ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)
Vol. 2, Issue 4 (Oct. - Dec. 2015)
International Journal of Advanced Research in Education & Technology (IJARET)

Masquerade Detection Using Data Driven Semi-Global Alignment Approach

[I]Amol C.Devkate, [II]Vaibhav G.Pagare, [III]Megha D.More, [IV]Sonali R.Misal, [V]Yogendra Patil
[I,II,III,IV,V]Computer, S.B.P.C.E.Indapur.

Abstract
Masquerade attackers embody a legal user to utilize the user services and advantages. The semi-global alignment algorithm (SGA) is one of the most effective and especially techniques to find out these attack but it has not stretch the accuracy and executions required by large scale, multiuser systems. To increase all the effectiveness and the execution of this algorithm, we suggest the Data-Driven Semi-Global Alignment, DDSGA approach. From the security operative view point, DDSGA increase the scoring systems by adopting different alignment arguments for each user. more ever, it allow little change in user command series by allowing little becoming different in the low-level showing of the command to ability to perform a task . It as well to make suitable changes in the client using technique by updating the pattern of the a user according to its current using technique. To perfect the runtime located, DDSGA to make as small the alignment overhead and parallelizes the find out and the update. After representing the DDSGA phases, we represent the experimental results. This result is show that DDSGA get the high hit ratio of 88.4 percent with a low false positive rate. It is increase the hit ratio of the enhanced SGA and reduces Maxion-Townsend cost . Hence, DDSGA results in increasing all the hit ratio and false positive rates with an capable calculation overhead.

Keywords
Masquerade detection, sequence alignment, security, intrusion detection, attacks.

I. Introduction

Masquerader is one of the attacker who easily finds the legal user which uses the services and immunity of the user. For this first SGA (Semi Global Alignment) algorithm can be used. It is most efficient and operative algorithm but the drawback of this algorithm is that it has not yet accuracy for the multiuser systems. For that view of point the DDSGA (Data Driven Semi Global Alignment) algorithm can be introduced. It is easily detect the attacks. It improves the effectiveness and efficiency more than the SGA algorithm. For the security system DDSGA improves the scoring system by adopting distinct alignment for each user. DDSGA minimizes the overhead of alignment and detects parallel and update it. After describing the DDSGA phases, we can got experimental results and this result shows the DDSGA can find the high hit ratio of 88.4% with low false positive ratio. DDSGA improves false positive rate and reduced Marion Townsend cost .

II. Related Work in Masquerade Detection

We studies some detection approaches for masquerade detection. The uniqueness approach consider that command that have not been seen in the training data point out a masquerader. Again, the chance that a masquerader has issued a command is controversially related to the number of the users that use such command. While performance of uniqueness is relatively poor. One-step markove is based upon one-step transitions from a command to the next. This method false alarm rate is not satisfactory. Sconlau et al. toggled between a Markove model and the and simple not dependent. This approach accomplish the good performance among the regard methods.

The main idea about the compression approach is that new and old data of same user should compress at the same ratio. Masquerading user will compress data in different ratio. For binary data classification Support Vector Machine(SVM) indicate set of machine learning algorithm SVM can gives a large set of pattern but it result in high false alarm and low detection rate. Maxion and Townsend .applied a Naïve Bayes classifier widely used in text classification task and also classify user command data sequences into masquerader. An episodes is introduce which is based on Naïve Bayes technique.

According to Naïve Bays algorithm these episodes are Masquerade or normal. Which is used to the number of command in Masquerade block. This technique improve the hit ratio but there is high false positive rates. So he does not update the user profile. In Naïve Bayes algorithm information on the probabilities of command one user over the other users. The WRBF similarity measure based on the frequency of commands f, The weight associated with the frequency vector.

