Identity Theft Prediction Using Game Theory

Nikita Chorghe[1,4,*], *Akshay* Jain[2,**], *Shraddha* Mali[3,***], and *Prathmesh* Gunjgur[4,****]

[1]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[2]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[3]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.
[4]Department of Computer Engineering, Ramrao Adik Institute of Technology Nerul, Navi Mumbai, India.

Abstract. Digital devices have become an integral part of every person's life. The range of use of these devices is increasing daily. Over the decades, the number of users has increased from thousands to millions and is still increasing. Due to the multi-functional features of digital devices, their importance is now being recognized more than ever. Initially, they were used only for calling and texting; however, nowadays, they are also being used to store relevant data such as account numbers, card numbers, credentials, private pictures, passport copies, etc. The most common form of Identity Theft attack is through stealing passwords. Once the password is stolen, user privacy is lost, and the data is compromised. Thus, a system consisting of a database that comprises of leaked passwords collected from various social sites and common passwords as a part of a dictionary attack used by hackers has been created by us. When a user enters his/her password, it runs it through the database and checks for a match. This document emphasizes on how game theory can be utilized in predicting the possibility of a successful attack and discusses essential concepts such as the various components of game theory and Nash Equilibrium.

1 Introduction

With widespread digital growth occurring globally, the threats associated with these digital devices are also evolving. All digital devices are prone to cyber-attacks. With India emerging as a digital nation, the safety of digital devices can be compromised due to multiple vulnerabilities, and they can be exploited.

Mobile devices connected to a network can quickly become a target of cyber-attacks. Hence, it is vital to secure these networks. The main motive of malware is to penetrate IOS, Windows, and Android defence system. The security mechanism must be an integral part of the system and must work in coordination to identify threats and safeguard user data. Cyber-attacks are carried out in the cyber-space with social or political gains as the primary objective. Digital devices, other than being used as a means of communication, are also used for storing personal and corporate data. Viruses, malicious codes, etc. can result in disruptive consequences, which can lead to data loss and endanger the privacy of its users.

Identity theft is one of the most common attacks used by an attacker to steal user data. It is also called as Identity Fraud wherein a hacker obtains important pieces of private, confidential data such as credentials of email accounts, pan card numbers, PINs, and passwords to assume someone else's identity.

This data can serve various purposes to the hacker, such as to obtain credit, goods, or other services registered in the name of the impersonated person. The hacker can also provide false identification to the police to avoid warrants or arrests and to prevent a criminal record.

As per the latest statistics, approximately 200 billion devices will be connected to the internet by 2020. This alarming rise in the number of devices connected to the internet gives increased opportunities for attackers to steal data and infect websites. Since 2013, there has been a theft of 3,809,448 records. In 2016, 95% of the attacks were focused on government, retail, and technology industries [6].

2 Related Work

K.Veena *et.al.* have proposed a new methodology that identifies the assorted identity of any user and determines if a synthetic identity theft attack has occurred. They took three styles of information: Input, Normal, and Target data-set. They used varied identities that could be text and string information. Various identities were classified in three classes as 100% that represented high identity with accurate data, 75% that represented medium identity with partial data, and 0% that represented low identity with wrong information. The expected values were 0% or 100% for

*e-mail: nikitachorghe98@gmail.com
**e-mail: aj2812@gmail.com
***e-mail: shraddha.mali1812@gmail.com
****e-mail: prathmesh.gunjgur@rait.ac.in

the various identities, and the normal value ranged from 0% to 100%. The neural networks were trained on top of values. The progress was obtained for the epoch values, time, performance, gradient, and validation checks [1].

Sharmistha Dutta *et.al.* have proposed a method that deals with credit card application crimes. Their techniques are used to remove identity theft. They have proposed new data mining techniques. Mainly two algorithms are used they are Communal Detection and Spike Detection for fraud detection. The communal algorithm identifies the communal data, and the Spike detection algorithm is used to detect spikes in the duplicates. The system uses the resilience concept, which is a multi-layered data mining based approach [2].

Philip A.K.Lorimer *et.al.* proposed a framework that focuses on the validation of social profile. Using the proof-of-concept study, their proposed framework detects abnormal behavior in social profiles. Their numerical value results show that if the matching threshold in a decision tree is appropriately set, then the system identifies compromised accounts by avoiding the heavy central processing [3].

