ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

Illegal Multimedia Content Redistribution Protection using Enhanced Traitor Tracing Protocol

D.Rajkumar[1], **A.Rajiv Kannan**[2]
[1]KSR College of Engineering, Tiruchengode
[2]KSR College of Engineering, Tiruchengode

ABSTRACT

Broadband home Internet access supports direct downloads of multimedia contents. Finger printing technique is used to avoid illegal content re-distribution. Fingerprinting consists of embedding an imperceptible mark in the distributed content used to identify the content buyer. The embedded mark allows the identification of the re-distributor by means of a traitor tracing system.

Anonymous fingerprint is used for the legal distribution of multimedia contents with copyright protection for privacy of buyers. Buyers identities are only revealed in case of illegal re-distribution. Recombined fingerprints model overcomes delay and scalability issues. Traitor tracing protocol is used to identify the illegal distribution of multimedia contents. The recombined fingerprint approach uses a complex graph search for traitor tracing. Traitor tracing requires the participation of other buyers and honest proxies in the P2P distribution scenario. P2P Distribution Protocol manages the Merchant, Seed Buyers, Proxies and Peer Buyer transactions. Proxies are used to handle anonymous communication between the peer buyers and higher authorities.

The traitor Tracing Protocol is improved with context information based redistribution analysis mechanism. Four party anonymous communication protocol is integrated with the system to verify misbehave proxies. The standard database search for buyer verification is enhanced with security and privacy factors. Mutual anonymity model is integrated with the system to ensure the communication privacy.

1. INTRODUCTION

P2P systems possess inherent challenges in terms of security, privacy and anonymity because of its loose peer management and extremely distributed working principles. Hence, devising security, privacy and anonymity protection mechanisms for P2P systems poses challenges for researchers and software engineers. According to Balfe et al. and Wasef and Shen, the main challenge in creating P2P systems stems from the perceived need of providing anonymity for users of the system and the growing need of offering robust access control, confidentiality and data integrity. Illegitimate attacks in which malevolent parties may assume multiple identities undermine the efficiency of P2P systems and characterize a fundamental security threat. This is because formulating essential security services is challenging in the absence of stable and verifiable identities. The concepts of security, privacy and anonymity used in this paper are defined in the context of providing a legal content distribution in a P2P system and are described as follows:

1. Security: A mechanism aimed to protect an intellectual property and provide trustworthiness.

2. Privacy: The protection of user-related information in such a way that no personal information of an end user is revealed, unless a user is found to be guilty of illegal re-distribution. It is also called a conditional privacy.

3. Anonymity: A method to protect the identity of provider and receiever and also to protect the contents of transferred data between them.

Xiaosong and Kai and Brinkmeier et al. studies highlight that the security state for P2P systems is worse unlike Client- Server system in which illegal or unauthorized data access can be prevented by a central authority which provides level of access to the clients, monitors and maintain logs of all the data requests and transfers made by these authorized users. In contrast to Client-Server system, P2P systems are not considered secure because of the absence of a centralized authority that can vouch for security parameters. According to Fan et al., the diverse nature of the multimedia material presents a severe challenge to the establishment of effective strategies that would foster secure systems.

P2P networking renders multimedia distribution channels vulnerable to various forms of attacks, e.g., potential of being involved with copyright infringement, the possibilities of downloading files infected with malicious codes and susceptibility to attacks. The security requirements of P2P content distribution systems cover mainly four aspects: copyright protection, data confidentiality, integrity and trust. These aspects are described as follows:

1. Data confidentiality: Confidentiality refers to limiting information access and disclosure to authorized users only.

2. Integrity: Data integrity implies that the data transferred between requesting and providing peer is an exact copy of an original version, i.e. the transferred data has not been corrupted during transmission between peers due to accidental or malicious altering.

3. Copyright protection: It guarantees that no additional replication is allowed other than the permitted copies.

4. Trust: Trust in P2P systems is a peer's belief in another peer's identity and reliability based on reputation.

In a P2P context, these characteristics become significantly more challenging than in the case of traditional domains due to lack of centralized control and distributive ownership. Robels et al. studies draw attention to the prosecution of P2P users for sharing pirated software or re-distribution of copyrighted material. For example, the Recording Industry Association of America (RIAA) has filed suits against more than 20,000 individuals in U.S. using P2P content distribution systems.

Apart from being a source for pirated content, P2P content distribution systems share files that pose severe security risks to end users. With millions of connected users and even more available files, there is no way to verify the legitimacy and safety of shared files. The downloadable files could contain malicious codes that can attack users′ computer with worms, malware, viruses and more. For example, a severe virus known as Antinny, affected the Japanese-based P2P content distribution system Winny. This virus led to the disclosure of a large amount of U.S. military base security codes along with private documents of a police investigator.

