ISSN 2347 - 3983

Volume 8. No. 10, October 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter538102020.pdf

https://doi.org/10.30534/ijeter/2020/538102020

The Effect of Bad Password Habits on Personal Data Breach

Praveen Raj Santhira Rajah[1], Omkar Dastane[2], Kinn Abass Bakon[3], Zainudin Johari

[1]Anglia Ruskin University, United Kingdom, praveen-raj-al.santhira-rajah@student.anglia.ac.uk

[2]Curtin University, Malaysia, omkar.dastane@postgrad.curtin.edu.my

[3] The National University of Malaysia (UKM), Malaysia, P108383@siswa.ukm.edu.my

[4]FTMS College, Malaysia, zainudin@ftms.edu.my

ABSTRACT

Users tend to utilize bad or weak passwords with memorable characteristics such as simple words from the dictionary and easy to remember sequence of numbers from birthdays. Poor or bad password habits lead to compromise of personal data privacy and allow hackers to gain unauthorised access to these passwords and use them for criminal and fraudulent cyber activities. The purpose of this research is to examine the impact of password habits among Malaysians on their personal data breaches. This study provides insights into the behaviour of users concerning their passwords use. This study used a positivism paradigm and applied a quantitative approach and used a convenience sampling technique to collect data from 297 respondents from Malaysian nationals. IBM SPSS AMOS 24 is used to conduct the analysis. The result from the study shows that “lacking the use of second-factor authentication’ have a significant and a positive impact on the personal data breaches. Based on this finding, it can be concluded that the lack of second-factor authentication is an essential factor that significantly impacts personal data breach. This finding provides a different perspective from the usual connection of bad password habits of weak password length and combination, easy to guess the password and common password reuse to be the main contributing factor of the personal data breaches by the previous literature. The contribution of this research is the provision of empirical evidence that emphasise the need to continually beef up own security by correctly using second-factor authentication across individual accounts. Doing so is crucial to curb personal data breaches.

Key words: Password habits, Bad passwords, Data breach.

1. INTRODUCTION

The European Data Privacy Services (EDPS) [1] defined personal data breach as a violation or breach of security that lead towards accidental or lawful damage, modification, destruction, loss or unauthorized disclosure or access to personal or individual data transmitted, processed or stored under the responsibility of the controller or owner of the data. The term data breach existed even before companies or associations begin storing data which are protected and confidential digitally. Data breaches have existed as long as individuals or companies store or maintain records of private information in any form, including paper [2].

A group of characters, symbols and numbers used for authentication, to gain access to a source or prove the identity of oneself is known as a password [3]. Poor or bad password habits are prone to lead to compromise of personal data privacy, criminal and fraudulent activities over online cyberspace [4]. It is typically common for users to follow bad or weak password security practices; this may result in their accounts being vulnerable or exposed to attack [5].

Users tend to utilize bad or weak passwords with essential characteristics for example using simple words from the dictionary, easy to remember number sequence, i.e. birthday dates or month and year [6]. It is common for a user to pick easy to remember password and yet meeting minimum password complexity as required by websites using a combination of name, date of birth or simple dictionary. This example illustrates how easily these passwords can be targeted for compromise.

Users apply this bad password habit in their multiple accounts [5]. Password Fatigue is another challenge that tends to affect many users due to the hardship in having to facilitate and remember multiple and numerous passwords [8]. It is a common and well known practise by attackers on the Internet who attempt to access other user accounts to commonly guess passwords or by retrieving data from a particular user, for example, his favourite soccer team, usually this data or information is easy to obtain from the user’s social networking sites like Facebook, Instagram or LinkedIn [9]. A research performed on American consumers showed that 61% respondents are likely to have a tendency to reuse similar passwords across multiple websites and 54 % were found to have less than five passwords [10].

Attackers and intruders are most often seen to exploit vulnerabilities to compromise a system account or access [11]. Vulnerability or weakness like weak passwords can be associated with either an internal or external security threat [12]. The multiple forms of password attacks are brute force attacks, e.g. looking or searching for poor hashes to crack weak passwords, dictionary attack and the case of rainbow tables attack to generate information data upfront to enable looking up for hashes [13]. Verizon in 2017 Data Breach Investigations Report (DBIR) highlighted that over 81% of current data breaches are attributed to hacked, stolen or weak passwords [14]; [15]. Why this problem or issue is relevant now is because data breaches are becoming frequent. We have heard of some of the most significant data breaches involving Yahoo with 3 billion records in 2013, Facebook in 2019 involving 540 million users and FriendFinder networks in 2016 with 412 million accounts [16]. This has become an important research topic and password habits in particular require attention of researchers.

The purpose of this research is to study and examine the impact of password habits among Malaysians on the personal data breach. This research will provide insights on the behaviour of users handling his or her online password security and primarily will be focusing on the context of Malaysian users. The objectives of this research are to examine; (1) the impact of the weak password length and combination on personal data breach; (2) the impact of easy to guess password dictionary on personal data breach; (3) the impact of everyday password use on personal data breach; (4) the impact of lacking use of second-factor authentication on personal data breach. Due to the research objectives above, the research questions for this study are as follows; (1) What is the impact of weak password length on personal data breach?(2). What is the impact of easy to guess password dictionary on personal data breach? (3) What is the impact of password use on personal data breach? (4) What is the impact of lacking use of second-factor authentication on personal data breach?

2. REVIEW OF KEY THEORIES AND CONCEPTS

2.1. Definition of Key Concepts and Terms

2.1.1. Password Habits

Password habits can be categorized and classified into four key concepts which are weak password length, easy to guess password dictionary, common password reuse and lacking use of second-factor authentication.

A. *Weak Password Length and Combination*

Password length is the right size of character sets that are measured or calculated to the proportional length of characters [5]. SANS Institute [17] being a renowned security research institute had described that a weak password is termed as having characteristics of containing less than fifteen characters. Most global organizations or standard practices by entities are inclined to setting passwords to having at least eight characters. Recent research has shown that focus on increasing password length is a more promising alternative than password complexity with minimal password length [18].

