Security Issues for Cloud Computing

Technical Report UTDCS-02-10
Department of Computer Science
The University of Texas at Dallas
February 2010
Kevin Hamlen, Murat Kantarcioglu, Latifur Khan and Bhavani Thuraisingham

Security Issues for Cloud Computing

Kevin Hamlen
Murat Kantarcioglu
Latifur Khan
Bhavani Thuraisingham

The University of Texas at Dallas

This work is supported in part by the AFOSR project on Secure Information Grid

ABSTRACT

In this paper we discuss security issues for cloud computing including storage security, data security, and network security and secure virtualization. Then we select some topics and describe them in more detail. In particular, we discuss a scheme for secure third party publications of documents in a cloud. Next we discuss secure federated query processing with map Reduce and Hadoop. Next we discuss the use of secure co-processors for cloud computing. Third we discuss XACML implementation for Hadoop. We believe that building trusted applications from untrusted components will be a major aspect of secure cloud computing.

1. INTRODUCTION

There is a critical need to securely store, manage, share and analyze massive amounts of complex (e.g., semi-structured and unstructured) data to determine patterns and trends in order to improve the quality of healthcare, better safeguard the nation and explore alternative energy. Because of the critical nature of the applications, it is important that clouds be secure. The major security challenge with clouds is that the owner of the data may not have control of where the data is placed. This is because if one wants to exploit the benefits of using cloud computing, one must also utilize the resource allocation and scheduling provided by clouds. Therefore we need to safeguard the data in the midst of untrusted processes.

The emerging cloud computing model attempts to address the explosive growth of web-connected devices, and handle massive amounts of data. Google has now introduced the MapReduce framework for processing large amounts of data on commodity hardware. Apache's Hadoop distributed file system (HDFS) is emerging as a superior software component for cloud computing combined with integrated parts such as MapReduce. The need to augment human reasoning, interpreting, and decision making abilities have resulted in the emergence of the Semantic Web, which is an initiative that attempts to transform the web from its current, merely human-readable form, to a machine processable form. This in turn has resulted in numerous social networking sites with massive amounts of data to be shared and managed. Therefore we urgently need a system that can scale to handle a large number of sites and process massive amounts of data. However, state of the art systems utilizing HDFS and MapReduce are not sufficient due to the fact that they do not provide adequate security mechanisms to protect sensitive data.

We are conducting research on secure cloud computing. Due to the extensive complexity of the cloud, we contend that it will be difficult to provide a holistic solution to securing the cloud at present. Therefore our goal is to make increment enhancements to securing the cloud that will ultimately result in a secure cloud In particular, we are developing a secure cloud consisting of hardware (includes 800TB of data storage on a mechanical disk drive, 2400 GB of memory and several commodity computers), software (includes Hadoop) and data (includes a semantic web data repository). Our cloud system will: (a) support efficient storage of encrypted sensitive data, (b) store, manage and query massive amounts of data, (c) support fine grained access control and (d) support strong authentication. This paper describes our approach to securing the cloud. The organization of this paper is as follows: In section 2 we will give an overview of security issues for Cloud. In section 3 we will discuss secure third party publication of data in clouds. In section 4 we will discuss how encrypted data may be queried. Section 5 will discuss Hadoop for cloud computing and our approach to secure query processes with Hadoop. The paper is concluded in section 7.

2. Security Issues

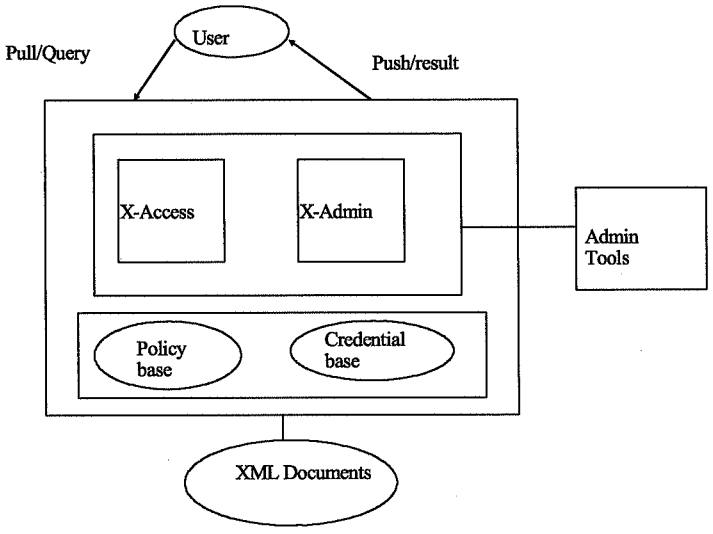
There are numerous security issues for cloud computing as it encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management. Therefore, security issues for many of these systems and technologies are applicable to cloud computing. For example, the network that interconnects the systems in a cloud has to be secure. Furthermore, virtualization paradigm in cloud computing results in several security concerns. For example, mapping the virtual machines to the physician machines has to be carried out securely. Data security involves encrypting the data as well as ensuring that appropriate policies are enforced for data sharing. In addition, resource allocation and memory management algorithms have to be secure. Finally, data mining techniques may be applicable to malware detection in clouds.

