IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 12, Issue 3 (Jul. - Aug. 2013), PP 14-19
www.iosrjournals.org

Game Theory Approach for Identity Crime Detection

J.R.Jayapriya[1], A.Karthikeyan[2]

[1]*(Department of Computer science and Engineering, Arunai Engg. College (Affiliated to Anna University), India)*
[2]*(Department of Computer science and Engineering, Arunai Engg. College (Affiliated to Anna University), India)*

Abstract : *To present game theory approach to detect identity crimes. Improve the adaptability of identity crime detection systems to real time application. Time constraints on the reactive time of the detection and fraud events need to be minimized. Identity crime has major thrust in credit application. Existing work presented multilayered detection system based on two layers named as Communal detection and Spike detection. Dynamic Time Warping algorithm is applied to minimize the time constraints on detecting fraudulent identity usage and reaction time. Performance analysis is carried out on CD and SD with real credit applications. Experiment is conducted on real time credit card application using UCI repository data sets with synthetic and real data sets.*

Keywords *- Data mining-based fraud detection, security, and data stream mining, and anomaly detection.*

Submitted date 15 June 2013 **Accepted Date: 20 June 2013**

I. INTRODUCTION

Identity crime detection is more demanded system for Credit card fraudulent. Synthetic identity fraud refers to the use of plausible but fictitious identities, effortless to create but more difficult to apply successfully. Real identity theft refers to illegal use of innocent people's complete identity details, harder to obtain but easier to successfully apply. In reality, identity crime committed with a mix of both synthetic and real identity details. Identity crime has become prominent as many real identity data available on Web. Confidential data accessible through unsecured mailboxes, easy for perpetrators to hide their true identities. Application of identity crime is in insurance, tax returns and telecommunications. Data breaches involve lost or stolen consumer's identity information lead to other frauds such as home equity and payment card fraud. As in identity crime, credit application fraud has reached a critical mass of fraudsters who are highly experienced, organized, and sophisticated.

Duplicates refer to applications which share common values. There are two types of duplicates, exact duplicates have the all same values, near duplicates have some same values, some similar values with slightly altered spellings, or both. Duplicates are hard to avoid from fraudster's point-of view because duplicates increase their success rate. The synthetic identity fraudster has low success rate, and is likely to reuse fictitious identities which have been successful before. The identity thief has limited time because innocent people can discover the fraud early and take action, and will quickly use the same real identities at different places.

In short new methods are based on white-listing and detecting spikes of similar applications. White listing uses real social relationships on a fixed set of attributes. This reduces false positives by lowering some suspicion scores. Detecting spikes in duplicates, on a variable set of attributes, increases true positives by adjusting suspicion scores appropriately. Throughout this paper, data mining is defined as the real-time search for patterns in a principled fashion. These patterns can be highly indicative of early symptoms in identity crime, especially synthetic identity fraud.

II. METHODOLOGY

1**.1Credit Application and Identity Crime:** Credit applications are Internet forms with written requests by potential customers for credit cards, mortgage loans, and personal loans. Credit application fraud is specific case of identity crime which involves synthetic identity fraud and real identity theft. Credit application frauds are different to each other constantly change persistent due to high financial rewards. Fraudsters use software automation to manipulate specific values and increase frequency of successful values. Duplicates are exact duplicates have all same values near duplicates some same values, some similar values with slightly altered spellings, both Identity crime is hiding true identity of Individuals, services and data.

Many real identity data available on Web. Confidential data accessible through unsecured mailboxes. Identity crime is prevalent and costly for countries do not have nationally registered identity numbers. Identity crime fraudster can be detected with characteristic pattern of behavior based on white-listing and detecting spikes of similar applications. White-listing uses real social relationships on fixed set of attributes which reduce false positives. Detecting spikes in duplicate based on variable set of attributes by adjusting suspicion scores appropriately.