B. *Easy to Guess Password Dictionary*

SANS Institute, [17] has defined that weak password as also having a word or more that can be obtained in a dictionary either in English or in other foreign languages. Narayanan & Shmatikov [19] explained that the distribution of letters in easily formed guessable passwords are likely to be similar to the distribution of letters in a user's language is native to that person. Thus, this becomes relatively simple when trying to exploit easy to guess passwords in a dictionary attack. Passwords with easily guessable diction are known to contain major weakness or vulnerability due to the possibility of automating scanning software that can be programmed to run ordered and systematic dictionary attacks [20].

C. *Common Password Use*

Common password use is the scenario when users maintain the same passwords between multiple websites or account logins to cope with difficulties or problem to remember too many passwords [5]. In current society, the required number of passwords or protected accounts with a password is growing increasingly. This cause's limitation to human memory capacity and capability to remember a multitude of passwords; thus, this behaviour tends to lead to the symptom of password reuse [21]. Password reuse brings on a form of security weakness and vulnerability as it enables intruders or attackers who have successfully compromised or exploited one of the victim's passwords. This gives the attacker an advantage to use the same compromised password in other protected accounts or website login, which are now an easy target for personal breach [22].

D. *Lack of use of second-factor authentication*

Second-factor authentication is a technology for consumers which have been around for a long time to improve digital security either optional or mandatory depending on the environment it is used for [23]. Two-factor authentication is defined as a means to authenticate users using two separate sets of information or way of identification, the first factor is normally your standard account password and the second factor being a one-time password generated from a third-party authenticator mechanism either soft or hard tokens.

2.1.2. Personal Data Breach

The European Data Privacy Services (EDPS) [1] has defined personal data breach as a violation or breach of security that lead towards accidental or lawful damage, modification, destruction, loss or unauthorized disclosure or access to personal or individual data transmitted, processed or stored under the responsibility of the controller or owner of the data. It can also be related to the personal breach of the three known security principles which are a breach of 'Confidentiality', 'Availability' and/or 'Integrity'. The breach may occur due to several possible reasons either through negligence, as a result of an accident or due to intentional wrongdoing or act by a person or threat actor [1]. There are laws enacted around the globe to protect personal or private data. In Malaysia, the Personal Data Protection Act 2010 came into full effect on Nov 15, 2013, which is intended to prevent misuse of individual's personal data for wrongful intention or commercial purposes [24]. The first-ever known data breach was known publicly in 2004, involving an AOL worker being arrested in stealing his company's subscriber list for selling it for personal financial gain [25]. In 2005, the first data breach to have compromised more than 1 million records of credit card numbers about DSW Shoe retail warehouse [26]. Statista [27] released statistics that the increase of personal data breaches over the years since 2005 have been significantly on the rise with 157 million records reported in 2005 and highest in 2017 with 1.6 billion records of personal data breaches.

2.2. Review of Key Theories

2.2.1 Markov Model

Markov model is used in studying the probability of stochasticity, which is useful in modelling random change of systems or behaviour of the subject [28]. Markov model is applied to learn the probabilistic occurrence of the state of future by recognizing patterns of the present status of with sequential data statistical analysis [29]. Through the years, the Markov model is a proven model that is demonstrated in password security. Narayanan & Shmatikov [19] has displayed the ability of password habits have led to cracking and disassembling of the passwords using the Markov model. Utilizing the construction using Markov model, password strengths can be tested by understanding the insecurity and potentiality of leakage of the password itself by adding on a finite amount of noise parameter for running a test [30]. The construction of the Markov model for checking password provides high accuracy with fine measurement of its password strength [31]. Tansey [32] explained a layered approach while using the Markov model.

2.2.2. Construal Level Theory

Construal Level Theory, or widely known as CLT, was developed by Trope & Liberman [33]. It is a theory used commonly in social psychology to describe the relationship between psychological distance and the extent of people's thinking. Psychological distance in CLT defines the spatial, temporal, hypothetical & social distance as the most common and understood dimensions [34]. Conceptual differences are referring to the data that is perceived to the mind, and perceptual differences otherwise are how the data or information is processed, this relationship may reflect the high or low abstract levels of construing the objects or events [35]. It is, therefore, a trade-off between feasibility and desirability, whereby strong emphasis is given on desirability while considering events in the distant future as compared to a stronger emphasis on feasibility while considering in the near future [36]. Thus, it is the context of password management or habits. Users are bound to place a stronger and higher emphasis on security that is desirability or the consequences of possible future events like a data breach or leakage as compared to an expectation of users applying weaker or bad passwords with the emphasis given on feasibility. This is considering the events of the near future, as no immediate or near-future threat of such consequences [37].

2.2.3. Agent-Based Modelling

Agent-based modelling in the form of multi-agent simulation. It tries to locate explanation and insights into the collective behaviour of agents, thus conforming to the rules in its natural system [38]. Agent-based modelling specific to password habits highlights the radical difference while reviewing and validating security in practice as compared to security in the abstract. Kothari et al. [39] highlight users have tendencies to circumvent password policy security by which application of the agent-based modelling. It provides a clearer and better understanding of the aggregated security that is measured to include circumventions, risks and costs for better judgement and decisions.

2.2.4. Critical Review of Theories

Markov model uses the stochasticity approach, which can be exemplified in password modelling. By understanding the user behavioural approach, this amplifies a more significant result when password cracking. Though it provides high accuracy in the n-gram model, it tends to affect usability when wrongly estimated [31]. Construal Level Theory, on the other hand, is used to determine the desirability of an event in the future state similarly in the event of password breach due to malformed practices within password habits. This theory is dependent on time frame effect but has weakness like not focus on the quality of the attribute, i.e. password quality [37]. The Agent-based modelling revolves around an agent's belief or cognitive burden [39], which is quite relevant to users' circumvention around password security due to limitation of cognitive capacity to perform stringent and complex passwords.

2.3. Research Gap

From previous research and empirical studies carried out, most studies are carried around password habits with variable factors, i.e. weak password length or combination, passwords re-use across multiple sites or easy password dictionary. They're also international and local scale studies in Malaysia concerning data breaches. The first gap identified on the international level of studies, no studies are linking the password habits and its variable factors towards personal data breach. Another research gap in the local Malaysian context, is no clear and distinct studies have been made significantly on password habits and as well as its relationship towards personal data breach. As identified in the above gaps, this provides an opportunity to conduct research linking password habits towards personal data breach among Malaysians that could prove beneficial in understanding these attributes and its construct in this research.