We will focus only on some aspects of secure cloud computing. One is to efficiently store the data in foreign machines. Another is to query encrypted data as much of the data on the cloud may be encrypted.

We are also using Hadoop distributed file system for virtualization and applying security for Hadoop which includes an XACML implementation. In addition we are investigating secure federated query processing on clouds over Hadoop. These efforts will be described in the subsequent sections.

3. Third Party Secure Data Publication Applied to CLOUD

Cloud computing facilitates storage of data at a remote site to maximize resource utilization. As a result, it is critical that this data be protected and only given to authorized individuals. This essentially amounts to secure third party publication of data that is necessary for data outsourcing as well as external publications. We have developed techniques for third party publication of data in a secure manner. We assume that the data is represented as an XML document. This is a valid assumption as many of the documents on the web are now represented as XML documents. First we discuss the access control framework proposed in [BERT02] and then discuss secure third party publication discussed in [BERT04].



In the access control framework proposed in [BERT02], security policy is specified depending on user roles and credentials (see Figure 1). Users must possess the credentials to access XML documents. The credentials depend on their roles. For example, a professor has access to all of the details of students while a secretary only has access to administrative information. XML specifications are used to specify the security policies. Access is granted for an entire XML document or portions of the document. Under certain conditions, access control may be propagated down the XML tree.

Figure 1. Access Control Framework

For example, if access is granted to the root, it does not necessarily mean access is granted to all the children. One may grant access to the DTDs and not to the document instances. One may grant access to certain portions of the document. For example, a professor does not have access to the medical information of students while he has access to student grade and academic information. Design of a system for enforcing access control policies are also described in [BERT02]. Essentially the goal is to use a form of view modification so that the user is authorized to see the XML views as specified by the policies. More research needs to be done on role-based access control for XML and the semantic web.

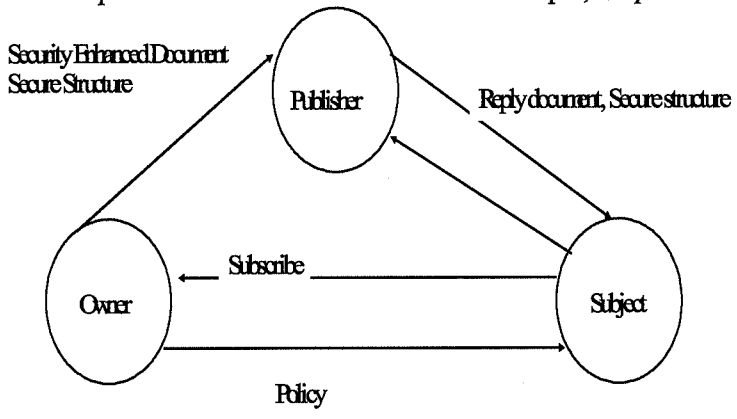


Figure 2. Secure Third Party Publication

In [BERT04] we discuss the secure publication of XML documents (see Figure 2). The idea is to have untrusted third party publishers. The owner of a document specifies access control polices for the subjects. Subjects get the policies from the owner when they subscribe to a document. The owner sends the documents to the Publisher. When the subject requests a document, the publisher will apply the policies relevant to the subject and give portions of the documents to the subject. Now, since the publisher is untrusted, it may give false information to the subject. Therefore, the owner will encrypt various combinations of documents and policies with his/her private key. Using Merkle signature and the encryption techniques, the subject can verify the authenticity and completeness of the document (see Figure 2 for secure publishing of XML documents).

In the Cloud environment, the third party publisher is the machine that stored the sensitive data in the cloud. This data has to be protected and the techniques we have discussed above have to be applied to that authenticity and completeness can be maintained.

