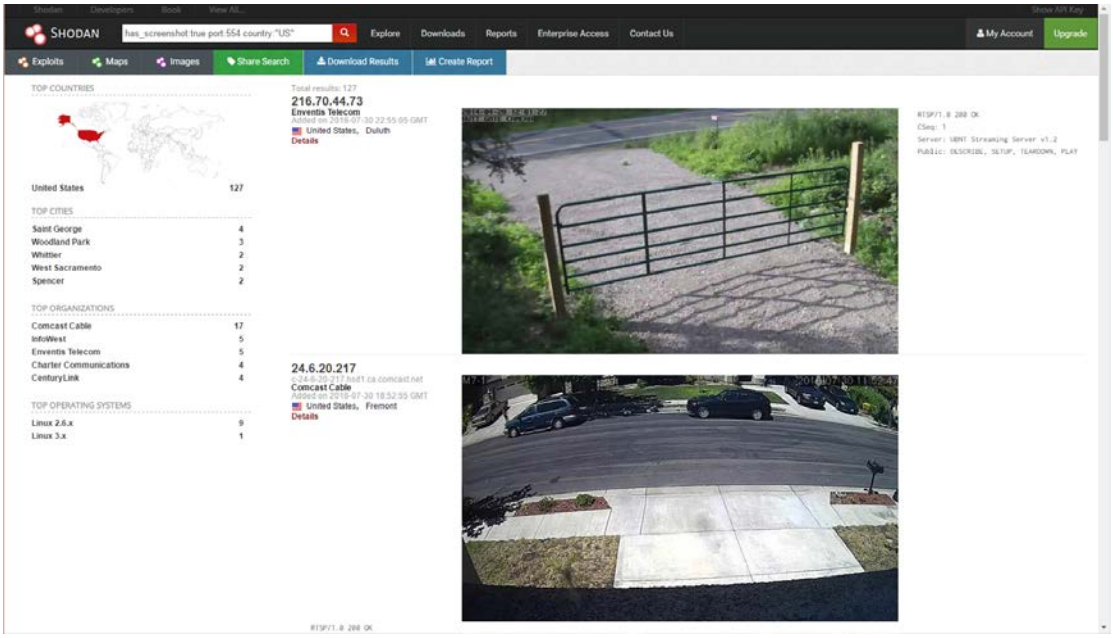

Fig.3 Vulnerable devices found by Shodan

Some of the major threats in Smart Home environments are mentioned as follows.

Table.1 Major types of threats in Smart Homes and their impact[7]

Type	***Impact of the threat***
Eavesdropping	Perform traffic sniffing and it is a threat to data confidentiality.
Masquerading	Claims authentic user, steals legal authorization information.
Replay attack	Gathers the valid info and later used it for exploitation.
Message modification	Fabricates the information between legal users and it is a threat to integrity.
Denail of service (DoS)	Interrupts the services to legal users and it is a threat to availability.
Malicious codes	To alter or damage the information and allow remotely unauthorized access.

2.2 Feature deviation based taxonomy[8]

Based on how the attacker deviates the features of compromised Smart Home devices from their functionality. Eyal et al. has proposed a new taxonomy of the attacks

1) Ignoring the functionality
2) Reducing the functionality
3) Misusing the functionality
4) Extending the functionality

In the first type of attacks, the attacker ignores the intended physical functionality of the IoT devices. These devices are only considered as computing devices with LAN or Internet access. Compromised devices may form a botnet used by the attacker to launch Distributed Denial of Services (DDoS) attacks or to mine bitcoins.

In the second type of attacks, the attacker tries to kill or reduce the normal functionality of the IoT devices and to prevent them from working normally. Financial loss or chaos and panic or even fatal threats may be caused by this kind of attack.

In the third type of attacks, the attacker uses the compromised devices in an incorrect or an unauthorized way. For instances, a hacker may reverse the policies inside a temperature control system to heat the room when the temperature is relatively high.

In the last type of attacks, attacker extends the designed functionality of the compromised IoT device, and uses it in order to achieve a completely unexpected physical effect. Compromised PWM driven LEDs may be manipulated to convey private or confidential information via frequency signals which is invisible to human eyes.

