PEEP: Passively Eavesdropping Private Input via Brainwave Signals

[1]Ajaya Neupane, [2]Md Lutfor Rahman *, [1]Nitesh Saxena

[1]University of Alabama at Birmingham, [2]University of California Riverside

Abstract. New emerging devices open up immense opportunities for everyday users. At the same time, they may raise significant security and privacy threats. One such device, forming the central focus of this work, is an EEG headset, which allows a user to control her computer only using her thoughts.

In this paper, we show how such a malicious EEG device or a malicious application having access to EEG signals recorded by the device can be turned into a new form of a keylogger, called PEEP, that passively eavesdrops over user's sensitive typed input, specifically numeric PINs and textual passwords, by analyzing the corresponding neural signals. PEEP works because user's input is correlated with user's innate visual processing as well as hand, eye, and head muscle movements, all of which are explicitly or implicitly captured by the EEG device.

Our contributions are two-fold. First, we design and develop PEEP against a commodity EEG headset and a higher-end medical-scale EEG device based on machine learning techniques. Second, we conduct the comprehensive evaluation with multiple users to demonstrate the feasibility of PEEP for inferring PINs and passwords as they are typed on a physical keyboard, a virtual keyboard, and an ATM-style numeric keypad. Our results show that PEEP can extract sensitive input with an accuracy significantly higher than a random guessing classifier. Compared to prior work on this subject, PEEP is highly surreptitious as it only requires passive monitoring of brain signals, not deliberate, and active strategies that may trigger suspicion and be detected by the user. Also, PEEP achieves orders of magnitude higher accuracies compared to prior active PIN inferring attacks. Our work serves to raise awareness to a potentially hard-to-address threat arising from EEG devices which may remain attached to the users almost invariably soon.

1 Introduction

Brain-computer interfaces (BCI), which extract physiological signals originated in the human brain to communicate with external devices, were once highly expensive and used only in medical domains. They were mainly used to develop neuroprosthetic applications which helped disabled patients to control prosthetic limbs with their thoughts alone [35]. However, these devices are now commercially available at low-cost and are becoming popular especially in gaming and entertainment industries.

* Work done while being a student at UAB

Electroencephalography (EEG) is the most commonly used physiological signal in the BCI devices due to its ease of use, high temporal resolution, and non-invasive setup. EEG measures the task related to electrical activity of the brain, referred to as event-related potentials. In the commercial domain, these EEG-based BCI devices have been used to improve the quality of user experience mainly in gaming and entertainment industries. Currently, EEG-based BCI devices from different vendors are available in the market (e.g., Emotiv [3], Neurosky [7], Neurofocus [5]). These devices also provide software developments kits to build applications, and have application markets (e.g. [2, 6]) in which the vendors host the applications developed by their own developers as well as provide a platform for third-party developers to share the applications developed by them. Recently, the BCI devices have been studied for building user authentication models based on user's potentially unique brainwave signals [17].

Given their interesting use cases in a wide variety of settings, the popularity and applicability of these devices is expected to further rise in the future. These devices may become an inevitable part of a users' daily life cycles, including while they use other traditional devices like mobile phones and laptop/desktop computers. In this light, it is important to analyze the potential security and privacy risks associated with these devices, and raise users' awareness to these risks (and possibly come up with viable mitigation strategies).

Our specific goal in this work is to examine how malicious access to EEG signals captured by such devices can be used for potentially offensive proposes. As the use of these devices becomes mainstream, a user may enter passwords or private credentials to their computers or mobile phones, while the BCI device is being worn by the user. To this end, we study the potential of a malicious app to capture the EEG signals when users are typing passwords or PINs in virtual or physical keyboards, and aim to process these signals to infer the sensitive keystrokes. The device to which the sensitive keystrokes are being entered could be the same device with which the BCI headset is "paired" or any other computing terminal. Several previous studies have used EEG signals to infer the types of mental tasks users are performing [36], to infer the objects users are thinking about [21], or to infer the limb movements users are imagining [33]. In line with these works, the premise of our presented vulnerability is that the user's keystroke input to a computer would be correlated with the user's innate visual processing as well as users hand, eye and head muscle movements, as the user provides the input all of which are explicitly or implicitly captured by the BCI devices.

Based on this premise, we demonstrate the feasibility of inferring user's sensitive keystrokes (PINs and passwords) based on their neural signals captured by the BCI device with accuracies significantly greater than random guessing. These BCI brain signals may relatively easily get leaked to a malicious app on the mobile device that is paired with the BCI headset since no extra permissions to access such signals is required in current mobile or desktop OSs. An additional avenue of leakage lies with a server, charged with the processing of brain signals in the outsourced computation model, which may get compromised or be malicious on its own.

Our Contributions and Novelty Claims: In this paper, we introduce a new attack vector called PEEP that secretly extracts private information, in particular users' private

input such as PINs and passwords, from event-related potentials measured by brain computer interfaces. Our contributions are two-folds:

- We design and develop PEEP, a new type of attack against keystroke inference exploiting BCI devices based on machine learning techniques. We study PEEP against a commodity EEG headset and a higher-end medical-scale EEG device
- We experimentally validate the feasibility of PEEP to infer user's PINs and passwords as they are being typed on a physical or virtual keyboard. We also validate the consistency of results across different BCI headsets.

Related to PEEP, Martinovic et al. [29]) studied the possibility of side-channel attacks using commercial EEG-based BCI to reveal the users' private information like user's familiar banks, ATMs or PIN digits. Their general idea is similar to a guilty knowledge test where items familiar to a user is assumed to evoke the different response as compared to the items unfamiliar to the user. Thus, when a person is shown images of many banks, the brain response to the image of the bank with which user has had more interaction or has opened an account will evoke higher event-related potential. However, their attack setup is intrusive and can be easily detectable as the users may notice the abnormality in the application when it shows the images of banks or ATMs related to her. In contrast, PEEP is highly surreptitious as it only requires passive monitoring of brain signals as users' type their PINs and passwords in regular use, not deliberate, and active strategies that may trigger suspicion and be detected by the user. In addition, PEEP achieves orders of magnitude higher accuracies compared to the active PIN inferring attack of [29].

