Privacy Preservation in Big Data using Web Log Analyzer and Attribute Based Encryption

VINIT KRISHNANKUTTY [1], Tanvir Sajal [2], and Jinan Fiaidhi [2]

[1]Lakehead University
[2]Affiliation not available

October 30, 2023

Abstract

Big Data faces many challenges with respect to the security while storing them on a cloud server. There is high chance of getting data viewed by the hacker and the server for performing various operations. In order to provide high level of confidentiality and integrity, a new Encryption technique is introduced known as Attribute Based Encryption (ABE), which instead of making use of the receiver's public key for encryption, uses various attributes. As a future scope, ABE can be combined with Homomorphic Encryption (HE) to provide a secure transfer of the identity. In order to provide more privacy of data, web log Analyzer is used to find out the loopholes and the unauthorized access to the web data

Krishnankutty, Vinit

Master's in Computer

Science

Student Id: 1096016Email ID: krishnankuttyv@lakeheadu.ca Lakehead University

Sajal, Tanvir Hasan

Master's in Computer Science

Student Id: 1117023 Email ID: tsajal@lakeheadu.caLakehead University

Dr. Fiaidhi, Jinan

Professor, Department of Computer Science Email ID: jfiaidhi@lakeheadu.caLakehead University

ABSTRACT

Big Data faces many challenges with respect to the security while storing them on a cloud server. There is high chance of getting data viewed by the hacker and the server for performing various operations. In order to provide high level of confidentiality and integrity, a

new Encryption technique is introduced known as Attribute Based Encryption (ABE), which instead of making use of the receiver's public key for encryption, uses various attributes. As a future scope, ABE can be combined with Homomorphic Encryption (HE) to provide a secure transfer of the identity. In order to provide more privacy of data, web log Analyzer is used to find out the loopholes and the unauthorized access to the web data.

Keywords: IBE, ABE, KGC, HE, RSA, SHA, MD5, IP,sql,xss

INTRODUCTION

Introduction to Big Data

Big data helps to grab knowledge by understanding the patterns, associations and trends. The data mentioned as a big data can be either structured data or unstructured data. In terms of big data, quality is given more importance than the quantity. The main idea behind gathering of data, processing them and perform analysis is to get important information about the pattern of data. In order to perform that big data analysis is performed. The definition of big data by the analysts is in terms of three V's, which are volume, velocity and variety. Volume indicates the quantity of data that are collected by different organizations and companies. These data include transaction data, machine and other device data. In order to handle huge volume of data, Hadoop is used in the industries. Velocity indicates the rate at which the data is getting transferred from one source to other. In order to control the data flow thereby minimizing the unauthorized access by the third party, data handling and security measures are to be incorporated. Variety is defined as different types of formats, which can be structured mathematical data, documents in the form of texts, video, audio, email.

Big Data Challenges

The challenges with the big data [3] are the methods and techniques that are framed for the security of small size data is

not practical for large size data. The most relevant challenges of big data in the modern society are volume, variety and combination of many datasets, velocity, relevance and quality, security, privacy, scalability. The volume of data in the recent years are getting increased in terms of geometric progression. Hence scientists use the word as data explosion. The data is expected to reach zeta bytes. The main source of data is generated by the social media and mobile phones. The most common types of data that are available for data analysis is in the form of unstructured or semi structured formats. The data is really hard to be obtained in the form of structured data. The complexity of data increases the increase in the quantity of data. The data come from different sources such web pages, documents which are in the text, audio, video formats, emails and multimedia data. One of the most common challenges is to analyze the source of data and how to control the flow of information. The data flow happens in excess and can be accessed by any person. In terms of quality, the data analysis can be performed in an effective manner if and only if the data has clarity. The machine learning algorithms such as supervised, unsupervised and reinforcement learning can be implemented only if the data that is analyzed has high quality. If quality of data is compromised, then the performance rate of data analysis will get affected. The data once released should be accesses by the authenticated users, hence encryption should be performed to prevent the unauthorized access of data by the third party. In order to maintain the integrity of data, hashing is to be implemented using different hash algorithms such Secure Hash Algorithms (SHA-1), SHA-512, Message Digest (MD5) etc. The scalability of data is another challenge faced by big data in today's world. Dynamically upon demand from the users, scaling up and scaling down of large volume of data is highly crucial. The projects based on big data should be able to analyze the direction of progress of data. Hence the other resources that are also to be included as a part of the project should have space.

