International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-11, November 2014

Distributed Versus Cloud Computing and data security issues and new trends- Fog Computing

Sachin R. Desale, Kadambari V. Vanmali, Brajendra Singh Rajput

***Abstract*— There is some confusion and misconception about two terms, in this paper we are going to discuss about what are the similarities and differences between ditributed computing and cloud computing. And furthermore we will discuss some security isssues and new trends about the same, hope so this paper will be helpfull to clarify the doubts about these two concepts or terms. Distributed computing comprises of multiple software components that belong to multiple computers. Whereas, a term Cloud computing is used to define a new class of computing that is based on network technology.**

Cloud computing have been proposed different ways of using computers and to access and store our personal and business information.Because of these new computing and communication technology there arise new data security challenges.Eventhough existing techniques use security mechanisms, data theft attacks prevention fails, it has been proved that these existing techniques are not enough to handle such attacks.We propose a completely different approach for securing data in the cloud using preventive decoy technology. We monitor data access in the cloud and detect abnormal data access patterns. When unauthorized access is suspected and then verified using challenge questions. We launch a disinformation attack by returning large amount of decoy information to the attacker. This protects against the misuse of the user's real data.

***Index Terms*— Distributed computing, Cloud computing, Fog Computing, User behavior profiling, Decoy technology.**

I. INTRODUCTION

Distributed Computing:- Distributed Computing is a field of computer science that studies distributed systems(DS). A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal. Distributed computing comprises of multiple software components that belong to multiple computers. It is the system that works or runs as a single system. The computers forming the distributed architecture may or may not be closely located. These systems are often preferred for the scalability factor. It is quite easy to add new components to the system without disturbing the existing system. A computer program that runs in a distributed system is called a **distributed program**, and distributed

programming is the process of writing such programs.There are many alternatives for the message passing mechanism, including RPC-like connectors and message queues. An important goal and challenge of distributed systems is location transparency.

Manuscript received November 18, 2014.

A term Distributed System (DS) is a collection of independent computers that appears to its users as a single coherent system. It consists of collection of autonomous computers, linked by a computer networks and equipped with distributed system softwares.These softwares enables computer to coordianate their activities and to share the system's hardware, software and data too .

The term Distributed Operating System (DOS) can be defined as an Operating System which manages a collection of independent computers and make them to appear to its users as a single sytem also, called "Single System Image".

Fig.1. Distributed Computing

The Open Software Foundation's Distributed Computing Environment (OSF'S DCE) is a rich software technology that enables the development of distributed applications across heterogeneous systems, taking the advantage of open, distributed computing. DCE is a suite of integrated services available from vendor-neutral source that enables organizations to develop, use and maintain distributed applications across heterogeneous networks.It is used in exploratory systems, World Wide Web, stock markets, and telecommunication services. Following figure shows typical architecture of DCE system.

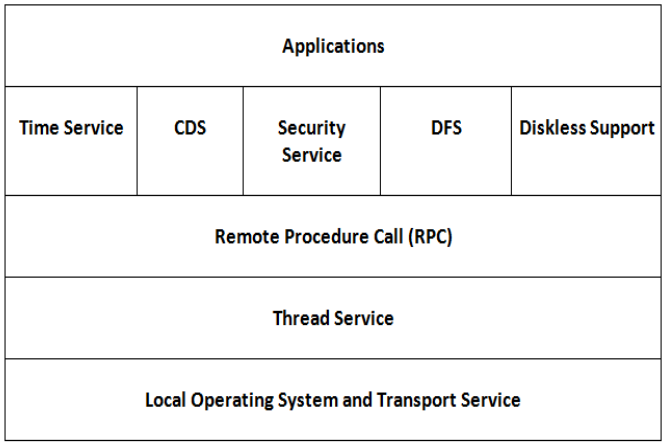

Fig. 2. DCE Architechture

Cloud Computing:- Cloud computing is entering our lives and changing the way people consume information dramatically. Clouds transform IT infrastructures with an emphasis on making them flexible, affordable, and capable of serving millions of users, satisfying their computing or storage demands. The design of early cloud computing systems has evolved from, and was dominated by, the concepts of cluster and grid computing. Currently, as the concepts of the cloud become advanced and mature, cloud networking and communication processes begin playing a central role. Cloud Networking has emerged as a promising direction for cost-efficient and reliable service delivery across data communication networks. The dynamic location of service facilities and the virtualization of hardware and software elements are stressing the communication network and protocols, especially when datacenters are interconnected through the Internet.