4. ENCRYPTED DATA STORAGE FOR CLOUD

Since data in the cloud will be placed anywhere, it is important that the data is encrypted. We are using secure co-processor parts cloud infrastructure to enable efficient encrypted storage of sensitive data. One could ask us the question; why not implement your software on hardware provided by current cloud computing systems such as Open Cirrus? We have explored this option. First, Open Cirrus provides limited access based on their economic model (e.g., Virtual cash). Furthermore, Open Cirrus does not provide the hardware support we need (e.g., secure co-processors). By embedding a secure co-processor (SCP) into the cloud infrastructure, the system can handle encrypted data efficiently (see Figure 3).

Basically, SCP is a tamper-resistant hardware capable of limited general-purpose computation. For example, IBM 4758 Cryptographic Coprocessor [IBM04] is a single-board computer consisting of a CPU, memory and special-purpose cryptographic hardware contained in a tamper-resistant shell; certified to level 4 under FIPS PUB 140-1. When installed on the server, it is capable of performing local computations that are completely hidden from the server. If the tampering is detected then the secure co-processor clears the internal memory. Since the secure coprocessor is tamper-resistant, one could be tempted to run the entire sensitive data storage server on the secure co-processor. Pushing the entire data storage functionality into a secure co-processor is not feasible due to many reasons.

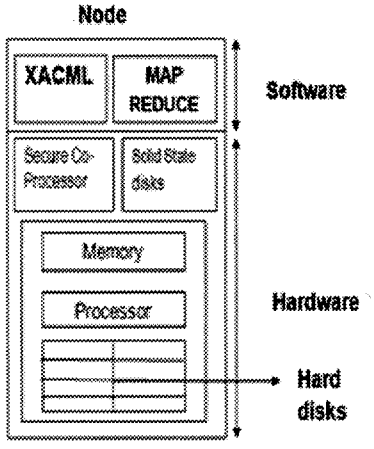


Figure 3. Parts of the Proposed Instrument

First of all, due to the tamper-resistant shell, secure co-processors have usually limited memory (only a few megabytes of RAM and a few kilobytes of non-volatile memory) and computational power [SW99]. Performance will improve over time, but problems such as heat dissipation/power use (which must be controlled to avoid disclosing processing) will force a gap between general purposes and secure computing. Another issue is that the software running on the SCP must be totally trusted and verified. This security requirement implies that the software running on the SCP should be kept as simple as possible. So how does this hardware help in storing large sensitive data sets? We can encrypt the sensitive data sets using random private keys and to alleviate the risk of key disclosure, we can use tamper-resistant hardware to store some of the encryption/decryption keys. (i.e., a master key that encrypts all other keys). Since the keys will not reside in memory unencrypted at any time, an attacker cannot learn the keys by taking the snapshot of the system. Also, any attempt by the attacker to take control of (or tamper with) the co-processor, either through software or physically, will clear the co-processor, thus eliminating a way to

decrypt any sensitive information. This framework will facilitate (a) secure data storage and (b) assured information sharing. For example, SCPs can be used for privacy preserving information integration which is important for assured information sharing [AAK06].

We have conducted research on querying encrypted data as well as secure multipart computation (SMC). With SMC protocols, one knows about his own data but not his partner's data since the data is encrypted. However, operations can be performed on the encrypted data and the results of the operations are available for everyone, say, in the coalition to see. One drawback of SMC is the high computation costs. However, we are investigating more efficient ways to develop SMC algorithms and how these mechanisms can be applied to a cloud.

5. SECURE QUERY PROCESSING WITH HADOOP

5.1 OVERVIEW OF HADOOP

A major part of our system is HDFS which is a distributed Java-based file system with the capacity to handle a large number of nodes storing petabytes of data. Ideally a file size is a multiple of 64 MB. Reliability is achieved by replicating the data across several hosts. The default replication value is 3 (i.e., data is stored on three nodes). Two of these nodes reside on the same rack while the other is on a different rack. A cluster of data nodes constructs the file system. The nodes transmit data over HTTP and clients access data using a web browser. Data nodes communicate with each other to regulate, transfer and replicate data.