Existing General Solutions to overcome Big Data Challenges

Big data and its Storage have created multiple privacy and security threats. The interruption from unauthorized third party and the server viewing the data each time for processing

them leads to a high chance of vulnerability. This creates a way for the data to get hacked. Hence there is no privacy and security for the data, which leads to various ethical issues. This needs to be handled legally. Big Data faces many challenges with respect to the security while storing them on a cloud server. There is high chance of getting data viewed by the hacker and the server for performing various operations. In order to provide high level of confidentiality and integrity, a new encryption technique is introduced known as ABE, which instead of making use of the receiver's public key for encryption, uses various attributes. In order to provide more privacy of data, web log Analyzer is used to find out the loopholes and the unauthorized access to the web data. Based on the base research works done on how to secure the data in an effective way, two solutions are put forward. First solution is, in order to provide security while data is stored and accessed by cloud users, encryption is performed. Encryption provides confidentiality of data. In order to perform a real time monitoring of web data access and to find the types of attack, there is web log analysis. There are certain general potential solutions for privacy and security challenges. The solutions are cloud providers have to be examined in a proper way. One of the most appropriate way to store large volume of data is in the cloud. But the cloud should be made secure so that only authorized users will be given the permission to access the data. In order to ensure that the cloud providers should be timely monitored, and enough security audits are to be conducted. The necessary control policies in terms of access should be included to maintain the integrity and confidentiality of data. The data should be protected using encryption mechanisms. Hence the different stages of data such as from the stage of data collection till data analytics will be secure. Data while getting transferred from source to destination should be protected. The access of data should be monitored on real time basis. The frequent threat analysis should be performed to make sure that data is not leaked. The mechanism of data anonymization should be included so that from the dataset, the most confidential information will be hidden from the common users. Threat intelligence mechanism is to be included to make sure that security monitoring in terms on real time basis is performed. Authentication and authorization methods are to be performed in to prevent illegal log in and access to the applications and the data associated. The effective key management scenarios are to be included such that while encryption using symmetric or public key technique is used, the keys can be shared in a proper manner. In normal cases for the ease of access the key will be stored in the disks drives which are local. The activity logs are to be analyzed on regular basis such that unwanted login attempts and unusual access to data will be registered in the activity log and it can be identified. In order to incorporate secure data communication, a secure network is highly essential. Secure Socket Layer or Transport Layer Security protocols should be used while creating a network for the secure communication. Restriction towards the data access and the data anonymization are the two effective approaches for protecting big data.

Out of these different mechanisms for protecting the big data, the research work mentioned in this paper focus on web log Analyzer to analyze different kinds of attacks that happened to the big data by considering the log files. This helps to get a better understanding of existing attacks and its patterns which helps to develop remedies from the same attacks happening again. In order to maintain the confidentiality of data while it is getting broadcasted to a group of users, by restricting the access only to genuine authorized group, ABE is implemented. Section 5 of this paper illustrates the different automated web log Analyzer tools to detect different attacks from the log file. Section 6 indicates ABE to provide data confidentiality in terms of encryption. Section 7 gives a detailed description about the implementation code of ABE.

LITERATURE SURVEY

The data volume is getting increased marginally due to the excess growth in the field of Internet of Things, cloud computing, mobile internet. Huge volume of data is generated from industries. In order to manage industrial data [1], the enterprise manager should take care of data in an efficient manner. Cloud computing provides a better solution for man- aging the industrial data. The main advantages of making use of cloud computing for the effective data management are configuration is highly flexible, purchase in the form of on demand, maintenance of the data in the cloud is easy. By making use of cloud computing, the enterprise is able to concentrate more on business rather than spending more time of data management. In order to provide confidentiality to the data, it should be encrypted before it is uploaded to the cloud. Even though data storage in cloud is really easy, there are many issues related to privacy and security. ABE is used

to provide a secure access mechanism by encrypting the data. ABE make use of attributes which can be any string of information related to the user. Any public key encryption system can be used to perform attribute-based encryption, which prevents from creating a database of public keys. The protection in terms of privacy at the time of the key generation is taken care by ABE. This is because Key Generation Center (KGC), knows all the attributes and the secret keys that are generated for each of the users. If the KGC is hacked, then the secret generation will be compromised. This research work indicates an ABE system, which is more secure in terms of key generation. The process of attribute auditing and generation of key modules are separated in order to make KGC about the key for each user it generates, and the audit mechanisms performed too.

More amount of data is getting stored and shared on the internet by the third parties, hence there is no guarantee that the data is secure and is accessed only by the genuine users. Hence an efficient encryption technique [2] is to be incorporated to ensure the data confidentiality and maintain the integrity by using hashing techniques. Considering the drawback of encrypting data is that it can shared only on selective basis at the level of course grain. The research works illustrates a new technique of ABE known as Key Based ABE.

