

An Efficient Technique for Data Security and Privacy During Data Mining

Shaik Mohammed Shafiulla¹ K.Srilatha²

¹Assistant Professor, ²M.Tech Student,

^{1,2}Department of CSE, Dhruva Institute of Engineering & Technology, Hyd.

Abstract- Database mining will be outlined because the method of mining for implicit, erst unidentified, and doubtless essential info from awfully vast databases by economical knowledge discovery techniques. The privacy and security of user information became vital public policy anxieties and these anxieties square measure receiving multiplied interest by the each public and government leader and controller, privacy advocates, and the media. during this paper we have a tendency to focuses on key on-line privacy and security problems and issues, the role of self-regulation and also the user on privacy and security protections, knowledge protection laws, regulatory trends, and also the outlook for privacy and security legislation. Naturally such a method could open up new assumption dimensions, notice new invasion patterns, and raises new knowledge security issues. Recent developments in info technology have enabled assortment and process of monumental amount of non-public knowledge, like criminal records, online shopping habits, on-line banking, credit and medical record, and driving records and nearly significantly the Govt concerned knowledge.

Keywords- Database mining, Database security, Data Privacy, Inferences, Intrusion Detection, Law. Worm detection, Masquerade detection, Anomaly detection.

I. INTRODUCTION

Security and Privacy protection are a public policy concern for many years. However, speedy technological changes, the rapid growth of the web and electronic commerce, and the development of additional refined strategies of aggregation, analyzing, and using personal data have created privacy a major public and government problems. the sector of information mining is gaining significance recognition to the supply of huge amounts of information, simply collected and hold on via laptop systems. Recently, the massive quantity of information, gathered from various channels, contains abundant personal data. When personal and sensitive information ar printed and/or analyzed, one important question to require into consideration is whether or not the analysis violates the privacy of people whose information is noted. The importance of data that may be wont to increase revenue cuts prices or each. data processing package is one

among variety of analytical tools for analyzing information. It permits users to research data privacy is growing perpetually. For this reason, many research works have centered on privacy-preserving information mining, proposing novel techniques that enable extracting knowledge whereas attempting to shield the privacy of users. Some of these approaches aim at individual privacy whereas others aim at corporate privacy. Data mining, popularly referred to as information Discovery in Databases (KDD), it's the nontrivial extraction of implicit, previously unknown and doubtless helpful data from data in databases. Data discovery is required to form sense and use of information. Though, data processing and information discovery in databases (or KDD) ar oftentimes treated as synonyms, data processing is truly a part of the information discovery method. [1,2,3]

II. DATA MINING AND THEIR FUNCTIONALITIES

Data mining is that the method of discovering patterns in giant information sets involving ways at the intersection of machine learning, statistics, and information systems.[1] data processing is associate degree knowledge base sub field of engineering with associate degree overall goal to extract data (with intelligent method) from a knowledge set and remodel the data into an obvious structure for any use. Data mining is that the analysis step of the "knowledge discovery in databases" method, or KDD. Aside from the raw analysis step, it conjointly involves information and information management aspects, information pre-processing, model and logical thinking concerns, power metrics, complexness concerns, post-processing of discovered structures, image, and on-line change.

The term "data mining" is in reality a name, as a result of the goal is that the extraction of patterns and information from giant amounts of information, not the extraction (mining) of information itself. It is also a buzzword[7] and is usually applied to any variety of large-scale information or information science (collection, extraction, deposit, analysis, and statistics) also as any application of laptop call network, together with computing (e.g., machine learning) and business intelligence. The book information mining: sensible machine learning tools and techniques with Java[8] (which covers largely machine learning material) was originally to be named

simply sensible machine learning, and also the term data processing was solely other for promoting reasons.[9] typically the a lot of general terms (large scale) information analysis and analytics – or, once relating actual ways, computing and machine learning – area unit a lot of applicable.

The actual data processing task is that the semi-automatic or automatic analysis of enormous quantities of information to extract antecedently unknown, fascinating patterns like teams of information records (cluster analysis), uncommon records (anomaly detection), and dependencies (association rule mining, ordered pattern mining). This sometimes involves exploitation information techniques like spacial indices. These patterns will then be seen as a form of outline of the input file, and should be employed in any analysis or, as an example, in machine learning and prophetic analytics. as an example, the data mining step would possibly determine multiple teams within the data, which might then be accustomed acquire a lot of correct prediction results by a choice network. Neither the info assortment, information preparation, nor result interpretation and news is a component of the info mining step, however do belong to the general KDD method as extra steps.

A. DATA MINING FUNCTIONALITIES

The data mining functionalities is used to specify different kinds of patterns extracted during data mining. Data mining are classified in descriptive and predictive tasks. Below are the functionalities in data mining

- Characterization and Discrimination
- Mining frequent patterns, Associations and Correlations
- Classification and prediction
- Cluster analysis
- Outlier analysis

III. DATA MINING FOR SECURITY

In recent years, intrusion detection technologies are indispensable for network/computer security as the threat of cyber terrorist act becomes a significant matter year by year. Most of typical technologies such as IDS (Intrusion Detection Systems) take a signature-based approach to that, during which variety of human-made rules describing pc worm/virus of known patterns are constructed and an alarm is made once a record matching one in every of the principles appears. The signature-based approach suffers from the following 2 important issues: 1) it's not capable to sight worms/virus of unknown varieties, and 2) it needs loads of computation time for signature matching.