2.4. Conceptual Framework

The Conceptual framework for this research displayed in Figure 1. Four critical components of password habits comprising of weak password length, easy to guess password dictionary, password reuse and lacking second-factor authentication are indicated as independent variables. A personal data breach is a dependent variable in this case.

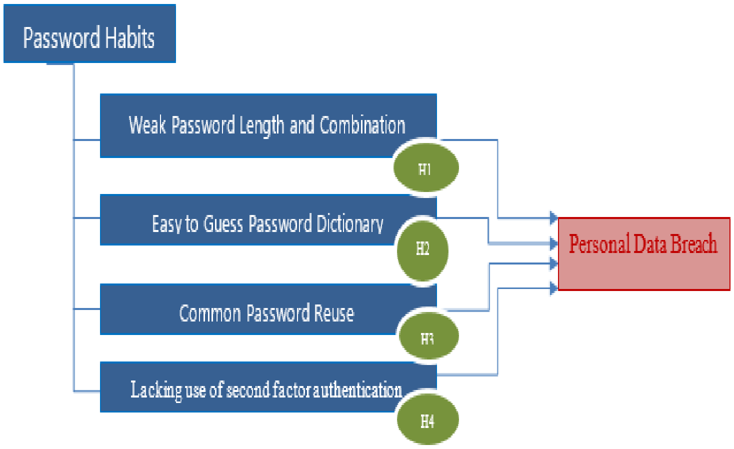
Password Habits
Weak Password Length and Combination
H1
Easy to Guess Password Dictionary
H2
Common Password Reuse
H3
Lacking use of second factor authentication
H4
Personal Data Breach

Figure 1. Password Habits Effect on Personal Data Breach Conceptual Framework

2.5. Hypothesis Development

A past study has shown that 81% of data breaches that occur are due to weak passwords; this is normally associated with user negligence [40]. A previous user study conducted based on 294 participants, 30% are reported to have at least one of their accounts breached or compromised due to poor or weak password length [18]. Ablon et al. [41]in a survey conducted across 2618 adults in the United States has stated that 26% of the participants or respondents received a notice on data breach infliction from prominent and popular online services that have implicated in a compromise like LinkedIn, MySpace, Adobe, Dropbox and Yahoo. Researchers conducting a study at Preempt Firewall Company had looked at the LinkedIn personal data breach and have discovered that 65% of the passwords leaked or compromised are attributed to weak and poor password combination and password lengths. They were discovered to be cracked easily using brute force technique from off-the-shelf cracking tools like Jack-the-Ripper and Hashcat [42]. InfoSecurity Group, [42] conducted a study and it revealed the top 10 list of bad or weak password lengths most common discovered with a number of credentials found in each ranking from the previous data breach of personal accounts of users involving Myspace [14]. Therefore, this study proposes this hypothesis:

H1: Weak password length and combination has a significant positive impact on personal data breach

Easy dictionary words are observed to be easily cracked or compromised by tools with dictionary listing. It is usually advised when constructing passwords to obfuscate dictionary words with numerals or special characters such as 'Winter' to 'W1nt3r' [43]. Based on empirical research conducted, common password dictionary that is considered weak are examples like using nouns, birthday dates, family, pet names, or even anniversary date [44]. In the past, a comparative analysis was performed based on different sets of password attacks. One of the main breach issues is dictionary attack versus other sets like brute force, shoulder surfing, replay, keylogging and phishing attacks [13]). Thus, therefore this study proposes following hypothesis:

H2: Easy to guess password dictionary has a significant positive impact on personal data breach

Das et al. [22] studied the password strategies for end-users who appeared or surfaced in multiple related credential leaks and estimated that 43% of passwords were re-used [22]. The research on the impact of password reuse across multiple sites and its weak practices leading towards security or personal data breaches was done by comparing same password reuse across 21 top universities in the United States [21]. An empirical study done on larger domino effects of password reuse has bigger implications. For example, a case study of the stolen or leaked credential from a rival in South Korea used to make illegal trade had amounted to $22 million in loss [45]. [46] supports the notion of the influence and impact common password reuse has on an individual's vulnerability towards data breach. Therefore, this study proposes this hypothesis:

H3: Common Password use has a significant positive impact on personal data breach

There have been known breaches in the past involving not just users but also organizations that forced the decision to move to second-factor authentication. Duo Security [47] reported that organizations or service providers like Bitly, Twitter, Buffer, Hootsuite, Tumblr and many more had turned into providing second-factor authentication not just for their internal users but their consumers too. The difference in adoptions to newer technology compared between 2010 and 2017 among US internet users to deal with newer advances in personal breaches exist [47]. It only proves from above depiction that second-factor authentication is becoming improvingly important to help users mitigate weakness in single password authentication or shield potential breach if a single password is compromised. Second-factor authentication is elevating the security of an individual's account as well as intended to circumvent shortfalls when the only single-factor password is used. Therefore, this study proposes this hypothesis:

H4: Lacking use of second-factor authentication has a significant positive impact on personal data breach

3. RESEARCH METHODOLOGY

T his research adopted positivism research as it involves realism of participative context, which is arguably a world perception, and it is more fulfilling and certain [48]. A quantitative approach was applied in this research, and convenience sampling technique was used to collect data from 297 respondents from Malaysian nationals and between the age group of 18 to 80. The questionnaire were designed in English as well as Bahasa Melayu The first part of the questionnaire was designed to measure the demographic makeup of the respondents and the second part of the questionnaire is made up of 4 subsections which were designed to investigate the variable of (a) Weak Password Length and Combination, (b) Easy to guess password dictionary, (c) Common Password Reuse, (d) Lacking use of second-factor authentication and (e) Personal Data Breach. In total, the questions were 32. The respondents were asked to select one of these seven Likert scale ranges; "1- Strongly Disagree", "2-Disagree", "3-Somewhat Disagree", "4- Neither Disagree nor Agree" and 5- Somewhat Agree, 6 - Agree and 7 - Strongly Agree. The four conventional statistical approaches include descriptive statistics, which is a method of describing the variables and used to measure central tendencies and variability [49]. The demographic analysis is also part of this deliverable. Next, reliability and normality testing are conducted whereby reliability assessment will look at findings to ensure stability determination and normality assessment is to test the distribution of the data according to a normal distribution [50]. Confirmatory Factor Analysis (CFA) is part of and validity assessment to address the truthfulness of the data findings. Finally, Structural Equation Model or SEM, is used whereby it combines factor and path analysis, also known as co-variance modelling structure [51]; [52]; [53]. Two main statistical tools used for analysis are IBM SPSS 24 for performing descriptive analysis, testing or assessment of reliability and normality and IBM SPSS AMOS 24 for performing the validity testing of CFA and SEM. Hypothesis testing is then conducted.