2 Background & Prior Work

2.1 EEG and BCI Devices Overview

Electroencephalography (EEG) is a non-invasive method of recording electrical activity in the brain, referred to as event-related potentials (ERPs), using electrodes on the surface of the scalp. EEG has higher temporal resolution and can depict changes within milliseconds. The electrical activity can be synchronized with the performed task to study changes in brain activation over time. ERPs are used as a tool in studying human information processing [20]. P300, a positive change in ERPs which appears around 300ms post-stimuli if the stimuli is a known target, is popularly used ERPs in studies involving EEG. Many devices, both consumer-based and clinical-based devices to measure the ERPs are currently available in market and are used in security studies (see Section 2.2).

In this study, we used two different EEG headsets for data collection, namely Emotiv Headset [3] and B-Alert Headset [1]. We use Emotiv as a representative instance of current commercial consumer-grade BCI devices, and B-Alert (a clinical-level Headset) as a representative instance of future devices.

Emotiv Epoch Headset: Emotiv Epoc headset is a wireless and lightweight EEG sensor to acquire and analyze 14 channels of high-quality EEG data. The sensors of this

EEG headset follow the 10-20 international system of placement. It uses the AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF42 sites to collect EEG data at 128Hz.

B-Alert Headset: The B-Alert headset is a clinical grade X10-standard wireless and lightweight system, developed by Advanced Brain Monitoring (ABM) [1], to acquire nine channels of high-quality EEG data. The headset also followed the 10-20 international system of electrode placement and used Fz, F3, F4, C3, Cz, C4, P3, POz, P4 sites to collect EEG data at 256 Hz with fixed gain referenced to linked mastoids. The tenth channel was used for measuring electrocardiogram signals. A portable unit is worn on the back of the head which amplifies and sends signals to the computer connected over Bluetooth.

2.2 Related Work

Information Retrieval using Brain Activations: EEG has been explored by researchers to develop user authentication model (for example, [10,17,25,30,37]). Ashby et al. [10] proposed an EEG based authentication system using a consumer grade 14-sensor Emotiv Epoc headset. Thorpe et [37] al. suggested pass-thoughts to authenticate users. Chuang et al. [17] used single-sensor Neurosky headset to develop a user authentication model based on ERPs collected during different mental tasks including pass-thoughts. Bojinov et al. [14] proposed a coercion-resistant authentication based on neuroscience based approach. Most relevant to our work, Martinovic et al. [29] used ERPs as a vector of side-channel attack to snoop into users private information. The authors showed images of numbers, banks, ATMs to the participants when their brain signals were measured. They used the brain signal to decrease entropy of information related to PIN, banks, ATMs by 23-40%. However, our attack is less intrusive and difficult to detect and our malicious app can run in background capturing EEG signals.

The BCI devices are also used to understand users' underlying neural processes when they are performing security tasks. Neupane et. al used fMRI [31] to study brain activations when users were subjected to phishing detection and heeding malware warnings. In another study, Neupane et al. [9] used EEG-based B-Alert Headset to measure mental states and mental workload when users were subject to similar security tasks.

Campbell et al. [16] used P300 ERP, originated when someone shows attention to specific stimuli, for developing neurophone, a brain controlled app for dialing phone number in mobile address book. The authors flashed a number of photos of contact persons in participants' address book, and when P300 potential amplitude for a photo matched the person the user thought of dialing, the app dialed the phone number.

Birbaumer et al. [13] proposed spelling device for paralyzed based on the P300 spikes. The alphanumeric characters were organized in a grid and were flashed to the patient. Whenever the patient focused on the target character, P300 was evoked. Tan et al. [36] asked users to perform different gaming tasks and used ERPs to classify what mental tasks users were performing. Esfahani et al. [21] used 14-channel Emotiv headset to collect neural data from 10 users when they were imagining cube, sphere, cylinder, pyramid or a cone. They were able to discriminate between these five basic primitive objects with an average accuracy of 44.6% using best features in Linear Discriminant Classifier (random guessing would have been 100/5 =20%).

Other Side Channel Attacks: Keystroke inference has received attention due to its potential consequences. Asonov et al. [11], Zhuang et al. [41] and Halevi et al. [22] used sound recorded from physical keyboards when users were typing passwords to infer keystrokes. Vuagnoux et al. [38] used electromagnetic waves emanated on users typing such keyboard. Song et al. [34] used inter-keystroke timing observations to infer keystrokes. Marquardt et al. [28] used accelerometer on a smartphone to record and interpret surface vibrations due to keystrokes to identify the user inputs on a nearby keyboard. All these side channel attacks exploited the physical characteristics of the keyboard, which became infeasible after the advent of smart phone with touch screen. However, new types of attacks to detect users' PINs, passwords and unlock patterns using motion sensors emerged on these smartphones [12, 15, 32, 40].

Unlike these attacks, we propose a new form of keylogger. We show how a malicious EEG device or a malicious application having access to EEG signals recorded by BCI device, can be used to elicit users' private information. We show the feasibility of our attack in both the physical keyboard and virtual keyboards.

3 Threat Model

The attackers' motive in this study is to passively eavesdrop on victim's neural signals, recorded by BCI devices, looking for sensitive information (e.g., PINs or passwords) entered on a virtual or a physical keyboard. The BCI devices provide APIs which allow easy access of raw signals recorded by the BCI devices to app developers. So a third-party developer can develop a malicious app with unfettered access to the ERPs measured by such BCI device. The app developed by attackers first captures the neural patterns of keystrokes to build a classification model (Training Phase) and later utilizes the model to infer the keystrokes only using the neural data (Testing Phase). Such malicious app developers are considered adversary of our system.