WRBF-NB similarity also increases the overall overhead by computing Naïve Bayes and WBRF and also integrate their results. It neglect the low level presentation of user commands. In Naïve Bayes algorithm both the command of legitimate user and those of an attacker may be different from the train signature. Due to the attacker one persists longer ,the deviation of legitimate the user is momentary.

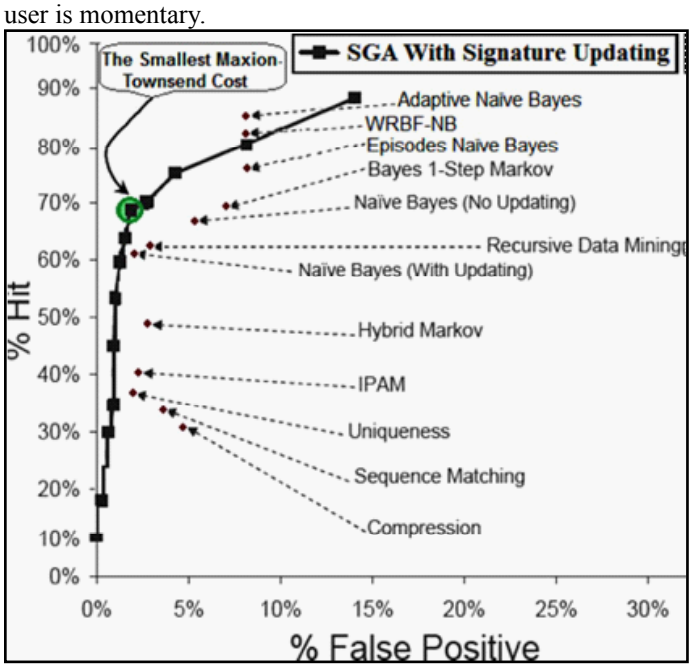

Fig. 1: ROC curves for detection techniques that use SEA Dataset.

ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)

Malek and Salvator used for the user os commands as bag-of-words without timing information. They used for the one-class support vector machine . The sequence alignment algorithm used to find area of similarity. Behavior of the normal user should be created by collecting sequence of audit data.

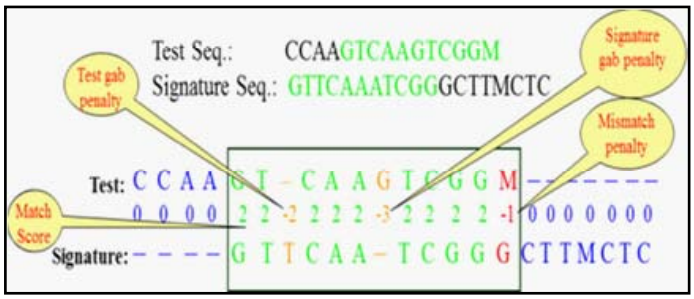

Fig. : SGA AND THE ENHANCE SGA

SGA is more accurate and efficient. It has low false positive and missing alarm rate and high hit ratio. SGA exploits dynamic programming. It initializes an m+1 by n+1 score matrix, M and then shows value of each position of M. In Below diagram there are three stages

1) Diagonal Transition
2) Vertical Transition
3) Horizontal Transition

This three transitions are used to fill each cell in the transition matrix.

The Enhance SGA

TO avoid same false positive, the signature is introduce a new behaviour is encountered by exploiting the ability of SGA .

The signature update scheme is augments the current signature sequence and the user lexicon. The modification heuristic aligning have been tested on the SEA data set for to simplify the comparison.