4. ANALYSIS

4.1. Demographic Analysis

In this sample size of 297 respondents, various demographic related inquiry, i.e. gender, age, ethnicity, education qualification, web surfing frequency and content, were acquired in the survey. To begin with, this survey was intended to target respondents who are Malaysians as to study the pattern of password habits among this nationality of users. From a total of 297 respondents, the majority refers to Malay ethnicity constituting of 36%, followed by Chinese 33%, Indians 26% and 5% making up the remaining. From a gender perspective, 38% or 113 respondents comprised females and 58% or 173 respondents are males. The biggest age group responding to this survey comes from the age group of 35-44 years old with a composition of 35% of total respondents or 105 out of 297. The second-largest group of respondents are from the age group of 25-34 years old, with 92 respondents, followed by 45-54 years old with 55 respondents. From the perspective of education qualification, 47% or close to half of the respondents come with Tertiary education background. The second-largest group are 81 respondents with Master's Degree qualification. From here, we can understand the majority of the respondents are qualified with an educational background with a high literacy rate. Majority of the respondents at about 76% claim confidently to use the Internet frequently. This behaviour is essential to support the idea of the importance of password security involving online accounts that may lead to unwarranted incidents, i.e. personal data breach. 38% of respondents spend more than 8 hours, and 27% spend between 6 to 8 hours. This is relevant to show high usage of the Internet each day to support further in this research analysis. Online services content which has associated online login account, signifying the use of password required. 91% of respondents use Email as the most common online content, followed by 78% on social media and 72% online or mobile banking. This information is useful to show the distribution of online services which require password or log in, which helps in this research to understand further password habits of the respondents.

4.2. Normality Assessment

The normality test if involving small to medium size samples (i.e. n < 300) would typically include formal normality test using skewness and kurtosis [54]. Skewness measures the asymmetry of the variable distribution and whereas kurtosis looks at peakedness of the distribution [55]. As a rule of thumb, Gravetter & Wallnau [56] illustrated that an acceptable range of skewness and kurtosis values are between -2 and +2. From the results for all question items fall into the range of -2 to +2, therefore rendering the results to be in an acceptable range. This further proves the fair normal distribution of the data collected for this research.

4.3. Reliability Assessment

Reliability testing is conducted to measure internal consistency; in this case, measuring the reliability of the questionnaire instrument using the Likert scale (1 to 7). Cronbach [57] introduced Cronbach's Alpha test, which is mainly used for reliability testing involving the development of scales for measuring attitudes and any other development constructs [58]. Cronbach's Alpha coefficient range was developed to measure the coefficients consistency for the items from the questionnaire [59]. An alpha value above 0.9 is deemed as excellent, whereas below 0.5 is regarded as unacceptable. Based on results, the overall measurement of Cronbach's Alpha for this research questionnaire items is 0.755, which is acceptable. The highest is being 0.877 for the first variable, and the lowest is 0.580 for the third variable. The smallest is, however, not deemed as unacceptable for this research. The other remaining variables are above 0.7, which are acceptable based on coefficient range. As per the Dependent Variable, the Cronbach's Alpha is rated at 0.682, which is deemed still within the safe level in the coefficient range when measuring for internal consistency and reliability.

4.3.4 Confirmatory Factor Analysis

CFA is performed in two phases, namely initial run by loading all items of the questionnaire and final run after seeing the best fit of the questionnaire items and removing any redundancy or noise. According to Hair et al. [64]; [60]. [61]. [62]). [63]. acceptable fit value for goodness of fit (GFI), comparative fit index (CFI), incremental fit index (IFI) and Tucker-Lewis index (TLI) are well above 0.900. Finally, the root mean square of error approximation (RMSEA) analyzes discrepancy between hypothesized model, parameter estimates and the population co-variance with RMSEA ranging from 0 to 1 with 0.08 or less as an acceptable indication [65]. The criteria for selecting unfit questionnaire items and dropping those items based on manifest variables with loading factor or value less than 0.5 is dropped [66]. Another criterion is also to review the Modification Indices (MI). High value, MI > 15 has an indication that redundant items exist in the model; therefore, it needs to be removed [67]. Figure 2 below depicts the new CFA Path Diagram with the revised mapping of the relationship with each variable.

The default model value of X2/DF or CMIN/DF above is 2.710, which is deemed acceptable as it is below the accepted value of 3.00. The GFI, CFI, IFI and TLI values in above are respectively 0.882, 0.914, 0.914 and 0.894. Two of the values are above 0.900, therefore are CFI and IFI. Consequently, they are accepted fit. However, GFI and TLI are just on the borderline to the value 0.900. Finally, the RMSEA value for CFA final run is measured at 0.076 in the table and is deemed accepted as fit since it is lesser than 0.08. The summary of results from the final run, as depicted in However, GFI and TLI are as indicated below in the borderline very close to 0.900.

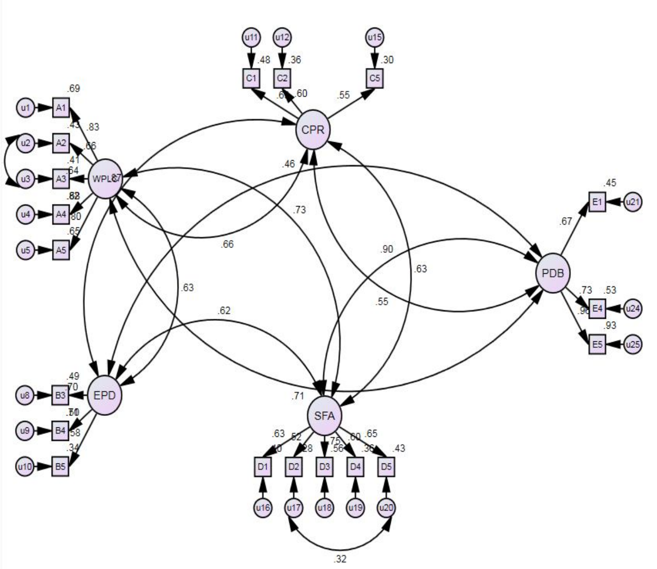

Figure 2 - CFA Path Diagram (Final Run)

4.4 Structural Equation Modelling

Structural Equation Modelling or SEM is related to CFA and often is used to analyse latent constructs using one or more observed variables and provides structural modelling to impute the relationship between the latent variables [68]. Figure 3 below depicts the SEM Path Diagram with the revised mapping of the relationship with each variable. This is also in sync with the list of questionnaire items dropped to improvise the modelling.

