A Simplified Rule Based Distributed Information Flow Control for Cloud Computing

Awani Joshi, Palak Purohit, Richa Jain

Department of Computer Science and Engineering,
Medi-Caps Institute of Technology and Management Indore, M.P, India

Abstract— Software development is now moving towards an efficient solution rather than achieving just in time goes. It focuses on the work effectiveness of individual components and measured by analyzing their interdependent behaviors. This behavior includes the nature of their responses, information flow between their components, isolation of the software and cloud based systems. It assures confidentiality and integrity, both with a clear isolation between the user interaction with the system. Mainly it separates the traffic generated by different users or components by using the tagging or labelling mechanism. By traffic, this categorization is made very easy and cost effective. After the categorization is achieved, the data is separated into multiple objects belonging to different subjects. Previously, there various mechanisms developed to provide effective IFC. But still there are some issues related with clear separation of a user's interaction, classification of flow information, categorizing the traffic classes, implicit tagging, covert channel, over and under labelling etc. This work suggests a simplified rule based distributed information flow control for cloud computing. Here the various rules are formed for guiding the information flow and achieves clear classification with higher accuracy. At the analytical level of evaluation, the approach is serving all the needs of effective flow control mechanism and later prototype will justify the same.

Keywords— Software System, Cloud Computing, Information Flow Control (IFC), Distributed IFC, Labelling, Classification, Rule Based.

I. INTRODUCTION

Software based system is a group of multiple instruction for performing the desired task. During this execution, the information flows between various modules of the software and the hardware installed with it. This information based data flow should be controlled for securing the exchanges of software systems. Securities of such flows have to perform two major tasks: first is to make the information secure from the outsider or attacker and second is to protect the distributional flow of information. From the above objectives only the prior one gets resolved effected using some authentication and confidentiality mechanism. Models of secure data access are often classified into Mandatory Access Control (MAC) or Discretionary Access Control (DAC) systems [1]. They only prevent the information updates and outcomes but does not control the information flows between the systems.

Nowadays, the software systems advancements are letting their support for distributed and parallel systems for supporting the market based requirements. One of such requirements is cloud computing which provides the various computing paradigms as a service to the users by provisioning and sharing. Here the information flows through various heterogeneous systems rather than a single system. Now for the cloud based system must require some modification in traditional flow control mechanism which only supports the formal access control. This access control is not sufficient for complete security. While the information flows handles how the information gets propagated during the complete operations of the system.

Understanding Information Flows Control (IFC)

Propagation model for information must follow two basic requirements: Confidentiality using cryptosystem and access control and integrity which protects the reading or writing of information. It aims towards making the information flow, secure using public output dependence, non interference and secret input processing. One of the important objectives of information flow integrity is to prevent its declassification from unintended user. Some of the outlines for getting a complete inner view of information flow some of the outlines are given with [2] are as follows:

- It controls the information dissemination using the propagation models from different heterogeneous objects.
- It frequently partitions the information into various classes to get the separate handling of objects and subjects for improving the flow security.
- The boundaries and conditions of security classes cannot be changed once created and hence the entity places respectively do not change their class.
- All the information flowing through the internal systems gets unambiguous paths which follow the security rules.

Both confidentiality and integrity policies can be staged into two parts. The policy language describing the possible classifications of information and how the different classifications relate, i.e., how information from one classification may flow with respect to the other and the semantic characterization describing the meaning of the
policy language in terms of the semantics of the programming language.

The former is predominantly described using a lattice model for information classification. Keeping in mind the above requirements or rules, an attacker’s model is required from which the comparison for attack confirmation has to be made. Traditionally, such attacker model is programmed centric, which has the control of the program completely and performs the operation to destroy or affects the normal working of the system. But as the web based computing and operation gets on operating between distant locations through clouds, these attacker models are inadequate for achieving the security goals. Some of the common model which works with the security improvements is:
- Access Control Lists (ACLs)
- Role-Based Access Control and Capability Systems

**RBAC are DAC systems**

In the later one (DAC) the owner of the data can modify access permissions to be taken. Systems trying to get protected by getting control over access to resources and infrastructures. Implementing them often focus on where access control checks are performed in the code of an application. The data is protected as a function of access control checks in the APIs provided to interact with that data. There are some problems also which the above models, especially the DAC faces are:

(i) Web-Based access control systems can be easily bypass.
(ii) Data can propagate or influence system behavior indirectly in ways that are discursive, detection of those loops holes may not be possible which access control barriers.

