ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)

Enhanced Tagging and Content Filtering Model for Web Portal

[I]Dr.P.Sumitra, [II]M.Kavinnela
[I]Assistant Professor, [II]M.Phil Research Scholar
[I,II]Dept. of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women (Autonomous), Elayampalayam, Tiruchengode, Tamil Nadu, India.

Abstract
Cloud computing offers dynamically scalable resources provisioned as a service over the Internet. The third-party, on-demand, self-*service, pay-per-use, and seamlessly scalable computing resources and services offered by the cloud paradigm promise to reduce capital as well as operational expenditures for hardware and software. Collaborative tagging is one of the most diffused and popular services available online. The important purpose of collaborative tagging is to loosely classify resources based on end-user's feedback. But the undefined semantics of tags, which are perse ambiguous and expressed in multiple languages, makes it difficult to enforce semantic interoperability and to grant a reasonable level of accuracy when determining the "meaning" of a tag. Tagging allows end user to loosely classify either offline/online resources based on their feedback, expressed in the form of tags. Although Tags may not be per se sensitive information, the wide use of collaborative tagging services increases the risk of cross referencing, thereby seriously compromising user privacy. In this thesis, a first contribution is made toward the development of a privacy-preserving collaborative tagging service, by showing how a specific privacy-enhancing technology, namely tag suppression, can be used to protect end-user privacy. In addition, it analyzes how the approach can affect the effectiveness of a policy-based collaborative tagging system that supports enhanced web access functionalities, like content filtering and discovery, based on preferences specified by end users. However, to achieve the enhanced use, the current architecture of collaborative tagging services must be extended by including a policy layer. The aim of this layer will be to enforce user preferences, intensionally denoting resources on the basis of the set of tags associated with them, and, possibly, other parameters concerning their trustworthiness (the percentage of users who have added a given tag, the social relationships and characteristics of those users, etc.). This is project make a first contribution in this direction by showing how a specific privacy-enhancing technology (PET), namely tag suppression, can be used to protect end-user privacy; and second, it analyzes how the approach can affect the effectiveness of policy-based collaborative tagging systems.*

Keywords
Cloud Computing, Collaborative Tagging, Privacy Tagging, Content Filtering, Security.

I. Objectives

I. To introduce a user-assisted friend grouping mechanism that enhances traditional group-based policy management approaches. Assisted Friend Grouping leverages proven clustering techniques to aid users in grouping their friends more effectively and efficiently.
II. To found measurable agreement between clusters and user-defined relationship groups. In addition, user perceptions of the improvements should be encouraging.
III. To introduce a new privacy management model that is an improvement over traditional group-based policy management approaches.
IV. To leverage a user's memory and opinion of their friends to set policies for other similar friends, which we refer to as Same-As Policy Management.
V. To make users associate the policy with an example friend and in doing so have this friend in the forefront of their mind. This allows users to be more selective and careful in assigning permissions.
VI. To detect user privacy sentiment that can be leveraged to further enhance privacy management models. For example, Unconcerned Users who author more open policies may leverage a less flexible coarse-grained privacy management approach.

II. Problem Definition

The main purpose of collaborative tagging is to loosely classify resources based on end-user's feedback, expressed in the form of free-text labels (i.e., tags). The novelty of such an approach to content/resource categorization has been seen, in recent years, as a challenging research topic.

In fact, collaborative tagging may be the basis for a semantic network connecting online resources based on their characteristics, and not only their URIs. At the same time, the undefined semantics of tags, which are per se ambiguous and expressed in multiple languages, makes it difficult to enforce semantic interoperability and to grant a reasonable level of accuracy when determining the "meaning" of a tag.

However, besides the support to policy enforcement, enhanced collaborative tagging requires another layer which addresses an issue so far not deeply investigated, i.e., privacy protection.