HDFS architecture is based on the Master-Slave approach (Figure 4). The master is called a Namenode and contains metadata. It keeps the directory tree of all files and tracks which data is available from which node across the cluster. This information is stored as an image in memory. Data blocks are stored in Datanodes. The namenode is the single point of failure as it contains the metadata. So, there is optional secondary Namenode that can be setup on any machine. The client accesses the Namenode to get the metadata of the required file. After getting the metadata, the client directly talks to the respective Datanodes in order to get data or to perform IO actions [HA]. On top of the file systems there exists the *map/reduce engine*. This engine consists of a Job Tracker. The client applications submit map/reduce jobs to this engine. The Job Tracker attempts to place the work near the data by pushing the work out to the available Task Tracker nodes in the cluster.

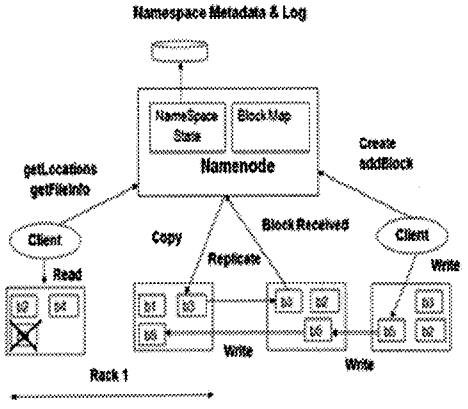


Figure 4. Hadoop Distributed File System (HDFS Architecture)

Inadequacies of Hadoop Current systems utilizing Hadoop have the following limitations:

(1) No facility to handle encrypted sensitive data: Sensitive data ranging from medical records to credit card transactions need to be stored using encryption techniques for additional protection. Currently HDFS does not perform secure and efficient query processing over encrypted data.

(2) Semantic Web Data Management: There is a need for viable solutions to improve the performance and scalability of queries against semantic web data such as RDF (Resource Description Framework). The number of RDF datasets is increasing. The problem of storing billions of RDF triples and the ability to efficiently query them is yet to be solved [MU06, TC07, RGK09a-c]. At present, there is no support to store and retrieve RDF data in HDFS.

(3) No fine grained access control: HDFS does not provide fine grained access control. There is some work to provide access control lists for HDFS [Zh09]. For many applications such as assured information sharing, access control lists are not sufficient and there is a need to support more complex policies.

(4) No strong authentication: A user who can connect to the JobTracker can submit any job with the privileges of the account used to set up the HDFS. Future versions of HDFS will support network authentication protocols like Kerberos for user authentication and encryption of data transfers. [Zh09].

However, for some assured information sharing scenarios we will need public key infrastructures (PKI) to provide digital signature support.

5.2 SYSTEM DESIGN

While the secure co-processors can provide the hardware support to query and store the data, we need to develop a software system to store, query, and mine the data. More and more applications are now using semantic web data such as XML and RDF due to their representation power especially for web data management. Therefore, we are exploring ways to securely query semantic web data such as RDF data on the cloud. We are using several software tools that are available to help us in the process including the following:

Jena: Jena is a framework which is widely used for solving SPARQL queries over RDF data [JE]. But the main problem with Jena is scalability. It scales in proportion to the size of main-memory. It does not have distributed processing. However, we will be using Jena in the initial stages of our preprocessing steps.

Pellet: We use Pellet to reason at various stages. We do real time query reasoning using pellet libraries [PE] coupled with Hadoop's map-reduce functionalities.

Pig Latin: Pig Latin is a scripting language which runs on top of Hadoop [GNC09]. Pig is a platform for analyzing large data sets. Pig's language, Pig Latin, facilitates sequence of data transformations such as merging data sets, filtering them, and applying functions to records or groups of records. It comes with many built-in functions but we can also create our own user-defined functions to do special-purpose processing. Using this scripting language, we will avoid writing our own map-reduce code; we will rely on Pig Latin's scripting power that will automatically generate script code to map-reduce code.

Mahout, Hama: These are open source data mining and machine learning packages that already augment Hadoop [MA, HM, MS08].

