International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 10
September 2017

Secure User Generated Content Sharing with Security and Data Processing in Big Data Application

T. Satya kiranmai & Swetha Koduri

Assistant Professor, Computer science and engineering CMR College of Engineering and Technology, Kandlakoya, Medchal, Telangana.
Assistant Professor, Information Technology, Malla Reddy College of Engineering and Technology, Maisammaguda, Dhulapalli,Telangana
tadepallikiranmai84@gmail.com & koduriswetha@gmail.com

Abstract— *The most important aspects in how computing infrastructures should be configured and intelligently managed to fulfill the most notably security aspects required by Big Data applications. One of them is privacy. It is a pertinent aspect to be addressed because users share more and more personal data and content through their devices and computers to networks and public clouds. So, a secure framework to social networks is a very hot topic research. In addition, the traditional mechanisms to support security such as firewalls and demilitarized zones are not suitable to be applied in computing systems to support Big Data.*

Keywords— Big Data, Data Security, Data Privacy, Cloud Computing;

INTRODUCTION

The Big Data is an emerging area applied to manage datasets whose size is beyond the ability of commonly used software tools to capture, manage, and timely analyze that amount of data. The quantity of data to be analyzed is expected to double every two years. All these data are very often unstructured and from various sources such as social media, sensors, scientific applications, surveillance, video and image archives, Internet search indexing, medical data, business transactions and system logs. Big data is gaining more and more attention since the number of devices connected to the so-called “Internet of Things” (IoT) is still increasing to unforeseen levels, producing large amounts of data which needs to be transformed into valuable information. Additionally, it is very popular to buy on-demand additional computing power and storage from public cloud providers to perform intensive data-parallel processing. In this way, security and privacy issues can be potentially boosted by the volume, variety, and wide area deployment of the system infrastructure to support Big Data applications. As Big Data expands with the help of public clouds, traditional security solutions tailored to private computing infrastructures, confined to a well-defined security perimeter, such as firewalls and demilitarized zones are no more effective.

BIG DATA CHALLENGES TO INFORMATION SECURITY AND PRIVACY

With the proliferation of devices connected to the Internet and connected to each other, the volume of data collected, stored, and processed is increasing everyday, which also brings new challenges in terms of the Data security. In fact, the currently used security mechanisms such as firewalls and DMZs cannot be used in the Big Data infrastructure because the security mechanisms should be stretched out of the perimeter of the organization's network to fulfill the user/data mobility requirements and the policies of Bring Your Own Device. Considering these new scenarios, the pertinent question is what security and privacy policies and technologies are more adequate to fulfill the current top Big Data privacy and security demands. These challenges may be organized into four Big Data aspects such as infrastructure security, data privacy, data management and, integrity and reactive security.

Solutions for Big Data Security and Privacy Issues

There is no single magical solution to solve the identified Big Data security and privacy challenges and traditional security solutions, which are mainly dedicated to protect small amounts of static data, are not adequate to the novel requisites imposed by Big Data services. There is the need to understand how the collection of large amounts of complex structured and unstructured data can be protected. Non-authorized access to that data to create new relations, combine different data sources and make it available to malicious users is a serious risk for Big Data. The basic and more common solution for this includes encrypting everything to make data secure regardless where the data resides. As Big Data grows and its processing gets faster, then encryption, masking and tokenization are critical elements for protecting sensitive data. Due to its characteristics, Big Data projects need to take an holistic vision at security.

Big Data need to take into consideration the identification of the different data sources, the origin and creators of data, as well as who is allowed to access the data. It is also necessary to conduct a correct classification to identify critical data, and align with the organization data security policy in terms of enforcing access control and data handling policies. As a recommendation, different security mechanisms should be closer to the data sources and data itself, in order to provide security right at the origin of data, and mechanisms of control and prevention on archiving, data leakage prevention and access control should work together.

In terms of security and privacy, it becomes mandatory that mechanisms that address legal requirements about data handling, need to be met. Secure encryption technology must be employed to protect all the confidential data like Personally Identifiable Information (PII), Protected Health Information (PHI) and Intellectual Property (IP) and careful cryptographic material (keys) access management policies, need to be put in place, to ensure the correct locking and unlocking of data – this is particularly important for data stored. In order to be successful these mechanisms need to be transparent to the end-user and have low impact of the performance and scalability of data.