Collaborative tagging requires the enforcement of mechanisms that enable users to protect their privacy by allowing them to hide certain user-generated contents (unless they desire otherwise), without making them useless for the purposes they have been provided in a given online service. This means that privacy-preserving mechanisms must not negatively affect the service accuracy and effectiveness (e.g., tag-based browsing, filtering, or personalization).

In this project a first contribution in this direction by showing how a specific privacy-enhancing technology (PET), namely tag suppression, can be used to protect end-user privacy; and second, we analyze how our approach can affect the effectiveness of policy-based collaborative tagging systems. Tag suppression is a technique that has the purpose of preventing privacy attackers from profiling users' interests on the basis of the tags they specify.

The data-preservative technology considered in this work is tag suppression, a technique that allows a user to refrain from tagging certain resources in such a manner that the profile resulting from this perturbation does not capture their interests so precisely. Our conceptually simple technique protects user privacy to a certain degree, but at the cost of the semantic loss incurred by suppressing

tags. Other approaches based on data perturbation include the submission of false tags.

III. Review of Literature

A. Related Work

1. Exploring Information Leakage in Third-Party Compute Clouds

In this paper [1] the authors THOMAS RISTENPART, ERAN TROMER, HOVAV SHACHAM and STEFAN SAVAGE stated that third-party cloud computing represents the promise of outsourcing as applied to computation. Services, such as Microsoft's Azure and Amazon's EC2, allow users to instantiate virtual machines (VMs) on demand and thus purchase precisely the capacity they require when they require it. In turn, the use of virtualization allows third-party cloud providers to maximize the utilization of their sunk capital costs by multiplexing many customer VMs across a shared physical infrastructure. However, in this paper, the authors showed that this approach can also introduce new vulnerabilities.

Using the Amazon EC2 service as a case study, they showed that it is possible to map the internal cloud infrastructure, identify where a particular target VM is likely to reside, and then instantiate new VMs until one is placed co-resident with the target. They explored how such placement can then be used to mount cross-VM side-channel attacks to extract information from a target VM on the same machine.

It has become increasingly popular to talk of "cloud computing" as the next infrastructure for hosting data and deploying software and services. In addition to the plethora of technical approaches associated with the term, cloud computing is also used to refer to a new business model in which core computing and software capabilities are outsourced on demand to shared third-party infrastructure.

While this model, exemplified by Amazon's Elastic Compute Cloud (EC2) [9], Microsoft's Azure Service Platform [10], and Rackspace's Mosso [11] provides a number of advantages—including economies of scale, dynamic provisioning, and low capital expenditures—it also introduces a range of new risks. Some of these risks are self-evident and relate to the new trust relationship between customer and cloud provider.

In particular, to maximize efficiency multiple VMs may be simultaneously assigned to execute on the same physical server. Moreover, many cloud providers allow "multi-tenancy" — multiplexing the virtual machines of disjoint customers upon the same physical hardware. Thus it is conceivable that a customer's VM could be assigned to the same physical server as their adversary. This in turn, engenders a new threat — that the adversary might penetrate the isolation between VMs (e.g., via a vulnerability that allows an "escape" to the hypervisor or via side-channels between VMs) and violate customer confidentiality.

This paper explores the practicality of mounting such cross-VM attacks in existing third-party compute clouds. The attacks they considered require two main steps: placement and extraction. Placement refers to the adversary arranging to place their malicious VM on the same physical machine as that of a target customer.

Using Amazon's EC2 as a case study, they demonstrated that careful empirical "mapping" can reveal how to launch VMs in a way that maximizes the likelihood of an advantageous placement. They found that in some natural attack scenarios, just a few dollars invested in launching VMs can produce a 40% chance of placing a malicious VM on the same physical server as a target customer. They showed preliminary results on cross-VM side channel attacks, including a range of building blocks (e.g., cache load measurements in EC2) and coarse-grained attacks such as measuring activity burst timing (e.g., for cross-VM keystroke monitoring). This point to the practicality of side-channel attacks in cloud-computing environments. Overall, their results indicated that there exist tangible dangers when deploying sensitive tasks to third-party compute clouds.