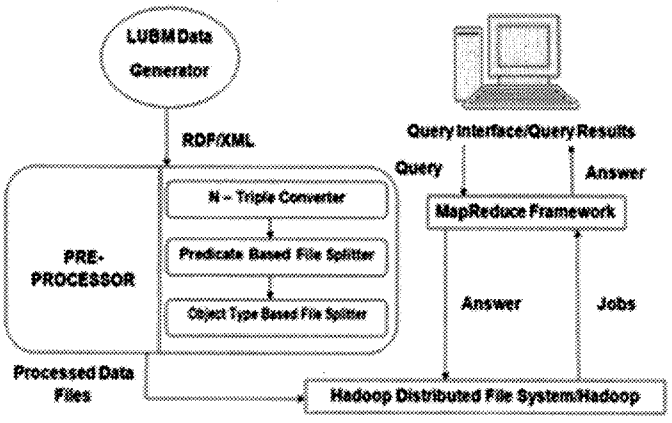


Figure 5. System Architecture for SPARQL Query Optimization

Our approach consists of processing SPARQL queryies securely over Hadoop. SPARQL is a query language used to query RDF data [SP]. The software part we will develop is a framework to query RDF data distributed over Hadoop [NHL08, MM07]. There are a number of steps to preprocess and query RDF data (see Figure 5). With this proposed part, researchers can obtain results to optimize query processing of massive amounts of data [HDK09]. Below we discuss the steps involved in the development of this part.

Pre-processing: Generally, RDF data is in XML format (see LUBM RDF data). In order to execute a SPARQL query, we propose some data pre-processing steps and store the pre-processed data into HDFS. We have an N-triple Convertor module which converts RDF/XML format of data into N-triple format as this format is more understandable. We will use Jena framework as stated earlier, for this conversion purpose. In Predicate Based File Splitter module, we split all N-triple format files based on the predicates. Therefore, the total number of files for a dataset is equal to the number of predicates in the ontology/taxonomy. In the last module of the pre-processing step, we further divide predicate files on the basis of the type of object it contains. So, now each predicate file has specific types of objects in it. This is done with the help of the Pellet library. This pre-processed data is stored into Hadoop.

Query Execution and Optimization: We are developing a SPARQL query execution and optimization module for Hadoop. As our storage strategy is based on predicate splits, first, we will look at the

predicates present in the query. Second, rather than looking at all of the input files, we will look at a subset of the input files that are matched with predicates. Third, SPARQL queries generally have many joins in them and all of these joins may not be possible to perform in a single Hadoop job. Therefore, we will devise an algorithm that decides the number of jobs required for each kind of query. As part of optimization, we will apply a greedy strategy and cost-based optimization to reduce query processing time. An example of greedy strategy is to cover the maximum number of possible joins in a single job [HDK09]. For the cost model, the join to be performed first is based on summary statistics (e.g., selectivity factor of a bounded variable, join triple selectivity factor for three triple patterns. For example, consider a query for LUBM dataset: "List all persons who are alumni of a particular university." In SPARQL:

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
SELECT ?X WHERE {
?X rdf:type ub:Person .
<http://www.University0.edu> ub:hasAlumnus ?X }

The query optimizer will take this query input and decide a subset of input files to look at based on predicates that appear in the query. Ontology and pellet reasoner will identify three input files (underGraduateDegreeFrom, masterDegreeFrom and DoctoraldegreeFrom) related to predicate, "hasAlumns". Next, from type file we filter all the records whose objects are a subclass of Person using the pellet library. From these three input files (underGraduateDegreeFrom, masterDegreeFrom and DoctoraldegreeFrom) the optimizer filters out triples on the basis of <http://www.University0.edu> as required in the query. Finally, the optimizer determines the requirement for a single job for this type of query and then the join is carried out on the variable X in that job.

With respect to secure query processing, we are investigating two approaches. One is rewriting the query in such a way that the policies are enforced in an appropriate manner. The second is query modification where the policies are used in the "where" clause to modify the query.

5.3 Integrate SUN XACML Implementation into HDFS

Current Hadoop implementations enforce a very coarse-grained access control policy that permits or denies a principal access to essentially all system resources as a group without distinguishing amongst resources. For example, users who are granted access to the Namenode (see Figure 4) may execute any program on any client machine, and all client machines have read and write access to all files stored on all clients. Such coarse-grained security is clearly unacceptable when data, queries, and the system resources that implement them are security-relevant, and when not all users and processes are fully trusted. Current work [Zh09] addresses this by implementing standard access control lists for Hadoop to constrain access to certain system resources, such as files; however, this approach has the limitation that the enforced security policy is baked into the operating system and therefore cannot be easily changed without modifying the operating system. We are enforcing more flexible and fine-grained access control policies on Hadoop by designing an In-lined Reference Monitor implementation of Sun XACML. XACML [MO05] is an OASIS standard for expressing a rich language of access control policies in XML. Subjects, objects, relations, and contexts are all generic and extensible in XACML, making it well-suited for a distributed environment where many different sub-policies may interact to form larger, composite, system-level policies. An abstract XACML enforcement mechanism is depicted in Figure 6. Untrusted processes in the framework access security-relevant resources by submitting a request to the resource's Policy Enforcement Point (PEP). The PEP reformulates the request as a policy query and submits it to a Policy Decision Point (PDP). The PDP consults any policies related to the request to answer the query. The PEP either grants or denies the resource request based on the answer it receives. While the PEP and PDP components of the enforcement mechanism are traditionally implemented at the level of the operating system or as trusted system libraries, we propose to achieve greater flexibility by implementing