Very few systems have been developed for Identity Theft attack detection. The process of identification of an Identity Theft attack is based on different parameters for which various databases have been compiled. These parameters can be easily manipulated.

1. They have proposed a method that deals with credit card information theft by using two algorithms: communal data and spike detection algorithm. A major advantage is that they have used a multi-layered data mining approach [1].

2. They have also developed a framework that focuses on the validation of social platforms [2].

3. A significant drawback of all the systems is that if by chance a user's password is hacked, then using that password, data can easily be accessed. Thus, there was a need to mitigate this risk by preventing users from already using passwords that have been leaked [3].

3 Proposed Work

The objective of developing the OSI model was to provide a set of design standards for equipment manufacturers to communicate with each other. It consists of 7 layers. Attackers target the vulnerabilities present in the network and physical layers.

Over the years, many developments have been made to improve the OSI model. However, the base structure remains the same. Hence, these problems exist to date. A cyber-attack is intentional exploitation of resources, hardware/software, and networks for personal or monetary gains. Using malware, infecting systems by viruses and worms, using backdoor traps, or brute force attacks are some of the common means and methods to conduct cyber-attacks. Attackers perform these attacks to gain access to confidential information or resources of an organization or to steal data that can be later sold on the dark web.

The main motive of attackers behind Identity Theft is to steal personal or financial information of a person with the sole purpose of obtaining that person's identity and making transactions or purchases.

Table 1. Issues in Physical and Network Layers

Network layer	**Physical layer**
Various data routing paths for communication are provided by the Network layer.	Actual physical connectivity is defined by the Physical layer
Data is transferred in the form of packets using different logical network paths in a sequential order controlled by the network layer.	Physical layer defines the hardware equipment, cabling, wiring etc.
Network layer utilises multiple routing protocols on the network.	It majorly consists of hardware equipment such as cables, routers etc.
Various attacks such as ICMP, packet sniffing and DOS attacks are performed majorly over the Internet.	DOS attacks are performed by cutting wireless network cables.

Table 1 refers to various issues present in the Physical and Network layer of the OSI model. These vulnerabilities are exploited by the attackers to gain access to confidential data. The different types of theft are shown in Table 2.

Table 2. Different Types of Thefts [4]

Theft Type	**Description**
Criminal	A criminal uses the identity of other person in order to evade an arrest and conviction records.
Medical	This type of theft is usually done in order to obtain free medical services.
Financial	Among all the types of thefts, this is one of the most common and frequently committed theft. Another person's identity is used for goods, credits and services.
Child Identity Theft	The perpetrator uses the child's name and other confidential information with the intention of avoiding arrests, obtaining loan or a job etc.

Identity theft is a crime in which an attacker steals the information from a legitimate user and uses that information to impersonate someone. This leads to the immediate impact of losing something valuable like money. Identity theft victims can face abstract costs such as defamation of character, prevention in securing credit cards or position in a company, etc. There are very fewer chances of recovering from Identity theft after such circumstances.

Participation in various events is only possible because of the identities assigned to individuals and the privileges granted to them. Some of these privileges

are driving, attending school, donating blood, using credits, taking loans, etc. In cases of account hacks, the hacker exploits the privileges granted to the victim by misusing resources such as writing cheques against the person's account or misusing stolen credit cards to buy utilities in the person's name. In other cases, hackers misuse an individual's identity for credit cards or loan applications. Fraudsters are coming up with new and varied methods to assume the identity of other individuals. Thus it is necessary to develop a preventive measure to reduce the disastrous effects of identity thefts [8].

Identity theft can occur in various forms, such as insurance, credit, and debit card or mortgage fraud. Most of these thefts are unreported because the victim is not aware of the occurrence of such an event. The risk to financial institutions and many such institutions continues to grow exponentially. Hence, security improvements that provide technical solutions and adjust as per human behavior are required.

Employees of businesses and organizations are targets of Identity thefts as it leads to defamation and breach of trust. Sometimes users keep a single password for multiple accounts. These can be social networking apps, email accounts, or even bank accounts. If the hacker gets access to any one of the passwords, the most common practice followed is to use the same password to hack other accounts linked to the victim. Thus, the hacker through a single password can gain access to a massive amount of personal data. Once the data is at hand, the hacker can demand a ransom or release the data on the internet. Apart from monetary loss, it can also lead to loss of life.