This helps to perform sharing of the cipher text on fine grain basis. The attributes set helps in labelling the cipher texts generated. The structures used for access are linked with private keys. Using these keys which are private, the users can decrypt the data. The scheme developed supports the audit log content sharing, and the encryption of data which is broadcasted. The paper also illustrates the details of IBE which is Hierarchy based.

HE [4] indicates a secure way of transferring data such that the third party who has to process the data to generate the results for the genuine users will be given a chance to see the data in the human readable form. The manipulation of the data to generate the outcome will on cipher text. In the cloud environment, four types of HE which are single encryption algorithm are abstracted. The entire HE is divided into fully HE and partial HE. Where the common operations performed are addition and multiplication. Five different types of fully HE is discussed in detail and the comparative table is illustrated to give a detailed explanation on how fully HE is better than any normal encryption scheme. Hence as a future progressive work, attribute based encryption can be combined with the HE to prevent the loss of integrity of data to the server or the third party and that the server can perform the operations as instructed by the user on the cipher text without even seeing the actual plain text.

WEB LOG ANALYZER

Web log analyzers are tools that are used to scan for vulnerabilities in web sites and their servers through the use of log files. It is a sort of log file analyzing mechanism that checks the log reports generated from the sites and visualizes them for the admin to look through and make decisions based on the reports. Traditionally, without these tools, people would had needed to go through tones of log records from a file and figure out the anomaly which sometimes might be too hard to find if done manually. But with the help of these log file analyzer tools, this has become easy as it is just a few step process and it can interpret various types of information just from one log dump file that can be found in the main root folder of the web server that is hosting the site. These are applications that are used on the server side of the system and needs proper access to the root directory of the servers that are hosting so that they can access the log files from within the servers. We have styled our solution with two applications where each does their own unique stuff and helps together to provide a better secured solution. The two tools that have been used for the web log analysis:

Snort

AWStats

A. Snort

Snort is an open sourced intrusion detection application that used on the server side of the system to detect any sort of intrusions that can occur and detect the system admin about the situation so that they can take

proper actions against them and

prevent the loss of any data. This is a small-scale application that runs on the networking device and is a console-based application, so no additional GUI is provided to the user for any other purpose. It uses a rule-based technique to prevent or alert the user where if the conditions are met, it will provide an alert prompt to the admin to take action. Snort keeps track on all the traffic whether its incoming or outgoing so that it can check whether any data transfer contains anything that might trigger one of the rule policies set by the admin. The following five types of action is performed by these rules:

- Alert: Generates an alert message
- Log: Logs the specified IP packet
- Pass: Ignores the specified IP packet
- Activate: Sends an alert message and then activates a dynamic rule
- Dynamic: Activated by an activate rule, this rule then acts as a log rule

Snort just doesn't only alert for intrusion detection but also helps to create log files which comes in handy in the later processes. They would log all the records of the incidents that happens from the intrusions and save it in a log file which can later be used to provide further action decided by the admin.

B. AWStats

It is an open sourced log file analyzer tool that helps to visualize the log files and represent them in a way that the admin will find it useful and can take some actions based on the information provided from the log files. While being an open sourced application, it has support for almost a lot of web server applications ranging from Apache to WordPress and etc. AWStats is written in Perl language and can only be installed on a server application that is hosting a website. Basic installation instruction for the application is to download the latest package from https://awstats.sourceforge.io and copy it to the root directory of the server and then run the script file from the folder using the command prompt in windows or terminal in MacOS. The code to be run on the command prompt is "perl awstats-configure.pl" which will guide you through the setup process and ask for some directories where your log files are and once you provide that then rest will be taken care of by the installing package. Once it is all done, the application would be ready to provide you guide on the report of the website. A full log analysis enables to show the following information:

- Dynamic reports using charts and graphs.
- Number of visits and unique visitors.
- Visit Durations.
- Visitor's OS used.
- Visitors Browsers Used.
- Robots/bots that has visited the site.
- Searched phrases.
- Worm Attacks.
- SQL Injection and XSS attacks.

ATTRIBUTE BASED ENCRYPTION SCHEME

Basics of ABE

The user attributes are used for encryption and decryption,

Scheme is that size (cipher text) is proportional to no (attributes used for the encryption). Attribute based encryption is based on two different schemes: key policy-based encryption and cipher text policy attribute-based encryption.

