International Journal of Computer Science and Engineering Communications,
Volume.4, Issue.2 (2016): Page.1402-1409
www.scientistlink.com/ijcsec

Decision Tree and Forensic Methodology for Detecting and Preventing Internal Attackers

M.Jayamagarajothi[1] and P. Murugeswari[2]
PG Scholar[1],Professor[2]
SriVidya College of Engineering and Technology, Virudhunagar, India
jothidct@gmail.com

Abstract: *At present many computer systems log in pattern as user id and passwords to authenticate users. In an organization the people share their log in with colleagues and request these colleagues to complete the co tasks, there by affect the computer domain security. Internal attackers are known users of the system, who attack the system internally and it's difficult to detect. In this paper, we propose Decision Tree and Forensic Methodology for Detecting and preventing internal attackers. Decision Tree and forensic Methodology creates user profile for users to keep track of their usage habits called as forensic features. Decision Tree and forensic Methodology work with intrusion detection mechanism. Network based attacks are identified by predefined algorithm. Decision Tree and forensic Methodology collected in the class limited system call list, as a key component of the system call monitor and filter. Decision Tree and forensic Methodology feasibility and accuracy are verified by Decisive rate threshold. User's Forensic features are identified and analyzed by corresponding system call sequence. Decision Tree and forensic Methodology able to port into parallel system to enhance the accuracy of attack detection and shorten the attacker's response time.*

Index Terms: *Decision Tree, forensic features, system call monitor and filter, system call list, system call sequence.*

I. INTRODUCTION

An intruder is a person who attempts to gain unauthorized access to the system to damage that system or to disturb data on the system. The intruder normally classified into three classes : Masquerader (outside) attack : They are typically outsiders from the trusted users and are not authorized to use the computer systems. These intruders penetrate the system protection by way of legitimate user accounts. Misfeasor (inside)attack : They are typically insiders and legitimate users who accesses resources that they are not authorized to use. Or they may be authorized but misuses privileges. Clandestine user (outside and inside) attack : These type of intruders gain supervisory access to the system. Intrusion detection and protection system (IDPS) have become a necessary addition to the security infrastructure of nearly every organization. An IDPS can provide some amount of quality control for security policy management. They maintain logs about the threats that they detect. IDPS technology adapts statistical method measurements are false positive and false negative. False positive indicate that the IDPS identifies malicious activity. False negative indicate fail to identify a malicious activity. The main objective of IDPS is to prevent or reduce the false negative. IDPS has a various detection methods Signature based detection comparing the signature with signifies a known threat against the events that are observed. Signature based detection is simple detection method, but ineffective against unknown threats. Grant Pannell et al [6] Anomaly based detection generally builds up the profile for the normal behaviors of the user's hosts or network connections or application and it matches the present activities with those profiles. Profile is generated over a period, this period being called a training period. Profiles classified into two types: static profile and dynamic profile. Static profiles are not changed for a long period. A dynamic profile constantly gets updated with additional events. Static profiles are not suitable as they get outdated soon . Dynamic profiles do not suffer from this deficiency.

1.1 DATA MINING

Data mining techniques play a vital role in intrusion detection system. Data mining is a procedure of mining knowledge from data. Data mining tools automatically predict future input and behavior of the system. Data mining techniques are classification, clustering, frequent pattern mining, mining data streams. Decision tree is one the data mining algorithm. In this paper analyze misuse and anomaly detection using decision tree algorithm.

1.2 DECISION TREE

Decision tree based on classification techniques. Decision tree implemented by decision theory and statistics, high effective tool for data mining, information extraction, machine learning and pattern recognition. From an intrusion detection perspective the classification algorithm characterize the system behavior. Classification algorithms create a decision tree for identifying pattern in data set. The input of algorithm is reclassified data. Decision tree provide a set of rules that can categorize new data. The decision tree techniques satisfy a minimum set of requirements to produce real results for an organization.

