IJE TRANSACTIONS A: Basics Vol. 30, No. 10, (October 2017) 1479-1486

International Journal of Engineering

Journal Homepage: www.ije.ir

An Architecture for Security and Protection of Big Data

F. Asadi Saeed Abad[a], H. Hamidi*[b]

[a] *Department of Information Technology Engineering, College of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Industrial Engineering, Information Technology Group, K. N.Toosi University of Technology, Tehran, Iran*

PAPER INFO

Paper history:
Received 13 June 2017
Received in revised form 04 July 2017
Accepted 07 July 2017

Keywords:
Big Data
Privacy
Security Model
Petri Net
Cloud Computing

ABSTRACT

The issue of online privacy and security is a challenging subject, as it concerns the privacy of data that are increasingly more accessible via the internet. In other words, people who intend to access the private information of other users can do so more efficiently over the internet. This study is an attempt to address the privacy issue of distributed big data in the context of cloud computing. One of the cases where data privacy is of great importance is the authentication and protection of ownership data. In this paper, this privacy issue is analyzed by Petri net modeling. What today's organizations need for their clouds are integrated comprehensive solutions that can deliver security intelligence. Advanced security intelligence solutions can close security gaps by using labor-saving automation to analyze millions of events occurring within the cloud, and discover system vulnerabilities through the normalization and correlation of these events. Using the proposed method, a model of security, including control of user access to databases of big data with RMS, the multiplicity and the virtual machine to prevent internal threats, deleting data, insecure or incomplete data protection and control of a third-party can be provided to improve the operation according to the rules of Petri net modeling and simulation.

doi: *10.5829/ije.2017.30.10a.08*

1. INTRODUCTION

The past decade has seen rapid progress in the internet, internet of things and cloud computing, and thereby a strong growth in data utilization for commercial and industrial applications. The term "Big Data" refers to massive data sets with large, diverse, and complex structures that are challenging to store, analyze, and handle with traditional processing approaches. For a big data handling approach to be effective, it should accommodate not only the size, velocity, and diversity of such data, but also their unique data protection requirements. Nevertheless, there is a clear conflict between the pervasive usage of big data and related security and privacy issues [1, 2].

Big data contains private information of people, who are generally very protective of their privacy. The way the term "privacy" is interpreted may vary with the country, culture, and legal sphere, but it is generally in conflicts with collecting, storing, using, processing and sharing of personally identifiable data. The primary objective of privacy measures is to ensure proper protection of private data in the course of processing or dissemination of sensitive information [3, 4].

Cloud computing is a service model where users are provided with computing and storage resources such as CPU, memory, or software over an online, high-speed and flexible network. This model allows the enterprises to deliver computing and storage resources to their clients on a massive scale. To use these services, users have to store their data in the physical environment hosting the cloud service, and in effect, entrust the physical control of their data to the cloud provider. In that case, users must be able to trust the cloud provider to store their data with adequate security and privacy protection [5, 6].

While the merits of cloud computing are widely acclaimed, information and database security in the cloud space remains a major concern in relation to this technology [7-9].

This study is an attempt to address the privacy issue of distributed big data in the context of cloud computing. One of the cases where data privacy is of great

*Corresponding Author's Email: h_hamidi@kntu.ac.ir (H. Hamidi)

Please cite this article as: F. Asadi Saeed Abad, H. Hamidi, An Architecture for Security and Protection of Big Data, International Journal of Engineering (IJE), TRANSACTIONS A: Basics Vol. 30, No. 10, (October 2017) 1479-1486

importance is the authentication and protection of ownership data. In this study, this privacy issue is analyzed by Petri net modeling.

2. DATA PRIVACY

Today, data privacy is an omnipresent issue in every field of computing and communication. It should be emphasized that although data reliability is an important aspect of secure storage of private data, this objective is also associated with other requirements, including management of users' permission to use their personal data, support for the use of sensitive data, and compliance with privacy-related regulations. Many methods from data encoding, encrypted data processing support, and data structures that hide data access patterns, to data anonymization and differential privacy protection techniques that provide private data transmission can be utilized to make it difficult to attribute a particular piece of data to a particular individual.

The recent research literature on privacy protection is more concentrated on specific areas such as location privacy [10-12], smartphones and social networks. So despite the breadth of research literature in this field, the problem of data privacy in big data is yet to be resolved.