Kp-abe: secret key generation is done based on access tree that defines the privilege of the users and encryption is based on set of attributes.

Cp-abe: the access tree is used for the encryption of data users secret key is generated over the set of attributes. Initial development of attribute-based encryption was Identity based encryption.

IDE

It is a one to one encryption scheme, but there is no need to maintain a public key directory. More bandwidth is needed for implementing the Identity based encryption, as the Cipher Text is to be send to multiple receivers, if all the receivers share the common receiving policy. The ID of the user can be used as the public key. The ID can be email id or biometric of the user. The user will send the ID to the key generation center. The key generation center will generate the secret key using certain public parameters and the Master Key. Once the Secret Key is generated, the Key Generation Center will make use of a secure channel may be TLS (Channel with Transport Layer Security) or SSL (Secure Socket Layer) to share the Secret to the corresponding user. If a data is to be send using Identity Based Encryption, the sender will request for the receivers Identity as the public key to encrypt the data. Once the cipher text is generated, using a secure channel it will be send to the receiver. The receiver will decrypt the cipher text using his/her own secret key. The secret is made using the identity and the public parameters, which indicates that the secret key is a transformation of the receiver's identity. Any kind of public key cryptosystem algorithm can be used to generate the cipher text and also for the decryption. But the main drawback with the Identity based encryption is that it can only be applied to a one to one communication. This is not apt for a server sharing data to multiple users or in the form of broadcasting to different groups.

Details of ABE

ABE scheme works as follows, instead of using the identity of a single user, a group id will be used as the public key. This group id will be used for encrypting the data to be sent. The group Id to which the user is belonging to, will be sent to the key generation center. They will generate the secret key for the group based on the group id. The secret key for the group is generated based on public parameters and the master key. This scheme is mostly used for a secure broadcast. If a data is to be broadcasted to a group. The data will be encrypted using the group identity satisfying certain mathematical combinations. The combination of the identities is done based on mathematical operations such as AND and OR. Once the cipher text is broadcasted, only the particular group whose secret key is a random combination of the group id with which the data was encrypted, can only decrypt the

data. The encryption and decryption can be done based on any public key cryptosystem scheme.

The figure number 01 illustrate how the secret key is generated by the key generation center. The user for example from the figure belongs to the department of mathematics and the course he is enrolled for is the MSc. Hence the identity of the user is MA and MSc. This attribute will be sent to the key generation center, to create the secret key. The Key Generation Center has certain Public attributes and Master Key. Using these two parameters and the identity send from the user, the Key Generation Center will generate the Secret Key. The transfer of the identities and the secret keys will be done via a secure channel. The channel can be encapsulated with TLS or SSL.

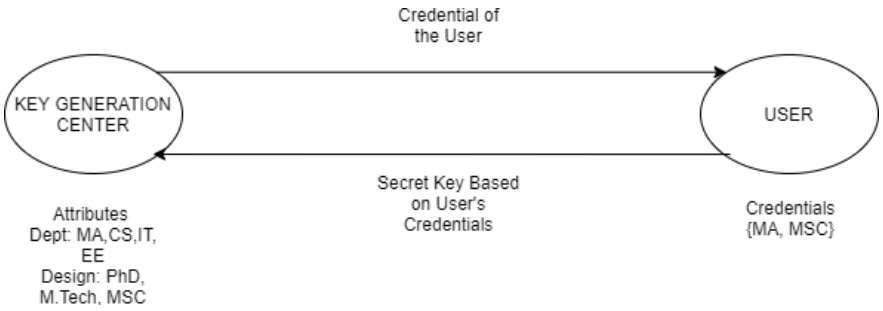


Fig. 1. Creation of Secret Key

The figure number 02 illustrates the process of encryption and decryption. The encryption is done at the sender end using the attributed received from the receiver. From the example below, the attributes with which the data is encrypted is based on the mathematical combination operations such AND and OR. The identities used for encryption are Mathematics department and that the person is either into PhD or MSc degree within the mathematics Department. Once the data is encrypted using the combination of attributes, the cipher will be send to the user (receiver).

