Big Data Security Policy Enforcement

[1]Abdullah Al-Shomrani , [2]Fathy Eassa, [3]Kamal Jambi
Computer Science
King AbdulAziz University, Jeddah, Saudi Arabia

Abstract - Protecting big data is a major challenge facing the industries. It becomes challenging as data grows and more accessible by more clients. Large-scale data storage is becoming a necessity for healthcare, business segments, government departments, scientific endeavors and individuals. Security policy is one of these challenge that we need to manage and enforce. Security policy needs to be integrated, flexible, context-aware and customizable. The framework receive data from customer and then analyze data received, extract privacy policy and then identify the sensitive data. In this paper we will present the enforcement of the privacy policy as part of our framework.

Index terms - privacy, security, big data, policy.

I. INTRODUCTION

Big Data is a new direction of technology in science, government, industry and business [1]. The size of the datasets for big data is beyond the capabilities of typical database software to handle and manage for capturing, storing, and analyzing data. The idea was limited to the current technology available during those times of processing and storage of huge volume of data. Data is always the most important resource in our society. For example, government agencies are collecting information that contains private data. Big data security has grown fast as a significant concern for clients over the past few years. 88% of the clients were substantially worried on their data privacy [2]. The security issues are increasing and it is happening because of the increasing usage of big data through adaptation of this technology. There are many benefits of big data. Although, it is vulnerable to attacks. Attackers are consistently trying to find loopholes to attack the big data storage.

Velocity, volume and variety of data magnify the big data [3]. Volume of data is increasing every second from different input resources. Big Data is huge volume of data that is structured or unstructured and itis very difficult to handle by normal databases technologies and software tools [4]. The exponential growth in volume of data with the speed of data gathering and processing for data inputs coming from large implementation of massive connected devices such as; cars, smartphones, RFID readers, webcams, and sensor networks require huge resources to handle this size of data and ensuring the security of data. Data streams are generated by these devices continuously without human intervention. With these different sources of unstructured data, it results to have diverse types of data. The massive volume, velocity, and variety of data have a tremendous impact to existing security solutions.

Security solutions for big data were not developed and intended to handle huge volume of data such as Big Data. The past two years have generated 90% of the world's data [5]. The amount of data is doubled every two years that require having different storage strategies with different storage media [6]. Big data security and privacy are big challenges for customers and service providers [3]. Eighty percent of large organization will endure suffering from major security issues with big data by 2016[7]. Most of these are not in standard forms which make it more difficult to analyze with the available tools today [8]. In 2013 Big Data has been introduced as one of the new technology trends, but facing big challenges with security and privacy issues which threatens to slow the momentum.

There are many challenges facing big data, security and privacy are just some of these challenges. It has been reported that big data advancement increases the threats to the existing security of the data. One issue with big data privacy is policy management and how to enforce it with this huge data and not affecting the performance.

Privacy policy needs to be integrated, flexible, context aware and customizable [9]. To protect the user's privacy, many privacy solutions have been proposed. One proposed solution is to use integration of privacy protection mechanism [4]. In most of the studies it concentrates on single mechanism while this study focuses on applying more than one mechanism to secure user privacy. eXtensible Access Control Markup Language (XACML) was developed to be used as uniform control policies and has been used for privacy. XACML is an XML-based language standardized by the organization for the Advancement of Structured Information Standards (OASIS)[7].

Due to its size and volume, especially as they attempt to maximize data efficiency and performance, protecting Big Data (while it is in storage) has been a challenge for the most of organizations. The optimal data protection is to ensure that when it falls into unauthorized hands, it is meaningless. Big data security is categorized in different levels which includes; communications, processing, authentication, and storage level. Current technology was not built for big data security. Data is stored in plain text; wherein critical information can be easily stolen by hackers. Logging to critical data is not logged in which means any abused of data cannot be identified. This technology is used by most businesses to store and analyze their own data and their customer's data which makes privacy and security very important in gaining the confidence of customers. Hence, there is a need for investing, studying, and understanding the challenges and providing better solutions to secure big data.