2. Cross-VM Side Channels and Their Use to Extract Private Keys

In this paper [2] the authors YINQIAN ZHANG and ARI JUELS details the construction of an access-driven side-channel attack by which a malicious virtual machine (VM) extracts fine-grained information from a victim VM running on the same physical computer. This attack is the first such attack demonstrated on a symmetric multiprocessing system virtualized using a modern VMM (Xen).

Modern virtualization technologies such as Xen, HyperV, and VMware are rapidly becoming the cornerstone for the security of critical computing systems. This reliance stems from their seemingly strong isolation guarantees, meaning their ability to prevent guest virtual machines (VMs) running on the same system from interfering with each other's execution or, worse, exfiltrating confidential data across VM boundaries.

The assumption of strong isolation underlies the security of public cloud computing systems [12] such as Amazon EC2, Microsoft Windows Azure, and Rackspace; military multi-level security environments [13]; home user and enterprise desktop security in the face of compromise [14]; and software-based trusted computing [15].

VM managers (VMMs) for modern virtualization systems attempt to realize this assumption by enforcing logical isolation between VMs using traditional access-control mechanisms. But such logical isolation may not be sufficient if attackers can circumvent them via side-channel attacks. In particular, they provided an account of how to overcome three classes of significant challenges in this environment:

I. Inducing regular and frequent attacker VM execution despite the coarse scheduling quanta used by VMM schedulers;
II. Overcoming sources of noise in the information available via the cache timing channel, both due to hardware features (e.g., CPU power saving) and due to software ones (e.g., VMM execution); and
III. Dealing with core migrations, which give rise to cache "readings" with no information of interest to the attacker (i.e., the victim was migrated to a core not shared by the attacker).

Specifically, they showed that the attacker VM's monitoring of a victim's repeated exponentiations over the course of a few hours provides it enough information to reconstruct the victim's 457-bit private exponent accompanying a 4096-bit modulus with very high accuracy—so high that the attacker was then left to search fewer than 10, 000 possible exponents to find the right one.

3. All Your Clouds are Belong To Us – Security Analysis Of Cloud Management Interfaces

In the paper [3], the authors JURAJ SOMOROVSKY, MARIO HEIDERICH, NILS GRUSCHKA and LUIGI LO IACONO

stated that cloud computing resources are handled through control interfaces. It is through these interfaces that the new machine images can be added, existing ones can be modified, and instances can be started or ceased.

In this paper, the authors provided a security analysis pertaining to the control interfaces of a large Public Cloud (Amazon) and widely used Private Cloud software (Eucalyptus).

Their research results are alarming: in regards to the Amazon EC2 and S3 services, the control interfaces could be compromised via the novel signature wrapping and advanced XSS techniques. Similarly, the Eucalyptus control interfaces were vulnerable to classical signature wrapping attacks, and had nearly no protection against XSS. As a follow up to those discoveries, they additionally describe the countermeasures against these attacks, as well as introduce a novel "black box" analysis methodology for public Cloud interfaces.

Cloud security discussions to date mostly focus on the fact that customers must completely trust their cloud providers with respect to the confidentiality and integrity of their data, as well as computation faultlessness. However, another important area is often overlooked: if the Cloud control interface is compromised, the attacker gains immense potency over the customer's data. This attack vector is a novelty as the result of the control interface (alongside with virtualization techniques) being a new feature of the Cloud Computing paradigm, as NIST lists On-demand self-service and Broad network access as essential characteristics of Cloud Computing systems.

In this paper, the authors refer to two distinct classes of attacks on the two main authentication mechanisms used in Amazon EC2 and Eucalyptus cloud control interfaces. The first class of attacks complies of the XML Signature Wrapping attacks (or in short {signature wrapping attacks) [10] on the public SOAP interface of the Cloud.