**IFC with Cloud**

Cloud is the recent technology, which aims to provide the applications and other resources using a computing paradigm through a service model to the user or intermediate provider with reduced managerial loads. It lets the provider and user the rapid adoption of application and other services using an online mechanism. As the number of users, their types, and requirements gets increased with the cloud their security controlling requirements also gets complex. Thus, achieving the confidentiality in cloud specifically with the data flow through tracking of boundaries for data is the recent field of work for researchers. It aims towards making the data of flow viewer which categorizes the data according to their privileges and authenticity ad forwards them accordingly to their destinations.

Here the data is first classified into various security classes with different access rights and flows the isolation rules using flow provisioning. To maintain the complete security boundary, the flow control will maintains the registry for each flow and give the categorization for protecting the sensitive data. Thus, here the information and its flow get security transited by isolating the data itself with low information. Users have to trust the efforts of both the third-party service provider and the cloud infrastructure provider for proper handling their private data as intended.

This works aims towards making the information flow control a complete a secure system using some of the novels filtering and label reading mechanism. Also the suggested models will help in acquiring the filtering schemes as per the system requirements. Traditionally, the information flow is working for only the static situations where the same types of software system is at both the end, but with this work, dynamic and heterogeneous supports for secure information flow with clear isolation is achieved in near future.

**II. Background**

Cloud computing is the recent area of advancements where the user is provided with their application in an effective service based delivery models. The ongoing shift of user in such technology giving rise to the significant concerns about security options over here. All it needs to increase the trust which should be more than traditional computing. Most of the companies are now working towards making their sensitive information secure form the attackers or must follow the access control according to their information visibility levels. Also, large scale companies are now moving towards integrating their solution with cloud computing which satisfies their business needs. But, the third party based access and data storage and exchanges always involves vulnerabilities related to their privacy and confidentiality constraints.

To add this objective with traditional developments of clouds lots of work had been done by the researchers. Most common situation faced by these organizations is multi-tenancy. It is problems here the users accessing their common or shared records form the different locations to some shared servers or data centers. It involves the risk of information dissemination. Here the attacker or some unintended user might disrupt their communication or might affects their data. It causes the compromises and gets reduction in trust towards cloud based applications.

Many security fears associated with cloud computing therefore revolve around incomplete isolation of these myriad users. Large category of cloud security research has therefore concerned the enforcement of various forms of data access control in clouds. The standard way to protect confidential data is (discretionary) access control: some privilege is required in order to access files or objects containing the confidential data [4]. Access control checks place restrictions on the release of information but not its propagation. Once the information access of contents gets open for attacker than some other changed information or denial request can be flowed in later communication causes assets damages for organizations. Thus the objective with the work is:

“To ensure that information is used only in accordance with the relevant confidentiality policies, it is necessary to analyse how information flows within the using program.”

Most the traditional security generators beliefs that the system is secure if the data is get converted to some encrypted content for achieving the confidentiality. But its flow towards various system modules and zones of information access should also be considered.
Confidentiality policies must be enforced to data as well as their information flow path also. The security analyst must control the flow of information through some confidentiality and integrity policies and should restrict their movements towards location which violates their rules or provisions. It is the best way to achieve the system design principle of end to end construct. Some of the researchers had made the secure programming languages which restricts the motion or flow of data according to strongly typed methodologies. Mainly, they have some policy based annotations which can be readout by the compilers and termed as security type checking for internal information flow to achieve complete isolation of data and their users. It provides a suitable tracking data flows across different services offers the cloud provider a way to log sensitive operations on tenant data frequently, so improving accountability.

How IFC Works: Labelling [5]

Information flow control is a data oriented approach works towards achieving protection of internal flowing information through security labels which tracks and limits the data flow and transitions. Here the labels are associated with the primary functionality of the developed system. It defines the provisions for permitting the secure exchanges of information based on some trust relationship between the labeled data and their requesters. That is, data protection policy checking can be based on comparing the labels associated with the data with the labels held by principals. Example: permitting unprivileged users to pass information to privileged users, but not read privileged information (so-called no read up, a write down”) with matching restrictions on the privileged users.