them in our system as In-lined Reference Monitors (IRM's). IRM's implement runtime security checks by in-lining those checks directly into the binary code of untrusted processes. This has the advantage that the policy can be enforced without modifying the operating system or system libraries. IRM policies can additionally constrain program operations that might be difficult or impossible to intercept at the operating system level. For example, memory allocations in Java are implemented as Java bytecode instructions that do not call any external program or library. Enforcing a fine-grained memory-bound policy as a traditional reference monitor in Java therefore requires modifying the Java virtual machine or JIT-compiler. In contrast, an IRM can identify these security-relevant instructions and inject appropriate guards directly into the untrusted code to enforce the policy.

Finally, IRM's can efficiently enforce history-based security policies, rather than merely policies that constrain individual security-relevant events. For example, our past work [JH09] has used IRMs to enforce fairness policies that require untrusted applications to share as much data as they request. This prevents processes from effecting denial of service attacks based on freeloading behavior. The code injected into the untrusted binary by the IRM constrains each program operation based on the past history of program operations rather than in isolation. This involves injecting security state variables and counters into the untrusted code, which is difficult to accomplish efficiently at the operating system level.

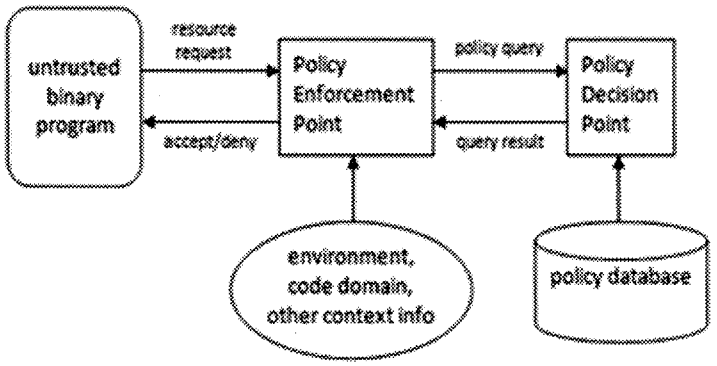


Figure 6. XACML Enforcement Architecture

The core of an IRM framework consists of a *binary-rewriter*, which statically modifies the binary code of each untrusted process before it is executed to insert security guards around potentially dangerous operations. Our proposed binary-rewriter implementation will be based on SPoX (Security Policy XML) [HJ08], which we developed to enforce declarative, XML-based, IRM policies for Java bytecode programs. In order to provide strong security guarantees for our system, we will apply automated software verification technologies, including type and model-checking, which we have previously used to certify the output of binary-rewriters [HMS06, DGH09]. Such certification allows a small, trusted verifier to independently prove that rewritten binary code satisfies the original security policy, thereby shifting the comparatively larger binary-rewriter out of the trusted computing base of the system.

5.4 Strong Authentication:

Currently, Hadoop does not authenticate users. This makes it hard to enforce access control for security sensitive applications and makes it easier for malicious users to circumvent file permission checking done by HDFS. To address these issues, the open source community is actively working to integrate Kerberos protocols with Hadoop [Zh09]. On top of the proposed Kerboros protocol, for some assured information applications, there may be a need for adding simple authentication protocols to authenticate with secure coprocessors. For this reason, we can add a simple public key infrastructure to our system so that users can independently authenticate with secure co-processors to retrieve secret keys used for encrypting sensitive data. We can use open source public key infrastructure such as the OpenCA PKI implementation for our system [OCA].

6. SUMMARY AND CONCLUSION

In this paper we first discussed security issues for cloud. These issues include storage security, middleware security, data security, network security and application security. The main goal is to securely store and manage data that is not controlled by the owner of the data. Then we focused on specific aspects of cloud computing. In particular, we are taking a bottom up approach to security where we are working on small problems in the cloud that we hope will solve the larger problem of cloud security. First we discussed how we may secure documents that may be published in a third party environment. Next we discussed how secure co-processors may be used to enhance security. Finally, we discussed how XACML may be implemented in the Hadoop environment as well as in secure federated query processing with SPARQL using MapReduce and Hadoop.