II. RELATED WORK

K. A. Garcia et al [13] proposed Markov chain model describe the user's normal operations based on system call analyzing. This model use pattern extraction model to identify the particular crime data. But this model not able to support remote login and also not able find the specific type of intrusion.A. KARTIT *et al [12]* proposed security policy at three levels. The three levels are external protection , functional security policies and operational security policies.These three levels combined and act against inside and outside the attack. New approach model use event reconstruction algorithm for identify the system weakness and improve the security policies and increase system performance. But these systems cannot detect any new type of intrusion. Third level of operational security policy require sensor. The sensor monitor the organization whether the particular person is present or not . if the person is absent the data's are rejected by the system and send a alarm to the administrator.Identifying the malicious nodes in network coding based peer to peer streaming network propose a novel approach to limiting the pollution attack by identifying the malicious nodes. Pollution attack describe the valid packets are mixed with invalid and malicious packets .So the whole networks are quickly polluted. This system uses two techniques on fly verification and error correction. On fly verification allow the intermediate node for verify the blocks. Error correction correcting the affected block based on redundancy. The advantage of the system is server broadcast authenticated message to all nodes via public key .so the authenticated message /information is accepted by sink. Disadvantage of the system is cannot find the suspicious neighbor that sent it a corrupted block.Fang-Yie Leu et al [5] proposed profile concept. A system keep track the user usage habit as forensic technique store in user profile. The intrusion detection system identified the underlying user is current account holder or not. The detection accuracy is 98.9% and response time is 0.45s.The first challenge of system is not able to manage the huge amount of data and also need a fast algorithm and grid computing to speed up data processing. Q. Chen, S. Abdelwahed[14] proposed self protection SCADA system monitor the industry environment condition of physical infrastructure. The application of scada is nuclear power plant and municipal water system.SCADA(Supervisory Control and Data Acquistion) gather the information from the system if any leakage in the system it just give the alaram to the central site.SCADA use methods supervisory control procdures. It involves HW[field solution],SW[set or control] and GUI procedures. SCADA mainly focus on system design, coding and verification control application.SCADA use signature based ids detection . It effectively resist the upcoming attacks. But SCADA detection only on log attack activities.Tarek Abbes Loria et al [11] proposed a concept as pattern matching and protocol analysis. Pattern matching and protocol analysis are combined together to effectively resist the network malicious. Protocol analysis monitor and analysis protocols used in the system. It also monitor and analysis the dynamic behavior and state of the protocol.J. T. Giffin, et al [9] proposed a method as model checking restrict program execution by system call sequence(SC). Application system call behavior and security critical operating system state , detect the actual attack. If any undetected attack present ,detection based on frequently used system call sequence .Attacks are discovered by system call sequence against sandboxed operating system.S. Gajek et al [10] proposed Compartmentalization of different trust level applications were isolated, Credentials and authenticating sensitive services stored by trusted wallet. Resolves Phishing sites of domain names by DNS server accessed by user. Trusted and entrusted domain may not sufficient for phishing scenario.F. Y. Leu, et al [3] Grid based platform proposed a method as dynamic grid based intrusion detection environment (DGIDE).DGIDE detect a large amount of intrusion packets and to manage the dynamic grid environment. Main advantage is balance detection workload among detectors. But detection process cannot continue for unfinished task.Z. Shan, X. Wang et al [4] Os level virtualization propose a concept as secom. Secom is a vm commitment system call.sec om naturally ignore malicious state changes. Secomm consists of three stages : Grouping state changes, distinguishing the malicious cluster, committing benign cluster. Malicious objects are recognized by coarse grained in a cluster fashion. It contains benign and malicious objects. But benign objects are only updated with in the vm and back to the host environment.

III. DECISION TREE AND FORENSIC METHODOLOGY FOR DETECTING AND PREVENTING INTERNAL ATTACKERS

MINING SERVER:

Mining server are run on the local computational grid to accelerate the IIDPS's online detection and mining speeds and enhance its detection and mining capability. If a user logs in to the system by using another person's login pattern, the IIDPS identifies who the underlying user is by computing the similarity scores between the user's current inputs, i.e., SCs, and the behavior patterns stored in different users' user profiles. IIDPS, the SCs collected in the class-limited-SC list, as a key component of the SC monitor and filter, are the SCs prohibited to be used by different groups/classes of users in the underlying system.