Cloud computing is used to define a new class of computing that is based on network technology. Cloud computing takes place over the internet.It comprises of a collection of integrated and networked hardware, software and internet infrastructure.These infrastructures are used to provide various services to the users.One of the biggest advantage of using cloud computing is that it hides the complexity and details of underlying infrastructure, and thus users can easily use the services through simple graphical interface.These systems are virtually managed and often distributed. Basically it is a step on from Utility Computing, and a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform).Using the Internet for communication and transport provides hardware, software and networking services to clients.

These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or Application Programming Interface(API). In addition, the platform provides on demand services, that are always on, anywhere, anytime and any place.It is like Pay for use and as needed, elastic way of using technology.The hardware and software services are available to general public, enterprises, corporations and businesses markets.In short we can say, Cloud computing is an umbrella term used to refer to Internet based development and services.

A number of characteristics define cloud data, applications services and infrastructure:

1. Remotely hosted: Services or data are hosted on remote infrastructure.
2. Ubiquitous: Services or data are available from anywhere.
3. Commodified: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity - you pay for what you would want!

In other words we can define cloud computing as , it is a shared pool of configurable computing resources, having on-demand access and provisioned by service-providers. And Public Clouds, Private Clouds , Community Clouds , Hybrid Clouds these are the various types of cloud system. Massive scale, Homogenity, Virtualization, Low Cost, Resilent computing, Geographic distribution, Service Orientation, On-demand service, Rapid elasticity, Broad network access, Resource pooling, Measured service, these are very important characteristics of cloud computing which makes it quite different from other technologies.Cloud Computing is “pay as much as used and needed” type of utility computing and the “always on!, anywhere and any place” type of network-based computing.The “flexibility and elasticity” allows these systems to scale up and down at will utilising the resources of all kinds including CPU, storage, server capacity, load balancing, and databases.

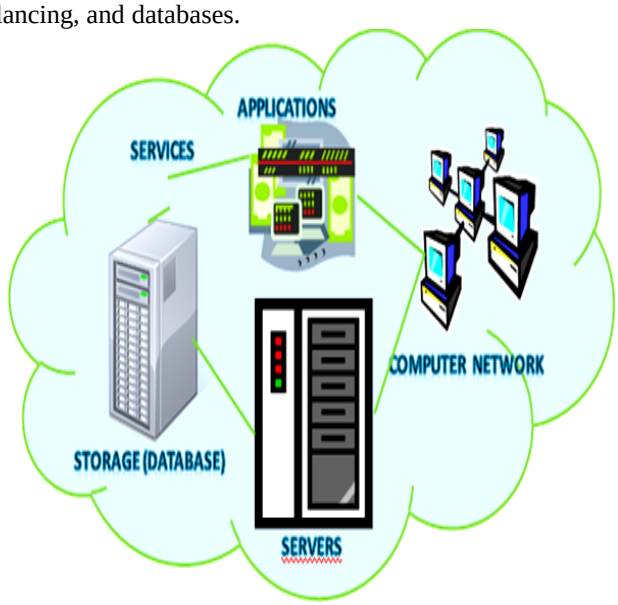

Fig. 3. Cloud System

The Three Layers/Models of Cloud Computing:-

As the delivery of IT resources or capabilities as a service is an important characteristic of Cloud Computing, the three architectural layers of Cloud Computing are illustrated in figure- 3 below.