One of the possibilities is that the leaked passwords could be of an organization's for centralized servers. Getting access to a server's password can be troublesome for an organization. The hacker can demand a ransom, can sell all the data to the organization's competitors, or sell their client's private data on the darknet, which can put the users at risk.

The proposed framework can be used to detect attacks and, thus, mitigate such threats and secure user data. We have developed a hacked password checker, which will work as an alert system for users. The system will check whether the password has been hacked and if so, it will report to the user.

3.1 Procedure

The proposed work has a hacked password checker, which will work as an alert system for users. It will check whether the passwords have been hacked or not. It examines the password, and if that password matches the list of passwords that have been leaked, it will report to the user. Algorithm 1 provides a series of steps to be followed to prevent the user from using an already leaked password. This, in turn, will prove to be beneficial for the user by eliminating the user as a possible target of identity theft.

Algorithm 1: Identity Theft Detection

1.Start
2.Create a data set of leaked passwords.
3.Store all the entries after converting them into a hash.
4.Take a user input.
5.Convert the input into hash value.
6.Check the user input with each value stored in the database.
7.**if** *The user input matches with any of the values stored in database* **then**
  Alert the user the password has been leaked;
else
  Notify the user that the password has not been leaked;
end

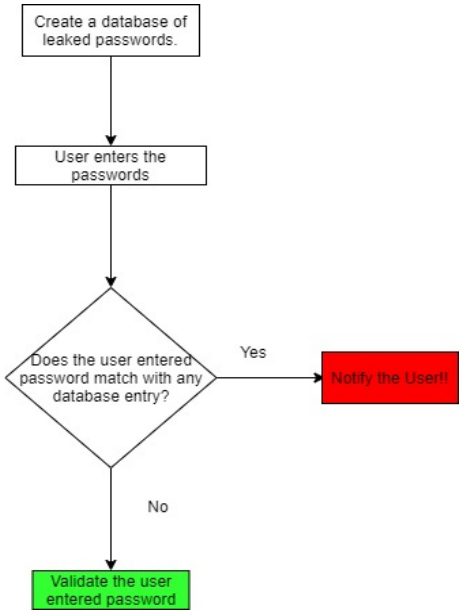

Figure 1. Identity Theft Flowchart

Figure 1 explains the flow of the prevention of Identity Theft attacks. A database of the leaked passwords is created. The user enters his/her password. The system will check if that password matches the user entered password; if that password matches the user will be notified, or else the user's password is not leaked.

3.2 Game Theory

Game theory is a theoretical framework that is used to analyze social situations among different competitors [5].

Game theory is used to study situations involving more than one participant in a scenario and in which not everything depends on one person. Game theory analyses the behavior of two or more participants involving rewards or punishments. E.g., in the game

of chess, the opponent's moves continuously influence your strategy. It is also used to analyze situations in which multiple companies are competing against one another. Let us suppose if one of the competitors reduces the price of its product, other companies might be forced to do the same. The modeling of strategic interaction between multiple competitors in a scenario containing rules and outcomes is called game theory [5].

Game theory can be a useful approach to solve cybersecurity problems that require rational decision making. An assumption is made in game theory that all the participants have complete information about each other and know all the rules and regulations. Game theory enhances the ability to anticipate the actions of hackers. Recently, there were a few attacks wherein the hackers stole sensitive data and demanded Bitcoins as a ransom to maintain integrity and data security. Once the data is leaked, the hackers have complete control over it. They can compromise the security of the data by threatening the availability and confidentiality of it. Game theory is preferable as it shares similar concerns with cybersecurity in various aspects of their application. The move of the hacker with respect to the defender is not only contingent of the hacker's decision but also depends upon the defender's behavior. Thus, game theory can be used as a mathematical tool to deal with cybersecurity problems based on multi-agent behavior. Moreover, the combination of game theory and cybersecurity can be used to solve real-time problems [7].

3.2.1 Nash Equilibrium

The concept is named after the American mathematician John Nash. It is a scenario of an optimal outcome wherein the player has no motivation to change the chosen strategy. It is a concept which states that an opponent's move provides no incentive to any of the players to deviate from the initially chosen strategy. The player, even after changing the course of action, will receive no incremental benefit provided all other players do not change their methods. A game can contain multiple Nash Equilibria or none at all. It helps in determining a set of actions that all players must take to secure the best possible outcome for themselves.