2.3 Related work

Denning *et al*. outlined a set of emergent threats to smart homes due to the swift and steadyintroduction of smart devices[9]. For example, there are threats of eavesdropping and direct compromise of various smart home devices. Denning *et al*. also discussed the structure of attacks that include data destruction, illegal physical entry, and privacy violations, among others.

Huichen *et al*. highlighted the dependency between the security of the domestic device network and the installation and configuration of the policies and mechanisms, and emphasized that for non-export users, vulnerabilities cannot be avoided. Huichen *et al*. proposed an auto-configuration and auto-update system to enhance system security in Smart Home systems. [5]

Earlence *et al*. analyzed the security problem among emerging Smart Home applications, mainly focusing on the software level i.e. programing framework of Smart Home systems and exploited framework design flaws to construct four proof-of-concept attacks: secretly planted door lock codes, stole existing door lock codes, disabled vacation mode of the home and induced a fake fire alarm.[10]

Other researches focusing on solution based on machine learning and statistics methods are further stated in *Sec. 3*.

3. Existing Machine-Learning based implementations

3.1 Smart Home Networks User Authentication using Neural Network[11]

Unlike some authentication systems which traditionally use a password table or verification table, in this authentication scheme, a RBF Neural Network (NN) is used to recall the relationship of username and password. This scheme can produce the corresponding encrypted password according to the entered username, replacing the hashed password table or verification table stored in common authentication systems.

This scheme can be divided into two phases: registration phase and authentication phase.

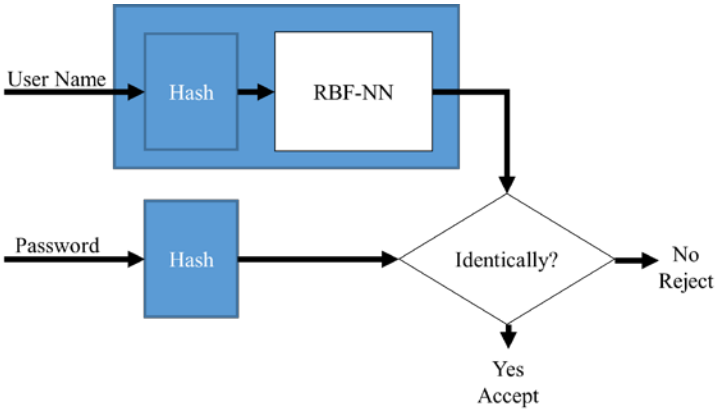

Fig. 4 The user authentication phase in the NN based system

In the registration phase, the system administrator obtains the training patterns from usernames and passwords to train the neural network, which is, more specifically, to feed the

NN algorithm with the hashed username as the input and the corresponding hashed and normalized password as the output and then to store the weights in the system database.

In the authentication phase, the system applies the same hash function on the entered username and password, inputs the hashed username to the trained NN and compares the result of the NN with the hashed password. If the two match, the user is authorized; if not, the user is rejected.

The three advantages regarding this mechanism are that an intruder cannot add a forged username and password pair to the neural network, the simple computation operations to produce results, and that the training time of RBF neural networks is short and acceptable.

3.2 Activity and Anomaly Detection in Smart Home[12]

Increasing use of Smart Home devices has motivated the research and development of the monitoring technologies for Smart Home systems.

As is surveyed in [12], recently more attention is given on anomaly activity detection, activity prediction, incorporating contextual aspects including temporal, spatial and health status, concept drift, irregular behavior adaptive learning and online data stream. The main approaches to detect anomalies are classification based, clustering based and statistical based methods.

Table.2 Comparison between algorithms used for anomaly detection

Algorithm	Pros	Cons
Gaussian mixture model	Could relate two data attributes for activity classification	Reducing the matching times and eventually improve the detection efficiency
Hidden Markov model	Simple; Handling sequential data; Having temporal dependency structure; A statistical model that handles noisy data	Need a full description of the big data; Requiring lots of trainings; Supervised Learning; Not fully capturing dependency structure of the data: a conditionally independent assumption
Conditional random field	Could capture long range dependency data structure; A statistical model that handles noisy data	Expensive training cost when capturing long-range dependency data structure; Supervised Learning; Requiring 1-class CRF for anomaly detection when anomaly data instance is rare or unavailable
Artificial neural network	Being able to add new rules	Complex network architecture; Not understandable logic and rules behind the trained model;

Support vector machine	Provides a good out-of-sample generalization data; Linearly separable	Requiring 1-class CRF for anomaly detection when anomaly data instance is rare or unavailable

4. Our proposal: Machine-Learning Based Instruction Authentication

4.1 Motivations

Authentication is a critical part of Smart Home system. Verification is the central part of user authentication in Smart Home systems.