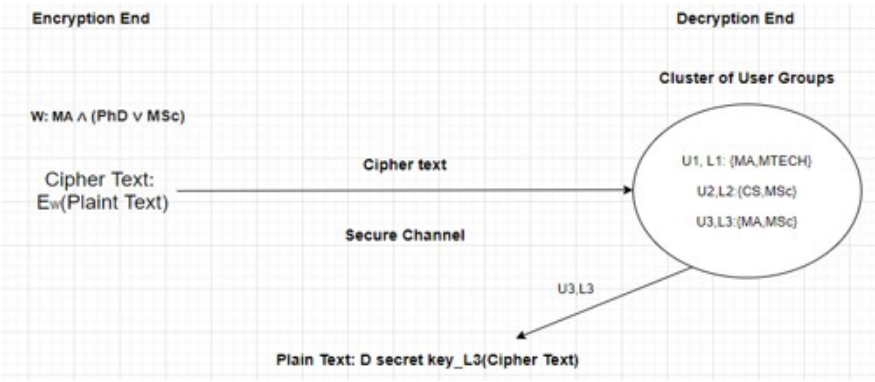


Fig. 2. Encryption and Decryption using ABE

The figure indicates a cluster of users who can act as decryptors. The three user groups with the attributes are U1 with credential as MA, MTECH, and U2 with CS, MSc, and U3 with MA, MSc. Since the data was encrypted using the credentials Mathematics and either PhD or MSc, only user U3 can decrypt the data. Hence even though the data was broadcasted, only the specific user group can decrypt the data based on the secret key which is generated as random

combination of the attributes of the authenticated user group. The major applications of ABE is in the fields of Cloud storage, distributed tolerant networks, wireless sensor networks, mobile Ad-Hoc networks, internet services, helps protecting IoT, ensures privacy in online data transfer.

Advantages of ABE over the other public key cryptosystem

1. One to one scheme, i.e. the encrypted communication can only be accomplished between a single sender and receiver.
2. Public Key Directory or the repository is not needed to be maintained.
3. Hacker cannot hack the channel of key distribution.
4. The usual public key cryptosystem restricts the encryption and the decryption to a single sender receiver system. In ABE, the server can encrypt the data using a group ID and the receivers can decrypt using the corresponding secret key of the group.
5. A random collection of characters can be used as the attribute for generating the Secret Key.
6. The attributes can be numbers, collection of numbers or a collection of raw data.
7. Collusion Resistance Property is guaranteed by Attribute Based Encryption Scheme.
8. Collusion Attack is prevented by using the login restriction to the users by Attribute Based Encryption Scheme.

HE for future Implementation

HE [5] enables the manipulations to be applied on Cipher Text. The third party or the server will be performing operations on cipher text without revealing the plain text. The operations that can be performed on the encrypted are addition and multiplication. There are two types of HE: Fully HE and Partially HE. Partially HE: allows only one of the operations, either addition or multiplication. Fully HE: provides the permission to apply both the operations. Both addition and multiplication. Consider two users A and B.

User A has to perform addition mathematical operations on two data but does not know how to perform the operation. But User A does not trust User B. Hence User A encrypts the two numbers say 1 and 2 into 33 and 54 respectively. Hence 1 is transformed into 33 and 2 is transformed into 54. This encrypted data will be sent to the User B along with the operation to be performed, which is addition (+). User B performs the addition on encrypted data and the result 87 will be send back to User

A. User A decrypts the data and gains the actual result, which is 3.

EXPERIMENTAL ANALYSIS

Implementation of Snort

Snort rule policies can be manually set by the admin or is also provided at a cost on various websites where other people make their own set of rules and sells it to whoever needs them most. So, as you can see, rules play a very important factor when it comes to the use of Snort. These rules are usually found on the root folder of the application inside a file called local rules. Whenever a rule needs to be added or

appended, the admin needs to access this file from the terminal or command prompt and edit it and save. Usually each rule has their own unique ID called SID (Snort ID). These SID can range from any numbers, but each rule has to have their own unique number otherwise the system will not allow you to store duplicate numbers. The format of the rules is very simple, it has the type of control protocol declarations, source IP address and the port, followed by the destination IP address and Port and inside the brackets, the functions of the rules that are supposed to be implemented. A l e r t t c p (s o u r c e i p a d d r e s s) (p o r t) -> (d e s t i n a t i o n i p a d d r e s s) (d e s t i n a t i o n p o r t) (msg: Msg D e t e c t e d ; c o n t e n t : login ; s i d: 1 0 0;) A l e r t t c p any any >1 9 2 . 1 6 8 . 1 . 0 / 2 4 any (f l a g s: A; ack: 0; msg: TCP p i ng d e t e c t e d; s i d: 1 0 1) A l e r t udp any any >any any (msg: Data t r a n s f e r r i n g on U D P; s i d: 1 0 2;) These rules help prevent and inform the admin about intrusions to the system. There are many types of intrusions that occurs every day among which some of the common and well-known examples are SQL injection and XSS attacks. SQL injection is the insertion of an unauthorized code that can make the computer to fetch data from the servers and view it without proper permission. Such attacks are very hard to catch as the computer operates it just like any other SQL command and due to the command having a true statement, it fetches and views the data for that SQL command. One such example is the 1=1 injection. SELECT * FROM Users WHERE UserID = 105 OR 1 = 1;