Training Phase: We assume the adversary has developed a malicious application to record neural signals and has fooled the victim to install the app on her device. The malicious application can be a gaming application which asks users to press different keys for calibration or enter particular numeric/alphabetical code before playing different levels of the game or resuming the game after a break. The developer can claim such codes will secure the game from being played by other users who has access to the computer. The attacker can then process the numeric/alphabetical code and neural signals corresponding to them to extract features and build a training model. The threat model is similar to the attack model studied in previous work [29]. However, our threat model is less intrusive and weaker as compared to their study as they propose explicitly showing images of ATMs or PINs to users, which users may eventually notice.

It is also possible for attacker to obtain keystroke-neural template may be leaked through servers. For example, a benign application may outsource these signals to some server for computations which may be malicious or can get compromised and can infer sensitive info.

We also assume a different threat model in which attacker does not have access to victims' keystrokes and corresponding brain signals. In this case, we assume the

attacker builds a training model using her brain data and keystrokes. The training model is then employed in PEEP.

Testing phase: We assume the attacker has now developed a training model to classify neural signals for each of the numbers and the alphabetic keys using one of the methods described in the previous section. The malicious app with training model is successfully installed in victims device and runs in the background stealthily recording the neural signals whenever victim enters sensitive information in the physical or virtual keyboard. We assume the attacker knows when the victim is entering private credentials in the device (e.g., in mobile devices, the keyboard shows up whenever the user starts to type). These neural signals recorded during the entry of these credentials will then be used by the app to infer the keystrokes which can then be exploited by attackers.

Apart from mobile and desktop apps, these devices also provide web APIs [4] which can be exploited to launch remote attacks. In this case, browser add-ons can be the malicious apps. In our threat model, we assume the victim only uses random numbers or random uppercase character-based passwords. We keep the length of the PIN to 4 and password to 6.

Practicality of Attacks: BCI devices are used by gamers to play games controlled by their mind. The game they are playing is malicious in nature. It asks users to enter predefined set of numbers (like captcha) to restart the game from the last position when they take a break. Doing this, the malicious app can record the ERPs related to each of the entered digits. The app can then be trained with these recorded datasets to predict keystrokes correctly. Now, when the gamers next take a break from the game and enter their login credentials in banking or social media websites, with the headset on, the app can listen to the brain signals and then run the classification model to predict keystrokes.

4 Experimental Design & Data Collection

4.1 Design of the Task

We followed the similar design for all of our experiments, while we varied different parameters, such as users, EEG devices (Emotiv vs. high-end), keypads (virtual vs. real), and data types (4-digit pin vs. 6-character password). Even though the experiments were conducted in controlled lab environment, we tried to simulate real-world PIN/Password entry methodologies. The design of the experiments remained same for both Emotiv and B-Alert headsets.

Virtual Keyboard PIN Entry (VKPE): The goal of this experiment was to assess whether the event-related potentials recorded using consumer-based EEG BCI device or B-Alert headset could be used to infer the numbers entered by the participant. We assume, visual and mental processing of digits, along with the head, hand, and eye movements while entering PIN may tell what key is being processed. For this task, we developed a virtual keyboard similar to the ones employed in login pages of websites (this layout is also similar to the numeric keyboards in smart phones in landscape view) (see Appendix A Figure 1(a)). We had a text box at the top of this virtual keyboard.

The participants were asked to enter 4-digit PIN codes using the mouse in the text box. When the user clicked a key on the virtual keyboard, the key was flashed in its frame for 500ms or till the next key was clicked, similar to the key press events in touch pads of smart phones. This was done to ensure the user that he had clicked on the right digit. When the user pressed a key, we put a trigger in the recorded event-related potentials to synchronize the neural data with key presses.

Virtual ATM PIN Entry (VAPE): Similar to the design of the virtual numeric keyboard, for this task, we implemented a virtual ATM keyboard with a text box at the top (see Appendix A Figure 1 (b)). The participants were asked to enter 4-digit PIN codes in the text box using the mouse. Like the previous designs, we assumed visual and mental processing of digits might tell what key is being processed. However, this design had the fewer number of keys in the keyboard compared to the virtual keyboard, so we expected the distraction while entering PINs to be lower and results of the prediction model to be higher for this task. This layout is also similar to the numeric keypad in smart phones in portrait view.

Physical Numeric Keypad PIN Entry (PNKPE): For this task, we developed a frame with a text box for entering PIN. Similar to the previous tasks, the participants were provided with random 4-digit numeric PINs and were asked to enter them in the text box. However, the mode of the key input, in this case, was a physical numeric keyboard, unlike virtual keyboard in previous tasks (see Appendix A Figure 2). In this task we assumed, the mental processing of digits, and the movements of facial muscles, eyes, head, hands, and fingers may create a digit-specific pattern in event-related potentials. These features may be eventually used to develop PEEP.

Physical Keyboard Password Entry (PKPE): In this task, we used a frame with the text box to enter the password. The participants were provided with random upper-case 6-character based passwords and were asked to enter the password in the text box using physical keyboard (e.g., laptop keyboard) (see Appendix A Figure 3). Like the previous task, in this task, we assumed the finger/hand movement to create a digit-specific pattern in event-related potentials, which may eventually be used to develop PEEP.

4.2 Experimental Set-up

For all the above mentioned tasks, we collected data in the lab environment using two different headsets, namely Emotiv and B-Alert Headsets. The basic set-up for both the experiments were similar, apart from the computer used for data collection. For Emotiv headset our experimental set-up comprised of a single laptop in which the Emotiv control panel, the virtual keyboards, and the text-input frames were installed. The Emotiv control panel was used to calibrate the headset for better signal-to-noise ratio. An in-house program, developed to record the neural data and the key press logs, was also installed in the stimuli computer (see Figure 1 left).