3. REVIEW OF LITERATURE

In this section, the previous work on the subject of privacy protection is reviewed.

Hamidi and Vafaei have explained that big data privacy faces many challenges, most of which are not technical but stem from organizational issues and regulations; however, it can be predicted that whether or not they can be overcome by technical measures [13].

Hamidi et al. have proposed a reliable privacy and security solution for data analysis of sensor network in smart homes [14]. The proposed method is the use of data protection software without conversion of stored data by encryption techniques. This method prescribes replacing the personal identifiers collected from sensor data with scrambled values before storage.

The methods of privacy protection in big data have been evaluated [15]. This article concludes that while differential privacy protection schemes perform well for big data, they require the analyst to know the inquiry before using the differential privacy model.

Wu et al. worked on the differential privacy, databases, data presented herein. In this approach, privacy is protected by adding noise to inter-dependent values of private data [16].

The Laplace distribution, which is typical for privacy differential noise and optimal distribution of independent data shows noise is used [17].

The privacy practices of differential and t-closeness to a branch or an extension of the k-anonymity method has been discussed [18]. Although the two are not completely equal, but are highly interdependent, so that k-anonymity pseudo-arguments and difference method maintain confidentiality random intervals t-closeness with the functions t, k and e's.

Hamidi and Kamankesh have introduced a method of data element classification based on data values, that is, use of three parameters (the storage, content type, and access control) to sort the data [17]. Figure 1 shows the classification and the relevant security measures.

Hamidi and Moradi have proposed a secure method of data storage in the cloud called "CloudSafe" [18]. This method improves the availability and confidentiality of cloud data storage by the use of AES for encryption and decryption. Wang and Li have introduced a framework consisting of various techniques and specialized methods, which can provide data protection from the origin (user) to the destination (cloud) [19]. This framework makes use of data classification based on three user-provided parameters: confidentiality, availability, and integrity. The proposed framework involves two steps: storage in the cloud, and data retrieval from the cloud. The proposed strategy is to use data encryption for evaluation of data integrity. This method first indexes and then encrypts the data and finally stores the encrypted file in one of the three classes: public, private and limited access as shown in Figure 2.

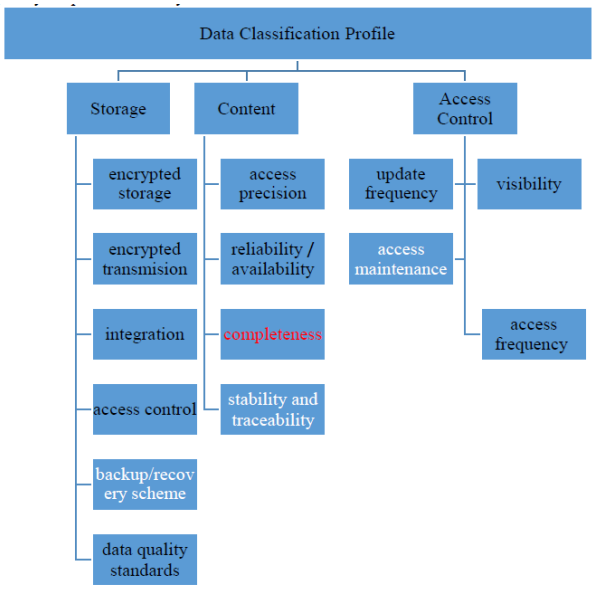


Figure 1. Data classification and related security measures

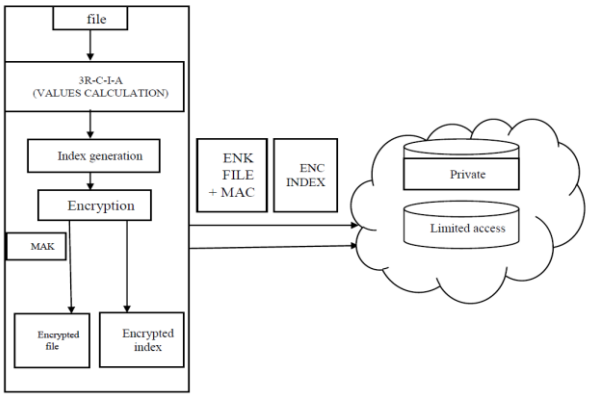


Figure 2. Diagram of data classification and transmission in the cloud environment

For a security model implemented for cloud computing to be effective, it should be able to deal with all types of possible threats while maintaining confidentiality and availability as well as system functionality and performance. In other words, the models should be able to fulfill all these goals to an acceptable level while exhibiting a level flexibility depending on the application [20, 21].