1.2 Communal Detection: CD is a white-list oriented approach on fixed set of attributes.CD algorithm works in real time by giving detection scores on exact or similar matches between categorical data. Reconstructs its white-list at end of the month, resets its parameter values at end of the day.CD process Multi-attribute link. Match current application against moving window. Determine single attribute similarity threshold.

Create multiple attribute links. Single link score matching multi-attribute link against link types in current white list. Single-link average previous score calculate every linked previous applications score for inclusion into current application's score and previous score act as established baseline level. Multiple-link score calculate every current application's score using every current application's score using every current application's score using every link and previous application score. Parameter's value change end of current micro-discrete data stream adaptive CD algorithm determines State-of-Alert (SoA) and updates random parameter's value tradeoff between effectiveness with efficiency. White list change end of current Mini-discrete stream's links. Increase tamper-resistance in white list.

1.3 Spike Detection: Spike Detection (SD) attribute-oriented approach on variable-size set of attributes.SD algorithm does attribute selection updates CD attribute weights at end of the month.SD process involves Single-step scaled count matches every current value against a moving window of previous values. Single value spike detection calculate every current value's score by integrating all steps to find spikes previous steps act as established baseline level. Multiple-values score calculate every current application's score using all values score and attribute weights.SD attributes selection end of every current Mini-discrete data stream selects attributes for SD suspicion score probe-reduction of selected attributes.CD attribute weights change end of every current Mini-discrete data stream.SD algorithm updates attribute weights for CD. Stand alone CD assumes all attributes are of equal importance. Resilient combination of CD-SD means CD is provided attribute weights by SD attributes; CD and SD scores are combined to give a single score.

1.4 Game Theory Approach: Individuals make strategic choices final outcome depend on what each person chooses to do viewed as a game. Game theory models seek to portray complex strategic situations in highly simplified setting. All games have three basic elements Players, Strategies activities, and Payoffs. Players make binding agreements in cooperative games not in non-cooperative games. Player is decision maker ability to choose among set of possible actions identity of play is the key. Strategy course of action available to individual access his confidential data simple or complex, non cooperative games each identity individual is uncertain about what the other will do. Fraudulent cannot reach agreements original identity individuals. Payoffs final detection rate of identity crime at the end of the transactional process measured in utility. Individual identity able to rank the payoffs from fraudulent intensity and density of breach the identity.

Methods proposed here to have assumed knowledge of link performance frequency distribution, information that is in many cases absent. In this paper, a two player non-cooperative game is envisaged between on the one hand the network user seeking a path to minimize the expected trip cost and on the other hand an "evil entity" choosing link performance scenarios to maximize the expected trip cost. Game theory is a study of strategic decision making.

Game theory is mainly used in economics, political science, and psychology, as well as logic and biology. The subject first addressed zero-sum games, such that one person's gains exactly equal net losses of the other participant. Today, however game theory applies to a wide range of class relations, and has developed into an umbrella term for the logical side of science, to include both human and non-humans, like computers. Classic users include a sense of balance in numerous games, where each person has found or developed a tactic that cannot successfully better his results, given the other approach. The normal game is usually represented by a matrix which shows the players, strategies, and pay-offs. More generally it can be represented by any function that associates a payoff for each player with every possible combination of actions.

Game theory is the systematic study of interdependent rational choice. It should be distinguished from decision theory, the systematic study of individual choice in parametric contexts. Decision theory has several applications to ethics. Game theory may be used to explain, to predict, and to evaluate human behavior in contexts where the outcome of action depends on what several agents choose to do and where their choices depend on what others choose to do. Game theory consequently is relevant to ethics, and it is used in moral and political in a variety of ways. Game theory was used to address one dimension of this resource management problem. The basic concept of three branches of game theory, leader-follower, cooperative and two-person nonzero sum games, are reviewed and applied to the study of the day-to-day pricing issue.