There are several other security challenges including security aspects of virtualization. We believe that due to the complexity of the cloud, it will be difficult to achieve end-to-end security. However the challenge we have is to ensure more secure operations even if some parts of the cloud fail. For many applications, we not only need information assurance but also mission assurance. Therefore, even if an adversary has entered the system, the objective is to thwart the adversary so that the enterprise has time to carry out the mission. As such, building trust applications from untrusted components will be a major aspect with respect to cloud security.

REFERENCES

[BERT02] Bertino, E., et al, Access Control for XML Documents, Data and Knowledge Engineering, Volume 43, #3, 2002.

[BERT04] Bertino, E. et al, Secure Third Party Publication of XML Documents, To appear in IEEE Transactions on Knowledge and Data Engineering, 2004.

[CKL07] C.-T. Chu and S. K. Kim and Y.-A. Lin and Y. Yu and G. Bradski and A. Y. Ng and K. Olukotun: Map-reduce for machine learning on multicore NIPS 2007

[DG04] Dean, Jeffrey and Ghemawat, Sanjay: MapReduce: simplified data processing on large clusters OSDI'04: Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation 10-10 2004

[DGH09] B. W. DeVries, G. Gupta, K. W. Hamlen, S. Moore, and M. Sridhar. ActionScript Bytecode Verification with Co-Logic Programming. In *Proc. of the ACM SIGPLAN Workshop on Programming Languages and Analysis for Security (PLAS)*. June 2009.

[GNC09] F. Gates, O. Natkovich, S. Chopra, P. Kamath, S. M. Narayanamurthy, C. Olston, B. Reed, S. Srinivasan and U. Srivastava. Building a High-Level Dataflow System on top of Map-Reduce: The Pig Experience. *To appear in* Thirty-Fifth International Conference on Very Large Data Bases (VLDB) (Industrial, Applications and Experience Track), Lyon, France, August 2009.

[HA] HADOOP: http://hadoop.apache.org; http://hadoop.apache.org/core/docs/r0.18.3/hdfs design.html

[HM] HAMA: http://cwiki.apache.org/labs/cloudsglossary.html

[HJ08] K. W. Hamlen and M. Jones. Aspect-Oriented In-lined Reference Monitors. In *Proc. of the ACM SIGPLAN Workshop on Programming Languages and Analysis for Security (PLAS)*. June 2008.

[HMS06] K. W. Hamlen, G. Morrisett, and F. B. Schneider. Certified In-lined Reference Monitoring on .NET. In *Proc. of the ACM SIGPLAN Workshop on Programming Languages and Analysis for Security (PLAS)*. June 2006.

[HPW06] Hurtado, C.A., Poulovassilis, A., and Wood, P.T. A Relaxed Approach to RDF Querying. *In Proceedings of International Semantic Web Conference.* 2006, 314-328.

[IBM04] IBM. IBM PCI cryptographic coprocessor, 2004. http://www.ibm.com/security/cryptocards.

[JE] Jena: http://jena.sourceforge.net

[JH09] M. Jones and K. W. Hamlen. Enforcing IRM Security Policies: Two Case Studies. In *Proc. of the IEEE Intelligence and Security Informatics Conference (ISI)*. June 2009.

[LUBM] Lehigh University Benchmark (LUBM) .http://swat.cse.lehigh.edu/projects/lubm.

[MA] Mahout: http://lucene.apache.org/mahout/

[MK09] James Mcglothlin, L. Khan, "RDFKB: Efficient Support For RDF Inference Queries and Knowledge Management," to appear in *International Database Engineering & Applications Symposium* (IDEAS) Cetraro (Calabria), Italy, 16-18 September 2009.

[MGK08] M. Masud, J. Gao, L. Khan, J. Han, and B. M. Thuraisingham, "A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data," In Proc. of *2008 IEEE International Conference on Data Mining (ICDM 2008),* Pisa, Italy, Page 929-934, December, 2008.

[MGK09] M. Masud, J. Gao, L. Khan, J. Han, and B. M. Thuraisingham, "Integrating Novel Class Detection with Classification for Concept-Drifting Data Streams," to appear in Proc. of *the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD),* Bled, Slovenia, September 7-11, 2009.