Fig. 1: (a)Experimental set up with Emotiv headset (b) Experimental set up with B-Alert headset (face masked for anonymity)

For the B-Alert headset, we used stimuli computer to present experimental tasks and a different data collection computer to record the neural data. The proprietary B-Alert data acquisition software installed in this data collection computer was used to calibrate and record brain data during the task. A signal was sent from stimuli computer to data collection computer using TCP/IP connection to mark the neural data on each key-press to synchronize the brain data and corresponding keystrokes. We could not install the B-Alert data acquisition software in stimuli computer as it was a proprietary software with the license for lab computer only (see Figure 1 right).

4.3 Study Protocol

Ethical and Safety Considerations: The study was approved by the Institutional Review Board of our university. The participants were recruited using flyers around the campus and on the social media (e.g. Facebook). The participation in our experiment was strictly voluntary, and the participants were provided with an option to withdraw from the research at any point in time. The best standard procedures were applied to ensure the privacy of the participants' data (survey responses, and neural data) acquired during the experiment.

Participant Recruitment and Pre-Experiment Phase: Twelve healthy members of our university (including students, housewives, and workers) were recruited for our study. Informed consent was obtained from these participants and were asked to provide their demographic information (such as age, gender and education level). Our pool was comprised of 66.6% male and 33.3% female, 55% were above the age of 24 and belonged to fairly diverse educational levels (e.g., computer science, civil engineering, business administration, etc.). Ten of these participants performed VKPE task. Rest of the three tasks were performed by two participants each. Some of these participants were among the ten participants who had performed VKPE task.

Task Execution Phase: We used the consumer-based 14-sensors Emotiv headset and 10-sensor B-Alert headset for the experiment. We prepared Emotiv headset and B-Alert headset for proper measurement of the electrical activity in the brain. We then placed the headset on the head of the participant. We calibrated the headset using Emotiv control panel and B-Alert software respectively, where we can validate the signal strength of

each electrode, for obtaining better signal-to-noise ratio. Once the headset was properly calibrated and the participants were seated comfortably to perform the task, we provided them with a sheet of paper with randomly generated thirty 4-digit random PINs or randomly generated thirty-six upper-case 6-character random passwords depending on the tasks they were performing.

We instructed participant to enter the PINs or passwords in the text box as if she was logging into her accounts. In case, she realizes to have entered the wrong digit; she was instructed to press the right digit again. The data was collected in four different sessions on the same day for each of the tasks. In every session, users were provided with a new set of randomly generated PINs or passwords. A break of 10-minutes was given to participant between each session of 4-minutes length.

5 Data Preprocessing and Feature Extraction

The APIs provided by the Emotiv headset and the B-Alert headset were used to collect the raw ERPs during the experiment. We then used EEGLAB [19] to process the raw data collected from both of these headsets. Before processing the brain data, we first segregated the samples related to each digit from the raw data and created a new file for each one of them. For each keystroke, we considered 235 ms of brain data (30 samples of data) before the key stroke and 468 ms of brain data during the key press (60 samples of data). The reason behind using 235ms before keypress is to include the ERPs generated when user thinks of the digit before pressing it.

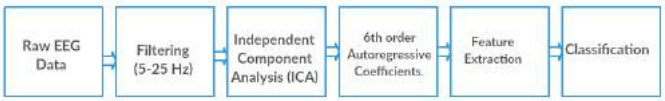

Fig. 2: Data Processing Flow Chart

We processed the raw data using band pass filter in EEGLAB to keep the signals with frequency above 5Hz and below 25Hz. The EEG signals measured by the electrodes from the scalp do not represent the electrical potential generated in the sources inside the brain [27]. Rather, they are the aggregation of several neurons' electrical activity in brain. So the filtered data was then processed using independent component analysis (ICA), a technique to separate subcomponents of a linearly mixed signal [24], to segregate the ERPs generated by statistically independent sources inside the brain.

A sample of recorded EEG data can be represented as $x(t) = (x_1(t), x_2(t), ..x_m(t))$, where m is the number of electrodes in the headset, and t is the time at which the neuron potential is measured. The ERPs recorded by each electrode at a time is the sum of the ERPs generated from n independent sources inside brain and can be represented as $(x_j(t) = a_{j1}s_1 + a_{j2}s_2 + ... + a_{jn}s_n)$, where n is the number of source components and a is the weight (like distance from the source) applied to the signal from a source. So we used ICA for identifying and localizing the

statistically independent electrical sources s from potential patterns recorded on the scalp surface by electrodes in the headset [27]. This process was repeated for the data collected for each of the digits for each session.

The data acquired after ICA was then processed using Autoregressive (AR) model for feature extraction. AR is commonly used algorithm for feature extraction from EEG data (e.g., [23]). An EEG signal recorded at any discrete time point t is given by $s(t) = \sum_{k=1}^{p} a_k x(t-k) + e(t)$, where p is the model order, $s(t)$ is the signal at the timestamps t, a_k are the real-valued AR coefficients and e(t) represents the error term independent of past samples [23]. We computed features from all 14-electrodes using sixth order Auto Regressive (AR) coefficients. Therefore, we had 6 coefficients for each channel giving a total of 84 features (6 * 14 channels) for each data segment for a digit. The feature extraction process was repeated for the brain data collected across different sessions for each of the digit (0-9).

Next, we used these features to build four classification models for predicting keystrokes based on the neural data. Two of the classification models were built using simple Instance Based Learning (IB1) [8] and KStar [18] algorithms. The other two were built using majority voting of two algorithms, *first*, IB1 with Naive Bayes (NB) [26] algorithm, and *second*, KStar and NB algorithm. We then used 10-fold cross validation for estimation and validation of these classification models on three different sets of data labeled with 10 different classes (0-9 digits).