Overall system architecture

In this architecture there are some elements that cooperate in order to provide the necessary functionalities to both the end-users and the network platform, in order to implement the necessary mechanisms to provide security and privacy to user generated content. Some of the services are deployed on the server-side while other are implemented on the user-side.

On the **user-side**, the **authorization service** handles the requests to render some type of content on the user device, processing the requests and matching them to existing credentials, licenses and permissions to render the content. Also, on the end-user side the **content rendering service** is responsible for verifying the necessary requirements to render the content and effectively renders the content for the end user.

On the **server-side**, is where a large part of the rights management responsibility lies. A set of components with a well-defined API that allows integration between the necessary ones to implement the specific content business model. These services are the following:

- **Content storage and distribution service**: this service is responsible for the storage and distribution of user generated content in a protected manner;
- **Content protection service:** the service is responsible for the protection of the content. The content is protected by specific protection tools and specific protection mechanisms that may change according to the content and the business model that is going to be implemented;
- **Content registration service:** this service is responsible for registering the content on the platform that will be used to uniquely identify the content on the system. This

unique identifier is used to identify the user generated content throughout the entire content lifecycle;

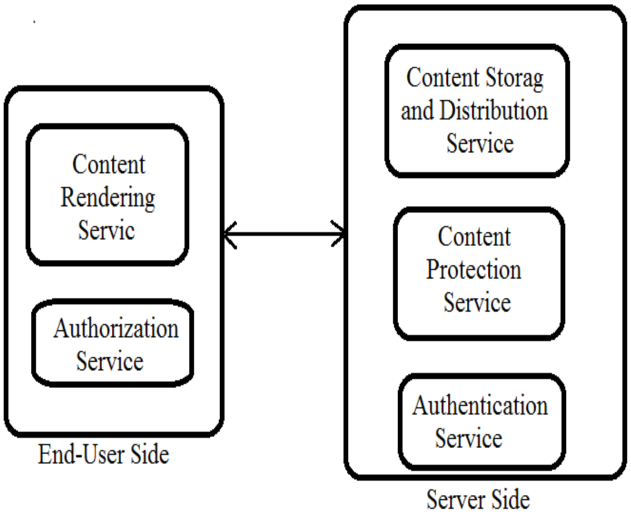


Fig:- Overview of the architecture integration

Registration on the platform:-

This novel platform presupposes that all the system services are initially registered on that platform. This means that each one of the different services, either server-side or client-side have to be individually registered at the platform. This registration process assigns unique credentials to each one of the services, ensuring that they are uniquely registered and that these credentials will be used to identify and differentiate the services in future interactions. This registration process is conducted by the authentication service that on its turn issues credentials to all the other services and acts as a central trustworthy mechanism.

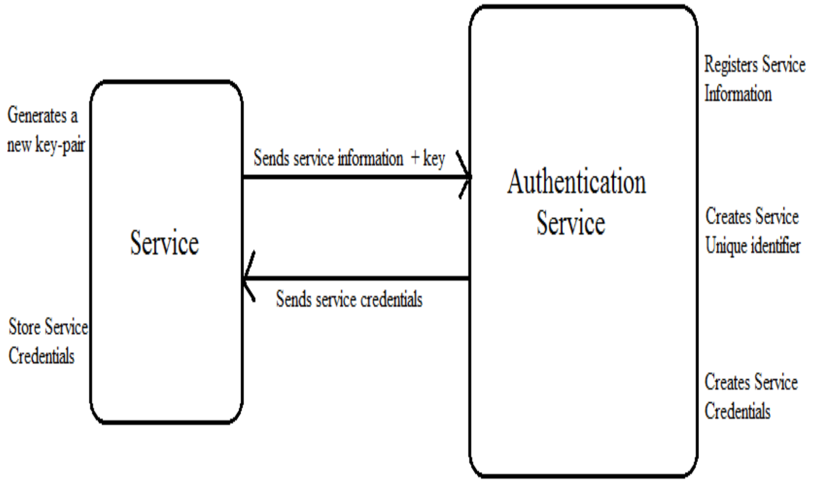

Generates a new key-pair
Store Service Credentials
Service
Sends service information + key
Sends service credentials
Authentication Service
Registers Service Information
Creates Service Unique identifier
Creates Service Credentials

