Intelligent and Active Defense Strategy of Cloud Computing

Hong-Jun Chen
Computer Department
Sichuan TOP Vocational Institute
Chengdu,China
2009missxiaochen@163.com

Xue-Qin Wu
Computer Department
Sichuan TOP Vocational Institute
Chengdu,China
doublesnowy1210@163.com

***Abstract*—Cloud's development has entered the practical stage, but safety issues must be resolved. How to avoid the risk on the web page ,how to prevent attacks from hacker, how to protect user data in the cloud. This paper discusses some satety solution :the credit level of web page ; Trace data ,analyze and filter them by large-scale statistical methods; encryption protection of user data and key management.**

Keywords-Cloud;safety solution;the risk on the web page;the credit level ;large-scale data;encryption;tracking;filtering; statistical methods;

I. CLOUD OVERVIEW

A. Concept of cloud

cloud is super-computing model based Internet.It is the development of parallel computing,distributed computing and grid computing,consisting of a series of interconnected and virtualized computer system.. Cloud integrates a large number of hardware,software resources and applications to work together, which is distributed in a variety of servers,personal computers and even other devices such as mobile phones[1].

B. Several forms of cloud

The seven kinds of cloud forms[2]：

SAAS (Software as a Service). This type of cloud computing pass program to thousands of users through the browser.SAAS is commonly used in Human resources and ERP .Typical example: Google's Google App Engine,which contains such as videos, calendars, presentations, attachments and text sharing.

Utility Computing,which creates virtual data center to integrate the memory, I / O devices, storage and computing power together into a virtual pool of resources,provides services for the entire network.

Providers are able to provide an API for developers to develop more Internet-based applications.

PAAS(Platform as a Service).It provides the development environment as a service.Developers develop their own programs and pass these to a user through the Internet and their servers by the brokers' equipment.

MSP (Management Service Provider).It is oriented with IT industry more than end-users,commonly used in e-mail virus scanning, process monitoring, and so on.

Business Services Platform.It mixs applications of SAAS and MSP,which provides a platform for interaction between users and providers.Such as personal expenses management system can manage its expenses and coordinate its subscription services based on user settings.

Network Integration.It integrates the companies which provide similar services on the Internet,so that users can more easily compare and choose their own service provider.

C. Cloud's current situation of development and security risks.

Cloud development has entered a practical stage.Some famous software vendors,such as Google,Microsoft,IBM,Amazon, have been paying attention to Cloud computing .They are promoting cloud computing researches and applications actively.They also propose solutions and achievements for cloud computing.Now the popular cloud applications such as: Google App Engine, IBM Blue Cloud, Amazon AWS, Microsoft Live Mesh, Sun Blaek Box [3] and so on.At the same time, cloud computing also face many safety issues to resolve, mainly security challenges are as follows:

1)How to avoid the risk on the web page

2)How to prevent hackers

3)How to protect user data in the cloud.

II. ACTIVE DEFENSE STRATEGY

A. Determine the level of web credit

There are many unknown risks on the web page,Such as viruses,Trojan,some of the plug-ins which will be automatically downloaded.High-risk web page will be the lower credit level. Analyze and organize massive source of information from internet by determining the level of web credit.Also build huge cloud server cluster,And saved high-risk sources of information into the cloud database.

How to determine the level of risk.This method can be used:give the credit score for each URL.The credit score can be calculated by the length of time of the website's

978-0-7695-4647-6/12 $26.00 © 2012 IEEE
DOI 10.1109/ICCSEE.2012.253

existence,the changs of the website's location and of history,and so on.[4]

Web credit can be constantly changing over time.When a user visits a website with potential risks,will be alerted or prevented timely.The techology can help user identify the target website security.Cloud identifys the new threats to provide real-time detection and timely common intelligence protection by checking the website credit .

B. Trace data ,analyze and filter them by large-scale statistical methods

How to identify the wasteful or unsafe data from massive data in cloud,such as junk mails and junk files , etc.Trace data,and analyze data by the large-scale statistical methods to determine the events and behaviors.

1) Junk mails filtering

Bayesian filtering algorithm[5] filters junk mails.Its basic steps is as follows:

a) Collect a lot of junk mails and non-junk mails,and build junk mails set named L(j) and non-junk mails set named L(nj).

b) Extract the independent string such as "sales","activity" from e-mail subject and message body to build a TOKEN string.And extract the number of occurrences of the TOKEN that named word-frequency.