SELECT UserID, Name, Password FROM Users WHERE UserID = 105 OR 1 = 1;

In the code above, which is a sql code with two conditions in an OR case is an example of such attack as 1=1 will be true and due to having OR condition, it will display all the records just because of that one true condition. So, when hackers inject a true condition to any sql, they can extract the data based on that true condition and display. So technically it they have access to the SQL codes then there is very less way to detect the problem in the leak of the data. One other form of attack is XSS attack which is the cross-site scripting attack. What they do in this process is they inject a JavaScript in the web forms, and they change the return location of the page and sometimes even change the looks of the webpage by changing the html code of the website. When such incidents occur, people who use those websites and key in data gets their data lost and also leaked at the same time. One famous technique for XSS attack is the stealing of cookies from websites as the cookies contains logged or stored data that are useful for the hackers when they are looking for information from web pages.

Implementation of AWStats

With the use of AWStats, one can see the number of bots and crawlers who have accessed the website as well just from the log files. Bots and crawlers have certain features that it does on a website which can be identified from the log and sometimes if it is a common bot then it can be identified on the list of the application as well.

Robots/Spiders visitors (Top 10) - Full list - Last visit

78 different robots*	Hits	Bandwidth	Last visit
MJ12bot	19,129+290	122.77 MB	31 Jan 2018 - 23:44
AhrefsBot	2,857+59	18.59 MB	31 Jan 2018 - 23:56
SemrushBot	2,376+15	15.05 MB	31 Jan 2018 - 03:13
bingbot	1,270+575	17.37 MB	31 Jan 2018 - 23:48
bubing	882+42	4.27 MB	31 Jan 2018 - 23:59
360Spider	678+5	3.72 MB	31 Jan 2018 - 22:08
empty user agent string	607+2	910.74 MB	31 Jan 2018 - 23:43
SeznamBot	271+333	19.88 MB	31 Jan 2018 - 23:05
Yahoo! Slurp	471+73	4.95 MB	31 Jan 2018 - 23:47
BingPreview	465	7.36 MB	31 Jan 2018 - 04:18
Others	3,122+397	74.65 MB	

* Robots shown here gave hits or traffic "not viewed" by visitors, so they are not included in other charts. Numbers after + are successful hits on "robots.txt" files.

Fig. 3. Sample of Bots and Crawler Log Report

One can even look for search phrases from the log of the website to see what visitors are mostly looking for on the search function of the website and maybe analyze that a little further to improve on the current system and develop something new or make those mostly searched after functions more to the front of the website so that the user's don't have to go looking for it every time.

The log analysis also displays the amount of time spent on a page by the users as well which sometimes helps to indicate whether the viewer is a human or a bot/crawler because the usual time for a human to spend on a page is much more than what crawlers and bots spends, so when the average visit duration is very less, almost in the range of few seconds then one can assume that most of the visitors are bots and crawlers rather than actual humans, for which they can implement one of those Google's verification methods of having to go through a few steps to prove that you are a human and not a bot whenever you try to access the website. Such information can help the admins to direct traffic a lot and manage the capacity of their servers as well since too much hits will lead to too requiring of huge amount of bandwidth which might be costly at times.

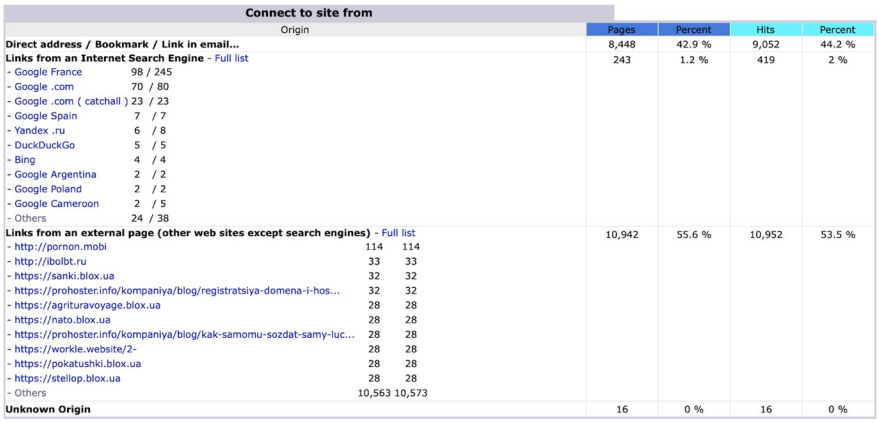
Visits duration

Number of visits: 8,200 - Average: 214 s	Number of visits	Percent
0s-30s	7,383	90 %
30s-2mn	111	1.3 %
2mn-5mn	101	1.2 %
5mn-15mn	86	1 %
15mn-30mn	77	0.9 %
30mn-1h	140	1.7 %
1h+	302	3.6 %