4. THE PROPOSED METHOD

This paper proposes a security model consisting of following components:

- Big data access control with RMS
- Multi-tenancy and virtualization
- Third-party control
- Data protection
- Prevention of insecure/incomplete data deletion
- Protection against internal threats

The proposed model is based on the following assumptions:

1- The modeled architecture is used for big databases (Figure 3).

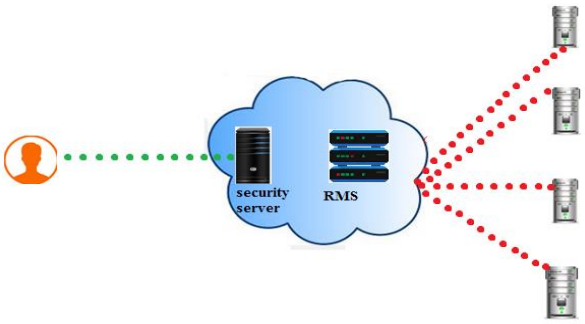


Figure 3. The modeled architecture for big databases

2- Users of big data submit their requests to the server.
3- The server forwards the requests to a security server which authenticates the request and the requesting user.
4- When verified by the security server, the request will be forwarded to RMS, where it will be divided into sub-tasks and assigned to resources.
5- Sub-tasks are independent of each other, and every resource begins to execute the assigned sub-task upon receiving its data from RMS.
6- RMS is very fast and completely reliable, so the time of task processing in RMS is negligible compared to the time of sub-task processing in resources.
7- Reliability is defined as the probability of a task being fully processed in less than specified time T (reliable).

4. 1. Multi-tenancy and Virtualization One of the main security issues of big data is the multi-tenancy. Multi-tenancy is a unique feature for sharing of resources in cloud computing. To provide one piece of data to multiple users, virtualization should be utilized to delimit the security boundaries preventing intentional or unintentional access to other user's share of data over a shared resource. This allows the users to share memory, software, network, and storage resources in an isolated and secure environment.

Figure 4 illustrates the Petri net diagram of security model proposed for multi-tenancy and virtualization. In this model, all users request arrive at the RMS of a virtual machine, which checks the license of requests and forwards the authorized requests to the server. The server first checks to see if the received request can damage the data. If authorized, the request will be forwarded to the execution unit, and once executed, will be returned to RMS.

4. 2. Third-party Control Another security issue is the possibility of a third-party gaining unauthorized access to data or mechanism of data processing. Clients have to trust the service provider to prevent such access.

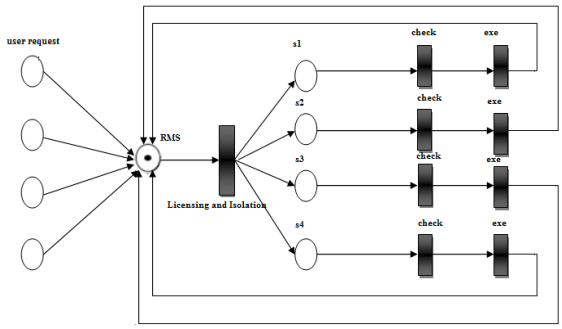


Figure 4. The security model for multi-tenancy and virtualization

In return for this trust, the service provider has to take sufficient measures to monitor and supervise the clients' security. There is a principle in this regard known as "trust-but-verify", according to which service providers are required to regularly update their security controls to accommodate the frequent changes in clients' security requirements.

Better management of critical enterprise data requires the existence of mutual relations based on trust between them. With the increasing value of information, it is possible for third-party to access the data, loss of intellectual property and corporate secrets to divulge.

Another challenge in this regard is the availability of client information to individuals within the organization, which expose this information and thus client security and rights to the agents of rival companies and intrusion of internal exploiters. In the proposed model, this problem is handled by the use of Extensible Authentication Protocol (EAP), which is a powerful protocol operating based on a request/response structure. The diagram of this architecture is shown in Figure 5.