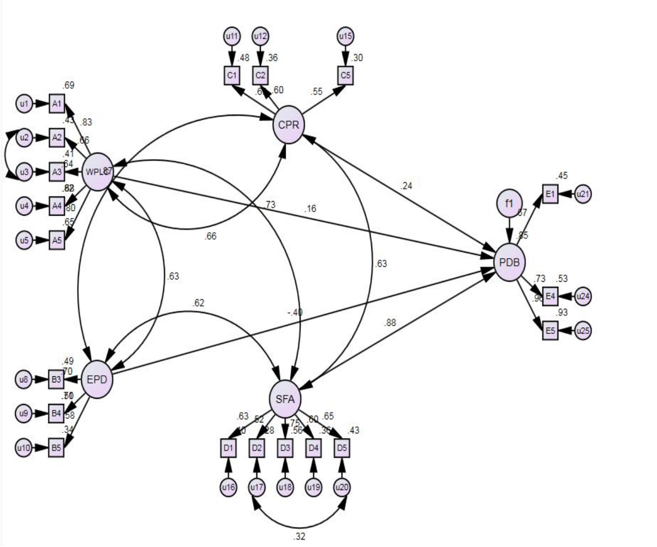

Figure 3 - Structural Equation Modelling Path Diagram

The default model value of X2/DF or CMIN/DF above is 2.710, which is deemed acceptable as it is below the accepted value of 3.00. The GFI, CFI, IFI and TLI values in above are respectively 0.882, 0.914, 0.914 and 0.894. Two of the values are above 0.900, therefore are CFI and IFI; consequently, they are accepted fit. However, GFI and TLI are just on the borderline to the value 0.900. Finally, the RMSEA value for SEM run is measured at 0.076 in the table and is deemed accepted as fit since it is lesser than 0.08. The summary of results from the SEM model fit analysis as depicted in Table 6 below shows 4 out of 6 criteria are met and acceptable within the level. However, GFI and TLI are as indicated below in the borderline very close to 0.900. It is also indicative that both CFA and SEM results are identical.

3. 4.7 Hypothesis Testing Result

Hypothesis testing is running from AMOS and the results from the regression weights and standardized regression weights. As illustrated in Table 7 below the results shows that hypothesis H1, H2 and H3 were negatively found to have a relationship with a personal data breach. However, hypothesis four (H4), was found to have a significant positive impact on the personal data breach.

Table 7. Summary of Hypothesis Testing Result

	Dependent Variable	Independent Variables	β	p	Decision
H1	Personal Data Breach (PDB)	Weak Password Length and Combination (WPLC)	0.159	0.088	Not Supported
H2	Personal Data Breach (PDB)	Easy to Guess Password Dictionary (EPD)	-0.403	0.057	Not Supported
H3	Personal Data Breach (PDB)	Common Password Reuse (CPR)	0.242	0.263	Not Supported
H4	Personal Data Breach (PDB)	Lack of Second Factor Authentication (SFA)	0.881	***	Supported

**p<0.05, **p<0.01, ***p<0.001*

5. RESULT AND DISCUSSION

H1: Weak password length and combination has a significant positive impact on personal data breach

The hypothesis above is not supported, as shown in Table 7 due to β = 0.159 and p = 0.088 (p >0.05, which is the minimum acceptable value). This means weak password length and combination for Malaysian Internet users are not significant impacts towards personal data breach due to insignificant p-value. In terms of the relationship between the two variables, there is an increase of only 15.9% over personal data breach when weak password length factor has a positive factor by 100%. However, any results can not be concluded if results are not significant. There might be another statistical reason behind such findings.

H2: Easy to guess password dictionary has a significant positive impact on personal data breach

The hypothesis is not supported due to β = -0.403 and p = 0.057 (p >0.05, which is minimum acceptable value). The results show easy to guess password dictionary has a negative impact on the personal data breach. The p-value is rather close to 0.05 but still not being significant due to lesser than 0.05. A personal data breach will decrease by 40.3% with the relationship of easy to guess a password, which is negative. This means that respondents do not necessarily agree to that easy to guess password dictionary greatly lead to individual or personal security breach. This could be possible due to split of respondents in two extremes for questions referring to simple diction and password that is containing a name of known people with top two majorities are either agreeing or disagreeing. It is therefore not evident enough to support the positive hypothesis, which resulted in the result is otherwise. This contradicts a previous empirical research done by [44] that found common password dictionary or passwords that are considered weak and uses nouns, family name, birth dates, pet names or even anniversary date significantly affect personal that breach.

H3: Common Password use has a significant positive impact on personal data breach

This hypothesis too, is not supported. As illustrated in Table 7, the value for this hypothesis is β = 0.242 and p = 0.263 (p >0.05 which is minimum acceptable value). This implies that common password reuse is not creating a significant positive impact on personal data breach due to insignificant p-value, which is more than 0.05. In terms of the relationship between the two variables, there is an increase of only 24.2% over personal data breach when common password reuse has a positive factor of 100%. This has been proven to be the opposite of the suggested past literature that common password reuse has been a common denominator to many reported security breaches. [69] reported that not many companies or individual in Malaysia are compelled to report or disclose data breaches which now is an increasing requirement to have laws that mandate such reporting. A 2018 Global Password Security Report shows a staggering 50% of users have the tendency of using the same password for their work or personal and this was further amplified by Google research identifying 65% doing so in 2019 resulting in compromised passwords responsible for 81% of hacking-related breaches [70]. This is probably not familiar to many Malaysia users in the impact of their perception of common password reuse towards personal data breach.

