ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

A Review on Cryptography, Data Mining &Machine Learning for Cyber Security

Shikha paanwaar

MS Scholar, Dept. of M.S. Cyber law & Information Security, Barakatullah University Bhopal (M.P.), India

ABSTRACT: Cryptography is a concept to protect network and data transmission over wireless network. Data Security is the main aspect of secure data transmission over unreliable network. Network security involves the authorization of access to data in a network, which is controlled by the network administrator Users choose or are assigned an ID and password or other authenticating information that allows them access to information and programs within their authority Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. In data mining, association rules are created by analysing data for frequent if/then patterns, then using the support and confidence criteria to locate the most important relationships within the data. Support is how frequently the items appear in the database, while confidence is the number of times if/then statements are accurate. The Internet is a large computer network, or a chain of computers that are connected together. This connectivity allows individuals to connect to countless other computers to gather and transmit information, messages, and data. Unfortunately, this connectivity also allows criminals to communicate with other criminals and with their victims.

KEYWORDS: Cryptography, Data mining ,Machine

I. INTRODUCTION

Network Security is the most vital component in information security because it is responsible for securing all information passed through networked computers. Network Security refers to all hardware and software functions, characteristics, features, operational procedures, accountability, measures, access control, and administrative and management policy required to provide an acceptable level of protection for Hardware and Software , and information in a network. Network security problems can be divided roughly into four closely intertwined areas: secrecy, authentication, no repudiation, and integrity control. Secrecy, also called confidentiality, has to do with keeping information out of the hands of unauthorized users. This is what usually comes to mind when people think about network security. Authentication deals with determining whom you are talking to before revealing sensitive information or entering into a business deal. Non repudiation deals with signatures. Cyber security is the set of technologies and processes designed to protect computers, networks, programs, and data from attack, unauthorized access, change, or destruction. Cyber security systems are composed of network security systems and computer (host) security systems. Each of these has, at a minimum, a firewall, antivirus software, and an intrusion detection system (IDS). IDSs help discover, determine, and identify unauthorized use, duplication, alteration, and destruction of information systems [1]. The security breaches include external intrusions (attacks from outside the organization) and internal intrusions (attacks from within the organization).There is three main types of cyber analytics in support of IDSs: misuse-based (sometimes also called signature based), anomaly-based, and hybrid. Misuse-based techniques are designed to detect known attacks by using signatures of those attacks. Message Integrity: Even if the sender and receiver are able to authenticate each other, they also want to insure that the content of their communication is not altered, either maliciously or by accident, in transmission. Extensions to the check summing techniques that we encountered in reliable transport and data link protocols. Cryptography is an emerging technology, which is important for network security. The widespread use of computerised data storage, processing and transmission makes sensitive, valuable and personal information vulnerable to unauthorised access while in storage or transmission. Due to continuing advancements in communications and eavesdropping technologies, business organisations and private individuals are beginning to protect their information in computer systems and networks using cryptographic techniques, which, until very recently, were exclusively used by the military and diplomatic communities.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Cryptography is a vital oftoday's computer and communications networks, protecting everything from business e-mail to bank transactions and internet shopping While classical and modern cryptography employ various mathematical techniques to avoid eavesdroppers from learning the contents of encrypted messages. Computer systems and networks which are storing, processing and communicating sensitive or valuable information require protection against such unauthorised access[1].This survey paper focuses on ML and DM techniques for cyber security, with an emphasis on the ML/DM methods and their descriptions. Many papers describing these methods have been published, including several reviews. In contrast to previous reviews, the focus of our paper is on publications that meet certain criteria. Google Scholar queries were performed using "machine learning" and cyber, and using "data mining "and cyber. Special emphasis was placed on highly cited papers because these described popular techniques. However, it was also recognized that this emphasis might overlook significant new and emerging techniques, so some of these papers were chosen also. Overall, papers were selected so that each of the ML/DM categories listed later had at least one and preferably a few representative papers. This paper is intended for readers who wish to begin research in the field of DM for cyber intrusion detection.