First, we used instances for each digit from single session for each individual (called as *Individual Model – Single Session*). *Second*, we vertically merged instances of each digit from all sessions of an individual (called as *Individual Model – Merging Sessions*). *Third*, we vertically merged features for each digit from all users for each session (called as *Global Model*). Global Model is a stronger model compared to individual model, where the attacker will train the classification model on the features extracted from her own neural and keystrokes data and use it to infer victims' keystrokes. Even though the brain signals are assumed to be unique among users, we presumed, there might be similarities in ERPs when numbers/alphabets are observed.

We report the average true positive rate (TPR) and the average false positive rate (FPR) for each digit. True positive rate is the ratio of total number of correctly identified instances to the total number of instances present in the classification model $TPR = TP/(TP + FN)$, where TP is True Positive and FN is false negative. False positive rate is the ratio of total number of negative instances incorrectly classified as positive to the total number of actual negative instances $FPR = FP/(FP + TN)$, where FP is false positive and TN is true negative. An ideal classification model has true positive rate of 100% and false positive rate of 0%.

6 Data Analysis and Results

In this section, we describe the results of the classification models built on the features extracted from the event-related potentials to infer the keystrokes.

6.1 Task 1: Virtual Keyboard PIN Entry (VKPE)

To recall, in this task, we had asked participants to enter thirty randomly generated 4-digit PIN in the virtual keyboard using mouse. Table 1(a) lists the results of different classification models on using datasets from individual sessions. We can observe that the best average true positive rate of predicting digits in this model is 43.4% (false positive 6.2%). Likewise, the best average true positive rate of predicting digits is 31.9% (false positive rate is 7.55) when data from all sessions are merged (see Table 1(b)). We can see that the results are relatively lower than the models trained on individual session because the amplitude of ERPs during the first session might have been different than the amplitudes towards the last session. Similarly, The results of global model are listed in Table 1(c). We can observe that the best average true positive rate of predicting digits is 31.3% (false positive rate is 7.6%). Since, in this model, the samples from all individuals are used, the overall prediction rate is lower than the previous models. The results from both models are significantly better than a random guessing classification model (10% for each digit) which verifies the feasibility of PEEP.

Table 1: VKPE Task: Average true positive rate and average false positive rate (a) Individual Model – Single Session (b) Individual Model – Merging Sessions (b) Global Model

Classifiers	Session 1		Session 2		Session 3		Session 4	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
IB1	41.1	6.5	39.9	6.6	38.9	6.7	42.2	6.4
KStar	42.4	6.4	40.1	6.6	38.9	6.7	42.8	9.7
IB1+NB	41.5	6.5	39.4	6.6	38.6	6.8	42.1	6.4
KStar+NB	43.4	6.2	42.4	6.4	39.0	6.7	42.4	6.4

Classifier	All Sessions	
	TPR	FPR
IB1	30.1	7.7
KStar	31.7	7.6
IB1+NB	30.0	7.8
KStar+NB	31.9	7.5

Classifier	All Sessions	
	TPR	FPR
IB1	28.4	7.9
KStar	31.3	7.6
IB1+NB	28.4	7.9
KStar+NB	30.7	7.7

6.2 Task 2: Virtual ATM PIN Entry (VAPE)

The participants in this task were asked to enter thirty randomly generated 4-digit PIN in virtual keyboard similar to the ones employed in ATM touch screens. Table 2(a) and (b) have the results of the classification models for individual single session and merged sessions datasets respectively. We can observe that on average the digits can be best predicted at true positive rate of 47.5% (false positive 5.8%) for single session and 32.6% true positive rate (false positive 7.5%) for merged session. Table 2 (c) shows the results for these classification models for grouped data and we can notice that on average the digits can be best predicted at 39.1% true positive rate (false positive rate is 6.7%). The results depict that these models are better than the random guessing model (10%) in predicting the keys entered by users.

In this task, we see that the overall true positive rate of the digit prediction is higher than the true positive rate in VKPE task (see Section 6.1. The virtual keyboard in VKPE task had many keys compared to the virtual keyboard in VAPE task. The higher number of keys might have caused higher distraction in processing of digits, reducing the strength of features representing the keys, resulting in lower prediction rate.

Table 2: VAPE Task: Average true positive rate and average false positive rate (a) Individual Model – Single Session (b) Individual Model – Merging Sessions (b) Global Model

Classifiers	Session 1		Session 2		Session 3		Session 4	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
IB1	47.0	5.9	47.0	5.9	47.5	5.8	44.5	6.1
KStar	42.5	6.4	40.0	6.6	42.5	6.4	39	6.7
IB1+NB	43.5	6.3	41.5	6.5	44.0	6.2	40.5	6.6
KStar+NB	39.5	6.7	42.5	6.4	39.5	6.7	43	6.3

Classifier	All Sessions	
	TPR	FPR
IB1	31.1	7.6
KStar	31.6	7.6
IB1+NB	31.1	7.6
KStar+NB	32.6	7.5

Classifier	All Sessions	
	TPR	FPR
IB1	39.1	6.7
KStar	39.3	6.7
IB1+NB	39.0	6.8
KStar+NB	37.3	6.9

6.3 Task 3: Physical Numeric Keypad PIN Entry (PNKPE)

In this task, the participants had to enter thirty randomly generated 4-digit PIN using physical numeric keyboard. The movement of the fingers measured using smart watch worn on victims' hand while typing password has been previously used to reveal victims' PIN [39]. Researchers have also translated thoughts about moving fingers into action in prosthetic hands [35]. So we assumed that, there might be unique neural signatures of typing the numbers you are thinking about, which might be used to predict the victims' PIN numbers. Table 3(a) and (b) displays the results on individual model – single session, and individual model – merged session respectively. We can observe that on average the digits can be best predicted in individual model – single session at 46.5% true positive rate (false positive rate is 6.0%), and at 28.4% true positive rate (false positive rate is 7.9%) in individual model – merged session datasets. Similarly, table 3 (c) reports that the digits can be best predicted at 33.6% true positive rate (false positive rate is 6.7%) in global model. All these models again have performance better than a random model (10%).