They demonstrated that these control interfaces are highly vulnerable to several new and classical variants of signature wrapping. For these attacks, knowledge of a single signed SOAP message is sufficient to attain a complete compromization of the security within the customer's account.

Those included actions such as starting or stopping virtual machines, downloading or uploading virtual machine image files, resetting the administrator's password for cloud instances, and so on. The second class are advanced XSS attacks on browser based Web front-ends.

They found a persistent Cross Site Scripting (XSS) vulnerability that allowed an adversary to perform an automated attack targeted at stealing username/password data from EC2/S3 customers. This attack was made possible by the simple fact the Amazon shop and the Amazon cloud control interfaces share the same log-in credentials, thus any XSS attack on the (necessarily complex) shop interface can be turned into an XSS attack on the cloud control interface. The Eucalyptus Web front-end was equally prone to these kinds of attacks. Their analysis has shown that in order to compromise this system, the attacker could easily use a simple HTML injection.

From a conceptual standpoint, cloud services need some form of cloud control which enables users to manage and configure the service, whilst also preserving access to the stored data. In IaaS-based clouds the control interface allows to, for example, instantiate machines, as well as to start, pause and stop them. Machine images can be created or modified, and the links to persistent storage devices must be configured. It is therefore quite undebatable that the security of a cloud service highly depends on robust and effective security mechanisms for the cloud control interfaces.

Technically, the cloud control interface can be realized either as a SOAP-based Web Service, or as a Web application. If the control interface is SOAP-based, then WS-Security [12] can be applied to provide security services. For the authentication purposes, security tokens (mainly X.509 certificates) and XML Signature can be employed. A problem that generally arises is that the WS-Security standard is vulnerable to signature wrapping attacks [13], which consequently may invalidate this authentication mechanism.

4. Amazonia: When Elasticity Snaps Back

In this paper [4], the authors SVEN BUGIEL, STEFAN NÜRNBERGER, THOMAS PÖPPELMANNY, AHMAD-REZA SADEGHI and THOMAS SCHNEIDER stated that cloud Computing is an emerging technology promising new business opportunities and easy deployment of web services. Much has been written about the risks and benefits of cloud computing in the last years.

The literature on clouds often points out security and privacy challenges as the main obstacles, and proposes solutions and guidelines to avoid them. However, most of these works deal with either malicious cloud providers or customers, but ignore the severe threats caused by unaware users.

In this paper they considered security and privacy aspects of real-life cloud deployments, independently from malicious cloud providers or customers. They focused on the popular Amazon Elastic Compute Cloud (EC2) and give a detailed and systematic analysis of various crucial vulnerabilities in publicly available and widely used Amazon Machine Images (AMIs) and show how to eliminate them.

Their Amazon Image Attacks (AmazonIA) deploy an auto- mated tool that uses only publicly available interfaces and makes no assumptions on the underlying cloud infrastructure. They were able to extract highly sensitive information (including passwords, keys, and credentials) from a variety of publicly available AMIs. The extracted information allows to

(i) Start (botnet) instances worth thousands of dollars per day,
(ii) Provide backdoors into the running machines,
(iii) Launch impersonation attacks,
(iv) Access the source code of the entire web service.

Their attacks can be used to completely compromise several real web services offered by companies (including IT-security companies), e.g., for website statistics/user tracking, two-factor authentication, or price comparison. Further, they showed mechanisms to identify the AMI of certain running instances.

Following the maxim "security and privacy by design" they showed how their automated tools together with changes to the user interface can be used to mitigate their attacks.

The high usability of today's cloud computing platforms makes this rapidly emerging paradigm very attractive for customers who want to instantly and easily provide web-services that are highly available and scalable to the current demands [14].