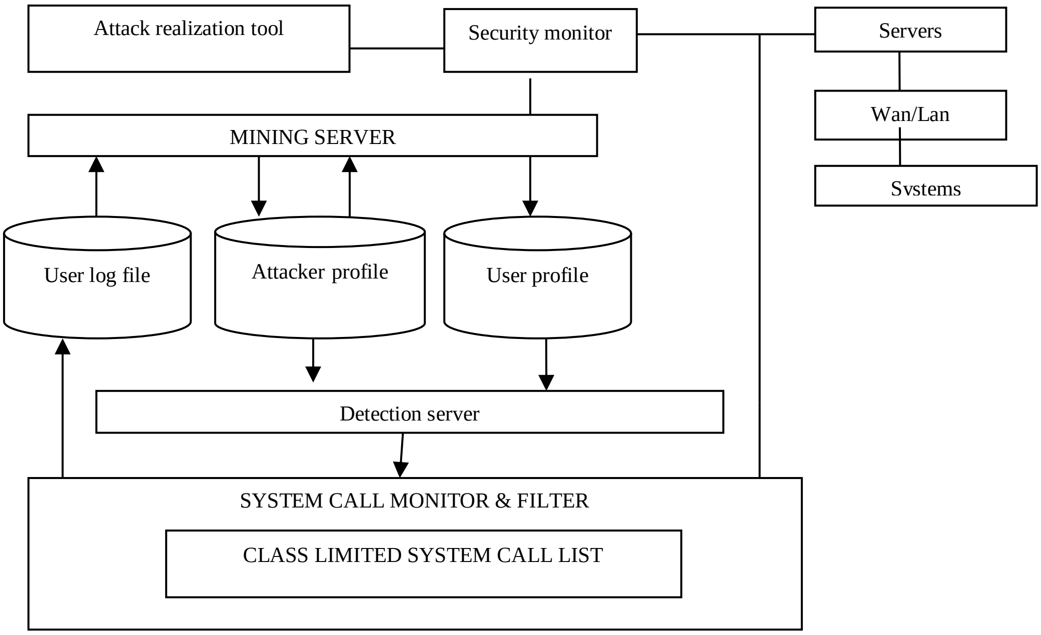


Fig 1. Architecture of intrusion detection system

ALGORITHM FOR MINING SERVER:

HABIT FILE GENERATION

Input: user log file

```
Output: user habit file
Habit = Extract long term common system call sequence from user log file.
Log sliding window[N] = user log file[N];
/* sliding window size is 10 */
G = Log sliding window  - user log file;
Log sliding window[10] = user log file [10];
k-window[k] = current user’s system call sequence ;
k’window[10] = k-window[10];
if(system call pattern already exists in the habit file)
        count = sytem call pattern + 1;
else
      insert system call pattern into the habit file.
```

SYSTEM CALL MONITOR AND FILTER

The SC monitor and filter, as a loadable module embedded in the kernel of the system. System calls are collected and submitted to the kernel .System calls are stores in the format of uid,pid,sc in the protected system where uid, pid, and SC represent the user ID, the process ID, and the SC submitted by the underlying user. It also stores the user inputs in the user’s log file, which a file is keeping the SCs submitted by the user Submitted sequence.

ATTACK REALIZATION TOOL

Attack realization tool maintain all types of attacks such as probe, Remote to User Attacks, User to Root Attacks, Distributed denial of attacks. Attack realizations create a attacker profile. The attacker pattern system calls are not commonly used by other users. So the attacker pattern system call has the high similarity weight.

SECURITY MONITOR

Security monitor is similar to administrator. Administrators allocate the task to users. The administrator monitor the all users behavior through wan or LAN. The user current input information is stored in user log file. The security monitor updates the attack realization when a new type of attack is discovered.

DETECTION SERVER:

Detection server gets the system call from underlying user. When users executing the system calls and stores these system calls in user log file. Then the detection server identified whether the user is authenticated user or not. Detection server identification based on system call pattern analysis. System call pattern analyzed by similarity score between the users’s currently generated system calls and user’s usage habit system calls. The similarity score between the current user’s system call and authenticated user system calls are ranked based on similarity threshold If threshold value is less than the similarity threshold value then the detection server identified the intruder is not present into the system. Detection server informs to the system call monitor and filter to concealed the current user from preserve system.