We observe that the results of PNKPE task are lower than the results of VAPE task (see Section 6.2). In VAPE task the keys flashing while typing the numbers, which might have triggered neural signals resulting better features in building classification model. However, from the results of this task, we find that the finger movement while typing a number leave a unique trace in the brain which can be used to infer the keystrokes.

Table 3: PNKPE Task: Average true positive rate and average false positive rate (a) Individual Model – Merging Sessions (b) Global Model

Classifiers	Session 1		Session 2		Session 3		Session 4	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
IB1	46.0	6.0	37.5	6.9	45.0	6.1	36.5	7.0
KStar	40.5	6.6	31.5	7.6	45.0	6.1	38.5	6.8
IB1+NB	46.0	6.0	3	6.9	44.5	6.2	37.0	7.0
KStar+NB	39.0	6.7	34.0	7.3	46.5	5.9	39.0	6.8

Classifier	All Sessions	
	TPR	FPR
IB1	28.4	7.9
KStar	27.5	8.0
IB1+NB	28.4	7.9
KStar+NB	27.6	8.0

Classifier	All Sessions	
	TPR	FPR
IB1	33.1	7.4
KStar	34.0	7.3
IB1+NB	32.7	7.4
KStar+NB	33.6	7.4

6.4 Task 4: Physical Keyboard Password Entry (PKPE)

To recall, in this task we had asked users to enter thirty-six randomly generated uppercase 6-character password in laptop keyboard. Using the brain and keystrokes data

recorded during the task, we built classification models to predict the users' keystrokes. Table 4(a) shows the results for the individual model - single session data. We can see that on average the digits can be best predicted at 34.7% true positive rate (false positive rate is 4.7%). Similarly, in this task , the classification models on merged sessions data can best predict the digits at 23.7% true positive rate (false positive rate is 5.4%) (Table 4 (b)). Table 4(c) reports that on average the digits can be best predicted at 30.1% true positive rate (false positive rate is 4.8%) in the group model. Like the previous tasks, we observe that the results are better than random model for keystroke detection (random prediction rate of a character is 3.8%). In this task, we see that the overall results for this classification model is lower than the results in previous tasks (see Sections 6.1 6.2 6.3). This task involved a physical keyboard with many keys on the keyboard. The numbers were not flashed on entering them and multiple fingers were used while typing passwords. Because of all these things, the features representing the digits might not have been strong enough for better detection of the keystrokes.

Table 4: PKPE Task: Average true positive rate and average false positive rate (a) Individual Model – Single Session (b) Individual Model – Merging Sessions (b) Global Model

Classifiers	Session 1		Session 2		Session 3		Session 4	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
IB1	27.1	5.2	28.7	5.1	34.7	4.7	37.3	4.5
KStar	30.7	5.0	31.3	4.9	28.7	5.1	37.3	4.5
NB+IBk	17.3	5.9	10.7	6.4	23.3	5.5	28.7	5.1
NB+kStar	28.7	5.1	28.9	5.1	34.7	4.7	36.7	4.5

Classifier	All Sessions	
	TPR	FPR
IB1	21.15	5.6
KStar	23.7	5.4
IB1+NB	17.75	5.7
KStar+NB	23.5	5.3

Classifier	All Sessions	
	TPR	FPR
IB1	27.8	5.1
KStar	30.1	4.8
IB1+NB	19.8	5.7
KStar+NB	29.0	5.1

6.5 High-End B-Alert Headset - VKPE Task

We used high-end B-Alert headset to collect data in VKPE task for one participant, to test the feasibility of our attacks on different categories of headsets used for recording the neural signals.

Table 5: B-Alert Headset VKPE Task: Average true positive rate and average false positive rate (a) Individual Model –Single Session (b) Individual Model – Merging Sessions

Classifiers	Session 1		Session 2		Session 3		Session 4	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
IB1	39.0	6.8	31.0	7.7	31.0	7.7	37.3	4.5
KStar	34.0	7.3	23.0	8.6	36.0	7.1	37.3	4.5
IB1 + NB	37.0	7.0	31.0	7.7	24.0	8.4	28.7	5.1
KStar +NB	25.0	8.3	25.0	8.3	38.0	6.9	36.7	4.5

Classifier	All Sessions	
	TPR	FPR
IB1	20.5	8.8
KStar	19.8	8.9
IB1+NB	17.5	9.2
KStar+NB	19.5	8.9

Table 5(a) shows the results of these classification models on single session data. We can see that on average the digits can be predicted at a true positive rate of 39.0% (false positive 6.8). The performance of the classification models on merged sessions data are presented in Table 5 (b). We can see that on average the digits can be best predicted at

20.5% true positive rate (false positive is 8.8%). These results are significantly better than a random guessing classification model (10% for each digit) which shows the feasibility of side-channel attacks using BCI devices.

7 Discussion and Future Work

In this section, we summarize and further discuss the main findings from our study. We also outline the strengths and limitations of our study.

7.1 Vulnerability of the Brainwave Signals

In this study, we focused on studying the vulnerability of BCI devices towards revealing the private information to malicious attackers. We designed PEEP to study the feasibility of brainwave side-channel attacks using such devices. PEEP stealthily monitors and records event-related potentials (ERPs) measured by BCI devices when users are typing their PINs or passwords on to physical or virtual keyboards. PEEP can then analyze the ERPs for extracting features representing each of the digit or character. These features are then used to build a training model which is later used to predict the keystrokes made by the users. We experimentally verified the feasibility of PEEP for both individual and global training models.