1.) Infrastructure as a Service (IaaS) : IaaS offerings are computing resources such as processing or storage which can be obtained as a service. Examples are Amazon Web Services with its Elastic Compute Cloud (EC2) for processing and Simple Storage Service (S3) for storage and Joyent who provide a highly scalable on-demand infrastructure for runningWeb sites and rich Web applications.

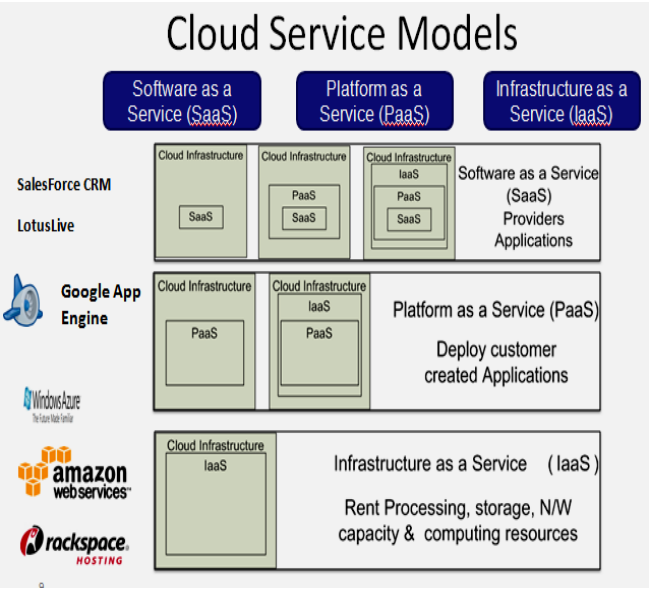

Fig.4. The Three Layers of Cloud Computing

2.) Platform as a Service (PaaS): Platforms are an abstraction layer between the software applications (SaaS) and the virtualized infrastructure (IaaS). PaaS offerings are targeted

at software developers. Developers can write their applications according to the specifications of a particular platform without needing to worry about the underlying hardware infrastructure (IaaS). Developers upload their application code to a platform, which then typically manages the automatic upscaling when the usage of the application grows. Examples are the Google App Engine, which allow applications to be run on Google's infrastructure, and Salesforce's Force.com platform.

3.) Software as a Service (SaaS): SaaS is software that is owned, delivered and managed remotely by one or more providers and that is offered in a pay-per-use manner .Saas the most visible layer of Cloud Computing for end-users, because it is about the actual software applications that are accessed and used. From the perspective of the user, obtaining software as a service is mainly motivated by cost advantages due to the utility-based payment model, i.e. no up-front infrastructure investment. Well known examples for SaaS offerings are Salesforce.com and Google Apps such as Google Mail and Google Docs and Spreadsheets.

New Trends:- Many proposals have been made to secure remote data in the Cloud using encryption and standard access controls. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including insider attacks, mis-configured services, faulty implementations, and buggy code. Building a trustworthy cloud computing environment is not enough, because accidents continue to happen, and when they do, and information gets lost, there is no way to get it back.

The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We posit that secure Cloud services can be implemented given two additional security features:

1. User Behavior Profiling:-It is expected that access to a user's information in the Cloud will exhibit a normal means of access. User profiling is a well known technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such 'normal user' behavior can be continuously checked to determine whether abnormal access to a user's information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often. These simple user specific features can serve to detect abnormal Cloud access based partially upon the scale and scope of data transferred.

2. Decoys: Decoy information, such as decoy documents, honeyfiles, honeypots, and various other bogus information can be generated on demand and serve as a means of detecting unauthorized access to information and to 'poison' the thief's ex-filtrated information. Serving decoys will confound and confuse an adversary into believing they have ex-filtrated useful information, when they have not.This technology may be integrated with user behavior profiling technology to secure a user's information in the Cloud. Whenever abnormal access to a cloud service is noticed, decoy information may be returned by the Cloud and delivered in such a way as to appear completely legitimate and normal.