This security model, which has been created using the Petri net, operates as follows:

1- The user will be asked to enter his username; if the username is valid, he will be asked to answer a security question.

2- The security question is dynamic in the sense that it has been created in the last login of the user.

3- If the security question is answered correctly, the user will be asked to enter the password.

4- If the password is correct, the user logs on.

5- Once logged in, the user sends a request to the cloud. The request will be checked, and if authorized, will be forwarded to RMS.

4. 3. Data Protection The use of distributed big data exposes both provider and client to a series of risks associated with data protection. One of such issues is how clients are ensured about the legality of actions of the provider on their data; a problem that exacerbates with the frequency of data processing and transfers.

The authors propose a credential classification and a framework for analyzing and developing solutions for credential management that include strategies to evaluate the complexity of cloud ecosystems [22, 23].

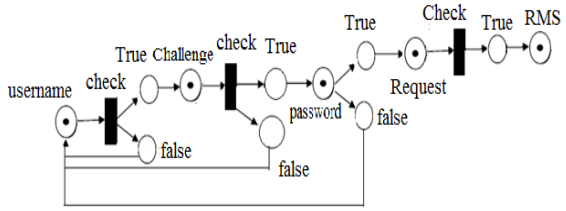


Figure 5. Diagram of security model with Extensible Authentication Protocol

This study identifies a set of categories relevant for authentication and authorization for the cloud focusing on infrastructural organization which include classifications for credentials, and adapt those categories to the cloud context. A trust relationship between a given user and SaaS domains is required so that SaaS users can access the application and resources that are provided. In a PaaS domain, there is an interceptor that acts as a proxy to accept the user's requests and execute them. From the consumers' perspective, cloud computing security issues, particularly data security and privacy protection concerns are the primary inhibitor for adoption of cloud computing services. In the current research we present a concise but all-round investigations on data security and privacy protection concerns associated with cloud computing across all stages.

Some providers provide their clients with a summary of processes, controls and security measures in the format of a certificate, which allows the client to verify the identity and function of provider's system. This aim can be achieved with the help of mutual authentication. In mutual authentication, both server and client should prove their identity to each other. After initial identification and agreeing on the connection quality, two parties define and establish a connection. Mutual authentication can be achieved with the help of digital signatures. The proposed digital signature modeled with Petri net is illustrated in Figure 6.

In this model, server and user both have a public and private key, which are generated at the time of sign in, and are used for mutual authentication in the event of suspicious activity.

4. 4. Prevention of Insecure/Incomplete Data Deletion A client's requests to delete his data from a big database may be impossible to accommodate adequately and in a timely fashion. This may be because of the existence of multiple copies of data which are unavailable or because the disk cannot be destroyed as it contains data belonging to others.

This security problem is more serious in the cases of multi-tenancy and reuse of hardware resources than in the cases where hardware is dedicated.

To avoid this problem, upon receiving a request for deletion of data, it will be checked whether this data is being utilized by someone else.

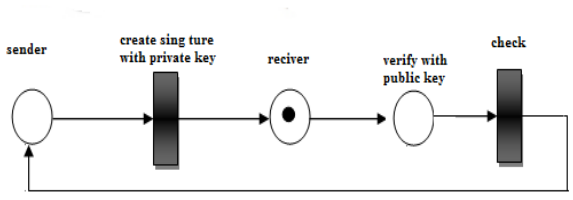


Figure 6. Data protection with mutual authentication

If currently free, the data will be removed, but if not, it will be suspended until it is no longer in use. When freed by the current user, data will be deleted upon system confirmation. The Petri model of above procedure is shown in Figure 7.

4. 5. Protection against Internal Threats Storing critical information in big databases, exposes clients to the risks of virtual machine attacks or misuse of information due to improper access of untrustworthy employees, which on the other hand, undermines the provider's credibility and reputation among the clients.

To avoid this problem, virtual machines need to be supervised by a security management system preventing any intrusion of boundaries set by the system. This aim can be achieved by task division.