III. Proposed System

To overcome the drawbacks of the existing system we proposed a new system which is algorithm called as the DDSGA. It totally based on the Enhanced-SGA. The main strategy is to align the user active session sequence to previous one of the same user & labels the misalign areas as anomalous. DDSGA tries to avoid small mutation in user command. The work flow of DDSGA can be shown as follows-

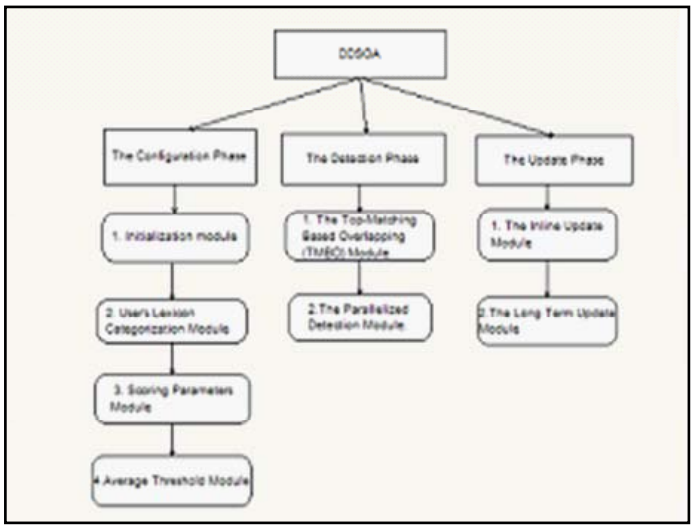

Fig: Phases and modules.

Above fig shows three main phases of DDSGA & In that first one is for configuration of user & other two phases based on alignment parameter. The phases of DDSGA can be explained as below-

1] The Configuration Phase-

There are some parameter which must be calculate for each user in this phase. The detailed description of each parameter can be given in following manner.

- Mismatch Score
  DDSGA calculates the mismatch score through two systems i.e 1) restricted permutation scoring system & free permutation one.
- Optimal gap penalties
  In some cases there is need to insert a gap into the test sequence & signature of user which is called as optimal gap penalties. In the Enhanced–SGA all the users share the same fixed penalties. DDSGA computes two different penalties for each user according to distinct behaviors.
- Average optimal threshold
  DDSGA find outs a distinct threshold value for each user according to change in behavior. It's necessary in both detection and update phase.
- Maximum factor of test gaps(mftg)
  This parameter is related to largest number of gaps inserted into the user test sequence to the length of that sequences. The detection phase uses this parameter to evaluate the maximum length of overlapped signature sequences.
- Initialization Module-
  We need to find out separate set of test & signature sequences for the configuration phase of each user. Here in this module we split the user signature into nt non-overlapped blocks each of length n & use this sequence as test sequence. This generated sequences show or represents all possible combination of user signature sequences & all the modules in the configuration. This sequences are different from those used in detection phase.

In contrast to non-overlapping property of test sequence. We need to generate a overlapped signature subsequence for that, We divide the user signature sequence into a set of overlapped groups of length m=2n. In that way, the last n symbols of a block also appear as the first n of next one. ns are the number of signature subsequence's which is equal to nt-1 groups to consider all possible adjacent pairs of the signature sequences of size n.

The overall procedure described above can be shown in simple format as shown in below.

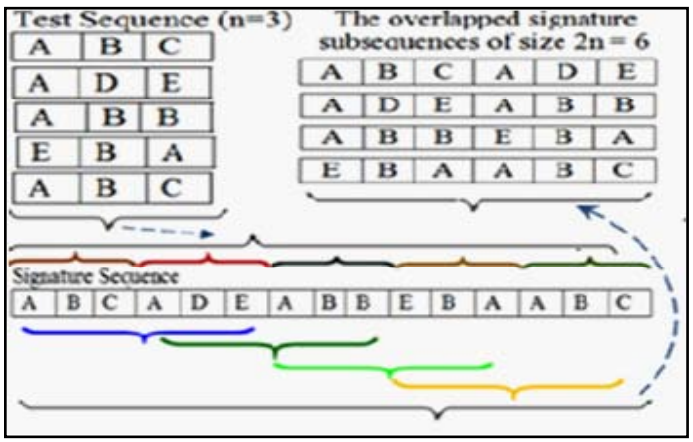

Fig : The non-overlapped test sequences and the overlapped signature subsequences.