II. MAJOR STEPS IN ML AND DM

There is a lot of confusion about the terms ML, DM, and Knowledge Discovery in Databases (KDD). KDD is a full process that deals with extracting useful, previously unknown information (i.e., knowledge) from data [2]. DM is a particular step in this process—the application of specific algorithms for extracting patterns from data. The additional steps in the KDD process (data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of DM) guarantee that useful knowledge is extracted from available data. However, there are many publications[e.g., Cross Industry Standard Process for Data Mining(CRISP-DM) [3] and industry participants who call the whole KDD process DM. In this paper, following Fayyad et al. [4],DM is used to describe a particular step in KDD that deals with application of specific algorithms for extracting patterns from data. There is a significant overlap between ML and DM. These two terms are commonly confused because they often employ the same methods and therefore overlap significantly. The pioneer of ML, Arthur Samuel, defined ML as a "field of study that gives computers the ability to learn without being explicitly programmed." ML focuses on classification and prediction, based on known properties previously learned from the training data. ML algorithms need a goal (problem formulation) from the domain (e.g., dependent variable to predict). DM focuses on the discovery of previously unknown properties in the data .It does not need a specific goal from the domain, but instead focuses on finding new and interesting knowledge. In reality, for most ML methods, there should be three phases, not two: training, validation, and testing. ML and DM methods often have parameters such as the number of layers and nodes for an ANN. After the training is complete, there are usually several models (e.g., ANNs) available. To decide which one to use and have a good estimation of the error it will achieve on a test set, there should be a third separate data set, the validation data set[5]. The model that performs the best on the validation data should be the model used, and should not be fine-tuned depending on its accuracy on the test data set. Otherwise, the accuracy reported is optimistic and might not reflect the accuracy hat would be obtained on another test set similar to but slightly different from the existing test set. For a multi-class problem (classification into more than two classes), usually the following metrics are used:

•Overall accuracy: Exemplars classified correctly, all exemplars.
•Class detection rate: Exemplars from a given class classified correctly, all exemplars from a given class.
•Class FAR or class FP rate: Exemplars from a given class classified incorrectly, all exemplars not from a given class.[6]

A. AIMS AND OBJECTIVES

- To determine the impact of cybercrime on networks.
- To determine the advent of cyber-crime.
- To determine the pros and corn of network security.
- To determine how network security reduces the treat of cyber-crimes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Understanding the nature and function of cyber-crimes and network security; the qualitative descriptive mechanism isthe most ideal means of collecting and analysing data due to the flexibility, adaptiveness, and immediacy of the topic.This brings inherent biases, but another characteristic of such research is to identify and monitor these biases, thus including their influence on data collection and analysis rather than trying to eliminate them. Finally, data analysis in an interpretive qualitative research design is an inductive process. Data are richly descriptive and contribute significantly to this research.

B. CRYTOGRAPHIC PRINCIPLES

A. Redundancy Cryptographic principle 1: The first principle is that all encrypted messages must contain some redundancy, that is, information not needed to understand the message. Messages must contain some redundancy.
B. Freshness Cryptographic principle 2: Some method is needed to foil replay attacks. One such measure is including in every message a timestamp valid only for, say, 10 seconds. The receiver can then just keep messages around for 10 seconds, to compare newly arrived messages to previous ones to filter out duplicates. Messages older than 10 seconds can be thrown out, since any replays sent more than 10 seconds later will be rejected as too old[7].