In the proposed task division scheme, RMS divides the tasks received from the user into multiple sub-tasks. Given the use of redundancy technique for resource-task assignment, this division should be such that the number of sub-tasks n remains below the number of available resources r. After this division, RMS assigns each sub-task to more than one resource. But each resource processes only one sub-task. To make maximum use of resources, RMS always utilizes all resources available for task execution. The Petri net diagram of the above operation is illustrated in Figure 8.

In this model, each task is assigned to a virtual machine, and how it will be executed is decided by RMS. With this system, virtual machines will no longer have any influence over each other.

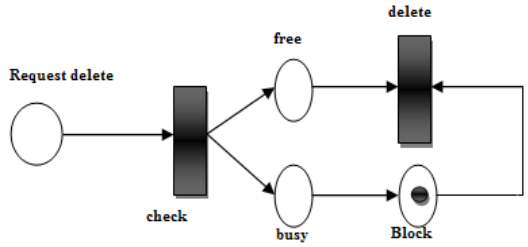


Figure 7. The proposed model for preventing insecure/incomplete data deletion

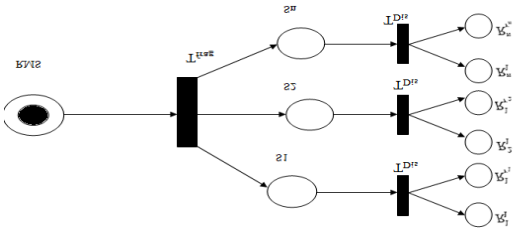


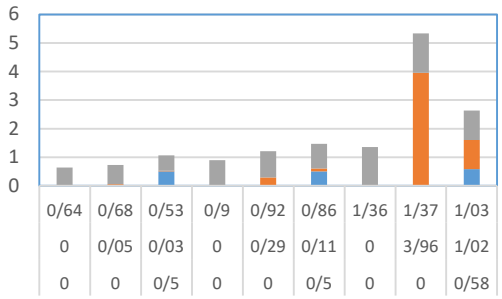
Figure 8. The task division procedure for protection against internal threats

4. 6. Preliminary Evaluation of CloudSafe In this section, we present a preliminary evaluation of CloudSafe. A comparison is conducted with another two scenarios commonly in real life. We did the test in three scenarios: first, Plaintext (P), a user only exploits one cloud storage provider, and the user stores the plain text of the data in clouds. This scenario represents the users who rarely concern the data security and availability because no efforts are taken to protect the data. The second scenario Encrypted Plaintext (EP) is that a user takes the privacy into concern and the user encrypts the data before storing it into the clouds. But this kind of user still relies on one cloud storage provider; the third group of users use CloudSafe (CS) that not only encrypts the data it-self, but also stores the data in multiple providers to provide data privacy.

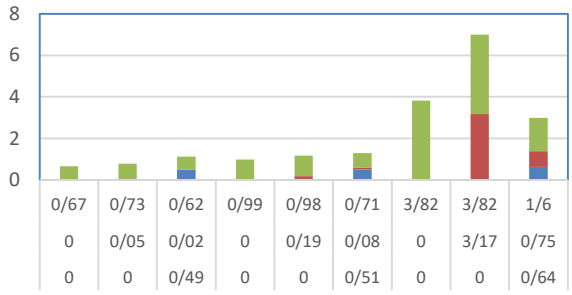
In the test, the client connects to the internet through a 20 Mbps network link. To test the read and write performance, a client performs read and write function to get and put data from the cloud providers. In P and EP scenarios, there is only one cloud storage provider.

For the CS scenario, all four providers are enabled. We chose Dropbox as the unique provider for P and EP scenarios because it has the smallest average Round Trip Time (RTT) among all four providers. According to the observation [23], 95 percent of the total data flow size between the user and Dropbox servers are less than 10 MB, which is observed among several European countries during 42 days. Furthermore, the phenomenon that the 95 percent the data size of upload and download operation are also less than 10 MB is also addressed in that paper. Based on this observation, we assume the users locating in US also have the same characteristics of their usage. Based on this interesting observation, we chose four kind of data sizes: 10 KB, 100 KB, 1 MB, and 10 MB, to do the read and write test, and we ran the test for 10 rounds in different time during one day. In order to show the time cost overhead, we count the time of operation (erasure coding, encryption, decryption, network transmission) separately.