2. RELATED WORK

The problem of protecting various types of multimedia content has attracted significant attention from academia and industry. One approach to this problem is using watermarking some distinctive information is embedded in the content itself and a method is used to search for this information in order to verify the authenticity of the content. Watermarking requires inserting

watermarks in the multimedia objects before releasing them as well as mechanisms/systems to find objects and verify the existence of correct watermarks in them. Thus, this approach may not be suitable for already-released content without watermarks in them. The watermarking approach is more suitable for the somewhat controlled environments, such as distribution of multimedia content on DVDs or using special sites and custom players. Watermarking may not be effective for the rapidly increasing online videos, especially those uploaded to sites such as YouTube and played back by any video player. Watermarking is not the focus of this paper.

The focus of this paper is on the other approach for protecting multimedia content, which is content-based copy detection (CBCD). In this approach, signatures are extracted from original objects. Signatures are also created from query objects downloaded from online sites. Then, the similarity is computed between original and suspected objects to find potential copies. Many previous works proposed different methods for creating and matching signatures. These methods can be classified into four categories: spatial, temporal, color and transform-domain. Spatial signatures are the most widely used. Their weakness is the lack of resilience against large geometric transformations. Temporal and color signatures are less robust and can be used to enhance spatial signatures. Transform-domain signatures are computationally intensive and not widely used in practice. For more details, see surveys for audio fingerprinting and 2-D video fingerprinting.

Youtube Content ID [9], Vobile VDNA and MarkMonitor [7] are some of the industrial examples which use fingerprinting for media protection, while methods can be referred to as the academic state-of-the-art. Unlike previous works, the contribution of this paper is to design a large-scale system to find copies that can be used for different types of multimedia content and can leverage multi-cloud infrastructures to minimize the cost, expedite deployment and dynamically scale up and down. That is, we design our system such that previous content-based copy detection methods for creating and matching signatures can be implemented within our system.

In addition to our cloud-based system, we propose a new method for 3-D video fingerprinting and a new design for the distributed matching engine. The works related to each of these components are summarized in the following subsections.

2.1. 3-D VIDEO SIGNATURES

Content-based copy detection of 3-D videos is a fairly new problem; we are aware of only two previous works [1]. The work computes SIFT points in each view and uses the number of matching SIFT points to verify matches. Comparing all SIFT points in each frame is not practical for large databases due to the storage overhead and search complexity. The work in [1] assumes that the depth maps are given or estimated. Estimating the depth map from stereoscopic videos is quite expensive. The method is suitable for 3-D videos encoded in the video plus depth format, but not for stereoscopic videos. Our proposed method in this paper captures the depth properties without calculating the depth map itself and it is computationally efficient because it does not compare all features in the frame. Although 3-D copy detection methods are scarce in the literature, there are many methods available for 2-D video copy detection. Hampapur et al. use the temporal features of the video as the signature. Similarly, Tasdemir et al. use motion vectors as the signature for each frame. Some methods use color histograms as signatures. The color histogram signature is prone to global variations in color which are common when

recoding video. Another group of methods use interest points of video frames as signature. For example, Liu et al. [4] use local SIFT features as the frame signature. Using gradient information has also shown to be robust to many 2-D transformations.

All of the above 2-D video fingerprinting methods can be implemented in the proposed system. In addition, while some of these methods can be used for 3-D video copy detection, they are designed for 2-D videos and they ignore the information in different views and the depth of 3-D videos. This information is important especially in the presence of 3-D video transformations such as view synthesis, where views from different viewpoints can be generated using the depth map of the 3-D video. When two new views are synthesized, the positioning of each pixel in the frame is changed and some areas are occluded while other areas become visible. The luminance, gradient, color and even the interest points in each block can change as well when a new view is synthesized. Thus, the extracted signature using any of the 2-D methods will change accordingly. When searching for similar signatures, manipulated versions may not be identified. The importance of using signatures that have some information from the depth signal has been shown in [1]. In addition, our experiments and comparisons in this paper show that the state-of-the-art copy detection system used by YouTube fails to detect many simple transformations made on 3-D videos such as re-encoding, conversion to row or column interleaved formats and creating new virtual views. Based on the available information from the patent describing the Content ID system [9] and our own experiments, we believe that the poor performance of Content ID on 3-D videos is because it does not consider any depth information.