ALGORITHM FOR GENERATING USER PROFILE

```
Input: Current user system call sequence(CSC)(each time multiple user system call sequence is input)
Output:  Finding an insider attacker
CSC = 0;
While (receiving user’s input system call sequence)
If (CSC >  | habit sliding window |)
   {
          Long Sliding Window = Right(CSC /habit sliding window);
    For( k = |CSC | - |k-window |; k>0;k--)
     K’window = mid(CSC,k,|k-window|);
 /* middle(x,y,z) return sliding window size(L) beginning at the position of y
 From x */
Compare k-window and k'window by using comparison method ;
For (each user <= N)
Calculate the similarity score .similarity(user,systemcall) between    current user and user profile )
If(CSC  mod input size ) == 0)
/* input size is 15 -> find whether the user is account holder or not in every 15 input system call.
}
Assign all users similarity scores are ascending order .
If(decisive rate of user profile < threshold)
{
      /* threshold is predefined average decisive rate of user's profile.*/
Attacker are not present into the system proceeding the work continuously.
}
Else
{
    Alert system call monitor and filter the attacker are entered into the system ;
}
```

IV. FLOW CHART

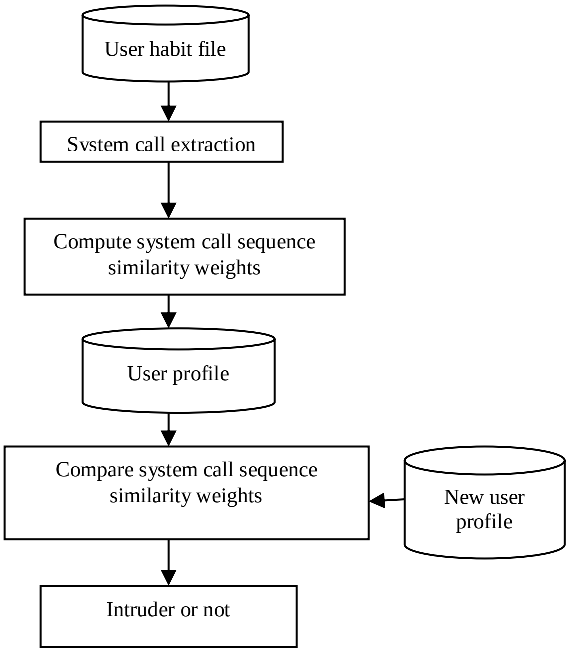

User habit file
Svstem call extraction
Compute system call sequence similarity weights
User profile
Compare system call sequence similarity weights
New user profile
Intruder or not

Fig 1.1 Flow chart for attacker identification

V. THEOREM:

Time complexity for generating a habit file is O(L^6) where L is the size of sliding window.

Proof :

Let H = | system call sequence | - | sliding window | - 1 ,number of sliding window based on given system call sequence. User habit file is generated by (log window ,comparison window) pair wise comparison when accessed by (h * (h - 1) / 2) times. Pair wise comparison based on

$$(|\text{Sliding window}| - G + 1) * (|\text{Sliding window}| - G + 1)$$

The total time of pair wise comparison denoted by

Total comparison

$$= (N – L + 1) (N – L) / 2 * (N – G +1) * (N – G' + 1)$$

$$= (N – L + 1) (N – L) / 2 * (L (L – 1)) * (L (L -1)/2)$$

= 1 / 8 (1 – L) ^ 2 L ^ 4.
Its shows the time complexity of pair wise comparison is O(L^6).

THEOREM : TIME COMPLEXITY

The time complexity of finding an insider attacker is O(HL^5) where l is the size of sliding window h is the number of authorized users.

Proof :

Every time the authorized user give the new inputs an system call and the number of system call is greater than size of sliding window, the final system call will be identified as an log sliding window. The current user input a total of N-L+1 sliding window present Sliding window comparison based on pairwise sliding window such as (L(L-1)/2) * (L(L-1)/2).

Total comparison = (N-L) * L(L-1)/2 *L(L-1)/2.

Total = H*Totalcomparison * (N-L) * L(L-1)/2 *L(L-1)/2.Its indicates the time complexity of algorithm is O(HL^5).

VI. EXAMPLE FOR SLIDING WINDOW

LOG SLIDING WINDOW

XYZW YZWX ZWXY WXYY XYYZ
YYZW YZWX ZWXX ZXXY XXYZ

Sliding window size is 10.