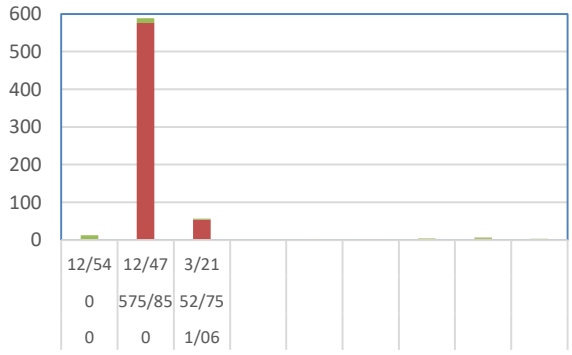
In Figure 9, we present the results of time consumption for read and write semantics in three scenarios mentioned before. Since the time of 10 MB data size is not from the same order as others, we split the results for 10 MB data out and use logarithmic relationship to show the results. Compared to P scenario, CloudSafe spent twice as much time as the original P scenario at most for both read and write operation when the data size is less than 1 MB, and on average, the increased time is 55 percent. Because the total access time for data size less than 1 MB is in the order of seconds, the access time overhead is relatively small and has little influence on user experience. However, using CloudSafe provides more data privacy and availability than using one provider with plain text in clouds.



(a) Accumulated access time of read operation for 10KB,100KB, and 1 MB data size



(b) Accumulated access time of write operation for 10KB, 100KB, and 1 MB data size



(c) Access time of read (left) and write (right) operation for 10 MB data size

Figure 9. Access time for read and write operations for different data size (The y-axis of (c) is logarithmic)

Specially, in some situations, such as write operation of 1 MB data in Figure 9(b), CloudSafe reduced the time by 21.6 percent of P scenario. This benefit mainly comes out from the parallel operations of cryptography and erasure coding.

While for the 10 MB data size, CloudSafe spends most of the time (92.5% of read and 80% of write) on cryptography, which boosts the access time heavily. This trend also occurs in EP scenario for 1 MB data size read and write. The reason of these observations is that the encryption and decryption time are not linearly proportional to the data size needed to encrypt or decrypt. This is more apparent in the EP scenario when the data size goes to 10 MB, the encryption or decryption time is around 10 minutes. When comparing to EP scenario, CloudSafe performs much better than EP scenario. For 10 KB and 100 KB data size, CloudSafe has a 31 percent overhead, but still a relatively small time. On the contrary, if the data size goes up to more than 1 MB, CloudSafe gains more benefits. For 1 MB data size, the read and write operation time is only around half of EP scenario. If the data size goes up to 10 MB, the access time of CloudSafe is only one tenth of EP's. It is time saving, which is benefited from parallel execution of cryptography. Overall, CloudSafe gains much more data security and availability than P and EP scenarios although in some situations the access time cost is twice the original or more.

Through the comparison, we see the benefits of parallel processing which can accelerate the access speed. However, the network transmission time of CloudSafe does not get too much time saving specially for small data size (10 KB and 100 KB). The reason is that CloudSafe uses HTTPS to communicate with cloud providers. The SSL connection overhead is from the same order as the data itself. Thus, when the data is small, the SSL connection overhead dominates the total transmission time even though the data size transmitted by CloudSafe is one third of P and EP scenarios. For a large data size (1 MB and 10 MB), the transmission time is reduced apparently. For the erasure coding, the time does not vary too much even if the data size is ten times or one tenth of the original for 10 KB, 100 KB and 1 MB data size scenarios. However, for small data sizes, the erasure coding time accounts for a large proportion of total operation time. For 10 MB data size scenario, the erasure coding time is doubled of 1 MB data size scenario, but the total erasure coding time is less than 2 percent of the total operation time both in read and write semantics. For the cost issue, because of that, the storage efficiency of CloudSafe is 1.33x, the average cost for per unit data is also 1.33x of the original cost. Although the cost increases 33 percent for unit data storage, the cost is worthy for the data privacy and availability provided by CloudSafe. One more limitation of CloudSafe is that compared to the Dropbox which using a delta synchronization algorithm to update the data only transmitting the data difference, CloudSafe needs to write all the data back to the storage servers for every data modification. However, the benefits of delta synchronization come from the plain text data storage, which can run the difference detection easily. If a data is encrypted, even a little change on the original data will make the encrypted data significantly different. Thus, the performance of delta synchronization may not be as good as the plain test.