Big data security is a major challenge in the field of research for scientists. Our research will focus on big data storage and enforcing the privacy policy. First, we will focus on how to protect data in storage. The use of encryption, fragmentation, authentication and secure communication within our framework will secure the big data storage which will be enforced based on the policy identified by client.

Today, business depends on customer's data protection. Data about health, shopping, purchasing and traveling iscontained collected digital information. The data are collected by businesses to aid in decision making and help to attain their goals. They analyze the data and use it to direct their customers in the way that achieve their goals. While this could be used legally, but sometimes it could be illegal. In 2014, 92% internet users in the US were worried regarding their online privacy while 89% avoid dealing with companies on the cloud as they feel that their data is not secure.

West in [10] defined privacy as "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." Big data security and privacy are big challenges for customers and service providers [3]. Eighty percent of large organization will endure suffering from major security issues with big data by 2016[6]. Most of these are not in standard forms which make it more difficult to analyze with the available tools of today [7].

Beneficial information can be extracted and analyzed from digital information collected from different sources such as credit card companies, government agencies, banking, and health care. Regardless of the usage of the information and the benefit, privacy and security is the main challenge in big data.

Privacy policy languages help in different phases of managing the privacy policy which from writing, reviewing, approving, issuing and finally enforcing policy [18].

II. Related Work

The privacy problem is main concern these days that comes because of the huge amount of personal information available on internet [11] as digital information. Managing security policy is a challenge that our framework will handle for big data. Privacy policy needs to be integrated, flexible, context-aware and customizable Hassan et al. [12] introduced Policy Management as a Service (PMaaS), a framework for cloud based policy management is designed, so the users can control their policies to access the resources of where they are stored. The framework consists of cloud user who manages the access policies on the policy management service provider. In addition, it also includes a policy management service provider which defines, edits and manages the access policies.

The cloud service provider which is the third component controls the access to the protected resources based on the policies identified by the cloud users. The last one is the requester which is an application which cooperates with a cloud service provider to be able to access the protected resource belonging to the specific cloud user.

Kudakwashe Zvarevashe et al. [5] listed security cases for big data starts with securing computations which may also return wrong result and at the same time, it is difficult to discover whether the data is faulty or not. The second one is the security for non-relational data stores where the security depends on the middleware. The next security case is securing data storage and how we can protect unauthorized access to data. Then they explain the importance of validating, the input devices. Real time security monitoring is a change in the area of big data security. Privacy of clients during data mining is another important case for big data security. Securing the data from end to end is one of the big data security cases.

In this paper they recommend solutions to these challenges to big data security which uses Kerberos system for authenticating, both users and services to ensure privacy between communicating protocol and users on the Internet,

the Transport Layer Security (TLS) is utilized. Data encryption is one way to secure data in storage with having a key encryption saved and available to reuse the data when needed. Secure and validate any software needed to be deployed.

A policy-compliant data processing framework with the following modules was proposed:

Policy-reasoning module which maps a policy to a specific set of tasks.

Data processing task rewriting module: enforces policies by considering different rewriting strategies. Pre-processing module: executes the pre-processing tasks on the underlying data.

Post-processing module: to process the final results to enforce policies.

Most of security tools fail to work with big data environment, so these tools need to be able to scale with the data [14]. The existing big data security solutions which are not developed and intended for huge volume, velocity, and variety cannot protect this amount of data [15]. Big data security is adapting the traditional security tools. They recommend placing the security closer to the data storage instead of having the security at the data center level. It is not bound to the number of policy sets, policies and rules and they showed the enhancement of performance by applying their way. Meanwhile, it still needs more work for better XACML performance.

In [16], for the convenience for the providers of cloud data storage, a security agenda has been offered to implement the security strategies and policies with the help of policy management module. This allows the system to assess the activities of users and analyze them for every system and finally categorized as fair" or "malicious". To make sure that system is going to work according to the requirements and set policy definitions, a strategy has been enforced for security maintenance.