Closely related to our study is the work done by Martinovic et al. [29]. They also studied the feasibility of side-channel attack with brain-computer interfaces. They showed the images of banks, ATMs, digits, months, etc., to participants to elucidate their private information related to banks, ATMs, PINs, and month of birth. They used the amplitude of P300 ERP, which appears in neuronal electrical activity for known artifacts, to infer such details. The participants in their study were asked to memorize 4-digit PINs and were shown the images of randomly permuted numbers between 0 and 9, one by one. Each number was shown 16 times, and the experiment lasted around 90 seconds. They were able to correctly predict the first digit of the PIN at 20% accuracy. In contrast, PEEP, on average, was able to predict digits at the true positive rate of 46.5% (FPR 6.0%) for PIN entered in the VAPE task (this is the task closely related to PIN study of Martinovic et. al). Also, their attack set-up is intrusive and can be easily detectable as the users may notice the abnormality in the app when it shows the images of banks or ATMs related to the user. In comparison, PEEP is highly surreptitious as it only requires passive monitoring of brain signals as users type their PINs and passwords in regular use of computing devices, not fraudulent strategies that may trigger suspicion and be detected by the user. By the passive nature of our attack, it can be used to learn private input from any (secondary) computing device, not necessarily the (primary) one to which the BCI device is connected like in [29].

7.2 Password Entropy

PEEP reduces the entropy of the PIN or textual passwords, making it easier to launch dictionary or brute force attacks. In our study, we assumed the passwords and PINs to be random. We used 0-9 digits to create 4-digit PIN and A-Z characters to create six

character-based passwords. If brute-force attack is launched, it will take 10^4 guesses to correctly identify the PIN and 26^6 guesses to correctly identify the password. The success of randomly guessing a digit of the PIN is 100/10 (10%) and the success of randomly guessing a character is 100/26 (3.84%). PEEP increases this accuracy of correctly identifying the digits of PIN to 47.5% and passwords to 34.7%. In case of non-random passwords, PEEP can be used in conjunction with dictionary-based password attacks, and further reduce the number of guesses in the brute-force attacks.

7.3 Possible Defensive Mechanisms

One of the possible strategies to mitigate the threat invoked by PEEP is to automatically insert noise in the neural signals when the user starts typing passwords or PINs (or other sensitive input). However, this might affect other benign applications dependent on brain signals during that time frame. Currently, the third-party developers are offered unfettered access to the neural signals captured by such devices. This access can be managed by operating systems to stop apps other than intended apps to listen on to brain signals while entering the private information in desktops or mobiles. The more sophisticated attacks are imminent with the technological advancements in these BCI devices. So it is important to study probable mitigations of such attacks in the future, especially given their potential hideous and powerful nature.

7.4 Study Strengths and Limitations

We believe that our study has several strengths. The study used randomly generated passwords which users knew at the time of the experiment. Despite the lack of pre-familiarity with the passwords/PINs, we were still able to predict them with true positive rate significantly better than random guessing. In real life, the password might remain in the memory for longer time, and the users might only be using certain fixed digits or characters in their PINs or passwords, which might provide better feature space and better prediction true positive rate. Further, we launched our side channel attacks using different categories of headsets (both consumer and clinical EEG headsets) and verified the feasibility of our attacks in a variety of contexts. Similar to any study involving human subjects, our study also had certain limitations. Our study was conducted in a lab environment. Although we tried to simulate the real-world scenarios of entering PINs or passwords, the layouts of the experimental tasks were simplistic. Also, the performance of the users might have been affected by the fact that their brain signal was recorded during the task. The EEG headsets we used in our experiment were quite light-weight, and the duration of the experiment was short (maximum four minutes for each task), however, the participants might have felt some discomfort that may have impacted their brain responses. Future work may be needed to assess the feasibility of our attacks in real-world or field settings. We believe that our work lays the necessary foundation that serves to highlight the vulnerability.

8 Concluding Remarks

The popularity of BCI devices is ever increasing. In not so distant future, these devices are going to be less costly and more sophisticated and will be integrated into many spheres of daily lives of users. In this light, it is important to study the possible security vulnerabilities of such devices and make people aware of such vulnerabilities. In this paper, we examined the possibility of one such side-channel attack for the purpose of inferring users' private information, in particular, their sensitive keystrokes in the form of PINs and passwords. We designed and developed PEEP, which successfully predicts the sensitive keystrokes made by the users just from the event-related potentials passively recorded during those keystrokes. PEEP predicts numbers entered in 4-digit PINs in virtual keyboard with an average TPR of 43.4%, virtual ATM keyboard with an average TPR of 47.5%, physical numeric keyboard with an average TPR of 46.5% and alphabets entered in 6-character passwords with an average TPR of 37.3%, demonstrating the feasibility of such attacks.

References

1. B-Alert X-10 Set-Up Manual. http://www.biopac.com/Manuals/b-alert%20x10%20setup.pdf
2. Emotiv app store. https://www.emotiv.com/store/app.php, accessed: 7-28-2016
3. Emotiv eeg headset. https://www.emotiv.com, accessed: 7-28-2016
4. Emotiv web apis. https://cpanel.emotivinsight.com/BTLE/document.htm#_Toc396152456, accessed: 7-28-2016
5. Neurofocus. http://www.nielsen.com/us/en/solutions/capabilities/consumer-neuroscience.html, accessed: 8-14-2016
6. Neurosky app store. https://store.neurosky.com/, accessed: 7-28-2016
7. Neurosky eeg headset. https://www.neurosky.com, accessed: 7-28-2016
8. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. Machine learning 6(1), 37–66 (1991)
9. Ajaya Neupane, Md Lutfor Rahman, Nitesh Saxena, Leanne Hirshfield: A Multimodal Neuro-Physiological Study of Phishing and Malware Warnings. In: ACM Conference on Computer and Communications Security (CCS), Denver, CO. ACM (2015)
10. Ashby, C., Bhatia, A., Tenore, F., Vogelstein, J.: Low-cost electroencephalogram (eeg) based authentication. In: Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on. pp. 442–445. IEEE (2011)
11. Asonov, D., Agrawal, R.: Keyboard acoustic emanations. In: IEEE Symposium on Security and Privacy. vol. 2004, pp. 3–11 (2004)
12. Aviv, A.J., Sapp, B., Blaze, M., Smith, J.M.: Practicality of accelerometer side channels on smartphones. In: Proceedings of the 28th Annual Computer Security Applications Conference. pp. 41–50. ACM (2012)
13. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature 398(6725), 297–298 (1999)
14. Bojinov, H., Sanchez, D., Reber, P., Boneh, D., Lincoln, P.: Neuroscience meets cryptography: designing crypto primitives secure against rubber hose attacks. In: Presented as part of the 21st USENIX Security Symposium (USENIX Security 12). pp. 129–141 (2012)