The true user, who is the owner of the information, would readily identify when decoy information is being returned by the Cloud, and hence could alter the Cloud's responses through a variety of means, such as challenge questions, to inform the Cloud security system that it has inaccurately detected an unauthorized access. In the case where the access is correctly identified as an unauthorized access, the Cloud security system would deliver unbounded amounts of bogus information to the adversary, thus securing the user's true data from unauthorized disclosure. The decoys, then, serve two purposes:
(1) Validating whether data access is authorized when abnormal information access is detected, and
(2) Confusing the attacker with bogus information.

II. Security issuses in Cloud Computing

Cloud computing promises to significantly change the way we use computers and access and store our personal and business information.With these new computing and communications paradigms arise new data security Challenges.Existing data protection mechanisms such as encryption have failed in preventing data theft attacks, especially those perpetrated by an insider to the cloud provider. Data theft attacks are amplified if the attacker is a malicious insider.This is considered as one of the top threats to cloud computing by the Cloud Security Alliance (CSA).

The Twitter incident is one example of a data theft attack from the Cloud. Several Twitter corporate and personal documents were ex-filtrated to technological website TechCrunch , and customers' accounts, including the account of U.S. President Barack Obama, were illegally accessed.The damage was significant both for Twitter and for its customers. While this particular attack was launched by an outsider,stealing a customer's admin passwords is much easier if perpetrated by a malicious insider. Rocha and Correia outline how easy passwords may be stolen by a malicious insider of the Cloud service provider.

We proposed a completley different approach to securing the cloud using decoy information technology , that we have come to call **"Fog Computing"**.Here the meaning of "Fog" is nothing but confusing to the attackers.We use this technology to launch **disinformation attacks** against malicious insiders or attackers , preventing them from distinguishing the real sensitive customer data from fake worthless data.In this paper, we propose two way of using Fog computing to prevent attacks such as Twitter attack by deploying decoy information within the Cloud by the Cloud service customer and within personal online social networking profiles by individual users.

III. Disributed versus Cloud environment

Following Table I will illustrate a complete and clear picture about similarities and differences between distributed and cloud computing environrnment.

Parameters	Distributed Computing	Cloud Computing
Definition	Distributed computing comprises of multiple software components that belong to multiple computers. The system works or runs as a single system. Cloud computing can be referred to as a form that originated from distributed computing and virtualization.	Cloud computing is used to define a new class of computing that is based on network technology. Cloud computing takes place over the internet. It comprises of a collection of integrated and networked hardware, software and internet infrastructure.
Goals	1.)Resource Sharing 2.)Openness 3.)Transparency 4.)Scalability	1.) Reduced Investments and Proportional Costs. 2.) Increased Scalability. 3).Increased Availability and Reliability
Types	1.)Distributed Computing Systems 2.)Distributed Information Systems 3.)Distributed Pervasive Systems	1.)Public Clouds 2.)Private Clouds 3.)Community Clouds 4.)Hybrid Clouds
Characteristics	1.) A task is distributed amongst different machines for the computation job at the same time. 2.) Technologies such as Remote Procedure calls and Remote Method Invocation are used to construct distributed computations.	1.) It provides a shared pool of configurable computing resources. 2.) An on-demand network model is used to provide access. 3.) The clouds are provisioned by the Service Providers. 4.) It provides broad network access.
Cons	1.) Higher level of failure of nodes than a dedicated parallel machine. 2.) Few of the algorithms are not able to match with slow networks. 3.) Nature of the computing job may present too much overhead.	1.) More elasticity means less control especially in the case of public clouds. 2.) Restrictions on available services may be faced, as it depends upon the cloud provider.

Table I. Distributed versus Cloud Computing

IV. EXISTING SYSTEM FOR CLOUD SECURITY

The existing mechanisms only facilitate security features to data and thereby don't allow for detection of invalid access and thereby its prevention to enable valid distribution of data. Businesses, especially startups, small and medium businesses (SMBs), are increasingly opting for outsourcing data and computation to the Cloud.

