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 2, February 2017

An Improved Shadow Honeypot with PCA Algorithm to Enhance Error handling on the Network

Shashank Singh Baghel[1], Vishal Singh[2], Manjit Jaiswal[3]

Student, Dept. of C.S.E, Institute of Technology Guru Ghasidas University Bilaspur, Chhattisgarh, India[1, 2]

Assistant Professor, Dept. of C.S.E, Institute of Technology Guru Ghasidas University Bilaspur, Chhattisgarh, India[3]

ABSTRACT: Nowadays the security issues of Network become more sharp and urgent, in order to improve the initiative of Network security protection and the validity. This paper presents a new proactive security algorithm named honeypot using PCA algorithm to expand the network topology space and confuse the attacker, Network is being confronted currently and the common attack tools, methods and rules, so as to amend the network security architecture according to specific situations, to revised security management principles of all levels, to adjust the firewall configuration to enhance the holistic security of Network.

Most current anomaly Intrusion Detection Systems (IDSs) detect computer network behaviour as normal or abnormal but cannot identify the type of attacks. Moreover, most current intrusion detection methods cannot process large amounts of audit data for real-time operation. In this paper, we propose a novel method for intrusion identification in computer networks based on Principal Component Analysis (PCA). PCA is employed to reduce the dimensionality of the data vectors and identification is handled in a low dimensional space with high efficiency and low use of system resources. The normal behaviour is profiled based on normal data for anomaly detection and models of each type of attack are built based on attack data for intrusion identification. Employment of PCA lowers the possibility of false alarm generation with better detection of false alarm. It lowers the unreliability of high-interaction production honeypot by two tier surveillance system. Using short basic level unreliability can be detected while PCA hold off the experience hacker by applying the concept of outlines

KEYWORDS:Honeypot; PCA; virtual topolog ; attack; snort

I. INTRODUCTION

With the ease of communicating with the world through Internet, came the threats that causes unexpected harm and damage to our security networks. To detect the black hat society it is necessary to keep up-to-date with the hackers innovations. Various security defence systems were introduced for the improvement of network security but could not detect attacks inside an organization network [1]. Also, in spite of the advances in technology, it does not recognizes the new attacks. But for making defensive as well as offensive strategies against such malicious attempts, we should be aware about the ever evolving hacker techniques and strategies. Honeypot systems are considered to be best machines for this purpose. With time honeypot systems have evolved. Principal Component Analysis has the capacity not only to make it compact but also to make it accurate. Algorithm will provide the dual features of features of intrusion detection system embedded in honeypot. PCA algorithm will not only cover the basic features of honeypot but also the useful features of intrusion detection system and anomaly detection system. It aims to increase the accuracy of the system in false alarm generations.

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 2, February 2017

II. HONEYPOT TECHNOLOGY

In computer terminology, a honeypot is a computer security mechanism set to detect, deflect, or, in some manner, counteract attempts at unauthorized use of information systems. Generally, a honeypot consists of data (for example, in a network site) that appears to be a legitimate part of the site but is actually isolated and monitored, and that seems to contain information or a resource of value to attackers, which are then blocked. This is similar to the police baiting a criminal and then conducting undercover surveillance, and finally punishing the criminal. [12] Types- Honeypots can be classified based on their deployment (use/action) and based on their level of involvement. Based on deployment, honeypots may be classified as

1. Production honeypots.
2. Research honeypots.

III. PRODUCTION HONEYPOTS

Are easy to use, capture only limited information, and are used primarily by companies or corporations. Production honeypots are placed inside the production network with other production servers by an organization to improve their overall state of security. Normally, production honeypots are low-interaction honeypots, which are easier to deploy. They give less information about the attacks or attackers than research honeypots. Research honeypots are run to gather information about the motives and tactics of the Black hat community targeting different networks. These honeypots do not add direct value to a specific organization; instead, they are used to research the threats that organizations face and to learn how to better protect against those threats[13]

IV. RESEARCH HONEYPOTS

Research honeypots are complex to deploy and maintain, capture extensive information, and are used primarily by research, military, or government organizations.

Based on design criteria, honeypots can be classified as:

1. Pure honeypots.
2. High-interaction honeypots.
3. Low-interaction honeypots.

Pure honeypots-are full-fledged production systems. The activities of the attacker are monitored by using a casual tap that has been installed on the honeypot's link to the network. No other software needs to be installed. Even though a pure honeypot is useful, stealthiest of the defence mechanisms can be ensured by a more controlled mechanism.

High-interaction honeypots- imitate the activities of the production systems that host a variety of services and, therefore, an attacker may be allowed a lot of services to waste his time. By employing virtual machines, multiple honeypots can be hosted on a single physical machine. Therefore, even if the honeypot is compromised, it can be restored more quickly. In general, high-interaction honeypots provide more security by being difficult to detect, but they are expensive to maintain. If virtual machines are not available, one physical computer must be maintained for each honeypot, which can be exorbitantly expensive. Example: Honeynet.