III. SOLUTION ARCHITECTURE

This section details the security policy management framework that efficiently enforces the sensitive data security policy. The policy enforcement framework must be highly flexible and must support different data processing requirements. To implement these modules efficiently, it needs to be specialized for different data types, computation types and policy requirements It is built

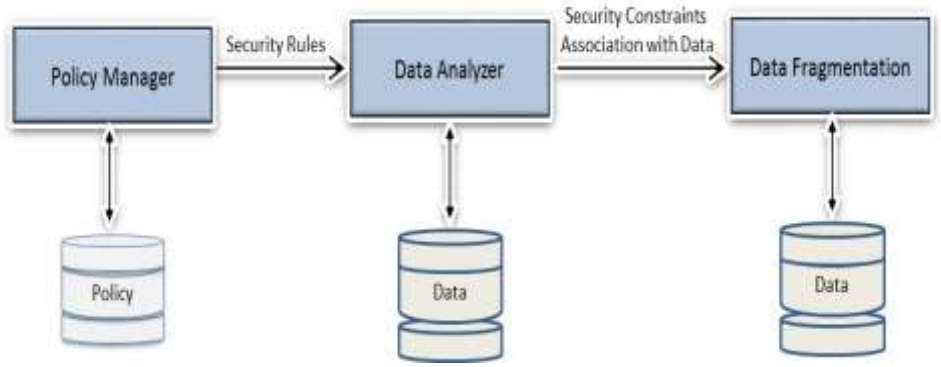

Fig1. the high-level view of the framework

on the concept of data distribution on mulita cloud service providers. Figure1 shows a high-level view of the framework which includes the main components. The first component is the "Policy Manager" which receives data from client and then extracts the security rules and then unify the policy and store the policy in data base. Figure2 illustrates the framework and its components in more detail.

Policy Manager is the manager of the security policies that enables the service provider to edit and manage the security policies, which have four components.

The policy manager is the core component of the security manager. It provides security by keeping track of promises the involved parties making access to data, along with controlling access to such data. The data is encrypted and is only accessible upon the acceptance and satisfaction of specified constraints and duties imposed by the policies.

Data owner uses different service providers to store big data. The data owner is in charge of defining security policies which will be used by service providers to identify sensitive data when a requester attempts to store data It is built on the concept of data distribution on multi cloud service providers. Once the user is granted then it will pass to the Policy Manager which have four components:

- Policy Administration Point (PAP) which is responsible for creating and updating policies.