H4: Lacking use of second-factor authentication has a significant positive impact on personal data breach

The hypothesis is Supported as, β = 0.881 and p <0.001, which is excellent. This signifies that there can be increase of 88.1% in personal data breach when lacking use of second-factor authentication is increased by 100%. Technology Visionaries [71] reported that even when users tend to have bad password habits which include easy to guess password or common password reuse for multiple accounts, second-factor authentication has significantly helped to protect users from stolen credentials or personal login being hacked. This is a significant insight among Malaysian respondents that second-factor authentication supersedes any other possible weak factors like password length or combination, easy to guess the password and/or common password reuse across multiple logins that have a severe impact towards personal data breach. The data represented from the standardized regression weights indicate a β value of 0.881 with desirable p-value under 0.01. This goes in line with a suggested study by Duo Security that between 2010 and 2017, US internet users are tech-savvy in moving up the advancement of second-factor authentication to deal with newer advances in personal breaches [47]. What is also implied from this research, is that most respondents in Malaysia agree to this as a single biggest factor contributing to securing from personal data breaches. This study confirms a finding by Albayram, et al., [72] that Internet users are willing to try second-factor authentication after they were exposed to both the Risk and Self-efficacy themes and correlates the experiment by Siadati et al.,[73] that found second-factor authentication prevent personal data breaches and social engineering attacks.

6. CONCLUSION

From the four proposed hypotheses, hypothesis 4 was found to have a positive impact on Personal Data Breach significantly. Based on this finding, it can be concluded that the lack of second-factor authentication is a major factor that significantly impacts personal data breach. This study provides a different perspective on the usual connection of bad password habits such as weak password length and combination, easy to guess the password and common password reuse to be the main contributing factor of personal data breach in the past literature. The contribution of this research is the provision of evidence that indicate that beefing up personal security using second-factor authentication across Internet online accounts, is very important to curb personal data breach. In regards to individual users, each individual should seriously consider enabling second-factor authentication wherever possible. This can be adopted by enforcing the use of One-Time-Pin (OTP). It can be enforced for online banking transactions, payment gateways, recovery of forgotten password, or lost account due to inactivity after a long period. For online accounts which are integrated with other second-factor mechanisms like Authenticator feature, i.e. Google, Microsoft authenticator, etc., users show start to activate or move to this feature to strengthen their credential access from password leaks of account hacks. For online accounts or websites with minimum feature like using email or other simple techniques for a way of second-factor authentication, it should be considered as a minimum viable security measure to securing login credentials.

Next, in regards to Internet Content Providers, not all websites or internet content providers with logon requirements have created a minimum viable solution using various features of second-factor authentication. This ought to be addressed by the providers to ensure that they provide a reliable and trusted avenue to their users to protect themselves from being a victim of a personal data breach. They are many free or commercial possible solutions to enable a second-factor authentication feature so that users can feel safe to use their services. Search engine providers like Google, Bing, etc. should consider tagging every single website in terms of reliability, including the availability of second-factor authentication features. This, in return, provides a confidence index to users to understand the risks of any websites which they entrust they login credentials to be safe or free from a personal data breach. Finally, looking at the role of regulations or law enforcement, there are current state-wise laws like Data Privacy Act, GDPR, National Security laws to protect the public and its people from criminalist or espionage activities like a breach of personal users. These laws should consider enforcing data processors like Internet Content Providers with stricter penal codes for enforcing second-factor authentication feature to provide public assurance of their data integrity, confidentiality and security.

The limitation of this research is; firstly, the limitation of the data sampling. The total; the number of respondents was 297. Future researchers should consider increasing the size of the data sample to increase the prediction power of the findings. This research only collected data from a single location in Malaysia; future researchers should try to collect data from all regions of Malaysia to provide wide geographical coverage and have a proper representation of the population. The final limitation of this study is the number of variables used in this study. Future researchers should explore variables such as back doors, application vulnerabilities, malware, social engineering, insider threats and physical attacks in their studies as that could lead to a better understanding of factor influencing data breach.

REFERENCES

1. The European Data Privacy Services or EDPS (2018) [0nline] Available at https://op.europa.eu/webpub/edps/2018-edps-annual-report/en/ [Accessed 15.02.2020]
2. J.De Groot, “The History of Data Breaches, Digital Guardian” 2019 [Online] Available at https://digitalguardian.com/blog/history-data-breaches [Accessed 30.01.2020]
3. Liu, Z., Hong, Y., and D. Pi, D. “**A Large-Scale Study of Web Password Habits of Chinese**

Network Users." *JSW*, vol 9, no 2, pp. 293-297, 2014.