Meanwhile, we tend to might use a policy-based approach so as to sight unknown pc worms/ virus. In it a general security policy is built and associate degree alarm is formed once some record violates the policy. However, it's not capable to sight new pc worm/virus just in case they eventually satisfy the

policy. Data mining-based anomaly detection could be a quite technology for police work pc worms/virus of un- known patterns additional adaptively and effectively than signature-based and policy-based ones. This is to learn applied mathematics regularities from past examples and to sight worms/virus as anomalies that square measure for the most part deviated from the learned regularities.

A. Intrusion knowledge generation

The Smartsifter will detect outliers and it will generate rules which are used to explain patterns. We can call that rule as Outlier filtering. The rule has been written as follows

If “src_byte<9.688 & flag=SF” then normal

else if “service=http” then outlier

else normal

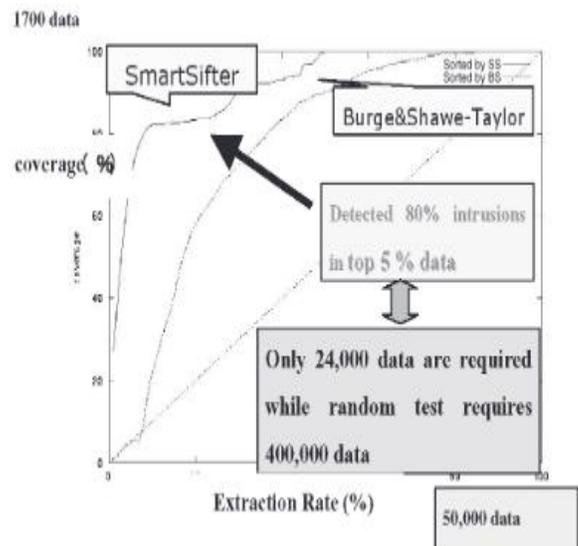


Fig.1: Intrusion detection using SmartSifter

B. Anomalous behavior detection engine

In previous sections we have a tendency to were involved with the issue of however abnormal a personal datum is. In other words, SmartSifter and ChangeFinder were designed to discover native anomalies in an exceedingly information set. However, there square measure some things wherever it's needed to detect international dynamics of anomalies, like anomalous behavior patterns, in an exceedingly set of your time series. Access-Tracer is meant to discover such a sort of anomalies. For example, it will be applied to police investigation masqueraders' behavior patterns from UNIX system command

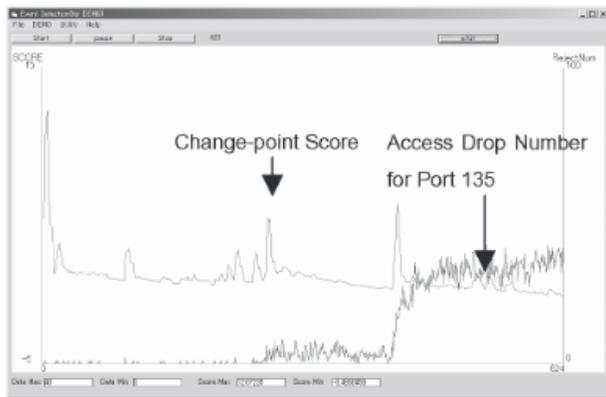


Fig.2: Anomalous detection using Change Finder.

IV. CONCLUSION

Data mining has become one amongst the key options of the many homeland security initiatives. usually used as a method for detecting fraud, assessing risk, and merchandise selling, data mining involves the employment of information analysis tools to get previously unknown, valid patterns and relationships in massive data sets. within the context of Homeland Security, data processing will be a possible means that to spot terrorist activities, such as money transfers and communications, and to spot and track individual terrorists themselves, like through travel and immigration records. whereas data processing represents a major advance within the sort of analytical tools presently offered, there are limitations to its capability.

One limitation is that though data mining will facilitate reveal patterns and relationships, it does not tell the user the worth or significance of those patterns. These varieties of determinations should be created by the user. A second limitation is that whereas data processing will establish connections between behaviors and/or variables, it doesn't necessarily establish a causative relationship. victorious knowledge mining still needs masterly technical and analytical specialists who will structure the analysis and interpret the output. Data mining is changing into more and more common in each the non-public and public sectors. Industries like banking, insurance, medicine, and selling normally use data processing to scale back costs, enhance analysis, and increase sales. within the public sector, data mining applications ab initio were used as a method to sight fraud and waste, however have mature to even be used for functions such as activity and rising program performance.

V. REFERENCES

- [1]. Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02- 5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.
- [2]. Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006
- [3]. Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.
- [4]. Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.
- [5]. L. Getoor, C. P. Diehl. "Link mining: a survey", ACM SIGKDD Explorations, vol. 7, pp. 3-12, 2005.
- [6]. Fayyad U.M., Piatetsky-Shapiro G., Smyth P. "From Data Mining to KDD : An Overview", AAAI/MIT Press, 1996.
- [7]. Han J.et Kamber M.,"Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Canada, 2002.
- [8]. Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999
- [9]. David Hand, Heikki Mannila, and Padhraic Smyth," Principles of Data Mining", MIT Press, Cambridge, MA, 2001.
- [10]. Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, and Alessandro Zanasi, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, NJ, 1998.
- [11]. Mafruz Zaman Ashrafi, David Taniar, Kate A. Smith, "Data Mining Architecture for Clustered Environments" , Proceeding PARA '02 Proceedings of the 6th International Conference on Applied Parallel Computing Advanced Scientific Computing, Pages 89-98, Springer- Verlag London, UK ©2002
- [12]. Clifton, C. and D. Marks, "Security and Privacy Implications of Data Mining", Proceedings of the ACM SIGMOD Conference Workshop on Research Issues in Data Mining and Knowledge Discovery, Montreal, June 1996