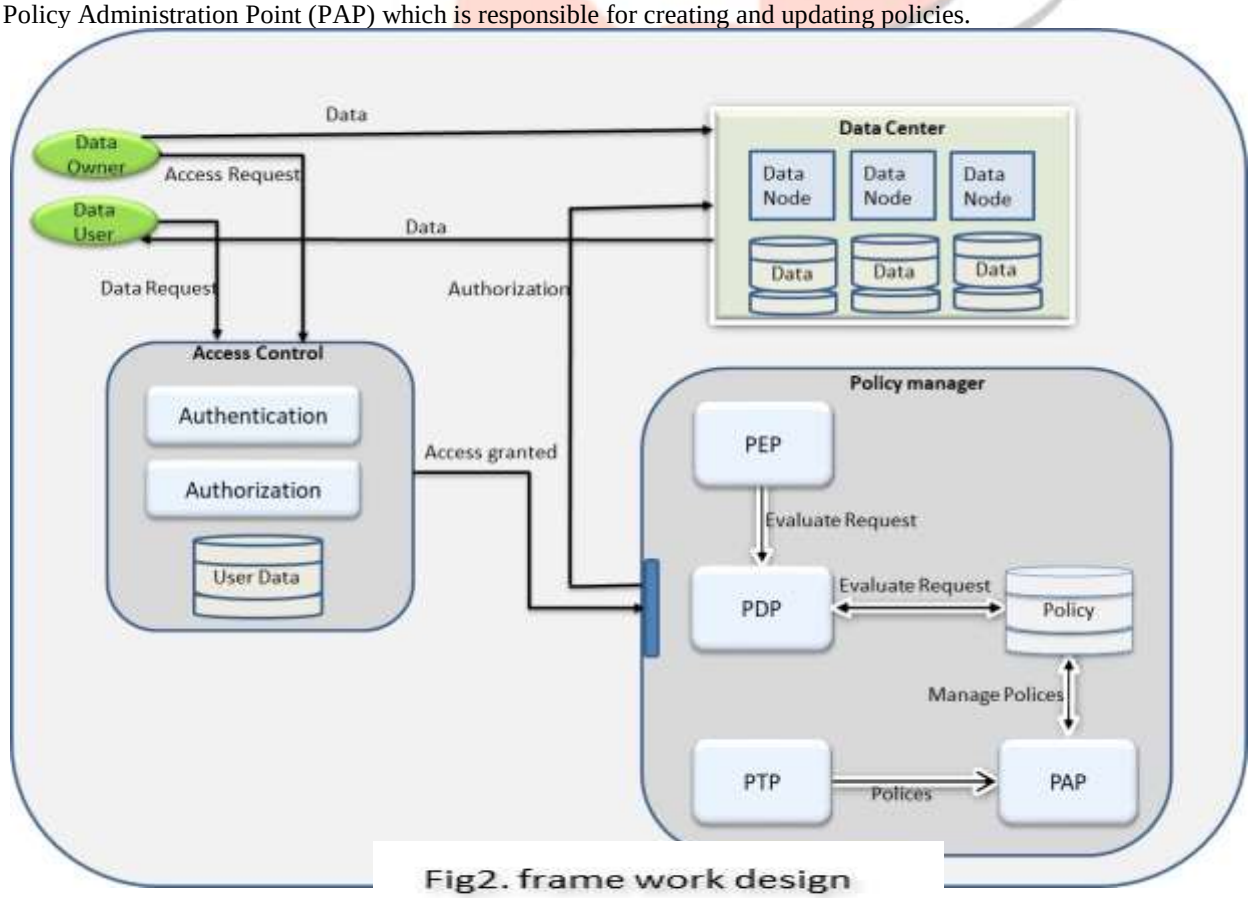

Fig2. frame work design

- Policy Enforcement Point (PEP) passes the request to Policy Decision Point and it applies the retrieved policy.
- Policy Decision Point (PDP) which searches for the policy and evaluates the request. It is in charge of making access decisions.
- Policy Translation Point (PTP) translates the owner policy to unify policy format.

The cloud users can specify their policies in controlled natural language which in turn are translated by the PTP into unify readable policy language. The output policy language is XACML.

At the beginning, the policy manager adapter is the entryway to the policy manager, receiving requests from the data owner/user and sending the response to them. The Policy manger receives the security constraints then analyzes the data and extracts the security rules. Security specification is passed to the Policy Translation Point (PTP) which translates the policy to unify format XACML. Security policy can be written based on the data type. It identifies the rules to be able to identify the sensitive and non-sensitive data. The client specifies the security rules, due to that; they know the data and the sensitivity of it. The sensitive association rules are a special group of association rules which represents the sensitive data. Once the 'Extract Security Policy' identifies the security policy, it will update the policy database.

These security rules are important for making decision and need to remain private (i.e., the rules are private to the company or organization owning the data). The Policy Decision Point (PDP) receives the request to lookup for the security policy for certain user. If it is found, it will be sent to the Policy Enforcement Point (PEP). The PEP will send the security constrains to the data analyzer which will identify sensitive and non-sensitive data based on the security constrains. It can specify their policies as part of their data which are translated by the policy management into XACML language. The framework does not inflict any restrictions on security policies specified by clients and our aim here is to allow clients to specify their policies without any restriction. Once the security policies are determined in client format, the PTP parses the policy, identifies policy elements, and transforms the policy into XACML. Access controls over the stored data are provided through a XACML policy-based security. The policies are separated based on client, using the association between client and the security policy. They do not need to deal with various policies beyond this system. The following algorithm is used to extract the security policy and identify the sensitive data:

Input:
D: Data to be store
C= {c1, c2,……ci}/ a set of security constraints*/*
Output:
*S ={s1,s2,…….sm} /*Security policy*/*
RS ={rs1,rs2,…..,rsk} / association between data and Security policy */*
Main:
get data
identify data type
extract security constraints (C)
associate Security policy to data (RC)
update meta data

This algorithm extracts policy rules which it is part of the data to be stored. Once the rules are extracted then the association between data and rules is done. Additionally, there are other components as part of the Policy Manager, which are:

Policy database is the database where the policy rules can be stored and retrieved. Each policy has an identifier (PID) mapping between the security policy and data.