4. E.H., Spafford, **"Preventing Weak Password Choices**", *Computers & Security*, vol.11, no 3, pp. 273-278, 1992
5. R.Wash, E, Rader, R. Berman, R., and Z .Wellmer, "**Understanding password choices: How frequently entered passwords are re-used across websites**", *Twelfth Symposium on Usable Privacy and Security, SOUPS*, pp. 175-188, 2016.
6. J.S. Vorster and R.P., and Van Heerdeen, "**A Study Of Perceptions of Graphical Passwords**", *Journal of Information Warfare* vol.14, no. 3, pp.75-85, 2015,
7. T.Hussain, K., Atta, N.Z Bawany, and T, Qaamr, "**Passwords and User Behaviour",** *Journal of Computers* vol.13, no. 6, pp.692-704,2018.
8. H.Sanchez, D.Sanchez, D. and J. Murray, "**Putting Your Passwords on Self Destruct Mode: Beating Password Fatigue**" 2016, [Online] Available at https://www.usenix.org/system/files/conference/soups2016/wsf_paper_sanchez_final.pdf [Accessed 20.02.2019]
9. A.MacGibbon, and N Phair, N., "Cyber White Paper", 2011, [Online], Available at http://www.canberra.edu.au/cis/storage/CentreforInternetSafetyCyberWhitePaperSubmission.pdf [Accessed 21.02.2019]
10. C. Gredler, "Consumer Survey: Password Habits" 2012, [Online] Available at https://www.csid.com/2012/09/consumer-password-habits-unveiled/ [Accessed 17.02.2019]
11. K.,Nayak, D., Marino, P. Efstathopoulos, and T. Dumitraş, "**Some Vulnerabilities Are Different Than Others, International Workshop on Recent Advances in Intrusion Detection**", *In International Workshop on Recent Advances in Intrusion Detection, Springer, Cham* pp. 426-446, T, 2014
12. G.Elahi., E. Yu,. and N. Zamone, "A Vulnerability-Centric Requirements Engineering Framework" 2010, [Online] Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.%20421.1569&rep=rep1&type=pdf [Accessed 21.02.2019]
13. M.Raza, M. Iqbal, M. Sharif, and W Haider "**A Survey on Password Attacks and Comparative Analysis on Methods of Secure Authentication**" *World Applied Sciences Journal* vol.19, no.4, pp.439-440, 2012.
14. K. Thomas, F.Li, A., Zand, J., Barrett, J. Ranieri, and L.Invernizzi and D. Margolis, **Data breaches, phishing, or malware? Understanding the risks of stolen credentials**. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* 2017, pp. 1421-1434.
15. Verizon **"Data Breach Investigations Report**" 2017 [Online], Available at https://www.knowbe4.com/hubfs/rp_DBIR_2017_Report_ execsummary_en_xg.pdf [Accessed 21.02.2019]
16. K. Kiesnoski, "**5 of the biggest data breaches ever"** *CNBC* 2019, [Online] Available at https://www.cnbc.com/2019/07/30/five-of-the-biggest-data-breaches-ever.html [Accessed 12.11.2019]
17. SANS Institute, "**Password Policy**" 2014 [Online] Available at https://www.sans.edu/student-files/projects/password-policy-updated.pdf [Accessed 19.04.2019]
18. R. Shay, I, Ion, R.W, Reeder, and S. Consolvo, " **My Religious Aunt Asked Why I Was Trying To Sell Her Viagra Experiences With Account Hijacking**". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 2657-2666).
19. A.Narayanan, and V. Shmatikov, "**Fast Dictionary Attacks On Passwords Using Time-Space Tradeoff**" *12th ACM conference on Computer and communications security*, 2005, pp. 364-372
20. B. Pinkas, and T.Sander, **Securing Passwords Against Dictionary Attacks,** *9th ACM conference on Computer and communications security*, 2002, pp. 161-170.
21. J., Abott , D.Calarco, and L. Jean Camp, " **Factor Influencing Password Reuse: A Case Study**", *Conference: TPRC46: Research Conference on Communications, Information and Internet Policy*, 2018
22. A.Das, J. Bonneau, M. Caesar,N. Borisov and X.Wang, "**The Tangled Web of Password Reuse**, *In NDSS*, vol. 14, pp. 23-26, 2014.
23. A.F. Pomputius, "**A Review of Two-Factor Authentication: Suggested Security Effort Moves to Mandatory**", *Journal Medical Reference Services Quarterly*, vol. 37, no. 4, pp. 397-402, 2019.
24. BorneoPost, "**Personal Data Protection Act comes into force on Nov 15**". 2013, [Online] Available at https://www.theborneopost.com/2013/11/27/personal-data-protection-act-comes-into-force-on-nov-15/ [Accessed 10.11.2019]
25. CNN "**AOL Worker Arrested In Spam Scheme, CNN Money**" 2004 [Online] Available at https://money.cnn.com/2004/06/23/technology/aol_spam/ [Accessed 30.01.2020]
26. NortonLifeLock, "**A Brief History of Data Breaches**" 2018 [Online] Available at https://www.lifelock.com/learn-data-breaches-history-of-data-breaches.html [Accessed 30.01.2020]
27. Statista, "**Personal Data Breaches and disclosures**". 2017,[Online] Available at https://www.statista.com/statistics/273550/data-breachesrecorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ [Accessed 30.01.2020]

28. P.A., Gagniuc, "**Markov Chains: From Theory To Implementation And Experimentation**" *USA, NJ: John Wiley & Sons,* 2017
29. L.R., Rabiner, "**A Tutorial On Hidden Markov Models And Selected Applications In Speech Recognition**", *Proceedings of the IEEE* , vol.77 no.2, pp.257-286, 1989.
30. C. D. Manning , and H Schutze, **"Foundations Of Statistical Natural Language Processing", *MIT Press, Cambridge, MA,* 1999,**
31. C. Castelluccia, M. Dürmuth, and D. Perito, "**Adaptive Password-Strength Meters from Markov Models**", *In NDSS,* 2012.
32. W.Tansey, "**Improved models for password guessing**" *University of Texas, Tech. Rep*. 2011.
33. Y. Trope, and N. Liberman, "**Construal-Level Theory Of Psychological Distance"**, 2010.
34. Y. Bar-Anan, N. Liberman, and Y. Trope "**The Association Between Psychological Distance and Construal Level: Evidence From An Implicit Association"** Test, *Journal of Experimental Psychology: General, vol*. 135, no.4, pp.609–622. 2006.
35. N. Liberman, and J. Förster "**The Effect Of Psychological Distance On Perceptual Level Of Construal**", *Cognitive Science: A Multidisciplinary Journal*, vol.33, no.7, pp. 1330–1341, 2009.
36. T. Eyal, N. Liberman, Y.Trope, and E. Walther, "**The pros and cons of temporally near and distant action"**. *Journal of personality and social psychology*, vol 86. no.6 781. 2004.
37. L.Tam, M. Glassman, and M.Vandenwauver "**The Psychology Of Password Management: A Tradeoff Between Security And Convenience**", *Journal of Behaviour & IT,* vol.29 pp. 233-244,2010.
38. V. Grimm and S.F. Railsback "**Individual-based Modeling and Ecology**", *Princeton University Press,* 2005
39. V. Kothari, J. Blythe, S.W.Smith, and R. Koppel "**Measuring The Security Impacts Of Password Policies Using Cognitive Behavioral Agent-Based Modeling**" *In Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, 2015,pp. 13
40. TraceSecurity, "**Data Breaches Due to Poor Passwords**" ,2017 [Online] Available at https://www.tracesecurity.com/blog/articles/81-of-company-data-breaches-due-to-poor-passwords [Accessed 23.04.2019]
41. L. Ablon, P.Heaton, D.C.Lavery and S. Romanosky " **Consumer Attitudes Toward Data Breach Notifications And Loss Of Personal Information**", *In Proceedings of the Workshop on Economics of Information Security (WEIS),*2016.
42. Infosecurity Group, "**Linkedin Breach: Weak Passwords Are The Norm"** 2017, [Online], Available at https://www.infosecurity-magazine.com/news/linkedin-breach-weak-passwords/ [Accessed 21.04.2019]
43. Forbes, "**The Worst Passwords Of 2018 Show The Need For Better Practices**" ,2018, [Online], Available at https://www.forbes.com/sites/kateoflahertyuk/2018/12/14/these-are-the-top-20-worst-passwords-of-2018/#4b17d3e14541 [Accessed 19.04.2019]
44. J.A. Cazier and B.D. Medlin "**Password Security: An Empirical Investigation Into Ecommerce Passwords And Their Crack Times**", *Information Systems Security: the (ICS)2 Journal*,vol.15, no.6, pp.45-55, **2006,**
45. **B. Ives, K. R.,Walsh and H. Schneider,"The Domino Effect Of Password Reuse."** *Communications of the ACM*, vol. *47, no.*4, pp.75-78, 2004.
46. K. Helkala, and T.H. Bakås, "**National Password Security Survey: Results**", *In EISMC*, pp. 23-33,2013,
47. Duo Security Submit, 2015. [Online] Available at https://duo.com/blog/2015-duo-security-summit-recap-adventures-in-sf
48. A.A. Aliyu, M.U,Bello, R. Kasim, and D. Martin, "**Positivist and Non-Positivist Paradigm in Social Science Research: Conflicting Paradigms or Perfect Partners**?" *Journal of Management and Sustainability*, vol. 4 no.3., pp.79-95, 2014.
49. P.Patel, "**Introduction to Quantitative Methods**", *In Empirical Law Seminar*, 2009.
50. D. L., Altheide, and J.M Johnson " **Criteria for Assessing Interpretive Validity in Qualitative Research**", *N. K. Denzin & Y. S. Lincoln (Eds.). Handbook of Qualitative Research*, pp. 485-499. 1994.
51. N.J. Gogtay, and U.M Thatte, "**Principles Of Correlation Analysis".** *Journal of the association of physicians of India*, vol.65, pp.78-81,2017
52. J.J.Hox, and T.M Bechger "**An Introduction To Structural Equation Modeling"** *Family Science Review* vol. 11,pp. 354-373, 1998.
53. R.Teclaw, M.C, Price, and K. Osatuke, "**Demographic Question Placement: Effect On Item Response Rates And Means Of A Veterans Health Administration Survey".** *Journal of Business and Psychology*, vol.27,no. 3, pp. 281-290, 2012.
54. H.Y. Kim, "**Statistical Notes For Clinical Researchers: Assessing Normal Distribution (2) Using Skewness And Kurtosi**s". *Restorative dentistry & endodontics*, vol. *38* no.1. pp. 52-54, 2013.
55. S.G.West, J.F.Finch, and P.J. "**Structural Equation Models With Nonnormal Variables: Problems And Remedies**" *In: Hoyle RH, editor. Structural equation modeling: Concepts, issues and*