This obviously supports better operational efficiency, but comes with greater risks, perhaps the most serious of which are data theft attacks.Data theft attacks are amplified if the attacker is a malicious insider. This is considered as one of the top threats to cloud computing by the Cloud Security Alliance. While most Cloud computing customers are well-aware of this threat, they are left only with trusting the service provider when it comes to protecting their data. The lack of transparency into, let alone control over, the Cloud provider's authentication, authorization, and audit controls only exacerbates this threat.

Much research in Cloud computing security has focused on ways of preventing unauthorized and illegitimate access to data by developing sophisticated access control and encryption mechanisms. However these mechanisms have not been able to prevent data compromise. Van Dijk and Juels have shown that fully homomorphic encryption, often acclaimed as the solution to such threats, is not a sufficient data protection mechanism when used alone.

V. PROPOSED SYSTEM – FOG COMPUTING

The proposed mechanism facilitates security features to data and thereby allows for detection of invalid access and thereby its prevention to enable valid distribution of data.

a) User Behavior Profiling: Legitimate users of a computer system are familiar with the files on that system and where they are located. Any search for specific files is likely to be targeted and limited. A masquerader, however, who gets access to the victim's system illegitimately, is unlikely to be familiar with the structure and contents of the file system. Their search is likely to be widespread and untargeted.
Based on this key assumption, we profiled user search behavior and developed user models trained with a oneclass modeling technique, namely one-class support vector machines. The importance of using one-class modeling stems from the ability of building a classifier without having to share data from different users. The privacy of the user and their data is therefore preserved.

We monitor for abnormal search behaviors that exhibits deviations from the user baseline.Accoording to our assumption, such deviations signal a potential masquarade attack.Our experiments on local file setting validate our assumption and demonstrated that we could reliably detect all simulated masquarade attacks using this approach with a very low false positive rate of 1.12%.

b) Decoy Technology: We placed traps within the file system. The traps are decoy files downloaded from a Fog computing site, an automated service that offers several types of decoy documents such as tax return forms, medical records, credit card statements, e-bay receipts, etc.The decoy files are downloaded by the legitimate user and placed in highly-conspicuous locations that are not likely to cause any interference with the normal user activities on the system. A masquerader, who is not familiar with the file system and its contents, is likely to access these decoy files, if he or she is in search for sensitive information, such as the bait information embedded in these decoy files. Therefore, monitoring access to the decoy files should signal masquerade activity on the system.The advantages of placing decoys in a file system are threefold:

(1) the detection of masquerade activity
(2) the confusion of the attacker and the additional costs incurred to distinguish real from bogus information.

(3) the deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers.

We posit that the combination of these two security features will provide unprecedented levels of security for the Cloud. No current Cloud security mechanism is available that provides this level of security.Experiments conducted in a local file setting provide evidence that this approach may provide unprecedented levels of user data security in a Cloud environment.

VI. ALGORITHMS

HMAC Algorithm for Key Security:- This standard describes a keyed-hash message authentication code (HMAC), a mechanism for message authentication using cryptographic hash functions. HMAC can be used with any iterative Approved cryptographic hash function, in combination with a shared secret key. This standard specifies an algorithm for applications requiring message authentication. Message authentication is achieved via the construction of a message authentication code (MAC). MACs based on cryptographic hash functions are known as HMACs.

The purpose of a MAC is to authenticate both the source of a message and its integrity without the use of any additional mechanisms. HMACs have two functionally distinct parameters, a message input and a secret key known only to the message originator and intended receiver.Additional applications of keyed-hash functions include their use in challenge-response identification protocols for computing responses, which are a function of both a secret key and a challenge message.An HMAC function is used by the message sender to produce a value (the MAC) that is formed by condensing the secret key and the message input.

The MAC is typically sent to the message receiver along with the message. The receiver computes the MAC on the received message using the same key and HMAC function as was used by the sender, and compares the result computed with the received MAC. If the two values match, the message has been correctly received, and the receiver is assured that the sender is a member of the community of users that share the key.