5. CONCLUSION

The issue of online privacy and security is a challenging subject, as it concerns the privacy of data that are increasingly more accessible via the internet. In other words, people who intend to access the private information of other users can do so increasingly more efficiently over the internet. Although people are generally very fond of their privacy, most people do not take necessary cautions when sharing their information over social media, and in some ways facilitate the unwanted and ill-intentioned access to their information.

What today's organizations need for their clouds are integrated, comprehensive solutions that can deliver security intelligence. Advanced security intelligence solutions can close security gaps by using labor-saving automation to analyze millions of events occurring within the cloud, and discover system vulnerabilities through the normalization and correlation of these events.

Using the proposed method, a model of security, including control of user access to databases of big data with RMS, the multiplicity and the virtual machine to prevent internal threats, deleting data, insecure or incomplete data protection and control of a third-party can be provided to improve the operation according to the rules of Petri net modeling and simulation. As future work, we are trying to change the model, the objective function in ways that are cost effective and better way. Also in a comprehensive study, we will discuss the types of algorithms that present and compare them.

6. REFERENCES

1. Hamidi, H. and Daraei, A., "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases", ***International Journal of Engineering-Transactions A: Basics***, Vol. 29, No. 7, (2016), 921-930.
2. Hamidi, H. and Hashemzadeh, H., "Using a data mining tool and fp-growth algorithm application for extraction of the rules in two different dataset", ***International Journal of Engineering (IJE), Transactions C: Aspects***, Vol. 29, No. 6, (2016), 1693-1700.
3. Gharagozlou, F., Saraji, G.N., Mazloumi, A., Nahvi, A., Nasrabadi, A.M., Foroushani, A.R., Kheradmand, A.A., Ashouri, M. and Samavati, M., "Detecting driver mental fatigue based on eeg alpha power changes during simulated driving", ***Iranian Journal of Public Health***, Vol. 44, No. 12, (2015), 1693.
4. Hamidi, H., "A combined fuzzy method for evaluating criteria in enterprise resource planning implementation", ***International Journal of Intelligent Information Technologies (IJIIT)***, Vol. 12, No. 2, (2016), 25-52.
5. Hamidi, H., "A model for impact of organizational project benefits management and its impact on end user", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 29, No. 1, (2017), 51-65.
6. Johnson, R.D., Li, Y. and Dulebohn, J.H., "Unsuccessful performance and future computer self-efficacy estimations: Attributions and generalization to other software applications", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 28, No. 1, (2016), 1-14.
7. Kakar, A.S., "A user-centric typology of information system requirements", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 28, No. 1, (2016), 32-55.
8. Liu, Y., Tan, C.-H. and Sutanto, J., "Selective attention to commercial information displays in globally available mobile application", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 2, (2016), 18-38.
9. DARAEI, A. and HAMIDI, H., "An efficient predictive model for myocardial infarction using cost-sensitive j48 model", ***Iranian journal of public health***, Vol. 46, No. 5, (2017), 682.
10. Hamidi, H., Vafaei, A. and Monadjemi, S.A.H., "Analysis and evaluation of a new algorithm based fault tolerance for computing systems", ***International Journal of Grid and High Performance Computing (IJGHPC)***, Vol. 4, No. 1, (2012), 37-51.
11. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "Analysis and design of an abft and parity-checking technique in high performance computing systems", ***Journal of Circuits, Systems, and Computers***, Vol. 21, No. 03, (2012), 1250017.
12. Shadloo, B., Motevalian, A., Rahimi-Movaghar, V., Amin-Esmaeili, M., Sharifi, V., Hajebi, A., Radgoodarzi, R., Hefazi, M. and Rahimi-Movaghar, A., "Psychiatric disorders are associated with an increased risk of injuries: Data from the iranian mental health survey (iranmhs)", ***Iranian Journal of Public Health***, Vol. 45, No. 5, (2016), 623.630.
13. Hamidi, H. and Vafaei, A., "Evaluation of fault tolerant mobile agents in distributed systems", ***International Journal of Intelligent Information Technologies (IJIIT)***, Vol. 5, No. 1, (2009), 43-60.
14. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "Evaluation and check pointing of fault tolerant mobile agents execution in distributed systems", ***Journal of Networks***, Vol. 5, No. 7, (2010), 800-807.
15. Nilchi, A.N., Vafaei, A. and Hamidi, H., "Evaluation of security and fault tolerance in mobile agents", in Wireless and Optical Communications Networks,. WOCN'08. 5th IFIP International Conference on, IEEE., (2008), 1-5.
16. Wu, J., Ding, F., Xu, M., Mo, Z. and Jin, A., "Investigating the determinants of decision-making on adoption of public cloud computing in e-government", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 3, (2016), 71-89.
17. Hamidi, H. and Kamankesh, A., "An approach to intelligent traffic management system using a multi-agent system", ***International Journal of Intelligent Transportation Systems Research***, (2017), 1-13.
18. Hamidi, H. and Moradi, S., "Analysis of consideration of security parameters by vendors on trust and customer satisfaction in e-commerce", ***Journal of Global Information Management (JGIM)***, Vol. 25, No. 4, (2017), 32-45.
19. Wang, Y. and Li, D., "Virtual space co-creation: The perspective of user innovation", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 28, No. 2, (2016), 92-106.
20. Chevers, D., Mills, A.M., Duggan, E. and Moore, S., "An evaluation of software development practices among small firms in developing countries: A test of a simplified software process improvement model", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 3, (2016), 45-70.
21. Bimonte, S., Sautot, L., Journaux, L. and Faivre, B., "Multidimensional model design using data mining: A rapid prototyping methodology", ***International Journal of Data Warehousing and Mining (IJDWM)***, Vol. 13, No. 1, (2017), 1-35.