It can be used as the fundamental of access control. For instance, user verification function can be activated sometimes or in different levels in case anyone in the room can modify the settings of Smart Home systems without limitation.

Also along with anomaly detection or activity monitoring mechanisms implemented on gateways or other relatively resource-rich terminals, it can be used to enhance the security of Smart Home system be means of continuous verification.

In *Sec. 4.2*, we propose a machine-learning based one-point password verification mechanism. Along with the anomaly detection and activity monitoring, a continuous verification security mechanism on network dataflow is proposed in *Sec. 4.3*.

4.2 Machine-Learning Based password verification

On condition that password based verification is widely used in a majority of aspects of people's daily life, due to its simple implementation, and that Smart Home devices are limited in hardware resources, we propose a one-point verification based on the password as well as the typing pattern, using machine-learning methods, to realize security enhancement in a resource-limited system.

On the hardware level, the devices should have basic computing power and storage capability to support a basic operating system (OS). The devices should also support touch screen operations.

A user is required to touch the screen to enter the password in some specific typing pattern such as the combination of touching, pressing and sweeping over the digits at different pace for several times. The program will record the time serious of digits and "press" and "release" operations, as is illustrated in *Fig. 5*.

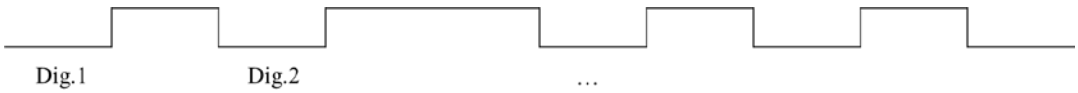

Fig. 5 The time series password data

The raw data will be processed through feature extraction process. The algorithm will then be trained based on this pattern. Once completed, the user is registered as a legitimate user corresponding to his typing pattern in the database.

The input data will always be recorded and used to enhance the algorithm and gradually fit the reasonable concept drift within a learning process. The recorded data are also used to reject the replay attack, which is a kind of attack mentioned in *Table. 1*.

4.3 Anomaly detection based instructions legitimacy verification

In a Gateway-based Internet Access Architecture Smart Home network, the gateway usually works as a firewall and has relatively rich hardware resources to monitor the activities and dataflow.

Along with the password verification, the gateway will monitor the dataflow and detect anomalies continuously. When anomalies like being accessed by abnormal address, irregular dataflow flooding, strange operations etc. are detected, the gateway will block the traffic or hold the operation by sending a "revoke" instruction and enter the password verification process proposed in *Sec. 4.2*.

Once the password and the typing pattern is authenticated, the gateway will approve the traffic or resume the operation.