- User's lexicon Categorization Module
  In this module we need to build a lexicon for each user according to their functionality. Lexicons are just like a commands to perform particular task.
  Suppose we take an example of command grep , we can be aligned it with find because both belongs to "searching".
- Scoring Parameter Module-
  It is necessary to calculate the score, It returns three parameters: optimal test penalty, optimal signature gap penalty, & mismatch score.

At first, the module puts into the list top-match-list, we select highest match scores for all the nt sequences. After that top-match-list sequences are aligned to the ns overlapped subsequence's by using any possible gap penalty.
The range of test gap penalty range from 1 to n, while the signature gap penalty range from 1 to n. The mismatch score is 0 & match score is +2

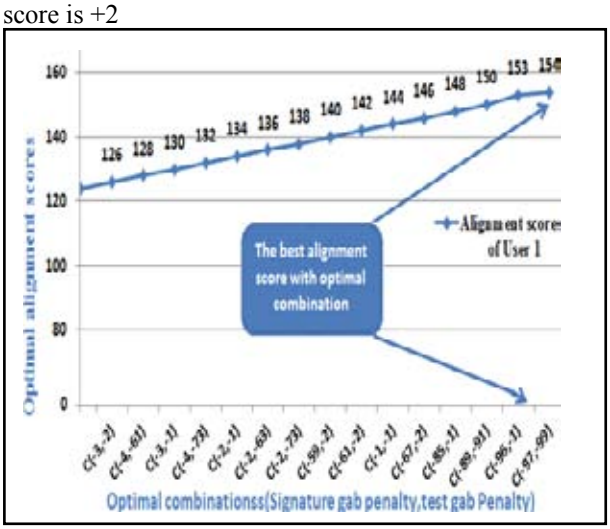

Fig: The best alignment score that corresponds to the optimal combination of gap penalties for user 1 in SEA Dataset.

- Average Threshold Module-
  This is special type of module in which computation of average threshold for each user to be used in the detection phase & that may be update in update phase. As In the phase If alignment score is lower than the threshold, then the behaviour is classified as masquerade attack. This module uses the same test & signature subsequence's of the initialization module & also it can adjust with test sequence of any length as in practical deployment user test session can be of any length.

  Avg-align-I = $(\sum_{j=1}^{ns} \text{score} - \text{align} - \text{ij})/\text{ns}/\text{max-score-align-i}*100$

- Maximum Test Gap Module-

  As we know the Enhanced-SGA Heuristic Aligning divides the signature subsequence's into 2n overlapped subsequence's because if subsequence's of length n are aligned , the maximum number of gaps that can be inserted into the test sequences is n for all users. The maximum test gap for each user varies according to level of similarity between the subsequence's in the user signature & to the length of test sequence. Even if the length of test sequence is long enough, the no of gaps is at most half of sequence of length. After dividing of signature sequence into 2n overlapped subsequence's the maximum test gap module can divide it as follow.

  L=n+[max $\{\sum_{k=1}^{nt}(\frac{ntgk}{itsk})\} * n]$

2]DETECTION PHASE :
-
We have run a complete alignment test on the basis of the test and signature blocks of the SEA data set to calculate the alignment parameters and the two scoring systems. To simplify a differentiation with other approaches, we prefer to use the ROC curve and the Maxion-Townsend cost function. Our Primary focus is on the effects of the alignment parameters on the false positive and false negative rates and on the hit ratio.