22. Esposito, C. and Ficco, M., "Recent developments on security and reliability in large-scale data processing with mapreduce", ***International Journal of Data Warehousing and Mining (IJDWM)***, Vol. 12, No. 1, (2016), 49-68.

23. Li, H., Yimin, Z., Mengshi, C., Xun, L., Xiulan, L. and Ying LIANG, H.T., "Development and validation of a disease severity scoring model for pediatric sepsis", ***Iranian Journal of Public Health***, Vol. 45, No. 7, (2016), 875-884.

An Architecture for Security and Protection of Big Data

F. Asadi Saeed Abad[a], H. Hamidi[b]

[a] *Department of Information Technology Engineering, College of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Industrial Engineering, Information Technology Group, K. N.Toosi University of Technology, Tehran, Iran*

P A P E R I N F O

Paper history:
Received 13 June 2017
Received in revised form 04 July 2017
Accepted 07 July 2017

Keywords:
Big Data
Privacy
Security Model
Petri Net
Cloud Computing

چکیده

مسئله حریم خصوصی و امنیت آنلاین، موضوع چالش برانگیزی است که به در دسترس بودن داده ها که به طور فزاینده ای در اینترنت در دسترس قرار دارند مربوط می شود. به عبارت دیگر، کسانی که تمایل به دسترسی به اطلاعات خصوصی دیگر کاربران دارند، می توانند از طریق اینترنت به این کار مبادرت ورزند. مطالعه حاضر به مسائل حریم خصوصی داده های بزرگ توزیع شده در فضای محاسبه ابری اشاره دارد. در این مقاله، مسئله حریم خصوصی با مدل شبکه پتری تحلیل شده است. چیزی که سازمانهای امروزی به آن احتیاج دارند، راه حلهایی یکپارچه و جامع است که بتواند هوش امنیت را ایجاد کنند.راه حلهای پیشرفته هوش امنیت می تواند شکافهای امنیتی را با استفاده از اتوماسیون ذخیره کار و با تحلیل میلیونها رخداد درون ابر، برطرف کند و آسیب پذیری سیستم را با نرمال سازی و همبستگی این رخداد ها کشف نماید. با استفاده از این روش، یک مدل امنیت، شامل کنترل دسترسی افراد به پایگاه داده های بزرگ با RMS، تعدد و ماشینهای مجازی، جهت جلوگیری از تهدیدهای داخلی، حذف داده ها، محافظت ناقص و ناامن داده ها و کنترل شخص ثالث، نکاتی است که در مدل پتری جهت بهبود عملیات توصیه می شود.

***doi**: 10.5829/ije.2017.30.10a.08*