It can be further extended to forcefully apply the general provisions of security through appropriate labelling and verification schemes. All its aims are to providing stronger defense than the formal once applied with DAC. Labels are mainly used to classify the data units flowing between the multiple system modules to identify the flow which is following the constraints of security towards integrity and confidentiality. Here the integrity provides the quality of data and confidentiality serves the security. Secrecy concerns where data is permitted to flow to, and integrity where it is allowed to come from it. Implementation of IFSs must ensure that labels can be allocated to principals but not be forged by them, can be allocated and “stuck” to data pieces, and that label checking enforces security policy regarding all aspects of information flow.

Labeling with Privileges

A system of privileges operates to introduce carefully controlled additional components into the Trusted Computing Base that can modify labeling contrary to the default restrictions. The privilege to override “secrecy” IFC restrictions is known as the declassification privilege. In this case, IFC labels such as publically accessible, secret and top-secret would be associated with data items and principals and used to enforce the required security policy.

Cloud Requirements Criteria’s for Implementing IFC

Cloud computing has particular needs in terms of information flow security. Here possible requirements for two cloud-hosted, interacting applications. Data isolation must be provided between compartments of the applications, flows tracked and/or enforced on input and output and inter-compartment and application communication [6]. It can be achieved by using following four criteria:

Criteria (i) When the system operates, (static, runtime, hybrid)

Criteria (ii) How the system isolates data, (e.g. hardware-assisted OS and virtualization mechanisms, programming language and library mechanisms)

Criteria (iii) How the system tracks data flow across isolated data, (e.g. domain level, process level, variable level, message level) and

Criteria (iv) How the system uses the output of data flow tracking to enforce data flow (how policy is specified, the structure of label metadata, and the declassification of security data)

Security engineering involves tradeoffs between security and efficiency. The designers of IFC systems will select their threat models to inform any compromises they need to make within the IFC design space.

III. LITERATURE SURVEY

During the last few years various authors had suggested several modifications in traditional information flow control models. Out of those some had worked towards achieving it a specific way to improve its performance factors and reduces overheads would be taken here as literature survey. Conflict of Interest (COI)

In the article [7], information flow control for objects and subjects are used to prevent the data dissemination of data using conflicts of interest (COI) phenomenon. Here the developed mechanism will improve the prior existing Chinese Wall Security Policy through levels wise separation of the industrial usable data. The highest level consists of conflict of interest classes which group all company datasets whose companies are in the competition together. All the subjects are allowed to access their data according to their interest. The paper resolves the specific issues of side channel vulnerability by its constrained regulations. Although efficient prevention of side channels is difficult within a single node in a network, there is a unique opportunity within a cloud. Our work proposes a low-overhead approach to cloud wide information flow policy enforcement through Cloud Flow information flow controlling tool. The approach identifies the side channels which could potentially be used to violate a security policy through run-time introspection at tone, and reactively migrating virtual machines to eliminate node-level side-channels.

Decentralized information flow control (DIFC)

In the work [8], a decentralized information flow control (DIFC) is suggested for improving the programmable writing of security controls. Here the DIFC runs on shared hardware and categorizes into language level and operating system. Traditionally the language level DIFC does not guarantees the security flows violations. Similarly the operating system based approaches are sometimes do not gives effective security in case of shared
resource and fine grained access. This paper gives an approach Laminar for flow control using set of abstraction for operating system and heap allocation policy based other objects. Here the programmers are defining these security labels which cover the aspects of both confidentiality and integrity both. Laminar enforces the security policies specified by the labels at runtime and limits dynamic security checks using the DIFC. It also supports the multithreaded monitoring model using heterogeneous labelling process.

Some of the papers also showed the cloud based flow control using some virtual machine monitoring and controlling functions.