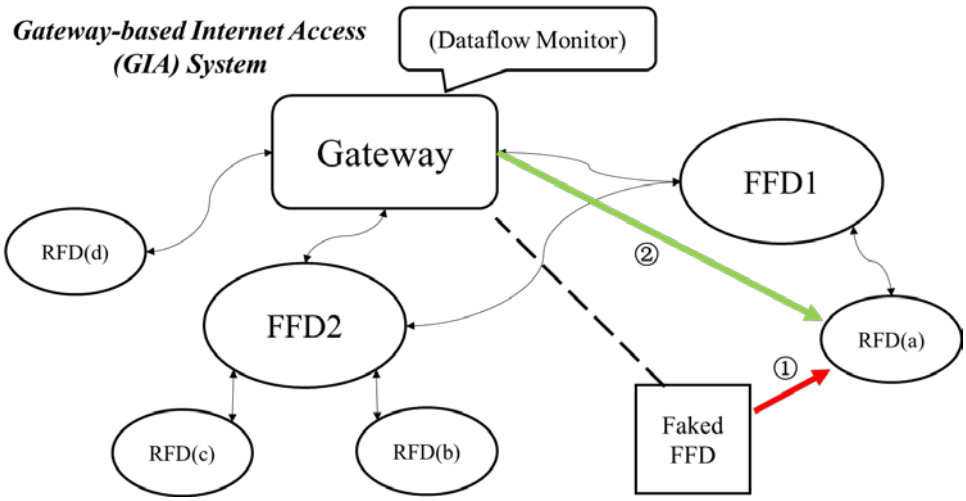

Fig. 6 Illustration of instructions legitimacy verification in GIS system

RFD – Reduced Function Device; FFD – Full Function Device; Faked FFD – attacker device.

① The faked FFD launches an instruction to RFD(a).

② After detecting the anomaly, the gateway issues a "revoke" instruction and enter the password verification process. Once authenticated, the previous instruction will be issued by the gateway again to resume the previous operation.

5. Preliminary Experiment

5.1 Experiments introduction

We have conducted a preliminary experiment on the first part of our proposal, which is stated in *Sec 4.1*.

We collect the PSW as well as the typing pattern of the legitimate user via a specifically developed Android APP, whose User Interface (UI) is shown in the screenshot below (Fig. 7).

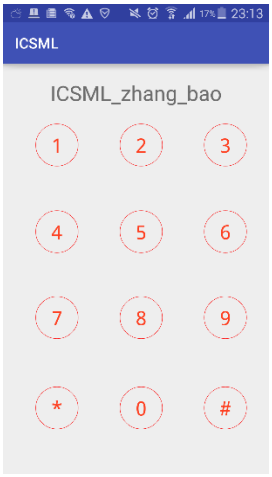

Fig. 7 The User Interface of the Android APP

This APP can collect data of the time stamps of the “Press” and “Release” events of the buttons. We use this APP to record the typing pattern of the password. Data of two different legitimate typing patterns are collected, one of which is normal and not specifically featured and the other of which is intentionally featured.

Examples of the pre-processed legitimate data is listed in *Table 1*. From *Table 1*, we can also draw the trivial conclusion that the than normal ones

Table.3 Examples of the pre-processed legitimate data

Down1	***Up1***	***Down2***	***Up2***	***Down3***	***Up3***	***Down4***	***Up4***	***Down5***	***Up5***	***Down6***	***Standard Deviation (Downs)***	***Standard Deviation (Ups)***	***Category***
104	586	82	563	95	608	83	541	61	529	85	13.23	28.86	Normal
84	651	84	484	107	451	75	408	63	683	64	14.89	110.58	Normal
85	528	74	650	107	506	86	428	86	538	87	9.78	71.341	Normal
731	639	161	934	64	351	43	319	65	450	76	244.82	227.02	Pattern
689	803	106	660	64	319	54	285	65	462	97	228.76	198.83	Pattern
690	494	86	550	75	341	42	310	51	451	65	233.82	90.83	Pattern

We then collect data from other users, which is labeled as “Negative(reject)”, to construct two relatively balanced datasets.

On Weka platform, we implemented eight mainstream classifier on the dataset, using 10-fold cross-validation method. The results and analysis is stated below in *Sec. 5.2*.

5.2 Results and Analysis

Generally, we obtained positive results during the preliminary experiments.

According to the results on non-featured normal password dataset, AdaBoost performed the best, with FPR = 0.057, AUC = 0.975 and competitive training time.

Table.4 Results of the experiments on non-featured PSW dataset

Algorithm	***TPR***	***FPR***	***Precision***	***Recall***	***F-Measure***	***AUC***	***Training Time(s)***
NN	0.76	0.2	0.8	0.76	0.76	0.893	0.03
Logistic	0.84	0.129	0.866	0.84	0.841	0.832	<0.01
Naïve Bayes	0.88	0.113	0.884	0.88	0.881	0.969	<0.01
SGD	0.88	0.087	0.907	0.88	0.881	0.897	<0.01
AdaBoost	0.94	0.057	0.941	0.94	0.94	0.975	<0.01
LogitBoost	0.92	0.084	0.92	0.92	0.92	0.984	0.01
C4.5(J48)	0.92	0.097	0.922	0.92	0.919	0.869	<0.01
Random Forest	0.9	0.099	0.901	0.9	0.9	0.971	0.01