C. CRYPTOSYSTEM TYPES

In general cryptosystems are taxonomies into two classes, symmetric or asymmetric, depending only on whether the keys at the transmitter and receiver are easily computed from each other. In asymmetric cryptography algorithm a different key is used for encryption and decryption. In the symmetric encryption, Alice and Bob can share the same key (K), which is unknown to the attacker, and uses it to encrypt and decrypt their communications channel[8].
A. Asymmetric cryptosystems there are practical problems associated with the generation, distribution and protection of a large number of keys. A solution to this key-distribution problem was suggested by Diffie and Hellman in 1976 [10]. A type of cipher was proposed which uses two different keys: one key used for enciphering can be made public, while the other, used for deciphering, is kept secret. The two keys are generated such that it is computationally infeasible to find the secret key from the public key. If user A wants to communicate with user B, A can use B's public key (from a public directory) to encipher the data. Only B can decipher the cipher text since he alone possesses the secret deciphering key. The scheme described above is called a public-key cryptosystem or an asymmetric cryptosystem[11]. If asymmetric algorithms satisfy certain restrictions, they can also be used for generating so-called digital signatures[12]. B. Symmetric cryptosystems In symmetric cryptosystems (also called conventional, secret-key or one-key cryptosystems), the enciphering and deciphering keys are either identical or simply related, i.e. 684 IEE PROCEEDINGS, Vol. 131, Pt. F, No. 7, DECEMBER 1984 one of them can be easily derived from the other. Both keys must be kept secret, and if either is compromised further secure communication is impossible. Keys need to be exchanged between users, often over a slow secure channel, for example a private courier, and the number of keys can be very large, if every pair of users requires a different key, even for a moderate number of users, i.e. n(n — l)/2 for n users. This creates a key-distribution problem which is partially solved in the asymmetric systems. Examples of symmetric systems are the data encryption standard (DES) [13] and rotor ciphers.

D. ATTACKS ON INFORMATION: WHAT ARE THETHREATS?

Not forgetting that the latter are always a combination of tools that have to do with technology and human resources(policies, training). Attacks can serve several purposes including fraud, extortion, data theft, revenge or simply the challenge of penetrating a system. This can be done by internal employees who abuse their access permissions, or by external attackers to remotely access or intercept network traffic .Another common attack on a computer system is the creation and distribution of malicious computer code, called “viruses”. Computer viruses are computer programs written specifically to damage other computer systems. Sometimes these malicious programs are contained within another program, known as a “Trojan horse,” and are copied by a user without his or her knowledge (Richards, Pp. 21-54).
After analysing the results of the study through qualitative analysis it can be said that computers and the Internet are now a familiar part of our lives. You may not see them often, but they are involved in some way in most of our daily activities in the business, educational institutions, and government. Without the support of any of these tools, we would be able to handle the overwhelming amount of information that seems to characterize our society. But the problem of security limits the integrity of information and computer systems. Cyber warfare has been defined as the process of nation-state to introduce computers of other countries or networks to cause damage or destruction. Cyber warfare is a

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

form of warfare that occurs on computers and the Internet, by electronic means rather than physical. Moreover, the Internets a means of easy access, where any person, remaining anonymous, can proceed with an attack that is difficult to associate, virtually undetectable and difficult to smuggle, let alone reaching a high impact such action directly hitting the opponent (network) by surprise. The term network security refers to protection against attacks and intrusions incorporate resources by intruders who are not allowed access to these resources[14].

Benefits of network Security
1.Prevents unauthorized users from accessing your network.
2.Provides transparent access to Internet-enabled users.
3.Ensures that sensitive data is transferred safely by the public network.
4.Help your managers to find and fix security problems.
5.Provides a comprehensive system of warning alarms attempt to access your network[15].

III. CONCLUSION

Network Security is the most vital component in information security because it is responsible for securing all information passed through networked computers. Network security consists of the provisions made in an underlying computer network infrastructure, policies adopted by the network administrator to protect the network and the network-accessible resources from unauthorized access, and consistent and continuous monitoring and measurement of its effectiveness (or lack) combined together. With advances in technology, no one is safe from an attack by "hackers.
Currently it is relatively easy to gain control of a machine on the Internet that has not been adequately protected. Companies invest a significant portion of their money in protecting their information, since the loss of irreplaceable data is a real threat to their business. The methods that are the most effective for cyber applications have not been established; and given the richness and complexity of the methods, it is impossible to make one recommendation for each method, based on the type of attack the system is supposed to detect. When determining the effectiveness of the methods, there is not one criterion but several criteria that need to be taken into account. They include(as described in Section VI, Subsection C) accuracy, complexity, time for classifying an unknown instance with a trained model, and understand ability of the final solution (classification)of each ML or DM method. Depending on the particularise, some might be more important than others.

