IFERP
connecting engineers... developing research

SSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

Combined Approach for Masquerade Detection Using Behavior Profiling and Decoys

[1] Mayuri M More, [2] Mangesh R More
[1] Asst. Prof. Department of Computer Engineering RDTC SCSCOE, Bhor , Pune
[2] Asst. Prof. Department of Computer Engineering RDTC SCSCOE, Bhor , Pune

Abstract- Insider data theft attacks are caused by a masquerader stealing a real user's credentials and using them to mimic the authenticated user and to carry out malicious activities. Prior work focuses on user behavior profiling techniques and baiting techniques, but profiling user behavior using single modeling technique suffers from a considerable number of false positives. Also, decoys are stored at noticeable locations rather than using automatically generated decoys which may not give significant accuracy to the detection system. The proposed system will extend prior work and presents an inbuilt detection mechanism where behavior profiling will be done by the combination of more than one classifier, each using the different modeling technique to decrease false positive rate. Along with this, the system will include a baiting approach based on an automated generation of demand decoy documents on the user's file system and user authentication by challenge questions, to provide more accuracy. The proposed system could give a powerful protection mechanism against malicious insider data theft attacks.

Key Words — Insider data theft, Behavior profiling, Naïve Bayes Classifier, One Class Support Vector machines, decoys.

I. INTRODUCTION

Much research in security is focusing on solutions of preventing malicious insider data theft. Cloud Security Alliance had considered it as one of the top threats to cloud computing. Lots of mechanisms are proposed to secure users data by encryption and typical access control mechanisms. But it is observed that these methods were unsuccessful to protect damage caused by masquerade attacks. A major challenge of insider attack detection research is the lack of real data which is hard to acquire from a masquerader or traitor while performing their malicious actions. So it is necessary to design system which works on real time data to increase the efficacy of anomaly detection. The anomaly detector built using existing algorithms suffers from low accuracy and particularly from high false positive rates. To overcome this limitation, proposed system presents a strong detection mechanism where user behavior is profiled by combination of more than one classifying techniques. Along with this, decoy documents are generated to bait attacker. The proposed approach could improve accuracy over prior mechanisms and will help to provide the superior and intelligent level of security in terms of insider attacks

II. LITERATURE SURVEY

Stolfo proposed a combined approach for detecting masquerade attacks [1]. The author pays attention on modeling user search behavior with a baiting technique to reveal an attacker's malicious purpose. They hypothesized and showed that a masquerader would engage in search activities different from those of the legitimate user in terms of their volume and frequency.

Maxion improved upon Schonlau's result by applying the Naïve Bayes classification algorithm using the "bag of words" features [2]. Naïve Bayes has been used in text classification for a long time and proved to be very efficient in this context as well. Maxion presents a detailed analysis of the origins of the classification error, revealing why some users are good masquerades and others are not. Godoy stated the profiling strategies for user profiling. In addition the author discusses the existing approaches and lines of research in the main dimensions of user profiling such as acquisition learning adaptation and evaluation are discussed. The author has discussed in detail the success of personal agents in satisfying user information which intensely relies on the learning approach to acquire user profiles as well as the adaptation strategy to cope with changes in user interests. A system composed of honey pots and network-level sensors for traffic profiling was proposed by Maybury [4]. The sensors monitored insider activities such as network scanning and file downloads. Pre-specified models of insiders and pre- attack indicators were used to infer the malicious intent of an inclusive insider.

Dzeroski empirically evaluated several state-of-the art methods for constructing ensembles of heterogeneous classifiers with stacking and have shown that they perform comparably good to select the best classifier from the ensemble by cross validation [5]. They had proposed a new method for stacking which uses multi-response model trees at the meta-level. McCallurn used two common models used in Naïve Bayes Classifier, one is the multi-variant Bernoulli model, and the other is the multinomial model [6].

In the multivariate Bernoulli event model, a vector of binary attributes is used to represent a document, indicating whether the command occurs or doesn't occur in the document. The multinomial model uses the number of command occurrences to represent a document, which is called "bag-of-words" approach, capturing the word frequency information in documents. According to McCallurn's result, multi-variants Bernoulli model performs better for small vocabulary size, and the multinomial model usually performs better at larger vocabulary size.