1.4.1 State of Equilibrium: Equilibrium occurred when all individuals had no incentive to change his or her behavior, when strategies are chosen. Equilibrium provides restriction for individuals to alter original identity and behavior. Equilibrium concept is to arrive the state of transparency maintained across different individual to access the credit application. Nash equilibrium pair of strategies in a two-player game a* is an optimal strategy for A against b*, b* is an optimal strategy for B against a*.Players cannot benefit from knowing equilibrium strategy of their opponents.

1.4.2. Dynamic Time Warping Algorithm:

Dynamic Time Warping algorithm is applied to minimize the time constraints on detecting fraudulent identity usage and reaction time.

Algorithm – Dynamic Time Wrapping

Step 1: To determining the Dynamic Time Wrapping distance dist (X , Y) between two time series X = ($x_1, \dots,$ x_M) and

Y = ($y_1, \dots,$ y_N), a matrix D € IR^{M*N}

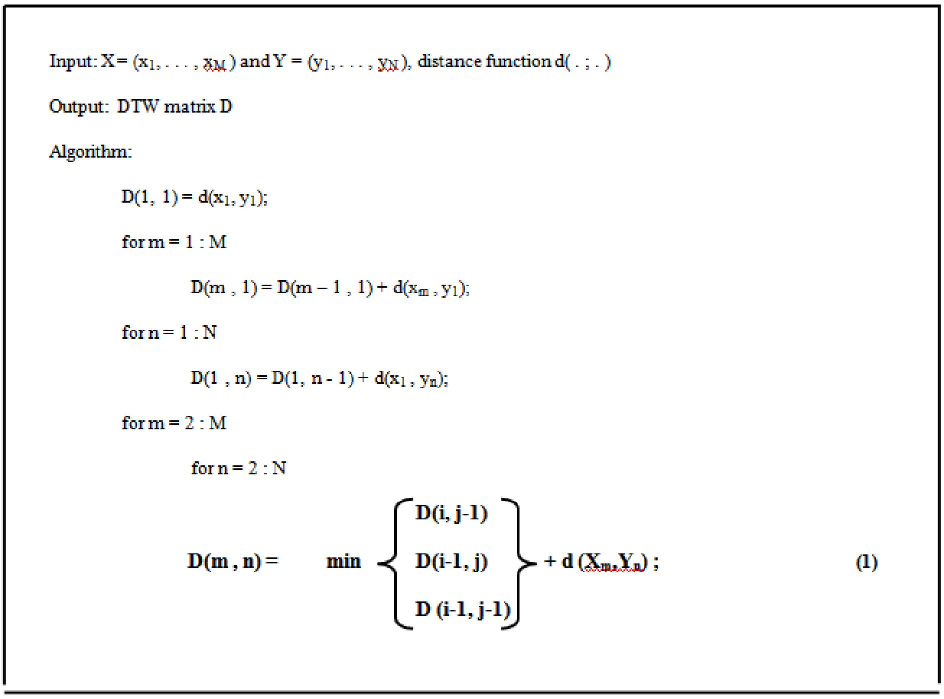
Step 2: Each entry D(i, j) is calculated from some D(i_0; j_0) plus an additional cost d, which is usually some distance between the samples x_i and y_i.

Step 3: warping path composed of index pairs ((i_1 , j_1), (i_2 ,i_2), .. ., (i_K, j_K)), which aligns corresponding samples in the input sequences X and Y.

Step 4: The values of D have been calculated, the warping path can be determined by backtracking the minimum cost path starting from (M, N).

Step 5: Dynamic Time warpping path is stored in D (m, n).

Pseudo code – Dynamic Time Wrapping



```
Input: X= (x_1, . . . , x_M) and Y = (y_1, . . . , y_N), distance function d( . ; . )
Output: DTW matrix D
Algorithm:
        D(1, 1) = d(x_1, y_1);
        for m = 1 : M
                D(m , 1) = D(m – 1 , 1) + d(x_m , y_1);
        for n = 1 : N
                D(1 , n) = D(1, n - 1) + d(x_1 , y_n);
        for m = 2 : M
                for n = 2 : N
                D(m , n) = min { D(i, j-1) ; D(i-1, j) ; D (i-1, j-1) } + d (X_m,Y_n) ;        (1)
```

III. Performance Results And Discussion

In this section, we give implemented results for decreasing crime rates in the credit cards. Identity Crime Detection scheme is identified with different performance metrics and compares it with other credit card frauds. The identity crime detection is used to detect the credit card frauds in the credit card users.