**H-One Approach**

In a way to do so, the paper [9] gives a administrative approach using the hypervisor process controlling and named as H-One. It is a new auditing mechanism which uses information flow tracking for effective privacy preserving solutions in cloud environments. The tool aims towards recoding all the types of flow starting or goes with the installed VM. Currently the H-One is working with the Xen hypervisors which will be extended for others also. The administrator has root privileges on the management stack and thus has the ability to use all privileges conferred on the administrative VM.

Some of the authors first led the basic understanding of all the cloud security controls which helps in further modification with traditional solutions. Once the clear and separated security requirement gets finalized the solution developments can be performed. The paper [10] provides a concise but all-round analysis on data security and privacy protection issues associated with cloud computing across all stages of data life cycle. It also addresses the major problem with cloud security handling is multi-tenancy. Now for achieving the complete isolation between its multiple users and applications various factors are analyzed such as scalability, massive processing, service delivery model, sensitive information, virtual resource sharing etc.

In the work [11], a distributed model for fine grained information flow control is presented which allows dynamic delegation and the revocation rights. The paper uses the defined Haskell libraries for such integrations for decentralized label allotments with individual integrity and confidentiality policies. It is a language based model which includes first-class references, higher-order functions, declassification and endorsement of policies, and user authority in the presence of global unrestricted delegation. The DLM allows each resource or collection of resources to be managed by a different principal. This principal owns the resources: it specifies the security requirements via a set of policies and is responsible for enforcing them. Here the policies use a label consists of exactly two policies: a confidentiality policy R and an integrity policy W. Integrity policy express which principals “trust” the data. Thus, dual to the ordering on confidentiality policies, an integrity policy that states that everyone must trust the data is the most restrictive, and an integrity policy that requires no one to trust the data is the least restrictive.

**Chinese Wall policy**

In the work [12] a information flow model based on Chinese Wall policy is used to protect the sensitive information. It is a kind of information disclosure policy with effective processing of information belongings to various organizations. Here the companies are gets separated according to their conflicts of interest classes based on their service types. The major components of the approach are query processing in cloud having CQL based executions. The paper suggests following functionality like multiuser server with specialized authentication, replicated query processing, trusted schedulers and operators, security level aware window and other options, unauthenticated flow categorization and blocking etc.

**The Interactive Fiction Database**

The work [13] presents IFDB based database managements system which provides security to the decentralized flow control using query label monitoring. It uses a novel abstraction of relational query based handling of flows. As the labeling and tagging is the first step for traffic and flow type detection. In this step an identifier is attached with the flow which defines the sensitivity of the data flowing between the systems modules. Here the labels are the set of tags that summarize the sensitivity of the information contained in the data. Controlling Information Flow IFDB ensures that the label of each object reflects the tags of all the data that produced it, and the label of each process reflects the tags of all the data the process read. It does this by enforcing the following standard rule. The Principle of Least Privilege Delegation of authority makes it possible to define who can declassify, but it doesn’t constrain how that authority may be used. At the evaluation points of view it is found that the IFDB reduces authority closures and calls. It gives good throughput for complex systems based applications when tested.

In the work [14], a traffic classification technique is well studied with some state of art modification is suggested for further improvements. It aims towards improving the current flow structure with statistical feature analysis using machine learning techniques. In this the classification performed is affected by the limited supervised flow of the system information. For effective classification a new method of handling and tagging the unknown application data is given with the paper also. The suggested method uses supervised information training the superior capability of detecting unknown flows generated by unknown applications and utilizing the correlation information among real-world network traffic to boost the classification performance.

**IV. PROBLEM DEFINITION**

Information flow is the behaviour analysis process for flowing data between the various system components or between other systems. Currently there is various solution of this effective flow analysis is suggested for both single and distributed systems. Also the market oriented computing with cloud deployments will face the isolation issues majorly with shared resources. In this situation, categorizing the data from different sources according to
their sensitivity of information is a quite complex task. Here the user outsources their data to any third party provider far apart from their trusted zones. Now if some user intentionally tries to fabricate this information at provider’s location, trust on the system gets reduced and losses occur. In this situation various flow based labelling and tagging approaches suggested over the last few years are used for isolating the traffic. But in case of cloud the similar types of virtual machine will generate same traffic and it is very tedious task to separate such traffic. Even though some of the problems which remain unaddressed is found out during the survey. These problems are given as:

**Problem (i)** Simultaneous multiple virtual machine access to the same system might not be separated because the labels assigned with that will be same and hence isolation violates.