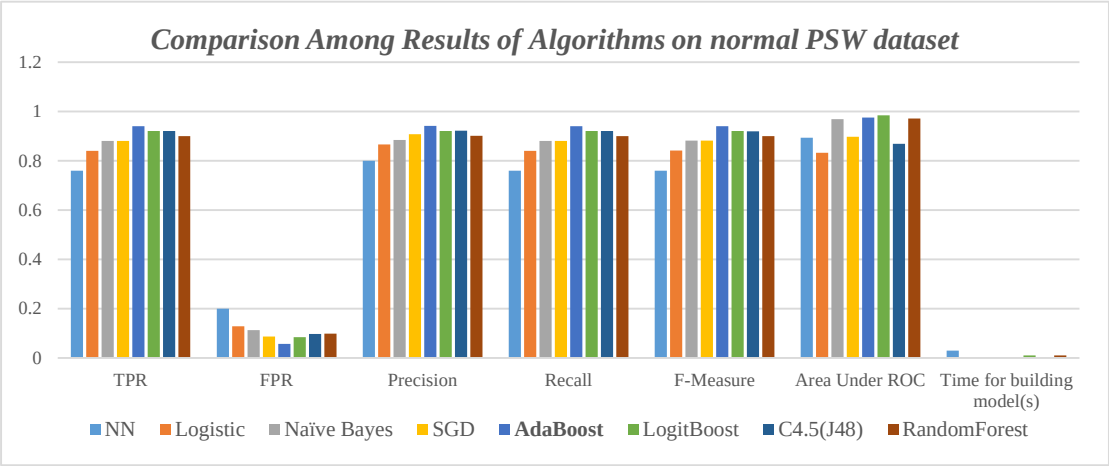

Fig. 8 Comparison among results on non-featured PSW dataset

According to the results on featured normal password dataset, Logistic performed the best, with FPR = 0.016, AUC = 1 and competitive training time.

Table.5 Results of the experiments on featured PSW dataset

Algorithm	***TPR***	***FPR***	***Precision***	***Recall***	***F-Measure***	***AUC***	***Training Time(s)***
NN	0.981	0.016	0.982	0.981	0.982	0.974	0.04
Logistic	0.981	0.016	0.982	0.981	0.982	1	<0.01
Naïve Bayes	0.926	0.08	0.928	0.926	0.926	0.982	<0.01
SGD	0.981	0.016	0.982	0.981	0.982	0.983	0.02
AdaBoost	0.981	0.016	0.982	0.981	0.982	1	0.01
LogitBoost	0.963	0.032	0.966	0.963	0.963	1	0.02
C4.5(J48)	0.944	0.048	0.95	0.944	0.945	0.966	<0.01
Random Forest	0.944	0.048	0.95	0.944	0.945	0.993	0.01