Low-interaction honeypots- simulate only the services frequently requested by attackers. Since they consume relatively few resources, multiple virtual machines can easily be hosted on one physical system, the virtual systems have a short response time, and less code is required, reducing the complexity of the virtual system's security. Example: Honeyd.

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 2, February 2017

Principal component analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables. [14]

PCA was invented in 1901 by Karl Pearson,[15] as an analogue of the principal axis theorem in mechanics; it was later independently developed (and named) by Harold Hotelling in the 1930s.[16] Depending on the field of application, it is also named the discrete Kosambi-Karhunen–Loève transform (KLT) in signal processing, the Hotelling transform in multivariate quality control, proper orthogonal decomposition (POD) in mechanical engineering, singular value decomposition (SVD) of X (Golub and Van Loan, 1983), eigenvalue decomposition (EVD) of XTX in linear algebra, factor analysis (for a discussion of the differences between PCA and factor analysis see Ch. 7 of[17]), Eckart–Young theorem (Harman, 1960), or Schmidt–Mirsky theorem in psychometrics, empirical orthogonal functions (EOF) in meteorological science, empirical eigenfunction decomposition (Sirovich, 1987), empirical component analysis (Lorenz, 1956), quasiharmonic modes (Brooks et al., 1988), spectral decomposition in noise and vibration, and empirical modal analysis in structural dynamics.

PCA is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute.[18] The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the component score).[19] PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a projection or "shadow" of this object when viewed from its (in some sense; see below) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

PCA is closely related to factor analysis. Factor analysis typically incorporates more domain specific assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.
PCA is also related to canonical correlation analysis (CCA). CCA defines coordinate systems that optimally describe the cross-covariance between two datasets while PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset. [20][21]

V. RELATED WORK

In [1] authors,developed a honeynet for trapping the attackers by analyzing their attacking techniques and a centralized repository is deployed where all logs are send and analyzed for better understanding of their techniques. In [2] authors explain as no productive components are running on the system, Honeypots have the big advantage of not generating false alerts as each observed traffic is doubtful. This fact enables the system to log every byte that flows through the network through and from the honeypot, and to relate this data with other sources to draw a picture of an attack and the attacker.

Here authors used the properties of intrusion detection system and anomaly detection system which are prebuilt in principal component analysis if used properly.Tests are performed under the lab conditions where virtual server is created and honeypot is deployed on it.With some ports intentionally open to interact without security check over the internet to allure the hackers to attack on the system.before going over the network ,the system based on principal

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 2, February 2017

component analysis was trained and fed with large number of authentic IP addresses and packets to train the system to form discrete sets of authentic rules and behavior based clusters.When this was deployed with an intrusion detection system and anomaly detection system ,it increases its accuracy.

VI. PSEUDO CODE

Step1: Initialize state =0
Step 2- Repeat for i =0 until n. packet length do
Do while g (state, a ;) =fail.
Step 3- Then store state <- f (state)
State<-g (state a ;)
Step 4- If output (state) =! Empty
Then repeat
Step 5- Else print i
Step 6- Read in tp addresses and computer gradients of ip addresses
Step 7- Call get addresses to read in training addresses and return array of gradient vector
Trn_ip=set of all vectors of n, y gradients of training addresses
Ip_set=set of all vectors of training addresses
Transpose ip_set;
Step 8- PCA decomposition of ip addresses gradients
Transpose trn_ip;
Fmatrix = trn_ip;
Step 9- Computer mean gradient vector of the set of ip address gradients
Fmean=mean (Fmatrix*2)
Step 10- Subtract the mean gradiant vector from each gradient vector
For every gradient vector in fmatrixFmatrix=fmatrix-fmean End.
Step 11- Perform singular value decomposition (SVD) on the transformed gradient of principal component as an input to the SVD function
(VSV)=SVD (fmatrix# of principal components)
Step 12- Prampt user to provide test ip address to find matches
(Filename, user_canceled) ipfile
Step 13- Read in specified test ip address
Test ip = read ip filename;
Step 14- Get gradient of ip address.
Grad_ip=gradient (test ip);
Step 15- Reshape ip address into column vector
Grad_ip=reshape (grad_ip);
Test_ip=grad_ip;
Step 16- Project test ip address gradient onto the PCA basis and compute its coefficient
Transpose test_ip;
Step17- Subtract mean gradient vector from each gradient vector
Test_ip=tst_ip_fmean;
Step18- Find the coefficient for each test ip address by multiplying the test ip address transpose in the
PCA decomposition of the training ip address gradient.
For each test ip address
Project test ip address=transpose (u)*test_ip;
End for;
Step19- Define the training ip address coefficients by multiplying the smatrix.
Step20- Singular value decomposition on the training ip address
Proj_tra_ip=S*transpose of V;
Step21- Compute the distance between the coefficient of the test ip address and each training ip address.
Step22- Retain training ip address associated with minimum data (deviation) for each test ip address.