Scholkopf proposed a method to adapt the SVM algorithm [7] for one-class SVM, which only use examples from one-class, instead of multiple classes, for training. The one-class SVM algorithm first maps input data into a high dimensional feature space via a kernel function and treats the origin as the only example from other classes. It then iteratively finds the maximal margin hyper plane that best separates the training data from the origin. Bowen concluded as masquerade attacks pose a grave security problem and detecting masqueraders is very hard [8]. In this paper, the author has investigated the use of such trap-based mechanisms for the detection of masquerade attacks. They evaluated the desirable properties of decoys deployed within a user's file space for detection. Existing algorithms used for modeling user behavior makes use of statistical features, such as the sequence of user commands or co-occurrence of multiple events combined through logical operators. The anomaly detectors built using these algorithms suffer from low accuracy and from high false positive rates. One way to overcome this limitation is to combine some base classifiers to create one ensemble of classifiers. If the real user baits the system with automatically generated decoy documents then a sophisticated masquerader can be trapped. So proposed system offers a good solution against this limitation where each classifier uses a different modeling algorithm to profile user behavior and decoy technology together.

III.PROPOSED SYSTEM

Figure 1 shows the Architecture of the insider data theft prevention system:

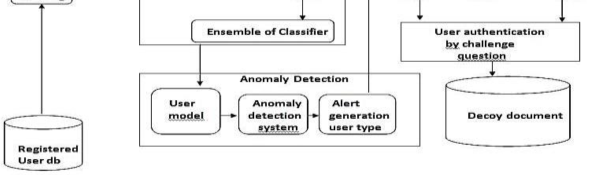


Figure 1: Architecture of the insider data theft prevention system

In the above system behavior profiling detects abnormal user behavior. Then it monitors for abnormal behaviors that show large deviations from the derived base. The system prepares a normal user model Nu that models the user's behavior by fetching distinguishing features and measures the difference between actual user behavior and the past user behavior as defined by the normal user model Nu. The distance Di is compared in order to determine whether there is enough proof for masquerade activity or not.

Insider data theft prevention system is implemented in the following modules:

Module 1: Validating user logins
The application is deployed to validate the system. User logins are the basic inputs for system which consists of necessary user details and at least three challenge questions at the time of account registration.

Module 2: User access behavior profiling
Abnormal behavior can differ from normal user behavior. Standard user model is prepared by selecting some of the distinguishing features such as speed of pressing keystrokes, mouse movements etc. The present work focuses to reduce feature set by selecting minimum distinguishing features among them. Also ensemble of classifier is prepared for reducing low accuracy of anomaly detection where most of the features are extracted using one-class support vector machine and remaining features are extracted using multinomial model of naïve bayes classifier.

Module 3: User authentication by challenge questions
If decoy documents are accessed by real user by mistake, even though authenticate user is treated as masquerader by the system. Hence it increases false positive rate. To overcome this limitation, a set of challenge questions are asked to the user whose answers are only known to the real user. This improves accuracy of overall detection system.

Module 4: Anomaly detection
Current user behavior is modeled by the system. If the current user behavior captured in feature vector v is similar enough with the user model u which captures the user's historical behavior, then the user behavior should be deemed normal. In other words, if the distance between the v and user model u is smaller than threshold, then no masquerader activity is suspected, and no alert gets generated. On the other hand, if feature vector u exhibits a high difference, then an alert is generated. The threshold is set to minimize the miss rate or false negative rate.

Module 5: Decoy document distribution

Whenever alert by anomaly detection is generated, decoy information may be served immediately on demand. In present work, decoy documents are supplied automatically on generation of first alert rather than keeping decoy traps so that adversary will not get any doubt of fake or worthless data is being served to him.

Module 6: Control system
The control system represents an account of administrator which monitors malicious insider accesses. Administrator blocks the masquerader by denying access to the system's real data. Administrator can view user details and uploading details. Admin generates different types of alerts such as suspicious, attacker and normal on the basis of results of behavior profiling and decoy distribution. Logs and records of insider data theft prevention system are stored.

IV.IMPLEMENTATION DETAILS

We provide a brief description of applying classifiers and decoys for masquerade detection in the following.

1) Behavior profiling using combined classifiers

Ensemble of classifier is created for reducing low accuracy of anomaly detection System. One class SVM and Naïve bayes classifiers are combined in this work.
Anomalous user behavior can vary from normal user behavior. According to this assumption standard user model is prepared by extracting most of the distinct features by using one class support vector machine. All features mentioned below are the key parameters to check whether the user behavior is normal or suspicious. If any of the features show variation from normal behavior, it is treated as indication of masquerade attack.

1. Attempts towards user login.
2. Timing of login.
3. The ip address of system from which user logs in.
4. Speed of pressing keys by logged user.
5. Habit of using mouse or key for submitting.
6. Attempts towards challenge questions.