Fig. 4. Visits duration from the log

Another function is helps is to provide the number of linkages with various sites that are forwarding their crowd towards the website. The more the linkage, the more visitors will come to the website and more load will be put onto the server to handle requests. Sometimes not all servers are capable in taking the huge load of requests.

So, if the admin ever feels that some linkages are unnecessary and they don't require those linkages to their site, then they can always block those IP addresses so that traffic doesn't come from those sites and servers don't get overloaded with requests.



Connect to site from					
Origin			Pages	Percent	Hits	Percent
Direct address / Bookmark / Link in email...			8,448	42.9 %	9,052	44.2 %
Links from an Internet Search Engine - Full list			243	1.2 %	419	2 %
- Google France	98 / 245					
- Google .com	70 / 80					
- Google .com (catchall)	23 / 23					
- Google Spain	7 / 7					
- Yandex .ru	6 / 8					
- DuckDuckGo	5 / 5					
- Bing	4 / 4					
- Google Argentina	2 / 2					
- Google Poland	2 / 2					
- Google Cameroon	2 / 5					
- Others	24 / 38					
Links from an external page (other web sites except search engines) - Full list			10,942	55.6 %	10,952	53.5 %
- http://pornon.mobi	114	114				
- http://ibolbt.ru	33	33				
- https://sanki.blox.ua	32	32				
- https://prohoster.info/kompaniya/blog/registratsiya-domena-i-hos...	32	32				
- https://agrituravoyage.blox.ua	28	28				
- https://nato.blox.ua	28	28				
- https://prohoster.info/kompaniya/blog/kak-samomu-sozdat-samy-luc...	28	28				
- https://workle.website/2-	28	28				
- https://pokatushki.blox.ua	28	28				
- https://stellop.blox.ua	28	28				
- Others	10,563	10,573				
Unknown Origin			16	0 %	16	0 %

Fig. 5. Linkages from different sites to the Server

ABE Implementation Code

The code below illustrates the detailed working of ABE using Rivest Shamir and Adleman (RSA) algorithm as the base public key cryptosystem.

```
import Crypto from Crypto.PublicKey import RSA from Crypto import Random import
os import time def rsaenc(text): randomgenerator = Random.new().read key = RSA.generate(1024, randomgenerator) publickey = key.publickey() print("####################") print("processbegins") print("####################") time.sleep(2) print("####################") print("Marks are entered and sent to KGC") print("####################") time.sleep(2) print("####################") print("Encryptionbegins") print("####################") encrypted=publickey.encrypt(text.encode('utf8'),32) time.sleep(2) print("ENCRYPTED TEXT",encrypted) f=open('encryption.txt','w') f.write(str(encrypted)) f.close() time.sleep(2) print("####################") print("studentpanel") print("####################") print("if you are student then enter you department") dept=input() print("pleaseenter you graduationlevel") gradlevel=input() if(dept=="MA" and gradlevel=="MSC" or gradlevel=="PHD"): f=open('encryption.txt','r') message=f.read() decrypted=key.decrypt(ast.literaleval(str(encrypted))) print('decrypted',decrypted) f=open('decryption.txt','w') f.write(str(message)) f.write(str(decrypted)) f.close() else: print("Sorry you cant view someone else marks") print("enter your marks") marks=input() rsaenc(marks)
```

The ABE algorithm is implemented using RSA algorithm for the encryption of data. The process illustrated for the code development is based on a scenario of a college professor sending the marks to a group of students belonging to the department of mathematics and the student should be either enrolled for the course MSC or PhD under the department of mathematics. The implementation begins with asking the professor to enter the marks. The Professor enters the mark and the process begins, the Key Generation Center will generate the Secret Key based on the Public attributes of the receiving group and a random number.