REFERENCES

[1] A. Mukkamala, A. Sung, and A. Abraham, “Cyber security challenges: Designing efficient intrusion detection systems and antivirus tools,” in Enhancing Computer Security with Smart Technology, V. R. Vemuri,Ed. New York, NY, USA: Auerbach, 2005, pp. 125–163.
[2] M. Bhuyan, D. Bhattacharyya, and J. Kalita, “Network anomaly detection: Methods, systems and tools,” IEEE Commun. Surv.Tuts., vol. 16, no. 1, pp. 303–336, First Quart. 2014.
[3] T. T. T. Nguyen and G. Armitage, “A survey of techniques for internet traffic classification using machine learning,” IEEE Commun. Surv.Tuts., vol. 10, no. 4, pp. 56–76, Fourth Quart. 2008.
[4] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges,” Comput. Secur., vol. 28, no. 1, pp. 18–28, 2009.
[5] A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, and B. Stiller, “An overview of IP flow-based intrusion detection,” IEEE Commun.Surv.Tuts., vol. 12, no. 3, pp. 343–356, Third Quart. 2010.
[6] DENNING, D., and DENNING, P.J.: 'Data security', ACM Comput. Surveys, 1979, 11, pp. 227-250
[7] A Role-Based Trusted Network Provides Pervasive Security and Compliance - interview with Jayshree Ullal, senior VP of Cisco.
[8] Dave Dittrich, Network monitoring/Intrusion Detection Systems (IDS), University of Washington.
[9] 'Data encyption standard', FIPS PUB 46, National Bureau of Standards,Washington, DC Jan. 1977
[10] Murat Fiskiran , Ruby B. Lee, ―Workload Characterization of Elliptic Curve Cryptography and other Network Security Algorithms for Constrained Environments‖, IEEE International Workshop on Workload Characterization, 2002. WWC-5. 2002.
[11] Casey, E. Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet. London: Academic Press, 2011: Pp. 5-19.
[12] Farmer, Dan. & Charles, Mann C. Surveillance nation.Technology Review; Vol. 106, No. 4, 2003: Pp. 46.
[13] Harrison, A. Privacy group critical of release of carnivore data. Computerworld; Vol. 34, No. 41, 2006:
Pp. 24
[14] Internet Tax Freedom Act of 1998: 112 Stat. 2681–2719. Retrieved from: (http://www.cbo.gov/doc.cfm?index=608&type=0). Accessed on : 29th January, 2012.
[15] Katz, Mira L. & Shapiro, Carl. Technology Adoption in thePresence of Network Externalities. Journal of PoliticalEconomy; Vol. 94, No.4, 1986: Pp. 822-841.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

[16] Ogut, Hulisi. Menon, Nirup. &Ragunathan, Srinivasia. CyberInsurance and IT Security Investment: Impact of IndependentRisk. Proceedings of the Workshop on the Economics ofInformation Security (WEIS), Cambridge, MA: HarvardUniversity, 2005: Pp. 14-28.
[17] Richards, James. Transnational Criminal Organizations,Cybercrime, and Money Laundering: A Handbook for LawEnforcement Officers, Auditors, and Financial Investigators.Boca Raton, FL: CRC Press, 1999: Pp. 21-54.
[18] Roland, Sarah E. The Uniform Electronic Signatures inGlobal and National Commerce Act: Removing Barriers toE-Commerce or Just Replacing Them with Privacy andSecurity Issues?. Suffolk University Law Review; Vol. 35,2001: Pp. 638-45.