VII. TECHNOLOGY MODULES

1. User Authentication:

The user is facilitated here to authenticate and thus, ensure that only valid users can access the application. But, it also tracks the user login operation and accordingly redirects the user to the decoy application.

2. Admin Module:

This module facilitates the admin to manage users, the data stored and the invalid activities occurring within the application. Thus, this user will be responsible for tracking the application functionalities. A set of valid access rules will also be defined by the admin for identification of invalid users.

3. File Search Module:

This module will enable to track whether the search operations executed by the user follow a valid set of operations or not. Accordingly, the system will decide whether the user should be redirected to the decoy environment.

4. Data Access Module:

The data available for user access will be authenticated using a separate user key specified by the application to the user during registration. Based on the validity of this user key the system will redirect the user to the Decoy Module for tracking and prevent invalid distribution of data. This key data will be secured in the system using HMAC mechanism.

5. Decoy Module:

This module will facilitate the system to redirect invalid users to a dummy set of modules wherein invalid data will be distributed to the invalid user and the user activities will be notified to the admin. Thus, the system will not notify the invalid user about the detection of invalid activity and prevent further attack on the system.

VIII. COMBINING THE TWO TECHNIQUES

The correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector's accuracy. We hypothesize that detecting abnormal search operations performed prior to an unsuspecting user opening a decoy file will corroborate the suspicion that the user is indeed impersonating another victim user. This scenario covers the threat model of illegitimate access to Cloud data. Furthermore, an accidental opening of a decoy file by a legitimate user might be recognized as an accident if the search behavior is not deemed abnormal.In other words, detecting abnormal search and decoy traps together may make a very effective masquerade detection system. Combining the two techniques improves detection accuracy.

We use decoys as an oracle for validating the alerts issued by the sensor monitoring the user's file search and access behavior. In our experiments, we did not generate the decoys on demand at the time of detection when the alert was issued. Instead, we made sure that the decoys were conspicuous enough for the attacker to access them if they were indeed trying to steal information by placing them in highly conspicuous directories and by giving them enticing names. With this approach, we were able to improve the accuracy of our detector. Crafting the decoys on demand improves the accuracy of the detector even further. Combining the two techniques, and having the decoy documents act as an oracle for our detector when abnormal user behavior is detected may lower the overall false positive rate of detector.

We trained twenty classifiers with computer usage data from 10 computer science students collected over a period of 5 days on average. The classifiers were trained using the **search behavior anomaly detection technique** .We also trained another 10 classifiers using a detection approach that combines **user behavior profiling with monitoring access** to decoy files placed in the local file system, as described above. We tested these classifiers using simulated masquerader data. Figure 5 displays the AUC scores achieved by both detection approaches by user model. The results show that the models using the combined detection approach

achieve equal or better results than the search profiling approach alone.

AUC Comparison by User

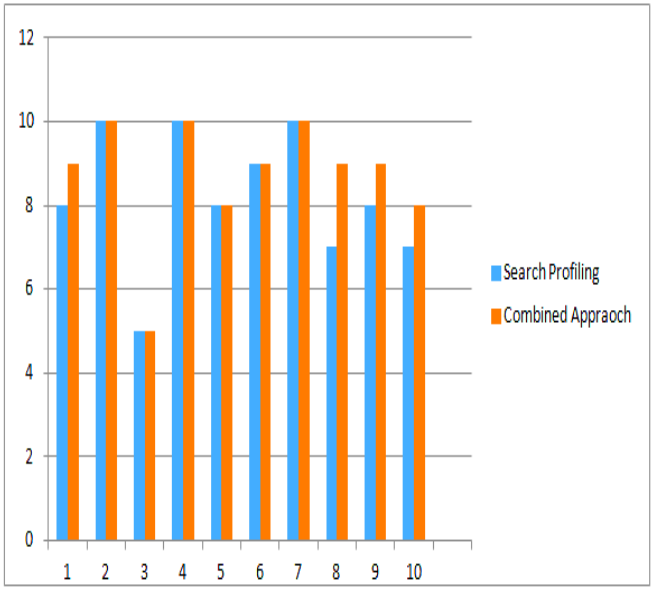

Fig. 4. AUC Comparison by User model for the Search Profiling and Combined Approach