1.1 Performance metrics: Performance metrics for Identity crime detection with game theory approach are Frequency of Identity Variation, Adaptive time gap, Number of fraudulent identities, False positive, True positive, Number of users and Identity Verification time.

Table1: Identity variation Frequency of crime detection

Identities	**Existing crime Detection**	**Proposed crime detection**
20	100	120
40	150	200
60	200	250
80	300	370

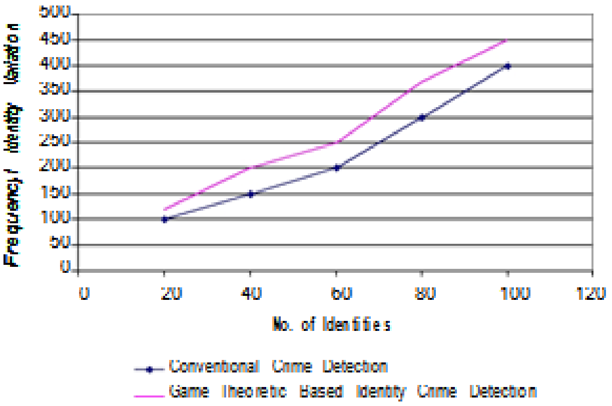


Fig 1: Identity variation frequency of crime detection

Fig 1 demonstrates the frequency identity variation. X axis represents the number of identities whereas Y axis denotes the frequency of identity variation using both the conventional crime detection method and with our proposed Game Theoretic Approach Based Identity Crime Detection. When the number of identities increased, frequency of identity variation also gets increases accordingly. The frequency of credit card frauds is illustrated using the existing conventional crime detection method and proposed Game Theoretic Approach Based Identity Crime Detection. Fig 1 shows better performance of Proposed Identity Crime Detection in terms of identities than existing crime detection and Proposed Identity Crime Detection. Identity Crime Detection achieves 2% to 5% less frequency of identity variation when compared with existing system.

Table 2: Fraudulent Identities of Credit Card Crime Detection

Identities	Existing Crime Detection	Proposed Identity Crime Detection
20	300	280
40	250	220
60	200	180
80	100	90
100	80	60

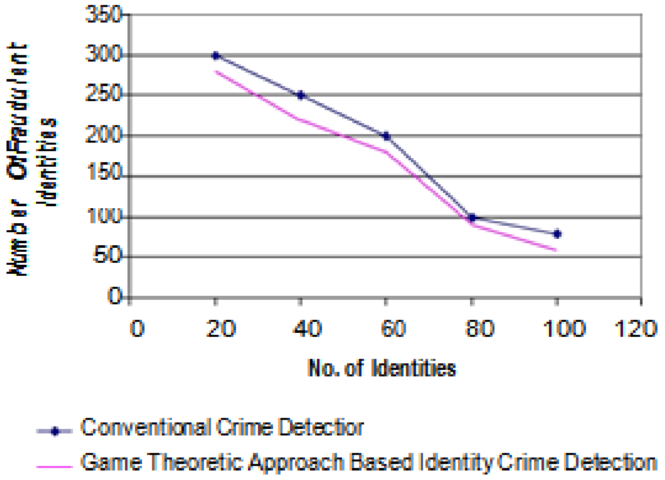


Fig 2: Fraudulent Identities of Credit Card Crime Detection

Fig 2 demonstrates the number of fraudulent identities. X axis represents number of identities whereas Y axis denotes the number of fraudulent identities using both the multilayered detection system and with our proposed Game theoretic approach based identity crime detection. When the number of identities decreased number of fraudulent identities also gets decreased. In the crime detection, minimum number of fraudulent identities in identity crime detection makes it easier to identify and maintain credit card frauds. All the curves show a more or less yet steady descendant when identity increases. Figure 2 shows better number of fraudulent identities performance of Game theoretic approach based Identity Crime Detection than existing Crime Detections. Identity Crime Detection achieves 5% to 15% less fraudulent identity result.