A set of challenge questions is asked to the user at the time of registration. Answers are stored using bag-of-words approach of multinomial model of naïve bayes classifier, creating vocabulary of each user. To improve accuracy of overall detection system random questions are asked rather than asking same question. The multinomial model checks whether entered answer is present or not in real users vocabulary. Incorrect answer to the first randomly asked question will lead you to the next question. Count of failed attempts towards the question is considered as the indicative for suspicious activity.

2) Decoy technology and honypots

Honey pot is placed in the file system to attract the masquerader. Honeypots are information systems used to deceive adversaries and trick them into thinking that they are dealing with real and authentic assets. Real user is not supposed to touch the honeypot as he knows that these are the fake files so touch to the honeypot is considered as an alert of suspicious activity. Whenever alert by anomaly detection is generated, decoy information is supplied immediately. The system maintains same file structure for the decoy file system and the original file system rather than providing unrelated fake data in the decoy files for confusing the attacker. The information contained in the decoy file is supplied in such a way that it will appear completely original to the attacker and attacker will not get any doubt of fake data is being supplied to him.

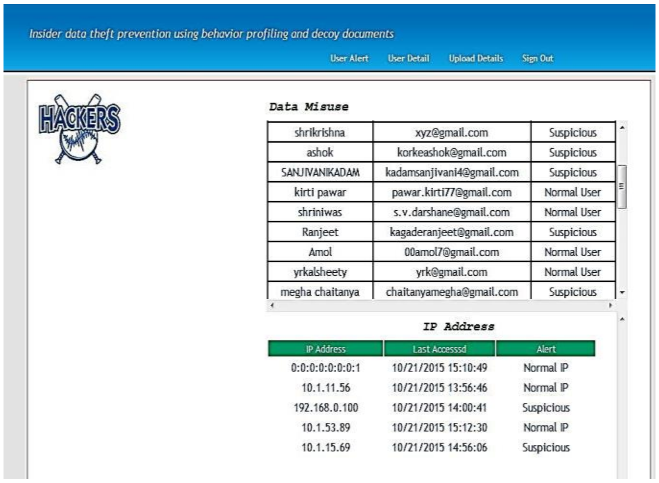
3) Using behavior profiling and decoy technology together

Profiling user behavior is a masquerade detection mechanism. But it may produce high FP rates, particularly if the user model is not created properly because of lack of training data. On the other hand, trap-based mechanisms such as decoys and honeypots are well known for their very low FP rates. So the combined approach can provide a strong detection mechanism which may improve overall detection accuracy. System asks random security questions rather than asking same question for user authentication. The current user behavior is compared with the standard behavior of that user, 'Normal' alert is generated if user behavior is same as standard behavior of that user. 'Attacker' alert is generated if user has acessed honeypots and decoys. After geration of attacker alert, administrator blocks that masquerader by denying access to the orignal data.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Our experimental results in a local file system show that combination of both detection techniques can perform well to gain better detection outcomes. We evaluate briefly the experimental results achieved by proposed system to prevent insider data theft attacks.

Figure 2.shows different alerts generated on admin account if user activity seems to be suspicious.


Insider data theft prevention using behavior profiling and decoy documents
User Alert
User Detail
Upload Details
Sign Out
HACKERS

Data Misuse

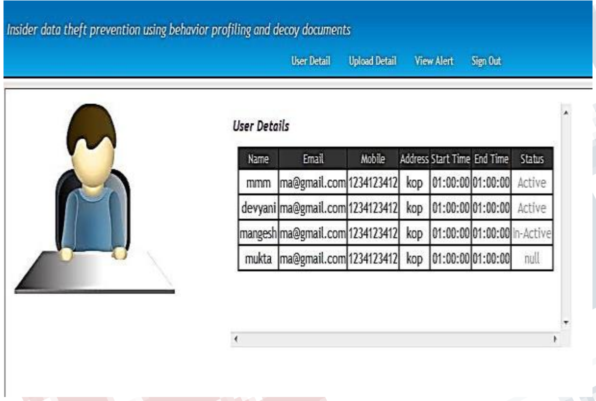
shrikrishna	xyz@gmail.com	Suspicious
ashok	korkeashok@gmail.com	Suspicious
SANJIVANIKADAM	kadamsanjivani4@gmail.com	Suspicious
kirti pawar	pawar.kirti77@gmail.com	Normal User
shriniwas	s.v.darshane@gmail.com	Normal User
Ranjeet	kagaderanjeet@gmail.com	Suspicious
Amol	00amol7@gmail.com	Normal User
yrkalsheety	yrk@gmail.com	Normal User
megha chaitanya	chaitanyamegha@gmail.com	Suspicious