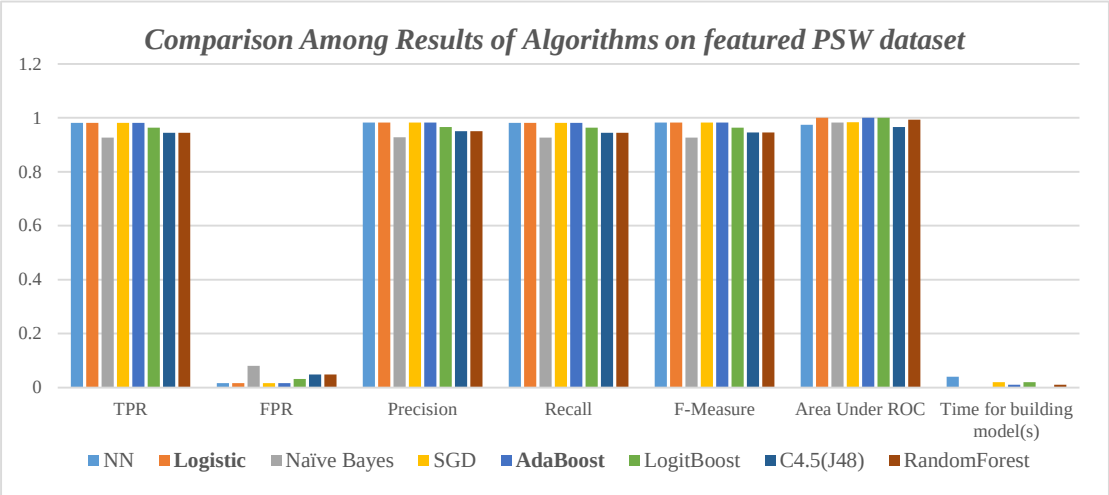

Fig. 9 Comparison among results on featured PSW dataset

From the comparison of the results on the two datasets, it is demonstrated that our proposed pattern, which is featured password verification based on machine learning, enjoys a high improvement on False Positive Rate (FPR) and a growth on Area Under Curve (AUC) as is illustrated below.

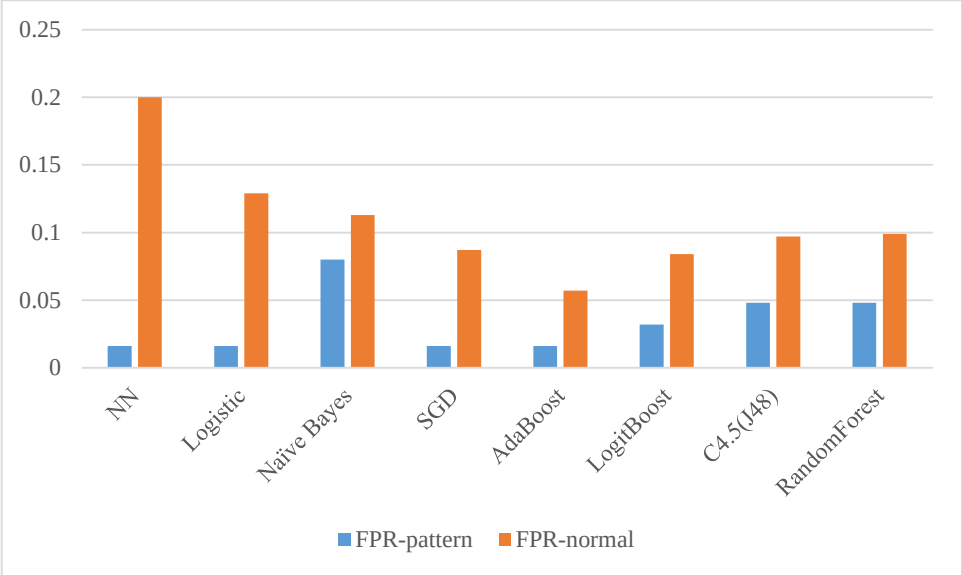

Fig. 10 Improvement on FPR measurement

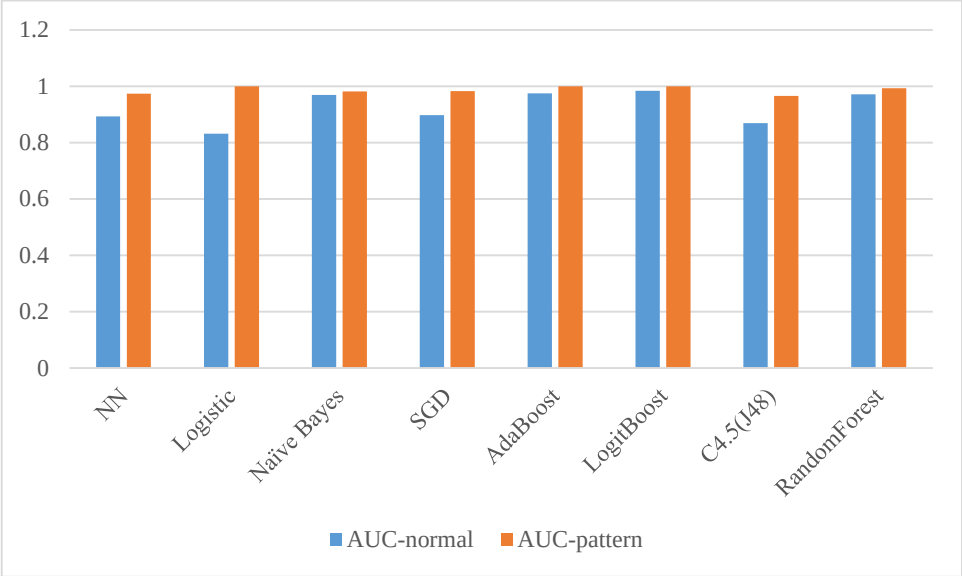

Fig. 11 Improvement on AUC measurement

The percentage of the improvement under FPR and AUC measurements is listed below in *Table.6*.