Total False Positive=$((\sum_{k=1}^{nu} fpk/nk)/\text{nu})*100$

Where:
- fp = No. of false positive alarms,
- n = No. of non-intrusion command sequence blocks,
- nu = No. of users (50 in our case)

Total False Negative=$((\sum_{k=1}^{nui} fnk/nik)/\text{nui})*100$

Where:
- fn = No. of false negatives,
- ni = No. of intrusion command sequence blocks,
- nui = No. of users who have at least one intrusion block

A. The Top-Matching Based Overlapping Module

Restricted permutation scoring system, Maximum Factor of Test Gaps (mftg) & scoring parameter of each user are used to align the session patterns to a set of overlapped subsequence's of the user signatures in these module. After splitting the signature sequence into a set of overlapped blocks of length L, it chooses the subsequence with the highest match to be used in the alignment process. We have proven that on average, the number of alignments is rather smaller because of the variation between the overlapped signature subsequence's.

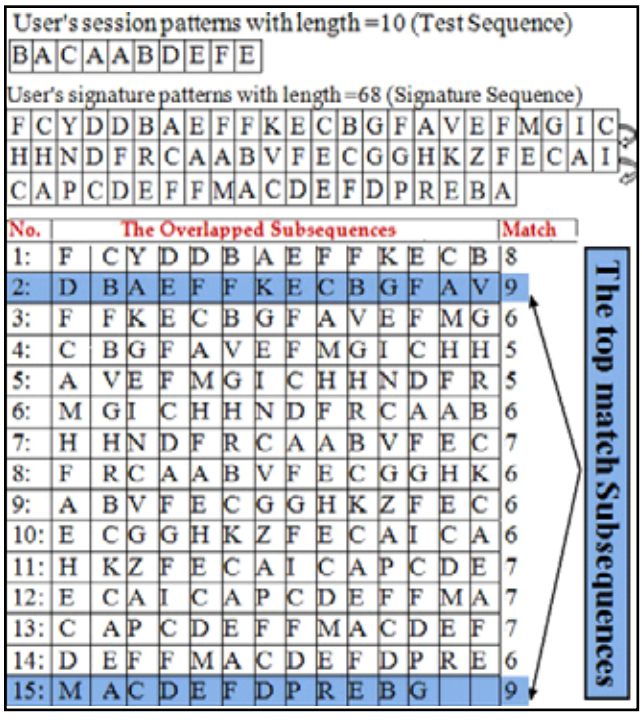

Fig: Overlapped Signature Subsequences of Size 14.

ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)

The working of the proposed TMBO method mainly depends on two parameters: (a) Number of average alignments for the detection process, (b) Theeffect of the TMBO on false alarm rates and hit ratio. The primary task of TMBO computes the following length of the overlapped subsequence's according to following equ.

L=(n+[mftg+n])

In the current phase the current overlapping runs with length L rather than 2n.

The secondary step computes the match corresponding to each subsequence as shown in the front of each subsequences.

The third step select the top match subsequence's, As in the fig subsequence's 2 and 15 , as the best signature subsequence's to be aligned against the test session patterns
of the user. To evaluate the reduction in the workload due to TMBO, consider the Number of Asymptotic Computations (NAC) computed.