The results of our experiments suggest that user profiles are accurate enough to detect unauthorized Cloud access . When such unauthorized access is detected, one can respond by presenting the user with a challenge question or with a decoy document to validate whether the access was indeed Unauthorized, similar to how we used decoys in a local file setting, to validate the alerts issued by the anomaly detector that monitors user file search and access behavior.

Abbreviations and Acronyms—

AUC-Authentication Certificates, API-Application Programming Interface, CSA-Cloud SecurityAlliance, DCE-Distributed Computing Environment, DS-Distributed Systems, HMAC- Hash Message Authentication Code, SMB-Small and Medium Business etc.

IX. Conclusion

In this paper, we discussed about similarities and differences regarding to two concepts, distributed computing and cloud computing. Distributed computing is a computing concept that, in its most general sense, refers to multiple computer systems working on a single problem. In distributed computing, a single problem is divided into many parts, and each part is solved by different computers.And Cloud computing is an umbrella term used to refer to Internet based development and services.

And in this paper, we also present a novel approach to securing personal and business data in the Cloud. We propose monitoring data access patterns by profiling user behavior to determine if and when a malicious insider illegitimately accesses someone's documents in a Cloud service. Decoy documents stored in the Cloud alongside the user's real data also serve as sensors to detect illegitimate access. Once unauthorized data access or exposure is suspected, and later verified, with challenge questions for instance, we inundate the malicious insider with bogus information in order to dilute the user's real data.

Acknowledgment

This work was partially supported by the European Union FP7/2007-2013 under project TCLOUDS (grant agreement 257243), and based on work supported by the Defense Advanced Research Projects Agency (DARPA) under the ADAMS (Anomaly Detection at Multiple Scales) Program with grant award number W911NF-11-1-0140 and through the Mission-Resilient Clouds (MRC) program under Contract FA8650-11-C-7190.

References

[1] Cloud Security Alliance, "Top Threat to Cloud Computing V1.0,March 2010. [Online]. Available : https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf.

[2] M. Ben-Salem and S. J. Stolfo, "Combining a baiting and a user search profiling techniques for masquerade detection," in Columbia UniversityComputer Science Department, Technical Report # cucs-018-11, 2011. https://mice.cs.columbia.edu/getTechreport.php?techreportID=1468.

[3] The permanent and official location for the Cloud Security Alliance Top Threats research is: http://www.cloudsecurityalliance.org/topthreats The Keyed-Hash Message Authentication Code (HMAC).pdf

[4] International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.1, January 2013 DOI : 10.5121/ijcnc.2013.5112171 A USER PROFILE BASED ACCESS CONTROL MODEL

[5] FOG Computing Mario Nemirovsky – ICREA/BSC With the Coloaboration Rodolfo Milito – CISCO Marcelo Yanuzzi – UPC.

[6] A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa. DepSky: Dependable and secure storage in a cloud-of-clouds. In Proceedings of the European Conference on Computer Systems (EuroSys), pages 31–46, April 2011.

[7] Cloud Security Alliance. Top threats to cloud computing v1.0, March 2010.

[8] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov. The Eucalyptus Opensource cloud-computing system. In Proceedings of the IEEE International Symposium on Cluster Computing and the Grid pages 124–131, 2009.

[9] E. Grosse, J. Howie, J. Ransome, J. Reavis, and S. Schmidt. Cloud computing roundtable. IEEE Security Privacy, 8(6):17 –23, 2010.J. Jones. (1991, May 10). Networks (2nd ed.)[Online].Available: http://www.atm.com