**Problem (ii)** Information flow security policy for infrastructure based outsourced environment is not yet achieved effectively.

**Problem (iii)** The decentralized solution of flow analysis and distributed approaches suffers from over labelling and under labelling. There is no such process which gives the exact labelling requires. If the information flow transits from multiple level of VM’s and physical machines and then through network, the single data is gets overloaded with multiple labels and makes the degradation in systems performances.

**Problem (iv)** The solution must able to handle the implicit tagging and coverts channel problems with reduced load.

**Problem (v)** Separation of objects and subjects must be classified clearly with some mining based approach for improving categorization of different data into different classes and service the complete isolations.

Apart from the other issues which this work had identified there are various direction available in the literature for improving the classification of flow and their filtering mechanism. But somehow, it’s a wide area and the work needs to restrict itself to achieve the time based goals. Hence to works aims towards improvements in distributed information flow control (DIFC) for cloud computing.

**V. OBJECTIVES OF WORK**

This works aims towards developing some new security controls for information flow in cloud based environments. It verifies the requirements first to serve the dynamic information flow control both at the pipelines and programming level of system. The process first categorizes the data and later on by effective labelling and partitioning of data into various classes the information isolation can be achieved.

Some of the defined direction of work is given here are:

(i) Before applying the information flow model for cloud first the service models where the flow analysis is required needs to be indentified first. Normally the infrastructure is analyzed first then platform and later on software.

(ii) Policy formation must be clearly defined along with the entities of the system and their privileges for accessing the flow.

(iii) Policy enforcements and flow analysis using labelling does not affects other process of the system by which degradation in performance can be avoided.

**VI. PROPOSED SOLUTION**

This work suggested a novel distributed information flow control model for handling of sensitive information in cloud computing environment. It works towards making the secure flow of information between the various cloud elements and shared resources. Primarily, the approach aims towards achieving the isolation between the users and providers of the cloud. Here in the figure1 below, the access control model blocks the information flow between the entities of the system if they are not belongs to the similar security groups. The sharing of data and other resources must be managed effectively maintaining the confidentiality and integrity of the data. This secure information flow model follows the set of guiding rules for monitoring the traffic flows. The traditional concept of Chinese wall security is completely followed here with some more rules for further improvements.

**Description:** Cloud is the shared medium where the user interacts with their data and applications at remote locations parallel with other users of the system. Now, with such a high interactions and heavily flowed traffic isolation and sensitivity of the data must be maintained. Now, the user first gets registered itself to the system with active creation of objects. For each user the interaction involves must be of object type. Each object must belong to the class whose data access nature and sources are common. Now, as an entity of the cloud, subjects will serve their resources to other entities by further spreading their objects boundaries. Each object of the subject will have permission or privileges with respect to that only the information up to a certain level of sensitivity is accessed by the user. Thus, a sharing of information will only be allowed for same group objects. The rules are maintained by the security administrator as its controlling part. Other than security rules, the administrator can also add, modify or delete rules and classes based on the conflicts of interest factors.

When the data enters into system or when a request of data access is generated these rules are enforced into the system by information flow modules of...
infrastructure as a service layer. Here the flow is managed using four basic components.

(i) Policy Enforce
(ii) Tagging and Reading
(iii) VM Instance
(iv) Classification

The policy enforcer will apply the security constraints on the data to create a sensitivity level access. It serves the isolation fulfillment by which later on the traffic and their sources can be identified. After the policy enforcer decides the data on which the rules are applied, the tagging module applies the tag accordingly. If the data is previously tagged with multiple labels than, they all are removed and which later on applies on a unified tag by which overall security interpretations can be made. The module also adds the tags form the traffic and separates them accordingly.

This separation of object based data is done in multiple classes formed according the properties of their subjects and actions of creators. These classes are security groups sharing common data and sources or devices. With cloud environment, the VM instance regularly monitors the types of information flowing from the VM to the provider or the users. If the flow violates the security rules, such data dissemination is blocked. After all the rules are satisfied then only the requester gets the data access accordingly. Flow can be in between various cloud providers and users makes the system operation very complex. Thus the rules which forces the security operations gets reduced and simplified with a unique rules formation to direct further flows of traffic.