15. Cai, L., Chen, H.: Touchlogger: Inferring keystrokes on touch screen from smartphone motion. HotSec 11, 9–9 (2011)
16. Campbell, A., Choudhury, T., Hu, S., Lu, H., Mukerjee, M.K., Rabbi, M., Raizada, R.D.: Neurophone: brain-mobile phone interface using a wireless eeg headset. In: Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds. pp. 3–8. ACM (2010)
17. Chuang, J., Nguyen, H., Wang, C., Johnson, B.: I think, therefore i am: Usability and security of authentication using brainwaves. In: International Conference on Financial Cryptography and Data Security. pp. 1–16. Springer (2013)
18. Cleary, J.G., et al.: K*: An instance-based learner using an entropic distance measure
19. Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. Journal of neuroscience methods 134(1), 9–21 (2004)
20. Donchin, E.: Event-related brain potentials: A tool in the study of human information processing. In: Evoked brain potentials and behavior, pp. 13–88. Springer (1979)
21. Esfahani, E.T., Sundararajan, V.: Classification of primitive shapes using brain–computer interfaces. Computer-Aided Design 44(10), 1011–1019 (2012)
22. Halevi, T., Saxena, N.: A closer look at keyboard acoustic emanations: random passwords, typing styles and decoding techniques. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. pp. 89–90. ACM (2012)
23. Huan, N.J., Palaniappan, R.: Neural network classification of autoregressive features from electroencephalogram signals for brain? computer interface design. Journal of neural engineering 1(3), 142 (2004)
24. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural networks 13(4), 411–430 (2000)
25. Johnson, B., Maillart, T., Chuang, J.: My thoughts are not your thoughts. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. pp. 1329–1338. ACM (2014)
26. Jordan, A.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes (2002)
27. Makeig, S., et al.: Independent component analysis of electroencephalographic data. Advances in neural information processing systems (145), 6 (1996)
28. Marquardt, P., Verma, A., Carter, H., Traynor, P.: (sp) iphone: decoding vibrations from nearby keyboards using mobile phone accelerometers. In: Proceedings of the 18th ACM conference on Computer and communications security. pp. 551–562. ACM (2011)
29. Martinovic, I., Davies, D., Frank, M., Perito, D., Ros, T., Song, D.: On the feasibility of side-channel attacks with brain-computer interfaces. In: Presented as part of the 21st USENIX Security Symposium (USENIX Security 12). pp. 143–158 (2012)
30. Monrose, F., Rubin, A.: Authentication via keystroke dynamics. In: Proceedings of the 4th ACM conference on Computer and communications security. pp. 48–56. ACM (1997)
31. Neupane, A., Saxena, N., Kuruvilla, K., Georgescu, M., Kana, R.: Neural signatures of user-centered security: An fMRI study of phishing, and malware warnings. In: Proceedings of the Network and Distributed System Security Symposium (NDSS). pp. 1–16 (2014)
32. Owusu, E., Han, J., Das, S., Perrig, A., Zhang, J.: Accessory: password inference using accelerometers on smartphones. In: Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications. p. 9. ACM (2012)
33. del R Millan, J., Mouriño, J., Franzé, M., Cincotti, F., Varsta, M., Heikkonen, J., Babiloni, F.: A local neural classifier for the recognition of eeg patterns associated to mental tasks. IEEE transactions on neural networks 13(3), 678–686 (2002)

34. Song, D.X., Wagner, D., Tian, X.: Timing analysis of keystrokes and timing attacks on ssh. In: Proceedings of the 10th Conference on USENIX Security Symposium - Volume 10. SSYM'01, USENIX Association, Berkeley, CA, USA (2001), http://dl.acm.org/citation.cfm?id=1251327.1251352
35. Sumon, M.S.P.: First man with two mind-controlled prosthetic limbs. Bangladesh Medical Journal 44(1), 59–60 (2016)
36. Tan, D., Nijholt, A.: Brain-computer interfaces and human-computer interaction. In: Brain-Computer Interfaces, pp. 3–19. Springer (2010)
37. Thorpe, J., van Oorschot, P.C., Somayaji, A.: Pass-thoughts: authenticating with our minds. In: Proceedings of the 2005 workshop on New security paradigms. pp. 45–56. ACM (2005)
38. Vuagnoux, M., Pasini, S.: Compromising electromagnetic emanations of wired and wireless keyboards. In: Proceedings of the 18th USENIX Security Symposium. pp. 1–16. No. LASEC-CONF-2009-007, USENIX Association (2009)
39. Wang, H., Lai, T.T.T., Roy Choudhury, R.: Mole: Motion leaks through smartwatch sensors. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. pp. 155–166. ACM (2015)
40. Xu, Z., Bai, K., Zhu, S.: Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors. In: Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks. pp. 113–124. ACM (2012)
41. Zhuang, L., Zhou, F., Tygar, J.D.: Keyboard acoustic emanations revisited. ACM Transactions on Information and System Security (TISSEC) 13(1), 3 (2009)

A Design of Experiments

Fig. 1: (a)VKPE Task: Virtual Keyboard (b) VAPE Task: Virtual ATM Keyboard

Fig. 2: PNKPE Task: (a) Layout to enter the PIN (b) Physical numeric keyboard used

Fig. 3: PKPE Task: (a) Layout to enter 6-digit character based password (b) Physical keyboard used