To better understand Nash Equilibrium, we can consider an example. Let us suppose there are two players: A and B. In this game, players can either choose strategy 1, which can win them Rs 100, or they can choose strategy 2 wherein they lose Rs 100.

It is common sense that both players A and B will choose strategy 1. Even after revealing the strategies of players A and B to each other, the likelihood of the players changing their currently selected strategy is zero. Thus, there is no deviation from the initially chosen strategy even after knowing the opponent's move. Thus strategy 1 is Nash Equilibrium [5].

3.2.2 Components of Game theory

The Table 3 represents various components of game theory and their description.

Table 3. Components of Game Theory [5].

Components	Description
Game	Any scenario where in the result is influenced by the actions of multiple players.
Player	Any entity making deliberate decisions based upon the rules of a game.
Strategy	Series of actions that players will perform under a given circumstance, which may or may not occur in the game.
Payoff	Payoff is anything that a player incurs by arriving at a consequence. It can be in any form such as money, utility etc.
Information Set	The data available at any given point in a game.
Equilibrium	The stage where all the players have reached an outcome after making their decisions.

3.2.3 Motivation to use Game Theory

Recently, 17 million user records were stolen from Zomato's database. This information included email id's and passwords of users. This confidential data was sold on the popular Dark web for a mere $1000. [11].

In the year 2016, the Internet service company Yahoo reported two significant breaches. Both the breaches are considered as one of the most massive breaches in the history of the Internet. The first breach affected roughly 500 million Yahoo users. Yahoo confirmed that the second breach affected all 3 billion users of Yahoo. Verizon, which was to buy Yahoo for $4.8 billion, bought it for a reduced price of $350 million, proving to be a loss for Yahoo [9].

Any user can become a victim of Identity theft. As per the US Department of Justice, 17.6 million people in the United States experience some form of identity theft every year. In 2014, victims experienced a combined average loss of $1,343. In total, victims lost a massive $15.4 billion [10].

In order to reduce such catastrophic events, it is obligatory to develop a solution to prevent thefts. The methods hackers use evolves with each security system developed to prevent them. They try to expose new vulnerabilities. Thus, an ideal approach to prevent these attackers is to be used. The move of the attackers may change based on the countermeasures taken by the defense system developed to prevent such attacks. Thus, game theory proves to be the most classic approach. The game theory takes into consideration the measures to be taken by the system software or the users to prevent identity thefts and how the attackers' reactions to such countermeasures.

3.2.4 Game Theory for Identity Theft

Consider an attacker 'A' and detector 'D'. The attacker has two possibilities of either attacking or not attacking using the obtained password. The detector will detect whether the password has been leaked or not. Consider the following example. Each scenario is associated with a penalty and reward as per the action taken.

Table 4. Terminologies in an ID Theft Game

Terminology	**Description**
q	probability that attacker attacks
r	probability that detector detects leaked password.
U_d	Detector's Utility.
U_a	Attacker's Utility.

Table 4 represents various terminologies used in an identity theft game and its description.

Table 5. Parameter

		Attacker	
		Attack	Don't Attack
Detector	Alert	0, -2	-4, 0
	Not Alert	-6, 6	2, 0

Table 5 refers to a scenario where there are possibilities of multiple events taking place depending upon the action taken by the attacker and defender. Consider the following parameter

1. **(Alert, Attack) = (0, -2)**
   In this scenario, the attacker attacks and is detected by the detector. Here, 0 denotes the reward the detector will gain, if it successfully detects that the password has been leaked and -2 denotes the value the attacker has been penalized for being caught.

2. **(Alert, don't attack) = (-4,0)**
   In this case, the attacker does not attack even though the system denotes that the password has been leaked. -4 represents detection of leaked password even though the attacker does not attack.

3. **(Not Alert, attack) = (-6,6)**
   In this situation, the attacker is successfully able to conduct an attack without triggering the detector. -6 is penalty faced by the detector for not being able to notify about the attack and 6 represents the reward gained by the attacker for a successful attack.

4. **(Not Alert, don't attack) = (2,0)**
   In this scene, neither the attacker attacks nor the system detects anything. 2 is used to represent that password is not leaked and 0 represents that the attacker does not attack.

