Information Sciences 258 (2014) 387–402

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

DocCloud: A document recommender system on cloud computing with plausible deniability

Juan Vera-del-Campo [a,*], Josep Pegueroles [a], Juan Hernández-Serrano [a], Miguel Soriano [a,b]

[a] Universitat Politècnica de Catalunya, Jordi Girona 1-3, C3, Barcelona, Spain
[b] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Av. Carl Friedrich Gauss 7, B4, Castelldefels, Barcelona, Spain

ARTICLE INFO

Article history:
Available online 16 April 2013

Keywords:
Recommender system
Doccloud
Social cloud
Plausible deniability

ABSTRACT

Recommender systems select the most interesting products for costumers based on their interests. The move of a recommender system to a cloud faces many challenges from the perspective of the protection of the participants. Little work has been done regarding secure recommender systems or how to cope with the legal liability of the cloud provider and any virtual machine inside the cloud.

We propose DocCloud, a recommender system that focused on the protection of all participants against legal attacks. We present the architecture of DocCloud and analyze the security mechanisms that the system includes. Specifically, we study the properties of plausible deniability and anonymity of the recommenders and intermediate nodes. This way, nodes can recommend products to the customers while deny any knowledge about the product they are recommending or their participation in the recommendation process.

1. Introduction

Many of the most successful shops on the Internet provide suggestions to their customers. These suggestions are usually based on the previous experience of users with the shop, or the features of the products the customers are currently viewing.

A *recommender system* is an automatic system that, given a set of available products and a model that captures the interests of a user, outputs a list of products that the system estimates will be of interest to the user. We identify five different roles in a recommender system: *merchants* that provide documents and document descriptions, or *profiles*; *customers* that request recommendations according to their *user profiles*; *indexers* that reply to queries; *repositories* that provide access to final documents and *intermediate nodes* that route messages from merchants and customers to indexers and repositories.

We explore the recommendation of documents that can be distributed electronically and may show high peaks of activity followed by long periods of silence, such as video or audio content. The dynamic structure that cloud systems show makes them a convenient approach for providing the service of recommendation and distribution of multimedia content. However, regardless of the cloud paradigm that the recommender system follows, the participants of a cloud may be accused or prosecuted of copyright infraction.

In this paper, we will use the different roles that exist in a cloud system and other additional mechanisms to protect requests and recommendations, and minimize the legal responsibility that a recommender systems faces.

* Corresponding author. Tel.: +34 699692582.
E-mail addresses: juanvi@entel.upc.edu (J. Vera-del-Campo), josep@entel.upc.edu (J. Pegueroles), jserrano@entel.upc.edu (J. Hernández-Serrano), soriano@entel.upc.edu (M. Soriano).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.04.007

1.1. Attacks against a recommender system

Recently, a new kind of attack against intermediate nodes has appeared. A new international treaty on copyright protection known as the Anti-Counterfeiting Trade Agreement (ACTA) is being negotiated at the moment of writing this paper. After some initial secrecy, the consolidated text is now public [38]. Article 2.15 deals with the liability of legal persons, and states that "the provisions of this section shall apply to *inciting, aiding and abetting* the offenses referred in article 2.14"[1]. The penalties proposed by this article "include imprisonment as well as monetary fines". Thus, not only is downloading or providing a protected document punishable under the ACTA, but so too is abetting downloading. ACTA was not ratified by the European Parliament in July, 2012 [24] However, it shows a trend among legislators throughout the world to make people who aid or even incite the downloading of copyrighted documents liable of copyright infringement. This is the case with the Stop Online Piracy Act (SOPA) and the Protect Intellectual Property Act (PIPA) in the United States, and the Sinde-Wert law in Spain. Most of these bills are currently under intense debate, but they have something in common with ACTA: they make companies liable for users' actions if the companies do not respond to a copyright infringement notification.

The reader should take note that the entity that makes documents available, the one that recommends documents, and the one that distributes them may be not the same. This way, a participant recommending or distributing a document may be unaware of its legal status. In this case, we believe that *recommending* is dangerously close to *inciting*, which is punishable according to ACTA. This situation is even worse for the recommender system if it includes mechanisms to upload and/or download the protected document.

Recently, the administrators of the direct download site Megaupload were arrested in New Zealand on behalf of the American FBI [40]. According to the FBI, this action "directly targets the misuse of a public content storage and distribution site to commit *and facilitate* intellectual property crime". In April, 2009, the administrators of The Pirate Bay, a popular document indexer that does not host any document, were found guilty of complicity in the provision of unauthorized access to copyrighted content and were sentenced to 1 year in jail and nearly 3 million euros in damages by a Swedish court [36]. The high cost of a lawsuit, the criminal charges against the defendants and the long wait time to get a ruling persuaded other site administrators (Filesonic, Fileserve, Uploaded.to, VideoBB, FileJungle, UploadStation, FilePost, UploadBox, x7.to, 4shared, etc.) either changed their policies or announced a voluntary shutdown [9].

Glorioso et al. [12] analyze some court decisions and conclude that, according to US jurisprudence, intermediate nodes in the communication channel, even at the network level such as ISPs, are threatened with legal attacks. The main argument for prosecutors is that the "making available" of copyrighted documents also includes the *indexers* of the system.

The security service that protects indexers from legal prosecution is plausible deniability. In the common law context, plausible deniability [41] refers to circumstances where a denial of responsibility or knowledge of wrongdoing cannot be proved as true or untrue due to a lack of evidence proving the allegation. Hermoni et al. [13] defined deniability for the field of information security as the property that the user has if they can claim that their actions are legitimate. Ref. [13] understands deniability as an absolute property that actors have by mixing up legitimate and not legitimate documents in the indexers. We introduce *plausible deniability* as the property that an actor achieves if he cannot make a probabilistic distinction between legitimate and not legitimate documents, or the process to make this distinction is too long and complex.

We focus our efforts on the protection of all participants of the system against legal attacks. As such, we are interested in the provision of plausible server deniability and document deniability as defined in [13], and we will provide plausible deniability to intermediate nodes and the cloud provider.

1.2. Our contribution

In this paper, we introduce DocCloud, a document recommender system offered by a service provider using the SaaS paradigm of cloud computing. This service involves running small virtual machines (PaaS/IaaS) that are controlled by individual users to actually provide recommendations. We offer protection in two layers: from the SaaS point of view, it will not be possible for the system provider to be accused of complicity since said party will not be aware of the documents that individual virtual machines are recommending; from the PaaS/IaaS point of view, we will protect individual virtual machines and their owners by means of hiding the identity of the node that outputs a recommendation, and providing mechanisms to calculate affinities without leaking all the information in the user profile.

In our security model, we will consider that any user of the recommender system or the virtual machines comprising the cloud may be attackers of the system. We are trying to protect intermediate nodes, indexers and normal users. To so, we will consider them as *attackers* of the system and check how much information they can learn. As such, we limit their power to the power needed for a fair use of the system. We call this power "reasonable efforts" and "fair use" of the system.

If an attacker using only reasonable efforts is able to learn something about the profiles of the documents or users, or the identity of customers or indexers, we consider that the legal system may assume that these attacks are possible and intermediate nodes and indexers can be accused of not performing them. At the same time, showing that under a fair use of the system it is not possible to attack the system, then nodes cannot be prosecuted for not trying to obtain information about

[1] In this citation and others that follow, emphasis is added.