Albeit the various advantages of cloud computing, serious concerns about security and privacy hinder many users from "going into the cloud". Most solutions to preserve security and privacy in the cloud proposed so far consider potentially faulty/malicious cloud providers or technical savvy/rogue customers. However, the much more serious and ubiquitous threat of unaware users who unintentionally harm their own or others' security or privacy is

often overseen.

The main goal of this paper is the investigation and evaluation of security and privacy threats caused by the unawareness of users in the cloud. Although the methods and techniques described in this paper are applicable to arbitrary IaaS providers, they focused on one of the major cloud providers, Amazon's Elastic Compute Cloud (EC2) [15] and adapt their terminology accordingly. In the following, they described the players involved in the (Amazon) Cloud App Store and the resulting security challenges.

IV. Conclusion and Future Works

Collaborative tagging is currently an extremely popular online service. Although nowadays it is basically used to support resource search and browsing, its potential is still to be exploited. One of these potential applications is the provision of web access functionalities such as content filtering and discovery. For this to become a reality, however, it would be necessary to extend the architecture of current collaborative tagging services so as to include a policy layer that supports the enforcement of user preferences. Collaborative tagging has been gaining popularity, it have been become more evident the need for privacy protection; not only because tags are sensitive information but also because of the risk of cross referencing. In addition to the existing system approaches, the proposed system takes care of multi language tagging.

A privacy preserving collaborative tagging if applied to content with multiple languages, and then it becomes more effective to fruitful to end users. Future work includes the development of a full prototype for the experimented system and it's testing and use in further scenarios.

References

[1] *Amazon Elastic Compute Cloud (EC2).http://aws.amazon.com/ec2/*

[2] *Amazon Web Services. Auto-scaling Amazon EC2 with Amazon SQS.http://developer.amazonwebservices.com/connect/entry.jspa?externalID=1464*

[3] *Amazon Web Services. Creating HIPAA-Compliant Medical Data Applications with Amazon Web Services. White paper, http://awsmedia.s3.amazonaws.com/AWS_ HIPAA_ Whitepaper_Final.pdf, April 2009.*

[4] *O. Acii, cmez, W. Schindler, and C. K. KO ,c. Cache based remote timing attack on the AES. In Topics in Cryptology – CT-RSA 2007, The Cryptographers' Track at the RSA Conference 2007, pages 271–286, February 2007.*

[5] *O. Acii ,cmez and J.-P. Seifert. Cheap hardware parallelism implies cheap security. In Workshop on Fault Diagnosis and Tolerance in Cryptography, pages80–91, September 2007.*

[6] *M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson,=A. Rabkin, I. Stoica, and M. Zaharia. A view of cloud computing. Commun. ACM, 53(4):50–58, 2010.*

[7] *Celera Assembler. http://wgs-assembler.sourceforge.net/*

[8] *E. Bangerter, D. Gullasch, and S. Krenn. Cache games–bringing access-based cache attacks on AES to practice. In 32nd IEEE Symposium on Security and Privacy, 2011.*

[9] *Eucalyptus. http://open.eucalyptus.com/. Akhawe, D., Barth, A., Lam, P. E., Mitchell,J. C., and Song, D.Towards a formal foundation of web security. In CSF (2010), pp. 290{304}.*

[10] *Balduzzi, M. New Insights Into Click jacking. In OWASP AppSec Research (2010).*

[11] *All Together Now: Amazon, we need those caps on billing. http://forums.aws.amazon.com/thread. Jspa? ThreadID=50075#jive-message-217130*

[12] *Amazon EC2: ec2-bundle-vol. http://docs.amazonwebservices.com/ AmazonEC2/dg/2006-10-01/CLTRG-ami-bundle-vol.html*

[13] *Amazon EC2 Pricing http://aws.amazon.com/ec2/pricing*

[14] *Amazon Elastic Block Store (EBS). http://aws.amazon.com/ebs/*

[15] *Amazon Elastic Compute Cloud (Amazon EC2). http://aws.amazon.com/ec2/.*