Fig:- Handling the registration of new services

1. The authentication service (AS) has cryptographic material (K_{pub}^{AS} , K_{priv}^{AS}) and credentials that were self-issued (C^{AS}_{AS}) or issued by other trustworthy entity (C^{CA}_{AS});
2. The service that needs to be registered generates a key pair (K_{pub}^{S} , K_{priv}^{S}) and sends a registration request to the AS, passing some information about the service (S_{info}) and the public key (K_{pub}^{S}) of the service: $S_{info}+ K_{pub}^{S}$;
3. AS receives this information, verifies it and then creates a unique service identifier (SUUID). After this verification the AS creates the service credentials that will identify this service globally and uniquely on the platform: $C^{AS}_{S[UUID]} = K_{priv}^{AS}\{S_{UUID}, K_{pub}^{S[UUID]},C^{AS}_{AS}\}^{2}$. These credentials, which are signed by AS, are then returned to the requesting service;
4. The requesting service, stores the credentials. This credential contains also the public key of the authentication service (K_{pub}^{AS}). This is used to prove this credentials to other entities that also rely on the same AS – services that trust AS, also trust on credentials issued by AS, presented by other services.

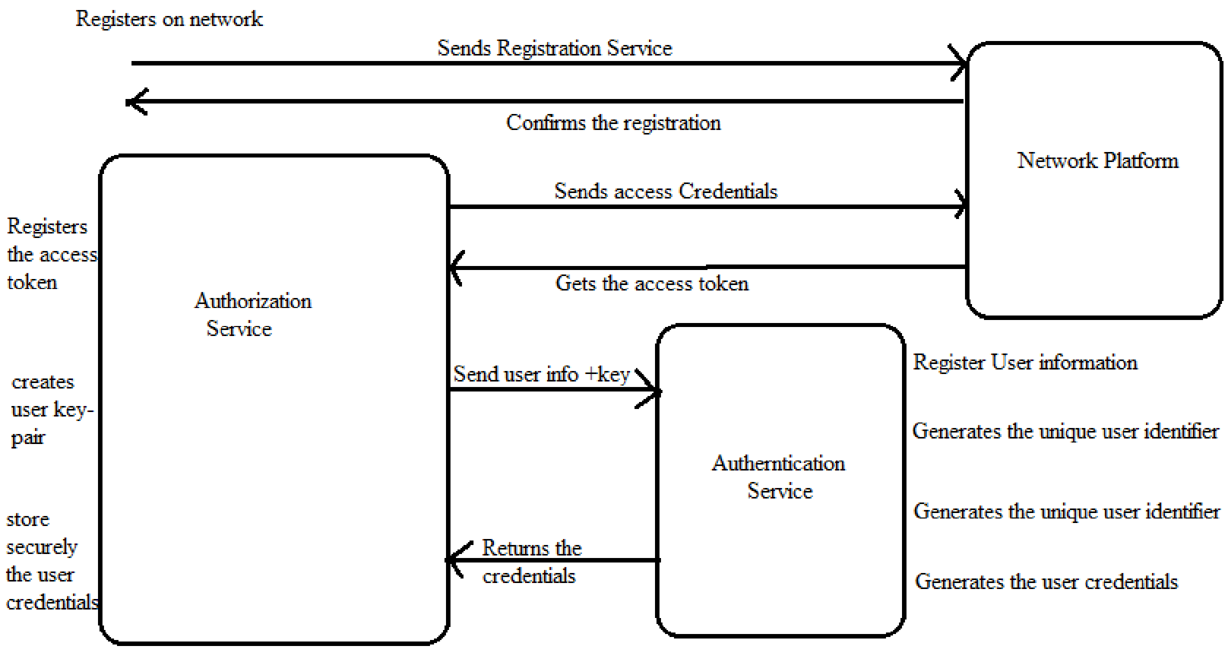
The service registration process, as described above needs to be repeated according to the number of services available within the social network platform. This enables the entire ecosystem of services to be trusted on that platform.