applications, Newbery Park, CA: Sage, pp. 56-75, 2013.

56. F.J. Gravetter, and L.B. Wallnau, "**Essentials Of Statistics For The Behavioral Sciences**". *Cengage Learning,*2020.

57. L. J., Cronbach, **"Coefficient Alpha and The Internal Structure of Tests"** *Psychometrika*, vol.16,no 3, pp. 297–334, 1951.

58. N.Schmitt, "**Uses And Abuses Of Coefficient Alpha"**. *Psychological assessment*, vol.*8*, no.4, 350, 1996.

59. K.S.Taber, "**The Use Of Cronbach's Alpha When Developing And Reporting Research Instruments In Science Education**" *Research in Science Education*, vol.48, no.6, pp.1273-1296, 2018.

60. C.Huber-Carol, N. Balakrishnan, M. Nikulin, and M. Mesbah, (Eds) "**Goodness-Of-Fit Tests And Model Validity**". *Springer Science and Business Media*, 2012.

61. P.M., Bentler "**Comparative Fit Indexes In Structural Models**" *Psychological bulletin*, vol.107 , no.2, pp.238,1990.

62. K.A. Bollen, "**A New Incremental Fit Index For General Structural Equation Models**" *Sociological Methods & Research*, vol.17, no.3, pp 303-316, 1989.

63. D.A.,Kenny, , "**Measuring Model Fit**", 2015 [Online] Available at http://davidakenny.net/cm/fit.htm

64. J. F., Hair Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. **"SEM: An Introduction". *Multivariate Data Analysis*** *: A Global Perspective"*, vol.*5, no*.6,pp. 629-686. (2010).

65. T. Brown, "**Confirmatory Factor Analysis for Applied Research",** *New York London: The Guilford Press* T., 2015.

66. W.W., Chin "**The Partial Least Squares Approach To Structural Equation Modeling**". *Modern Methods For Business Research*, vol.295, no.2, pp. 295-336, 1998.

67. L. M., Ahmad, S. A. Ahmad and Z. Smith, "**U.S. Patent No. 9,299,238**." Washington, DC: U.S. Patent and Trademark Office,2016.

68. M.C., Shelley, "Structural Equation Modeling" *Encyclopedia of Educational Leadership and Administration,*2006.

69. TheStar, "**Making It Mandatory To Declare Data Breaches**".2018.[Online] Available at https://www.thestar.com.my/tech/tech-news/2018/07/02/making-it-mandatory-to-declare-data-breaches/ [Accessed 15.02.2020]

70. HelpNetSecurity, "**The Password Reuse Problem Is A Ticking Time Bomb**", 2019. [Online] Available at https://www.helpnetsecurity.com/2019/11/12/password-reuse-problem/ [Accessed 10.02.2020]

71. Technology Visionaries, "Why Two-Factor Authentication Can Significantly Reduce Your Chance of a Data Breach" 2019. [Online] Available at https://www.technologyvisionaries.com/two-factor-authentication-reduce-breach/ [Accessed 20.02.2020]

72. Y. Albayram, M. M. H, Khan, and M.Fagan, **A study on designing video tutorials for promoting security features: A case study in the context of two-factor authentication** (2FA). *International Journal of Human–Computer Interaction*, vol.*33,no*.11, pp 927-942, 2017.

73. H.,Siadati,T. Nguyen, P. Gupta, M. Jakobsson, and N. Memon "**Mind your SMSes: Mitigating Social Engineering In Second Factor Authentication**". *Computers & Security*, pp.*65*, 14-28, 2017.