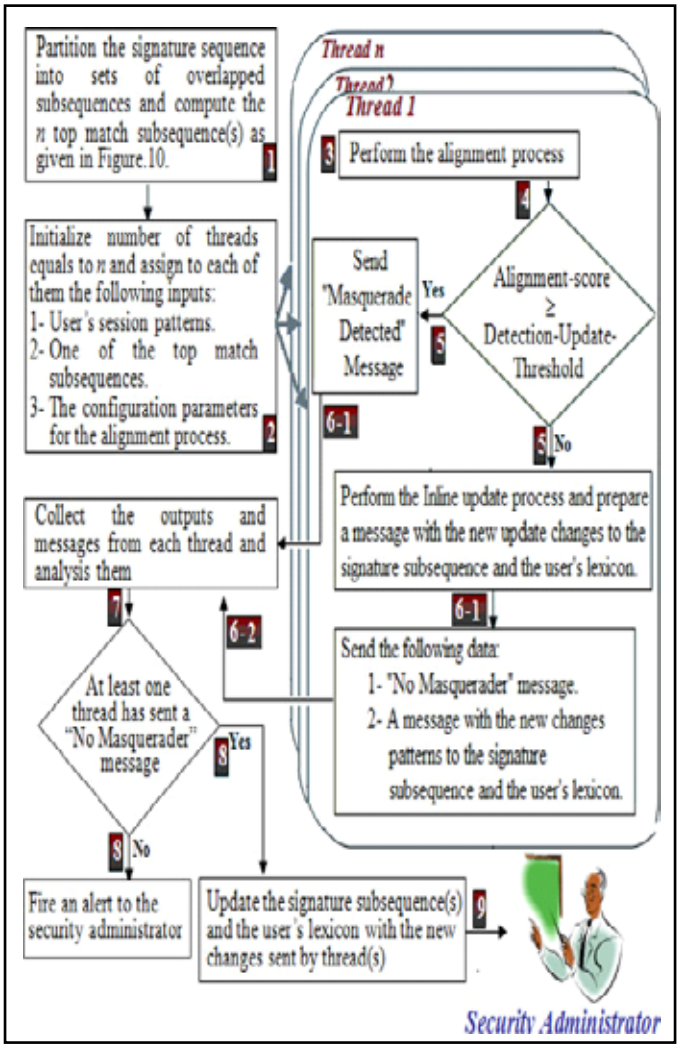

Fig: The processes of the parallelized detection module.

3] The Update Phase :

Its mandatory phase when user is not masquerade. Update of user signature is not an easy task because any IDS should be automatically update to the new legal behavior of user. The update is taken place by two modules: the inline update module and the long term update one.

NAC=Avg_n_align*sig_len*telt_len

As explained in the update phase, if at least one of the previous eight alignments has a score larger than or equal to the detection_ update_threshold, then a process of inline update should be executed for the signature subsequence and the user lexicon.

B. The Parallelized Detection Module

As TMBO partitions the user signature in a set of overlapped subsequence's, we can parallelize the detection algorithm because it can align the commands in the user test session to each top match signature subsequence separately. In this phase we tries to find out whether or not the masquerade is detected. For that we need to perform a simple operation. We just whether Alignment score is less than the Detection_Update_Threshold. If result of this test is yes then thread raises a "Masquerader Detected " alert & if not then perform the signature update by inline update process. This overall process can be shown as follows.

- The Inline Update Module

This module has two primary functions:

- Searching areas in user signature subsequence's to be updated and accumulate with new behaviour pattern.
- Update the lexicon of user with inserting new commands.

Three cases are possible in the TBA that are as follows-

- The test sequence pattern matches the corresponding signature subsequence pattern,
- A gap is inserted into either or both sequences
- There is at least a mismatch between the patterns in the two sequences.

Which can be shown by fig below

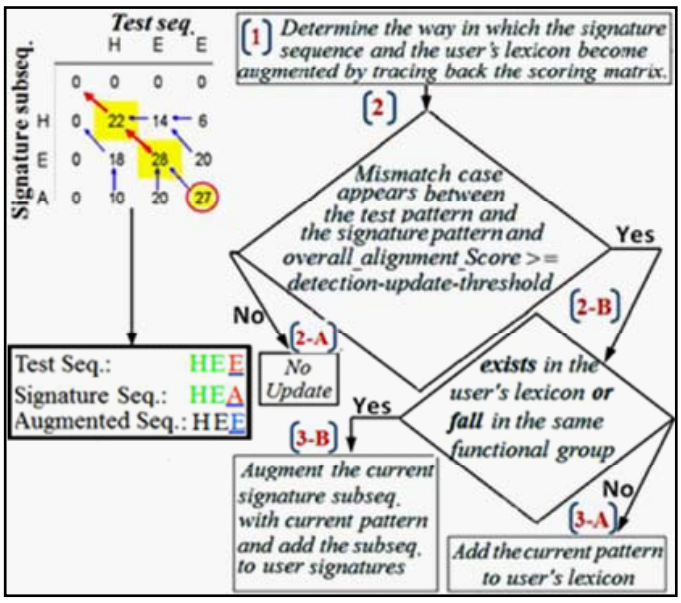

Fig: The inline update steps.