Another important aspect of the registration process concerns the registration of the users on the rights management platform. The registration of the user on the rights management platform can be dependent or independent of the social network platform. In the example that is presented here, it is assumed that this registration process is performed fully integrated with the social network platform.

6.

This process performs in the following manner:

1. Assuming that the user still has no account on a social network platform, the user starts the registration process on the social network. In order to do that the user needs to supply its email address (as username) and a password;
2. The registration process on the social network platform finishes and a confirmation message is sent to the end-user;
3. Next, the user, using the client-side rights management authorization service (AUTS), initiates the registration process in the rights management platform. The AUTS presents several registration options to the end-user (integrated with some social network platforms -using either Oauth- or an independent mode). For this case, the user will use registration options by using the mode of integrated authentication;
4. The user introduces the social account credentials (email, password) on the AUTS that starts the authentication process on the social network platform. If successful, the social network returns an access token that has a specific validity and a set of permissions to conduct different operations on the social network on behalf of the user;
5. AUTS, using the user credentials (email, password) creates a secret key that is used to initialize a secure storage on the authorization service: $S_{k}^{SStorage}$= SHA1[email + password];


Registers on network
Sends Registration Service
Confirms the registration
Network Platform
Sends access Credentials
Gets the access token
Registers the access token
Authorization Service
creates user key-pair
store securely the user credentials
Send user info +key
Autherntication Service
Returns the credentials
Register User information
Generates the unique user identifier
Generates the unique user identifier
Generates the user credentials

7. The AUTS securely stores the user information, and the social network access token. Additionally, the AUTS creates a key-pair for the user (K_{pub}^{U} , K_{priv}^{U}) also storing it in a secure manner: $S_{k}^{SStorage}$ (K_{pub}^{U} , K_{priv}^{U} , user info, token);
8. AUTS contacts the AS to register the user on the platform. This is performed using the $C^{AS}_{S[AUTS]}$ that contains the K_{pub}^{AS}. $C^{AS}_{S[AUTS]}$ is also sent to ensure that the AUTS has been previously registered: K_{pub}^{AS}(email, K_{pub}^{U} ,$C^{AS}_{S[AUTS]}$);
9. The AUTS receives all this information and after deciphering it, and validating the AUTS credential, registers the user information, generates a unique identifier for the user and creates credentials for the user: $C^{AS}_{UUID}= K_{priv}^{AS}$ {UUID, K_{pub}^{U}};
10. The credentials are returned to the AUTS and are securely stored: $S_{k}^{SStorage}$ (C^{AS}_{UUID}). The user is notified about the result of the registration operation.

Sharing content on the platform:-

The other important functionality on the system is the sharing of user generated content (UGC) on the social network. This sharing mechanism is performed through the rights management platform, and the content is stored securely on a configured location (it can be on a specific storage location, on the social platform or on the rights management platform). When the user uploads user generated content, the content is protected and the rights, permissions and restrictions about the content can be defined by the user.

The user generated content is uploaded to the rights management platform, the access rights and permissions are defined by the user, the content is protected, and a URI is returned to be shared on the social network platform. The novel content sharing process, using the mechanisms described in this chapter, can be now defined in the following steps:

1. The user sends the user generated content (UGC) that it expects to share on the social network. This UGC is uploaded through the content rendering service (CRS). This service requires the user to enter its credentials (email and password), if the user is not yet authenticated.

These credentials are used to access the secure storage: $S_k^{SStorage}$ = SHA1[email+ password];