2.2. DISTRIBUTED MATCHING ENGINE

Unlike many of the previous works, e.g., [3] which designed a system for image matching, our proposed matching engine is general and it can support different types of multimedia objects, including images, 2-D videos and 3-D videos. To achieve this generality, we divide the engine into two main stages. The first stage computes nearest neighbors for a given data point and the second stage post-processes the computed neighbors based on the object type. In addition, our design supports high-dimensionality which is needed for multimedia objects that are rich in features.

Computing nearest neighbors is a common problem in many applications. Our focus in this paper is on distributed techniques that can scale to large datasets such as [5], [6], [3], [2]. Liao et al. [5] build a multi-dimensional index using R-tree on top of the Hadoop distributed file system (HDFS). Their index, can only handle low dimensional datasets—they performed their experiments with two dimensional data. They solve the nearest neighbors over large datasets using MapReduce. Lu et al. construct a Voronoi-like diagram using some selected pivot objects. They then group the data points around the closest pivots and assign them to partitions, where searching can be done in parallel. The system in [6] is also designed for low dimensional datasets; it did not consider data with more than 30 dimensions. In contrast, in our experiments we used images and videos with up to 128 dimensions. Aly et al. [3] propose a distributed system for image retrieval. A major drawback of this system is using a single machine that directs all query points, which makes it a single point of failure as well as a bottleneck that could slow down the whole system. Our system does not use a central machine and thus it is more robust and scalable.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 10, March 2016

The closest work to ours is the RankReduce system [2], which implements a distributed LSH (Locality Sensitive Hashing) index on a computing cluster using MapReduce. Rank Reduce maintains multiple hash tables over a distributed cluster, which requires storing multiple replicas of the datasets in hash tables [10]. This incurs significant storage cost and it increases the number of I/O operations. In contrast, our system stores the dataset only once. We compare the proposed matching engine against Rank Reduce and we show that our system returns more accurate neighbors and it is more efficient.

3. PRIVACY-PRESERVING MULTIMEDIA DISTRIBUTION UNDER P2P NETWORKS

Legal distribution of multimedia contents is a recurrent topic of research. Broadband home Internet access has enabled the sustained growth of e-commerce, including direct downloads of multimedia contents. Copyright infringement is one of the most relevant threats to the content industry.

Fingerprinting emerged as a technological solution to avoid illegal content re-distribution. Basically, fingerprinting consists of embedding an imperceptible mark—fingerprint—in the distributed content to identify the content buyer. The embedded mark is different for each buyer, but the content must stay perceptually identical for all buyers. In case of illegal re-distribution, the embedded mark allows the identification of the re-distributor by means of a traitor tracing system, making it possible to take subsequent legal actions. Although fingerprinting techniques have been available for nearly two decades, the first few proposals in this field are far from nowadays' requirements such as scalability for thousands or millions of potential buyers and the preservation of buyers' privacy.

Most fingerprinting systems can be classified in three categories, namely symmetric, asymmetric and anonymous schemes. In symmetric schemes, the merchant is the one who embeds the fingerprint into the content and forwards the result to the buyer; hence, the buyer cannot be formally accused of illegal re-distribution, since the merchant also had access to the fingerprinted content and could be responsible for the re-distribution. In asymmetric fingerprinting, the merchant does not have access to the fingerprinted copy, but he can recover the fingerprint in case of illegal re-distribution and thereby identify the offending buyer. In anonymous fingerprinting, in addition to asymmetry, the buyer preserves her anonymity and hence she cannot be linked to the purchase of a specific content, unless she participates in an illegal re-distribution [10]. Anonymous fingerprinting is the most convenient strategy to protect both the buyers' privacy and the owner's rights, since it guarantees the following properties: 1) only the buyer obtains the fingerprinted copy of the content, making it impossible for the merchant to accuse her of unlawful redistribution and 2) it preserves the anonymity of the buyers' identities with respect to the merchant.

As scalability is concerned, the unicast approach in which the merchant establishes a connection with each single buyer is not a convenient strategy. Broadcast distribution is not suitable for fingerprinting applications since different fingerprints are required for different buyers in order to guarantee traceability. Peer-to-peer (P2P) distribution can be the answer to this difficulty, as proposed in this paper, since this technique blends some of the advantages of the unicast and multicast solutions. In fact, some content distributors are already operating under the P2P paradigm.

Many anonymous fingerprinting schemes exploit the homomorphic property of public-key cryptography. These schemes allow embedding the fingerprint in the encrypted domain in such a way that only the buyer obtains the decrypted fingerprinted content after using her private key. Developing a practical system using this idea appears difficult, because public-key encryption expands data and substantially increases the communication bandwidth required for transfers. Homomorphic encryption constrains the type of mathematical operations which can be performed on the content for embedding, making it difficult to use the more advanced and robust techniques in the data hiding literature. In addition, the application of this idea in a distributed scenario is not simple, since embedding would have to be performed by peer buyers, requiring a complex and supervised protocol.