CURRENT SLIDING WINDOW :

XYZW YZWX ZWXY WXYY XYYZ
YYZW YZWX ZWXY ZXYX XYXY

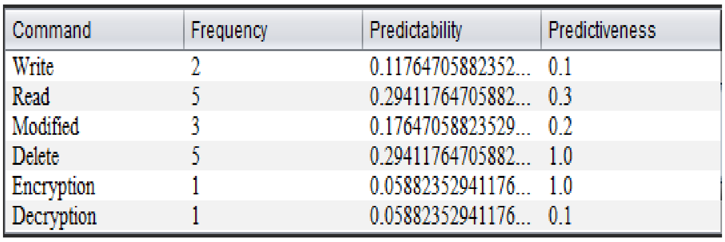
→	(XYZW YZWX) 1	(XYZW YZWX) 2
→	(ZWXY WXYY) 3	(ZWXY WXYY) 4
→	(XYYZ YYZW) 5	(XYYZ YYZW) 6
→	(YZWX ZWXX) 7	(YZWX ZWX) 8
→	(ZXXY XXYZ) 9	(ZXYX XYXY)10

In the above comparison the anomaly is detected on 7, 8,9,10 th comparison.

VII. EXPERIMENTAL RESULTS

HABIT FILE GENERATION

Admin analyze the user’s log file. System call sequences are extracted from user’s log file. The system call appear in the file should be counted. System call pattern similarity weights are calculated by using team frequency inverse document frequency (TF- IDF).Frequently used system call pattern has high similarity weight. Common similarity patterns are eliminated and store it in the user habit file.

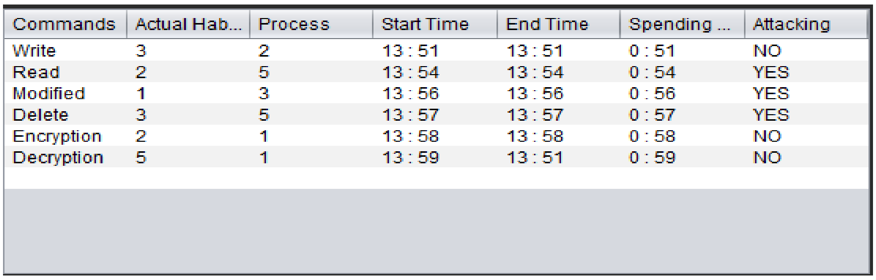


Command	Frequency	Predictability	Predictiveness
Write	2	0.11764705882352...	0.1
Read	5	0.29411764705882...	0.3
Modified	3	0.17647058823529...	0.2
Delete	5	0.29411764705882...	1.0
Encryption	1	0.05882352941176...	1.0
Decryption	1	0.05882352941176...	0.1

Figs 1.2 Habit file generation

ATTACKER DETECTION

Detection server get the input from user profile and current user profile. Detection server analyze the attacker specific pattern .If the attacker specific pattern present in the user profile the detection server sent the signal to system call monitor . The system call monitor isolate the system from unauthorized user for prevent the system with 0.45 seconds.



Commands	Actual Hab...	Process	Start Time	End Time	Spending ...	Attacking
Write	3	2	13 : 51	13 : 51	0 : 51	NO
Read	2	5	13 : 54	13 : 54	0 : 54	YES
Modified	1	3	13 : 56	13 : 56	0 : 56	YES
Delete	3	5	13 : 57	13 : 57	0 : 57	YES
Encryption	2	1	13 : 58	13 : 58	0 : 58	NO
Decryption	5	1	13 : 59	13 : 51	0 : 59	NO

Fig 1.3 Attacker detection

PERFORMANCE EVALUATION

Decision tree and forensic methodology handle the multi user and detect the intruder very fastly,the response time of the user is 0.28seconds.