The rules created by the security administrator will make the process simple and less overhead involved. Now the rules will maintain the conflict of interest, sensitivity of information, reduced tagging, prioritize the abstracted data with high sensitivity and maintains the value of information. Thus after clear formation of approach, it looks to satisfies all the security requirements for information flow controls. Also, the suggested approach will works towards resolving the identified issues of tagging and overheads. Hence it proceeds towards achieving its objective with an implementation prototype in near future.

VII. CONTRIBUTION AND BENEFITS

The suggested information flow control for distributed environments in cloud computing will serve following contributions:

(i) The suggested work aims towards clear data flow tracking and portioning based on simplified rules which are maintained by security administrator. These rules can be audited and confined with all the requirements.
(ii) Here the tag size can also be fixed so that instead of over tagging or under tagging, a unified single tag must be sufficient for achieving complete security.
(iii) Fewer rules will reduce the loads and hence the tag reading will not delay the information flows and maintains its value.
(iv) Automated tagging read and writes does not require manual interventions which will also identify the untagged information’s.
(v) Information sensitivity varies with respect to time, thus the work will also have possibilities for rules modification for COI through administrative privileges.
(vi) As, this tagged flow traffic is analyzed at the compile time, if the system calls priority gets higher than tag, then it could be processed or forwarded its and vice versa is also true. Thus, the tag priority helps in normal working of the system.
(vii) Hierarchical tagging or annotation can also be supported by the approach and is emulsifying with unified tagging scheme.
(viii) Dynamic handling and modification of rules is provided for wide workability.
(ix) Secure compiled code may be transferred from a remote site and run locally with less concern that it might leak information. Code transfer is useful both for clients, which download applications from servers, and for servers, which upload code and data from clients for remote evaluation.

VIII. CONCLUSION

As Cloud computing is getting popular day by day, Cloud service providers need to update their systems with the policies which may lead to better performance as well as flow control. This work is focusing on a simplified rule based distributed information flow control for cloud computing. Here the various rules are formed for guiding the information flow and achieves clear classification with higher accuracy. At the analytical level of evaluation, the approach is serving all the needs of effective flow control mechanism and later prototype will justify the same.

ACKNOWLEDGMENT

I am especially grateful to my guide Mrs. Richa Jain, Assistant Professor, Department of Computer Science and Engineering, Medi-caps Institute of Technology and Management, Indore for her comment and direction in my research work. My sincere thanks are due to Dr. S. K. Somani, Director, Medicaps Institute of Technology and Management, Indore for individual encouraging thoughts. I express my heartfelt gratefulness to Dr. Promod Nair, HOD, Computer science and Engineering Department, Medicaps Institute of Technology and Management, Indore, for his stimulating supervision.

REFERENCES

"Collaboration-Based Cloud Computing Security Management
Framework" published in 4th International Conference on Cloud
Computing IEEE 2011.

Balasubramaniam, Radu Sion, and Donald E. Porter "CloudFlow:
Cloud-wide policy enforcement using fast VM introspection"
published in 2011.

[8] Pardeep Kumar, Vivek Kumar Sehgal, Durg Singh Chauhan, P.
K. Gupta and Manoj Diwakar "Effective Ways of
Secure, Private and Trusted Cloud Computing" published in IJCSI
International Journal of Computer Science Issues, Vol. 8, Issue 3,
No. 2, May 2011.


[10] Deyan Chen, Hong Zhao "Data Security and Privacy Protection
Issues in Cloud Computing" published in International Conference
on Computer Science and Electronics Engineering IEEE 2012.

with Restricted Delegation and Revocation in Haskell" published
in ACM Sep 2012.

[12] Xing Xie, Indrakshi Ray, Raman Adaikkalavan, Rose Gamble "
Information Flow Control for Stream Processing in Clouds"


[14] Jun Zhang, Chao Chen, Yang Xiang, Wanlei Zhou, and
Athanasios V. Vasilakos "An Effective Network Traffic
Classification Method with Unknown Flow Detection" published
10, No. 2, June 2013.