[MM07] Mcnabb, Andrew W. and Monson, Christopher K. and Seppi, Kevin D.: MRPSO: MapReduce particle swarm optimization GECCO '07: *Proceedings of the 9th annual conference on Genetic and evolutionary computation* 177 ACM Press 2007

[MO05] T. Moses, ed. eXtensible Access Control Markup Language (XACML) Version 2.0, OASIS Standard, February 2005. http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

[MS08] Christopher Moretti and Karsten Steinhaeuser and Douglas Thain and Nitesh V. Chawla: Scaling Up Classifiers to Cloud Computers, IEEE ICDM 2008
[MSB08] Stocker, Markus and Seaborne, Andy and Bernstein, Abraham and Kiefer, Christoph and Reynolds, Dave: SPARQL basic graph pattern optimization using selectivity estimation WWW '08: Proceeding of the 17th international conference on World Wide Web 595-604
[MU06] A.Muys, "Building an Enterprise-Scale Database for RDF Data", 2006,
[NHL08] Newman, A. and Hunter, J. and Li, Y. F. and Bouton, C. and Davis, M.: A Scale-Out RDF Molecule Store for Distributed Processing of Biomedical Data Semantic Web for Health Care and Life Sciences Workshop WWW 2008
[OCA] OpenCA Implementation, (http://www.openca.org/projects/openca/)
[PAK08] J. Partyka, N. Alipanah, L. Khan, B. M. Thuraisingham, and S. Shekhar, "Content-based ontology matching for GIS datasets," *ACM International Symposium on Geographic Information Systems, ACM-GIS 2007,* November 7-9, 2008, Seattle, Washington, USA, Page: 407-410.
[PE] Pellet: http://clarkparsia.com/pellet
[PKT09] J. Partyka, L. Khan, B. Thuraisingham, "Semantic Schema Matching Without Shared Instances," to appear in *'Third IEEE International Conference on Semantic Computing*, Berkeley, CA, USA - September 14-16, 2009.
[RDF] W3C Recommendation, "RDF Primer", Feb, 2004. http://www.w3.org/TR/rdf-primer/.
[RGK09a] S. Ramanujam, A. Gupta, L. Khan, S. Seida, B. Thuraisingham, "R2D: A Bridge between the Semantic Web and Relational Visualization Tools," to appear in *'Third IEEE International Conference on Semantic Computing,* Berkeley, CA, USA - September 14-16, 2009.
[RGK09b] Sunitha Ramanujam, Anubha Gupta, Latifur Khan, Steven Seida, Bhavani M. Thuraisingham: Relationalizing RDF stores for tools reusability. *18th International Conference on World Wide Web, WWW 2009,* Madrid, Spain, April 20-24, 2009, Page: 1059-1060.
[RGK09c] S.Ramanujam, A.Gupta, L.Khan, and S.Seida, "R2D: Extracting relational structure from RDF stores", *ACM/IEEE International Conference on Web Intelligence,* September, 2009, Milan, Italy.
[RKE07] Rohloff, Kurt and Dean, Mike and Emmons, Ian and Ryder, Dorene and Sumner, John: An Evaluation of Triple-Store Technologies for Large Data Stores On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops 1105-1114 2007
[SAK07] G. Subbiah, A. Alam, L. Khan, B. M. Thuraisingham, "Geospatial data qualities as web services performance metrics," in *Proc. of ACM International Symposium on Geographic Information Systems, ACM-GIS 2007,* November 7-9, 2007, Seattle, Washington, USA
[SW] Semantic Web: http://challenge.semanticweb.org
[SW99] S.W. Smith and S.H. Weingart. Building a high-performance, programmable secure coprocessor. Computer Networks (Special Issue on Computer Network Security), (31):831–860, 1999
[TC07] W.Teswanich, S,Chittayasothorn, "A Transformation of RDF Documents and Schemas to Relational Databases", IEEE Pacific Rim Conferences on Communications, Computers, and Signal Processing, 2007, pp. 38-41.
[W3a] RDF: http://www.w3.org/RDF
[W3b] SPARQL: http://www.w3.org/TR/rdf-sparql-query
[WKB08] Weiss, C, Karras, P, and Bernstein, A. Hexastore: Sextuple Indexing for Semantic Web Data Management. In Proceeding of VLDB. 2008.
[Zh09] K. Zhang, Adding user and service-to-service authentication to Hadoop, https://issues.apache.org/jira/browse/HADOOP-4343,
http://www.netymon.com/papers/muysa06buildforrdf.pdf