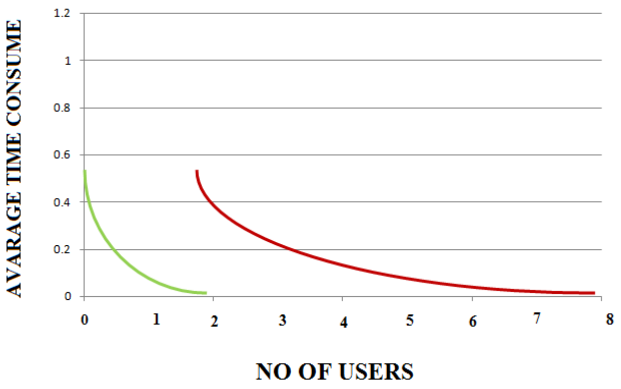


Fig 1.4 performance chart

VIII CONCLUSION

In this paper we have proposed Decision tree and forensic methodology for detecting and preventing internal attackers to analyze the corresponding system call pattern for each user. That time the system call patterns were counted and habit file was generated. User common system call pattern behaviors are filter out and find the similarity score between the system call pattern created the user profile. With help of decisive rate threshold to identify whether the user is authorized account holder or not. The further study will be improving the Decision tree and forensic methodology for detecting and preventing internal attacker's performance and response time to effectively oppose the insider attacker.

ACKNOWLEDGMENT

I would like to thank my project advisors Dr.P.Murugeswari phd.,Head of Department of srividya college of engineering and technology ,virudhunagar,tamilnadu. . I have been wonderful fortunate to have advisors who gave me the freedom to explore on my own and at the same time the guidance to recover when my failure. They taught me how to question thoughts and express ideas. Their patience and support helped me overcome many crisis situations and finish this work.

REFERENCES

[1] M. Jayamagarajothi and P. Murugeswari, "Decision tree and forensic methology for detecting and preventing internal attackers", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 4, Issue 11, November 2015.

[2] Fang-Yie Leu, Kun-Lin Tsai, *Member, IEEE, Yi-Ting Hsiao, and Chao-Tung Yang "*An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques*"* IEEE Systems journal ,21 Apr ,2015

[3] F. Y. Leu, M. C. Li, J. C. Lin, and C. T. Yang, "Detection workload in a dynamic grid-based intrusion detection environment," J. Parallel Distribute. Comput, vol. 68, no. 4, Apr. 2008, pp. 427–442.

[4] Z. Shan, X. Wang, T. Chiueh, and X. Meng, "Safe side effects commitment for OS-level virtualization," in Proc. ACM Int. Conf. AutonomicComput., Karlsruhe, Germany, 2011, pp. 111.

[5] Fang-Yie Leu , Kai-Wei Hu "A Real-Time Intrusion Detection System using Data Mining Technique "systemics, cybernetics and informatics volume 6 - number 2 ,2012.

[6] Grant Pannell and Helen Ashman " Anomaly Detection over User Profiles for Intrusion Detection " in proc . *8th Australian Information Security Management Conference* ,2010.

[7] Al-Khanjari, Z.Alanee, A.Kraiem, N.vJamoussi, Y. "proposing a real time internal intrusion detection system towards a secured development of e-government web site "in Proc . European Scientific Journal vol.3, December 2013.

[8] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," Appl. Soft Comput., vol. 10,no. 1, Jan. 2010, pp. 1–35.

[9] J. T. Giffin, S. Jha, and B. P. Miller, "Automated discovery of mimicry attacks," Recent Adv. Intrusion Detection, vol. 4219, Sep. 2006, pp. 41–60.

[10] S. Gajek, A. Sadeghi, C. Stuble, and M. Winandy, "Compartmented security for browsers—Or how to thwart a phisher with trusted computing," in Proc. IEEE Int. Conf. Avail., Rel. Security, Vienna, Austria, Apr , 2007, pp. 120–127.

[11] Tarek Abbes LORIA/INRIA-Lorraine, Adel Bouhoula SUP'COM , Michaël Rusinowitch LORIA/INRIA-Lorraine "Protocol Analysis in Intrusion Detection Using Decision Tree " in proc. International Conference on Information Technology Coding and Computing (ITCC'04) ,2004.

[12] A. Kartit, A. Saidi, F. Bezzazi, M. El marraki, A. Radi " a new approach to intrusion detection system "journal of theoretical and applied information technology , 29th february 2012. Vol. 36 no.2

[13] K. A. Garcia, R. Monroy, L. A. Trejo, and C. Mex-Perera, "Analyzing log files for postmortem intrusion detection," *IEEE Trans. Syst., Man,Cybern., Part C: Appl. Rev.*, vol. 42, no. 6, Nov. 2012, pp. 1690–1704.

[14] Q. Chen, S. Abdelwahed, and A. Erradi, "A model-based approach to self-protection in computing system," in *Proc. ACM Cloud Autonomic Comput. Conf.*, Miami, FL, USA, 2013, pp. 1–10.