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 2, February 2017

If for each test ip address
For each training ip address;
Delta = distance between projtstip&projtrnip
End for
Matching ip = training ip address with minimum delta
End for
Else
Outserip address
Step23- Show resultant ip address
Step24- Show output ip address and corresponding test ip address.
Display training ip address attempting to find match for;

VII. SIMULATION RESULT

Table1: Attacks origin on a low interaction honeypot (Windows XP)

country	Attacks observed
United states	89
Germany	24
Korea	30
France	17
Canada	4
Italy	6
Belgium	2
Taiwan	1
Russia	1
India	1

Table 2:Attacks on a low interaction honeypot (Windows XP)

Service	Observed attacks
HTTP	138
FTP	29
POP3	7
Telnet	5
SMTP	4
Blaster	3
Sub-7	1

Table 3: Top 5 attempted username and passwords used for hacking

Username	Attempts	Percent	Password	Attempts	Percent
Root	2889	13.67	Username	10983	46.69
Admin	462	1.83	Username123	2479	8.85
Test	502	1.15	123456	2143	7.43
Guest	324	0.83	Password	498	2.65
Info	198	0.68	1234	298	1.67

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 2, February 2017

VIII. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed algorithm performs better than the factor analysis algorithm in terms of totalnumberof false alarm generation. The proposed algorithm provides independency to the shadow honeypot system as it can perform the work of intrusion detection system which makes the shadow honeypot system compact maximizes the lifetime of entire system.As the performance of the proposed algorithm is analyzed on the basis of negative false alarm generation and positive false alarm generation, in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other efficient algorithm. We have used a small virtual network system under laboratory conditions, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

REFERENCES

1.ParamjeetRawat, SakshiGoel, Megha Agarwal and Ruby Singh, " ECURING WMN USING HYBRID SYSTEM", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.3, May 2012YSTEM", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.3, May 2012

2.Srivathsa S Rao,VinayHegde , BoruthalupulaManeesh, Jyothi Prasad N M, SuhasSuresh,"Web Based Honeypot networks"International Journal of Scientific and Research Publications, Volume 3, Issue 8, August 2013 1 ISSN 2250-3153

3 Wei Wang, Roberto Battiti ,"Identifying Intrusions in Computer Networks with principal component analysisfComputer Science and Telecommunications University of Trento Via Sommarive 14, 38050 POVO, Trento, Italy

4 Ritu Tiwari, Abhishek Jain," Design and Analysis of Distributed Honeypot System ", Graphic Era University Dehradun, Uttrakhand India. (International Journal of Computer Applications (0975 – 8887) Volume 55– No.13, October 2012)

5 Wilhelm, Douglas (2010). "2". Professional Penetration Testing. Syngress Press. p. 503. ISBN 978-159749-425-0.

6 EC-Council. eccouncil.org

7. Moore, Robert (2005). Cybercrime: Investigating High Technology Computer Crime. Matthew Bender & Company. p. 258. ISBN 1-59345-303-5.Robert Moore

8. O'Brien, Marakas, James, George (2011). Management Information Systems. New York, NY: McGrawHill/ Irwin. pp. 536–537. ISBN 978-0-07-752217-9.

9. Moore, Robert (2006). Cybercrime: Investigating HighTechnology Computer Crime (1st Ed.). Cincinnati, Ohio: Anderson Publishing. ISBN 978-1-59345-303-9.

10.Thomas, Douglas (2002). Hacker Culture. University of Minnesota Press. ISBN 978-0-8166-3346-3.

11. Andress, Mandy; Cox, Phil; Tittel, Ed (2001). CIW Security Professional. New York, NY: Wiley. p. 638. ISBN 07645-4822-0.

12. "Blue hat hacker Definition". PC Magazine Encyclopedia. Retrieved May 31, 2010. A security professional invited by Microsoft to find vulnerabilities in Windows.

13. Fried, Ina (June 15, 2005). "Blue Hat summit meant to reveal ways of the other side". Microsoft meets the hackers. CNET News. Retrieved May 31, 2010.

14. Markoff, John (October 17, 2005). "At Microsoft, Interlopers Sound Off on Security". The New York Times. Retrieved May 31, 201

BIOGRAPHY

Manjit Jaiswal is an Assistant professor in the Computer Science and Engineering Department, Institute of Technology, Guru GhasidasUniversity. His research interests are Parallel computing, data structure, Algorithms etc.

Vishal Singh is a student in the Computer Science and Engineering Department, Institute of Technology, Guru Ghasidas University. His research interests are Fuzzy mathematics, data structure, Algorithms etc.

Shashank Singh Baghelis a student in the Computer Science and Engineering Department, Institute of Technology, Guru Ghasidas University. His research interests are logic designs, internet security, Algorithms etc.