IP Address

IP Address	Last Accesssd	Alert
0:0:0:0:0:0:0:1	10/21/2015 15:10:49	Normal IP
10.1.11.56	10/21/2015 13:56:46	Normal IP
192.168.0.100	10/21/2015 14:00:41	Suspicious
10.1.53.89	10/21/2015 15:12:30	Normal IP
10.1.15.69	10/21/2015 14:56:06	Suspicious

Figure 2: Alerts generated on admin account

System maintains logs of user activity which includes session timings along with status whether the user is active or inactive as shown in Figure 3.

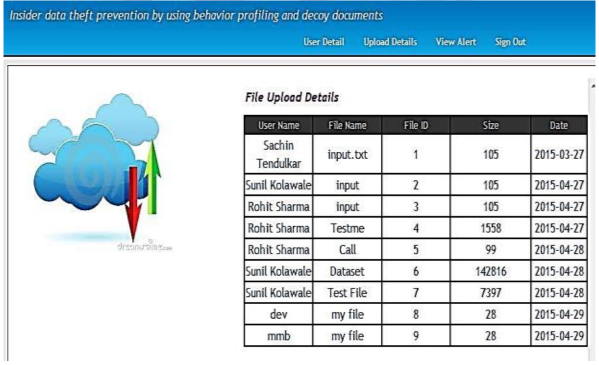

Insider data theft prevention using behavior profiling and decoy documents
User Detail
Upload Detail
View Alert
Sign Out

User Details

Name	Email	Mobile	Address	Start Time	End Time	Status
mmm	ma@gmail.com	1234123412	kop	01:00:00	01:00:00	Active
devyani	ma@gmail.com	1234123412	kop	01:00:00	01:00:00	Active
mangesh	ma@gmail.com	1234123412	kop	01:00:00	01:00:00	In-Active
mukta	ma@gmail.com	1234123412	kop	01:00:00	01:00:00	null

Figure 3: Logs maintained on admin account

Admin can view uploading details which consists of file ID, file size and date of uploading as shown in Figure 4.


Insider data theft prevention by using behavior profiling and decoy documents
User Detail
Upload Details
View Alert
Sign Out

File Upload Details

User Name	File Name	File ID	Size	Date
Sachin Tendulkar	input.txt	1	105	2015-03-27
Sunil Kolawale	input	2	105	2015-04-27
Rohit Sharma	input	3	105	2015-04-27
Rohit Sharma	Testme	4	1558	2015-04-27
Rohit Sharma	Call	5	99	2015-04-28
Sunil Kolawale	Dataset	6	142816	2015-04-28
Sunil Kolawale	Test File	7	7397	2015-04-28
dev	my file	8	28	2015-04-29
mmb	my file	9	28	2015-04-29

Figure 4: Upload details on admin account

Implemented system is trained using 7 classifying features with computer usage data from 25 computer science students collected over a period of 20 days on average. The classifiers were trained using classification and regression techniques described in a prior report. We had given passwords of above 25 users to another 100 computer science students. These classifiers were tested using users real time data in order to check whether the system is accurate enough to identify masquerade activity.

Scenario 1:

In first scenario of experimentation, passwords of real users were given to fake users and their activity was monitored. The combined approach achieves a 89% detection rate or True Positive (TP) rate.

Scenario 2:

In second scenario, experimentation was done on real authenticate users of the system. Normal input is given to all the deciding features. 'Normal' alert is expected to as the real user is operating on the system. Still implemented system is suffering from 4% False Positive (FP) rate which is better than previous anomaly detection schemes.

The Area Under Curve (AUC) score is calculated for each user. Figure 5 displays the AUC scores achieved by insider data theft prevention system. The results show the combined detection approach achieves a higher or equal AUC score, i.e. equal or better accuracy results than the user model based on the search profiling approach alone. The best accuracy improvements were achieved for users 5, 10, and 23. These user models had the top three FP rates amongst all user models. This confirms the efficacy of using this combined approach to limit the number of false positives and improve the accuracy of the masquerade attack detector.