2. The CRS contacts the AUTS, which reads from the secure storage the user rights management system credentials: C^{AS}_{UUID};
3. The CRS uploads to the content protection service (CPS) the UGC and sends the user credentials, obtained in the previous step: UGC_{UUID}, C^{AS}_{UUID};
4. The CPS, after retrieving some metadata information about the UGC (such as the type, the format, the encoding, among others), contacts the protection tools service (PTS), requesting a list of available protection tools, that can be suitable to protect the UGC. The PTS sends its credentials and some information about the content: C^{AS}_{CPS}, UGC_info;
5. The PTS also returns a list of protection tools that match the request made by the CPS. This information is signed by PTS: K_{priv}^{PTS}{protection_tools_list};
6. The CPS returns the list of protection tools to the CRS, and presents it to the user. The user selects the most appropriate protection tools, adjusting the parameters of applicability of the tools to the UGC and submits its request about the necessary protection tools;
7. The CPS requests the selected protection tools from the protection tools service. The PTS returns the requested tools to the CPS;
8. Next, the CPS requests to the content registration service for the UGC to be registered. For this, the CPS send its credentials, the UCG metadata and the content hash: C^{AS}_{CPS}, UGC_info, SHA1[UGC];
9. The content registration service (CRGS), stores the received information, and generates a unique content identifier that is returned to the content protection service: K_{priv}^{CRS}.{ UGC_{UUID} };
10. The CPS generates one or more content encryption keys (CEK[1], CEK[2] … CEK[n]) that are applied over the UGC, using the selected protection tools, in order to ensure the appropriate content protection;
11. Following this protection process, the CPS sends the content encryption keys for registration at the licensing service. Each of the content encryption keys is protected with the user key, and the entire message is protected by the CPS key: C^{AS}_{CPS}, K_{pub}^{CPS} (K_{priv}^{u} (CEK[1], CEK[2] … CEK[n]), UGC_{UUID});
12. The licensing service (LS) after validating all the received information, returns a list of licensing templates to the content protection service. The CPS returns the list of licensing templates to CRS, and the user can select the most appropriate license template, modify it and adapt it, or simply create a new one;
13. The license template ((LIC_{TPL}) is sent to the CPS that after sends it to the licensing service and associates it with the identifier of the UGC: (LIC_{TPL}, UGC_{UUID}. The licensing service returns the license template identifier (LIC_{TPL} [UUID]);
14. In the next stage, the CPS sends the protected UGC to the content storage and distribution service that stores the encrypted content: C^{AS}_{CPS}, K_{priv}^{CPS} {CEK[n](UGC), UGC_{UUID} };
15. The content storage and distribution service returns a URI for the location of the stored encrypted UGC. This URI is returned to the user that can share it on the social network platform afterwards.

CONCLUSION

The most important security and privacy challenges that affect Big Data projects and their specificities. Although the information security practices, methodologies and tools to ensure the security and privacy of the Big Data ecosystem already exist, the particular characteristics of Big Data make them ineffective if they are not used in an integrated manner. This chapter also presents some solutions for these challenges, but it does not provide a definitive solution for the problem. It rather points to some directions and technologies that might contribute to solve some of the most relevant and challenging Big Data security and privacy issues. Next, two different use cases were presented. Both of the use-cases present some directions that contribute to solving part of the large Big Data security and privacy. First, it was presented an approach that tries solving Data security and Data privacy problems on network user data. In current approach, an open an interoperable rights handler system was suggest as a way to increase the data privacy of data users that share data over the networks. The current management system keeps the data users on the handle of their own user generated content, and how they prevent misuse from either other data users or the network platform itself. A novel security proposes a novel feedback loop to increase the handle of defense mechanisms of network architecture, and is centered on five core aspects: Data protect, sense, adjust, data access, counter.

REFERENCES

[1] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Security assurance for averting information over-gathering in savvy city," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1339–1350, 2015.

[2] M. Qiu, L. Chen, Y. Zhu, J. Hu, and X. Qin, "Online information designation for crossover recollections on inserted tele-wellbeing frameworks," in 2014 IEEE eleventh Intl Conf on Embedded Software and Syst (ICESS), Aug 2014, pp. 574–579.

[3] M. Qiu and E. H.- M. Sha, "Cost minimization while fulfilling hard/delicate timing limitations for heterogeneous implanted frameworks," ACM Transactions on Design Automation of Electronic Systems, vol. 14, no. 2, p. 25, 2009.

[4] J. Niu, C. Liu, Y. Gao, and M. Qiu, "Vitality productive undertaking task with ensured likelihood fulfilling timing limitations for implanted frameworks," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 8, pp. 2043–2052, 2014.

[5] K. Gai and S. Li, "Towards distributed computing: a writing survey on cloud figuring and its advancement patterns," in 2012 Fourth Int'l Conf. on Sight and sound Information Networking and Security, Nanjing, China, 2012, pp. 142–146.