Deal with all the mails in set set L(j) and set set L(nj) in accordance with the method[5].

c))Each mail set corresponds to a hashtable. set L(j) to hashtable_j,set L(nj) to hashtable_nj.The mapping of TOKEN string to the word-frequency is stored into the corresponding hashtable.

d) Calculated the occurrence probability P of TOKEN string. Probability P = (word-frequency of a TOKEN) / (the length of the hashtable of the corresponding hashtable).

e) Consider hashtable_nj and hashtable_j,and conclude the probability when a TOKEN string appears in a new mail,the new one as spam . Mathematical expression is: A event--- the message is spam;t1, t2 ... . tn behalf TOKEN string.So P (A | ti) is bahalf on the probability that a message is spam,when string TOKEN ti appears in the mail.Let P1 (ti) = (ti in hashtable_nj values) P2 (ti) = (ti in hashtable_ j values),then

P (A | ti) = P1 (ti) / [(P1 (ti) + P2 (ti)]

f) Create a new hashtable hashtable_probability to store the mapping of TOKEN string to word-frequency.Thus,the learning process of the set of Junk mail and non-spam is to the end.According to the established hashtable hashtable_probability can estimate a new e-mail as spam possibility. Now when the new mail is coming,follow step (2) to generate TOKEN string.Query hashtable_probability to get the TOKEN's keys. Assuming that the mail obtained n TOKEN strings .t1, t2 ... . Tn, the corresponding hashtable_probability is P1, P2,. . . . . PN,then P (A | t1, t2, t3 ... ... tn) shows that probability of the mail as spam,when many TOKEN string appear in the mail at the same time. From the compound probability formula, get the expression:

$$P(A \mid t1, t2, t3 \ldots \ldots tn) = (P1 * P2 * \ldots PN) / [P1 * P2 * \ldots . PN + (1-P1) * (1-P2) * \ldots (1-PN)] \quad (1)$$

when P (A | t1, t2, t3 ... ... tn) exceeds a predetermined threshold value,we can determine the e-mail as a junk mail.

2) Analysis technology of behavior associated

Integrated activities linked to the threat by analysis technology of behavior associated, to determine whether they are malicious behaviors.A single web threat did not seem to be a harm.But it could lead to some malicious results if a number of activities occur at the same time.Therefore following the heuristic view,it can check the potential threats by the relationship of different components.Associating different parts of the threat and continuously updating threat database can protect email and Web automatically in a timely manner.For Example,an IP, corresponding to 17 domain names,then the IP will be attended.Trace the corresponding content,and analyze them .Also analyze all the related page links, determine their the level of web credit.

3) Establish an automatic feedback mechanism

Establishing an automatic feedback mechanism, that is to build uninterrupted communication between the end-user and cloud cluster.And update every new threat timely by the user's check and sharing.

A. Encryption protection of user data and key management

Remote user data and personal information is stored in the cloud, the provider must ensure that the user's security.Confidential data stored in the cloud must be protected combined with the measures such as access control, encryption and other contractual obligations,etc.Among them, the encryption provides the following benefits: the minimum dependence of the cloud service providers,the minimum dependence of operational failures.

Strong encryption and key managemen is a central and important mechanism to protect data for cloud computing systems.Because the encryption itself does not guarantee against data loss.Encryption provides a resource protection, and key management provides a protected resource access control [6].In addition to data access to a strict limit,encryption can avoid malicious users to access some sentive data easyly.Because of the great strength of the encryption algorithm,decryption becomes very difficult.So that once data theft, the important information is not easy to be stealed and used.

Separate the cloud service provider of data storage from key management to establish series of "separation".This is to

protect the cloud provider, but also to protect the user.Avoid conflicting when providing data as required by law .

OASIS Key Management co-Protocol (KMIP) is the new standard of co-cloud key management.IEEE1619.3 standard covers storage encryption and key management, especially for storage IaaS.But there are many key management-related issues and challenges in cloud computing[7].

III. CONCLUSION

Cloud computing is the future of the Internet,its applications are developed rapidly.Cloud security research is necessary and urgent. some new technogies protecte cloud computing at a great degree.These technogies are such as determining the level of web credit,tracing data ,analyzing and filtering them by large-scale statistical methods,encryption protection of user data and key management,etc.But these is not enough. There are still some challenges:How to manage key better ,how to avoid the monopoly of cloud provider, and how to ensure the quality of the provider's service.

[1] HUI-DONG.body and digital libraries [M]. Wuhan: Wuhan University Press, 2008: 121-12http://stor.zol.com.cn/128/1288536.html

[2] HANSEN C D, JHNSON C R· The visualization handbook [M].[S· 1·]: Elsevier Inc, 2005.

[3] http://sec.chinabyte.com/458/8546458.shtml

[4] http://scholar.googleusercontent.com/scholar?q=cache:HcU-q5IrTgEJ:scholar.google.com/+Bayesian+filtering+algorithm&hl=zh-CN&lr=lang_zh-CN%7Clang_zh-TW&as_sdt=0,5&as_vis=1

[5] http://scholar.googleusercontent.com/scholar?q=cache:mTnBuMIml0UJ:scholar.google.com/+key+management&hl=zh-CN&lr=lang_zh-CN%7Clang_zh-TW&as_sdt=0,5&as_vis=1

[6] http://news.ccidnet.com/art/32859/20100802/2137337_1.html