Table.6 Improvement on FPR and AUC between the two datasets

Algorithm	***Enhancement on FPR***	***Enhancement on AUC***
NN	92.0%	9.07%
Logistic	87.6%	20.19%
Naïve Bayes	29.2%	1.34%
SGD	81.6%	9.59%
AdaBoost	71.9%	2.56%
LogitBoost	61.9%	1.63%
C4.5(J48)	50.5%	11.16%
RandomForest	51.5%	2.27%

The improvement of performance under FPR measurement is of significant value in practical fields.

The results show that using password with specific typing pattern can significantly improve the verification effectiveness under FPR measurement. Even the digits of the password are leaked, the typing pattern is still hard to mimic, which enhance the security level of passwords.

Given the verification is not continuous, the mechanism can be implemented to devices with constraints on hardware resources in Smart Homes.

5.3 Future works

Given the data collected in this mechanism is a form of time-series data, our next step is to implement Dynamic Time Warping (DTW) algorithm, which is an algorithm for measuring similarity between two temporal sequences which may vary in speed, along with a Nearest Neighbor Classifier.

We will implement the verification mechanism independently on ARM devices to evaluate performance and time latency.

Simulation of the second part of our proposal will also be conducted and evaluated in future works.

6. References

[1] (2015) Unlocking the potential of the internet of things. [Online]. http://www.mckinsey.com/business-functions/business-technology/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world

[2] Wade, V.; Soar, J.; Gray, L. Uptake of telehealth services funded by Medicare in Australia. *Aust. Health Rev.* 2014, 38, 528–532.

[3] Shelby, Z.; Hartke, K.; Bormann, C. *The Constrained Application Protocol (CoAP)*; RFC 7252; Internet Engineering Task Force: Fremont, CA, USA, 2014.

[4] Alam, M.R.; Reaz, M.B.I.; Ali, M.A.M. A Review of Smart Homes—Past, present, and future. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* 2012, 42, 1190–1203.

[5] Huichen Lin; Neil W. Bergmann. IoT Privacy and Security Challenges for Smart Home Environments. *Information* 2016, 7, 44;

[6] Patton, M.; Gross, E.; Chinn, R.; Forbis, S.; Walker, L.; Hsinchun, C. Uninvited connections: A study of vulnerable devices on the Internet of Things (IoT). *IEEE Joint Intelligence and Security Informatics Conference (JISIC)*, 2014; pp. 232–235.

[7] Shafiq Ul Rehman; Selvakumar Manickam. A Study of Smart Home Environment and its Security Threats. *International Journal of Reliability, Quality and Safety Engineering*, Vol. 23, No. 3 (2016)

[8] Eyal Ronen; Adi Shamir. Extended Functionality Attacks on IoT Devices: The Case of Smart Lights. *IEEE European Symposium on Security and Privacy*. 2016

[9] T. Denning; T. Kohno; H. M. Levy, Computer security and the modern home, *Commun. ACM*, Vol. 56, No. 1, pp. 94–103, Jan. 2013. [Online]. Available: http://doi.acm.org/10.1145/2398356.2398377

[10] Earlence Fernandes; Jaeyeon Jung; Atul Prakash. Security Analysis of Emerging Smart Home Applications. *IEEE Symposium on Security and Privacy*. 2016.

[11] Shahbaz Zahr Reyhani; Mehregan Mahdavi. User Authentication Using Neural Network in Smart Home Networks", *International Journal of Smart Home*, Vol. 1, No. 2, July, 2007

[12] U.A.B.U.A. Bakar; Hemant Ghayvat; S.F. Hasanm; S.C. Mukhopadhyay. Activity and Anomaly Detection in Smart Home: A Survey. *Next Generation Sensors and Systems*, Vol. 16, pp.191-220, July 2015.