IV. CONCLUSION

We presented a technique for the privacy policy as part of the big framework for big data security. The security framework for big data consists of many techniques which are security policy manager, fragmentation approach, encryption approach and security manager. The future work will present the remaining part of our framework.

REFERENCES

[1] Yuri Demchenko, Canh Ngo, Peter Membrey; "Architecture Framework and Components for the Big Data Ecosystem" ; SNE technical report SNE-UVA-2013-02; 2013.

[2]] "Top ten big data security and privacy challenges" Cloud Security Alliance, November 2012, https://www.isaca.org/Groups/Professional- English/big- data/GroupDocuments/Big_Data_Top_Ten_v1.pdf

[3] Venkata Narasimha Inukollu1 , Sailaja Arsi1 and Srinivasa Rao Ravuri, " Security Issues Associated with Big Data in Cloud " International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014

[4] Kudakwashe Zvarevashe1, Mainford Mutandavari2, Trust Gotora3; "A Survey of the Security Use Cases in Big Data ";International Journal of Innovative Research in Computer and Communication Engineering; Vol. 2, Issue 5, May 2014.May 2014

[5] Klaus Engelhardt, "Secure data storage an overview of storage technology", http://primera.eu/en/brochures/wp_ke_080816_uk.pdf, 2008.

[6] Chris Marrison ,"Gartner warns of big data security problems", Network Security, Volume 2014, Issue 6, June 2014.

[7] Elmustafa Sayed Ali Ahmed1 and Rashid A.Saeed, "A Survey of Big Data Cloud Computing Security," international Journal of Computer Science and Software Engineering (IJCSSE), Volume 3, Issue 1, December 2014 Page: 78-85

[8] Neha Upadhyay , Ajay Kumar, "A Framework based on Authentication and Authorization to ensure Secure Data Storage in Cloud", International Journal of Computer Applications (0975 – 8887), Volume 90 – No 15, March 2014.

[9] P.Kamakshi1 , "Survey on big data and related privacy issues", International Journal of Research in Engineering and Technology, Dec- 2014,Volume: 03 Issue: 12.

[10] A. Westin, "Privacy and Freedom", Washington and Lee Law Review, Volume 25, Issue 1.

[11] R. Agrawal, R. Srikant, "Privacy-preserving data mining", In: Proceedings of the 2000ACM-SIGMOD on management of data, Dallas, TX, USA, May 15-18, 2000

[12] Hassan Takabi, James B. , D. Joshi "Policy Management as a Service:An Approachto Manage Policy Heterogeneity in Cloud Computing Environment", 2012 45th Hawaii International Conference on System Sciences.
[13]] K. W. Hamlen, L. Kagal, and M. Kantarcioglu, "Policy enforcement framework for cloud data management.," IEEE Data Eng. Bull., vol. 35, no. 4, pp. 39–45, 2012.
[14] Adrian Lane, Securosis ,"Securing Big Data: Security Recommendations for Hadoop & NoSQL Environments", October 12, 2012, http://securosis.com
[15] The Big Data Security Gap;http://www.zettaset.com/wpcontent/uploads/2014/04/zettaset_wp_security_0413.pdf
[16] C. Basescu, A. Carpen-Amarie, C. Leordeanu, A. Costan, and G. Antoniu, "Managing data access on clouds: A generic framework forenforcing security policies," in Advanced Information Networking and Applications (AINA), 2011 IEE
[17] Fujitsu. Personal data in the cloud: A global survey of consumer attitudes. Technical report, Fujitsu Research Institute, 2010
[18] Moses, T. eXtensible Access Control Markup Language (XACML) Version 2.0. Tech. rep., Oasis, Retrieved June 17, 2005, http://xml.coverpages.org/XACMLv20CDCoreSpec. pdf, 2004.