- The Long Term Update Module

In that module we reconfigures system configuration the system parameters through the outputs of the inline update module. There are three main strategies to run the module: Periodic, idle time, threshold. The periodic strategy completes the reconfiguration step with a Static frequency, i.e. 3 days or 1 week. To reduce the complications, the idle time strategy runs the reconfiguration step anytime when the system is idle. This solution is applicable in highly overloaded systems that require an sophisticated use of

the network and computational resources. The threshold strategy completes the reconfiguration step as soon as the number of test patterns inserted into the signature sequences reaches a threshold that is distinct for each user and frequently updated

IV. Conclusion

Masquerading means the attack intentionally . It is one of the most critical attack. So attacker can easily enter into the system with wrong intension and can control the system.SGA is a model based on sequence alignment and it is used to detect the different sequential audit data means checked and observed data but the SGA has very low false positive rate and missing alarm rates .low accuracy even its new version or achieved the correct accuracy and also not given the performance for practical deployment .So the overcome from SGA problem we have DDSGA model this model is security perspective and with more accuracy .DDSGA keep the consistency by providing different parameter to different user and then it offers two level scoring system that tolerate means avoids change in the low level commands functionality of user command and aligning commands in the same class but without reducing the alignment score . The scoring systems also allow all to carry out of its commands and changes in the user behavior extra time. All features strongly degrade false positive and missing alarm rates and increase the detection hit ratio. In the SEA data set, the performance of DDSGA is always better as compare to the one of SGA. Top-Matching Based Over-lapping approach reduces the computational load of alignment by reducing the pattern sequence into a smaller set of overlapped subsequences. Furthermore, the detection and the update processes can be parallel with no loss of accuracy.

References

[1] *Hisham A. Kholidy, Fabrizio Baiardi, and Salim Hariri DDSGA: data-driven semi-global alignment approach for detecting masquerade attack.*

[2] *M. Schonlau, W. DuMouchel, W. Ju, A. F. Karr, M. Theus, and Y. Vardi, "Computer intrusion: Detecting masquerades," Statist. Sci. vol. 16, no. 1, pp. 58–74, 2001.*

[3] *S. E. Coull, J. W. Branch, B. K. Szymanski, and E. A. Breimer, "Intrusion detection: A bioinformatics approach," in Proc. 19th Annu. Comput. Security Appl. Conf., Las Vegas, NV, USA, Dec. 2003, pp. 24–33.*

[4] *A. H. Phyo and S. M. Furnell. "A detection-oriented classification of insider it misuses," in Proc. 3rd Security Conf. 2004.*

[5] *A. H. Phyo and S. M. Furnell. "A detection-oriented classification of insider it misuses," in Proc. 3rd Security Conf. 2004.*

[6] *A. Sharma and K. K. Paliwal, "Detecting masquerades using a combination of Na€ıve Bayes and weighted RBF approach," J. Comput. Virology, vol. 3, no. 3, pp, 237–245, 2007.*

[7] *S. Malek and S. Salvatore, "Detecting masqueraders: A comparison of one-class bag-of-words user behavior modeling techniques," in Proc. 2nd Int. Workshop Managing Insider Security Threats, Morioka, Iwate, Japan. Jun. 2010, pp. 3–13.*

[8] *A. S. Sodiya, O. Folorunso, S. A. Onashoga, and P. O. Ogundeyi,"An improved semi-global alignment algorithm for masquerade detection," Int. J. Netw. Security, vo1. 12, no. 3, pp. 211–220, May 2011.*

[9] *Hisham A. Kholidy, Fabrizio Baiardi, and Salim Hariri, Member, IEEE Computer Society,"*