Table 3: Verification Time of Identities

Identities	Existing Crime Detection	Proposed Identity Crime Detection
20	80	100
40	100	120
60	140	200
80	230	250
100	300	340

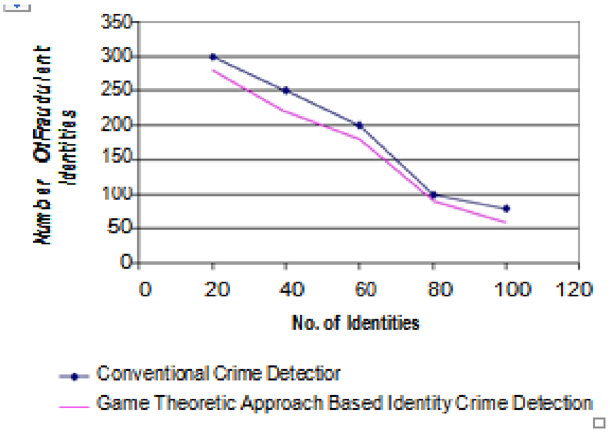


Fig 3: Variation Time of Identity

Fig 3 demonstrates the identity variation time. X axis represents number of identities whereas Y axis denotes the identity variation time using both the multilayered detection system and with our proposed Game theoretic approach based identity crime detection. When the number of identities increased identity variation time also gets increased. Figure 3 shows better performance of identifying Frauds in terms of latency than existing Crime Detection and Identity Crime Detection. Identity Crime Detection achieves 10% to 25% more identity variation time when compared with existing schemes.

IV. CONCLUSION

The main focus of this paper is Resilient Identity Crime Detection, in other words, the real-time search for patterns in a multilayered and principled fashion, to safeguard credit applications at the first stage of the credit life cycle. This paper describes an important domain that has many problems relevant to other data mining research. It has documented the development and evaluation in the data mining layers of defense for a real-time credit application fraud detection system. In doing so, this research produced three concepts which dramatically increase the detection system's effectiveness. These concepts are resilience, adaptively, and quality data. These concepts are fundamental to the design, implementation, and evaluation of all fraud detection, adversarial-related detection, and identity crime-related detection systems.

REFERENCES

[1] Clifton Phua, Member, IEEE, Kate Smith Miles, Senior Member IEEE, Vincent Cheng- Siong Lee, and Ross Gayler"ResilientIdentity Crime Detection", Ieee Transaction On Knowledge And Data ENGINEERING*(references)*A. Bifet and R. Kirkby "Massive Online Analysis, Technical Manual, Univ. of Waikato, 2009.

[2] R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, vol. 17, no. 3, pp. 235-255, 2001.

[3] P. Brockett, R. Derrig, L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs," The J. Risk and Insurance, vol. 69, no. 3, pp. 341-371, 2002, doi: 10.1111/1539-6975.00027.

[4] R. Caruana and A. Niculescu-Mizil, "Data Mining in Metric Space: An empirical Analysis of Supervised

[5] Learning Performance Criteria," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge discovery and Data Mining (KDD '04), 2004, doi: 10.1145/1014052.1014063.

[6] Exper Detect: Application Fraud Prevention System, Whitepaper, http://www.experian.com/products/pdf/ experian_detect.pdf, 2008.

[7] T. Fawcett, "An Introduction to ROC Analysis," Pattern Recognition Letters, vol. 27, pp. 861-874, 2006, Doi: 10.1016/j .patrec. 2005.10.010.