For this game, there is no pure Nash Equilibrium. However, we can derive a mixed strategy equilibrium.

Let us consider:
q $\Longrightarrow$ probability that attacker attacks.

For 'q' to be in equilibrium, the detector needs to be indifferent in detecting leaked and not leaked passwords. Let U_d be Detector's Utility (Expected Utility). When attacker attacks with probability 'q' and detector will detect that password is leaked.

$$U_d(q, Alert) = [q \times (0) + (1-q) \times (-4)] \quad (1)$$

$$U_d(q, Not\ Alert) = [q \times (-6) + (1-q) \times (2)] \quad (2)$$

For 'q' to be in Nash equilibrium, Equation 1 and 2 must be equal.
Thus equating both equations we get Equation 3:

$$\begin{aligned} (1-q) \times (-4) &= q \times (-6) + (1-q) \times (2) \\ -4 + 4q &= -6q + 2 - 2q \\ -4 - 2 &= -6q - 2q - 4q \\ -6 &= -12q \\ q &= 1/2 \end{aligned} \quad (3)$$

Now for attacker:
r $\Longrightarrow$ probability that detector detects leaked password.

$$\begin{aligned} U_a(Attack, r) &= [(-2) \times (r) + (1-r) \times (6)] \\ U_a &= r(-2) + (1-r)(6) \\ U_a &= -2r + 6 - 6r \\ U_a &= 6 - 8r \end{aligned} \quad (4)$$

$$U_a(Don't\ attack, r) = 0 \quad (5)$$

For r to be in Nash equilibrium, Equation 4 and 5 must be equal.
Thus equating both equations we get Equation 6:

$$\begin{aligned} 6 - 8r &= 0 \\ 6 &= 8r \\ r &= 3/4 \end{aligned} \quad (6)$$

Thus, Nash equilibrium is solved for the game. Hence, when attacker attacks 1/2 of the times, the detector alerts about leaked password for $3/4^{th}$ of the time.

4 Conclusion

Setting a password is not the only protective measure one must undertake to safeguard data. There is always a possibility of the password getting hacked or the password already being leaked. Hence, one can lower the risk of data theft by using passwords that have not been previously leaked or those which are not a part of common dictionaries used by hackers to perform brute force attacks. Thus, a system has been developed wherein a database of passwords

leaked over the years from multiple social sites and those commonly used in brute force attack are compiled. This system notifies its user if their password is a part of the database. Analysis of the possibility of a successful data theft attack has been conducted with the help of Game Theory. The key concepts, such as Nash equilibrium and mixed strategy equilibrium, have been applied to predict the possibility of an attack in multiple scenarios.

References

[1] Veena, K., & Meena, K. "Determination of performance to verify the synthetic identity theft by training the neural networks," IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017.

[2] Sharmistha Dutta, Ankit Kumar Gupta and Neetu Narayan."Identity Crime Detection Using Data Mining," IEEE International Conference on Computational Intelligence and Networks (CINE), 2017.

[3] Philip A. K. Lorimer, Victor Ming-Fai Diec,and Burak Kantarci."Participatory detection of identity theft on mobile social platforms," IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017.

[4] Tajpour, Atefeh, "Identity Theft and Fraud Type," International Journal of Information Processing and Management (IJIPM), 2013.

[5] Raoof, Omar & Al-raweshidy, Hamed."Theory of Games: an Introduction" 10.5772/46930.(2010).

[6] Identity Theft - 1 [Online]. Available https://www.cybintsolutions.com/cyber-security-facts-stats/

[7] Annapurna P Patil, Bharath S and Nagashree M Annigeri. "Applications of Game Theory for Cyber Security System: A Survey." International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 12987-12990, 2018.

[8] Helser, Susan, "Identity theft education: Comparison of text-based and game-based learning" (2016). Graduate Thesis and Dissertations. 15930. https://lib.dr.iastate.edu/etd/15930

[9] Identity Theft - 2 [Online]. Available: https://en.wikipedia.org/wiki/Yahoo!_data_breaches

[10] Identity Theft - 3 [Online]. Available https://www.csid.com/2016/09/real-cost-identity-theft/

[11] Identity Theft - 4 [Online]. Available bit.ly/zomatoleaks