Then the encryption starts, the encryption is done based on the public key crypto system. The public key cryptosystem can be any, in this implementation of attribute-based encryption, RSA Cryptosystem is chosen. Based on public attributes already shared and the random number generated by the random generator, the encryption will be performed based on attribute-based encryption. RSA algorithm makes use of modulus

operation to perform encryption and decryption. The encryption is done using the public key of the receiver and the decryption is performed based on the matching private key of the receiver, whose key was used for encryption. encrypted = publickey.encrypt(text.encode('utf-8'),32), indicates the conversion of the plain text into a cipher text. The encoding scheme helps to make the cipher text in the human non-readable form. Encryption is done and the cipher text generated which is human non-readable and in the encoded format. From the student panel, the legitimate authenticated group of students who has the rights to view the marks that was send by the Professor. It asks the student to enter

the group details. The group details are the department they below to and the course they are enrolled for. The sample group considered in the implementation is the student should belong to Mathematics Department, i.e. MA department and the student should be either an MSC or PHD Student. If and only if the student belongs to these categories the decryption can be performed, and the student can view the data which is the marks. The student is asked to enter the department, the student belongs to and as a next level of security, the student is again asked to enter the graduation level of studies. Hence after the student enters the details, based on mathematical operations such as AND and OR, the decryption will be done using the secret key generated by the Key Generation Center. Hence the decryption is done, and the marks is visible to the student. If the student enters the wrong combination of decryption factors the marks will not be decrypted and will not be visible.

CONCLUSION

As a part of research work done with respect to ensuring privacy in big data, Snort and AWStats automated web log analyzing tool is studied and explained in detail. This tool helps to identify the different kinds of attacks that happened to a particular application or web server. Hence upon understanding the pattern of different attacks, necessary precautions have to be made to prevent the same attack from happening again. these tools not only help us notify about certain security issues and loopholes of a system but also helps to achieve some other objectives as well. Both the tools can help provide real time monitoring of the system to help improve security and privacy and at the same time keeps track of the network from any unwanted intrusions which might cause trouble to the system. They help to provide the admin of the system with valuable information like bad IP addresses and helps them block or blacklist those IP addresses while also informing the admin about what is coming through certain IP addresses and who are sending those. These applications help secure the system from bots and crawlers, while also provides certain information like bandwidth control and hits which can help the admin to improve their system and figure out scope of improvement which sometimes can be of much more value for research and development. In order to protect data and provide data confidentiality while broadcasting the information from the cloud to a group of users, ABE cryptosystem is discussed. Different public key cryptosystem is mentioned with reference to the drawbacks. To rectify the security issues with public key cryptosystem, identity-based encryption scheme is introduced where user identity is used for encryption and the secret key is generated by the key generation center based on public attributes and user identity. Recognizing the shortcoming of Identity based encryption scheme, which is one to one secure communication scheme, attribute-based encryption scheme is identified. This enables a secure broadcast communication. The encryption is done using group identity and the decryption is done using the secret key generated based on the group identity.

As future scope, HE technique can combined with the attribute-based encryption, which enables the transfer of Identity from the user to the key generation center in the cipher text.

ACKNOWLEDGEMENT

We would like to express our sincere appreciation to our beloved professor Dr. Jinan Fiaidhi from Lakehead University for her constant support and guidance without which this would not had been possible at all. We are truly grateful for her unwavering support throughout this whole period of time and for giving us this opportunity to work with her under her guidance and achieve what we have done today. With her help, we have come to learn of new technologies and research methods which we were not aware of before and this has helped us with our own research where we came to know many new ideas and concepts. We would also

like to thank Lakehead University for providing us a platform of such manner where we could work at our fullest and get all these opportunities given to us through the course of our studies.

REFERENCES

1. Yujiao Song, 1 HaoWang, 1, 2 XiaochaoWei, 1 and LeiWu 1, 3, "Efficient Attribute-Based Encryption with Privacy-Preserving Key Generation and Its Application in Industrial Cloud," Hindawi Security and Communica- tion Networks.
2. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM Conference on Computer and Communications Security, (CCS '06), pp. 89–98, Alexandria, VA, USA, November 2006.
3. Khushboo Wadhwani, "Big Data Challenges and Solutions", Technical Report Bradley University.
4. Min Zhoa E, Yang Geng, "Homomorphic Encryption Technology for Cloud Computing", 8th International Congress of Information and Com- munication Technology, ICICT 2019.
5. Aderemi A. Atayero*, Oluwaseyi Feyisetan, "Security Issues in Cloud Computing: The Potentials of Homomorphic Encryption", VOL. 2, NO. 10, October 2011, Journal of Emerging Trends in Computing and Information Sciences
6. https://www.nltechno.com/awstats/awstats.pl? month=01year=2018output=mainconfig=destailleur.frframename=index