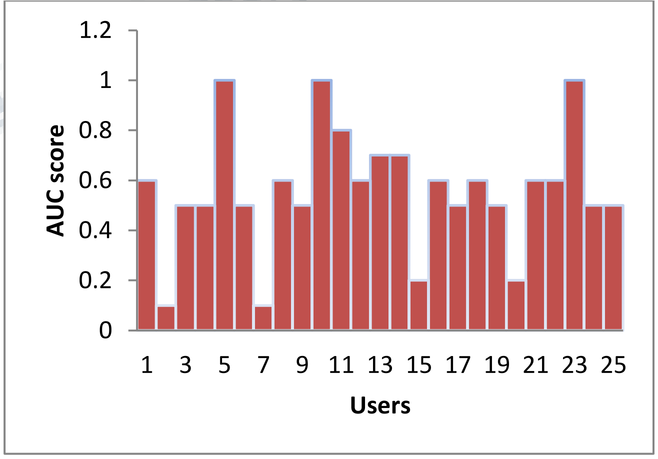


Figure 5: Area Under Curve (AUC) for each user

VI. CONCLUSION

Insider attack is very difficult to diagnose, so the proposed system helps to provide the higher and intelligent level of security in terms of insider attacks. The approaches are based on the predefined user behaviours and using decoy technology. System is implemented in such a way that it could provide an integrated detection approach where profiling user search behavior is done with combination of two classification techniques and decoys are used in order to increase overall efficiency of detecting malicious insider data theft attacks. A major challenge of insider attack detection research is that most of the prior methods work on already created datasets which suffer from lack of the real data. However implemented system works on real time data which increases the efficacy of anomaly detection. Insider data theft prevention system achieved a false positive rate 4% with a 92% masquerade detection rate which makes the system more secure.

REFERENCES

[1] Salvatore J. Stolfo,Malek Ben Salem, Angelos D. Keromytis,“ Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud”, IEEE Symposium on Security and Privacy Workshops, July 2012.

[2] Maxion, Roy A. and Townsend, Tahlia N, “Masquerade Detection Using Truncated Command Lines”, International Conference on Dependable Systems and Networks (DSN- 02), Washington, D.C, June 2002.

[3] D. Godoy and A. Amandi, “User profiling in personal information agents: a survey,” Knowl. Eng. Rev., Dec. 2005.

[4] Maybury, M., Chase, P., Cheikes, B., Brackney, D., Matzner, S., Hether- ington, T., Wood, B., Sibley, C., Marin, J., and Longstaff, T. Analysis and detection of malicious insiders. In Proceedings of the International Conference on Intelligence Analysis, jun 2005.

[5] Dzeroski S., and Zenko B. “Is combining classifiers better than selecting the best one” In Proceedings of the Nineteenth International Conference on Machine Learning San Francisco, CA, USA, 2002.

[6] Ben-Salem, M., and Stolfo, S. J., “ Detecting masqueraders: A comparison of one class bag-of-words user behavior modeling techniques.” In MIST '10: Proceedings of the Second International Workshop on Managing Insider Security Threats, Japan, June 2010.

[7] M. Ben-Salem and S. J. Stolfo, “Combining a baiting and a user search profiling techniques for masquerade detection,” In Columbia University Computer Science Department, 2011.

[8] Lingaswami,G. Avinash Reddy, “Offensive Decoy Technology For Cloud Data Attacks.”,International Journal of P2P Network Trends and Technology, Nov 2013.

[9] Cloud Security Alliance, “Top Threat to Cloud Computing”, March 2010.

[10] A. McCallurn, K. Nigam, “A Comparison of Event Models for Naive Bayes Text Classification”, Workshopon Learning for Text Categorization, 1998.

[11] T. M. Mitchell, Bayesian Learning, in Machine Learning, McGraw-Hill, 1997.

[12] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson, “Estimating the support of a highdimensional distribution”. Technique report, MicrosoftResearch, 1999.

[13] Bowen B. M., Hershkop S., Keromytis A. D., and Stolfo S. J. , “ Baiting inside attackers using decoy documents.” In SecureComm'09: Conference on Security and Privacy in Communication Networks, 2009.

[14] ShlomoHershkop et al “A survey of insider attack detection research”, In Insider Attack and Cyber Security: Beyond the Hacker, Springer, 2008.

[15] MaloofM.A.and Stephens et al, “ Detecting Insider Data Theft of Trade Secrets” ,published by the ieee computer and reliability societies , november/december 2009 .

[16] Chawla, N. V., Eschrich S. and Hall L. O., “ Creating ensembles of classifiers.” In Proceedings of the 2001 IEEE International Conference on Data Mining Washington, DC, USA, 2001.