Other approaches for anonymous fingerprinting do not exploit homomorphic encryption in this way, but either 1) require highly demanding technologies such as public-key encryption of the contents, secure multiparty protocols, commitment protocols or zero-knowledge proofs, among others, incurring prohibitive computational and communicational costs; or 2) are based on theoretical secure embedding algorithms for which no proof of existence is available.

Very few anonymous fingerprinting schemes with P2P distribution have been suggested. Game theory is applied to develop a fingerprinting scheme where embedding occurs between peer buyers requires multi-party secure protocols between buyers which may be difficult to apply in a real scenario. The proposal is more attractive, since embedding occurs only for a few seed buyers and the fingerprint of the other buyers are automatically generated as a recombination of the fingerprints of their "parents" in a graph distribution scenario. The traitor tracing protocol presented in those references requires an expensive graph search and disturbs a few honest buyers who must cooperate with the authority to identify the source of an illegal re-distribution. This is a relevant inconvenient, not only for the associated computational cost and the nuisance caused to honest buyers, but also because it represents a weakness. The participation of other buyers in the tracing protocol can lead to situations in which some illegal re-distributors may be untraceable even if no malicious behavior occurs. In addition, the distribution protocols rely on the honest behavior of proxies.

This paper reviews the main features of the proposal suggested and suggests several significant improvements to achieve a more efficient and practical system, especially as traitor tracing is concerned, since it avoids the situations in which illegal redistributors cannot be traced. Furthermore, better security properties against potentially malicious proxies are obtained.

Although the system proposed in this paper uses publickey encryption in the distribution and traitor tracing protocols, it must be taken into account that this encryption is only applied to short bit strings, such as the binary fingerprints and hashes, not to the content. The fragments of the content are encrypted using symmetric cryptography, which is much more efficient.

4. PROBLEM STATEMENT

Anonymous fingerprint is used for the legal distribution of multimedia contents with copyright protection for privacy of buyers. Buyers identities are only revealed in case of illegal re-distribution. Recombined fingerprints model overcomes delay and scalability issues. Traitor tracing protocol is used to identify the illegal distribution of multimedia contents. The recombined fingerprint approach uses a complex graph search for traitor tracing. Traitor tracing

requires the participation of other buyers and honest proxies in the P2P distribution scenario. P2P Distribution Protocol manages the Merchant, Seed Buyers, Proxies and Peer Buyer transactions. Proxies are used to handle anonymous communication between the peer buyers and higher authorities. The following problems are identified from the current multimedia data security methods.

- Standard database search process is not secured
- Proxy misbehavior identification is not supported
- Spatial and temporal factors are not adapted
- Limited privacy on communication process

5. MULTIMEDIA CONTENT REDISTRIBUTION PROTECTION USING ENHANCED TTP

The traitor Tracing Protocol is improved with context information based redistribution analysis mechanism. Four party anonymous communication protocol is integrated with the system to verify misbehave proxies. The standard database search for buyer verification is enhanced with security and privacy factors. Mutual anonymity model is integrated with the system to ensure the communication privacy. Media content distribution scheme is designed to manage the legal distribution process. Security and privacy factors are integrated with the communication process.

Context based data verification process is integrated with the Traitor Tracing Protocol. The system is partitioned into six major modules. They are Merchant Application, Redistribution Proxy, Buyer Application, Recombined Finger Print Management, Communication Security and Traitor Discovery Process. Merchant application is designed to sell multimedia contents. Redistribution proxy is used to handle the data transmission process. Buyer application is designed to market the media data. Recombined finger prints are used to verify the legal content distribution process. Data anonymization tasks are carried out under the communication security. Illegal data releases are identified in traitor discovery process.

The merchant application is build to manage media data market services. Redistribution proxies and buyers are connected with the merchant applications. Media content verification is initiated by the merchant. Finger prints are maintained under the merchant environment. The redistribution proxy is deployed to manage media contents. Merchant uploads its media content to the proxy environment. Media data values are transferred to the buyers. Data distribution operations are updated into the log files. Buyer application is divided into two types. Seed buyer manages the content reselling operations. Peer buyer handles the communication between the customers. P2P communication protocol is used to manage the data communication process. Copyright protection details are updated with the finger print information. Recombined finger prints are generated with seed user details. Illegal data release is discovered with the recombined finger prints. The finger prints are updated with location and time information

Communication security is provided to manage data transmission over the applications. Security and privacy are provided for the graph search process. Mutual anonymity schemes are used to protect the identity in the communication process. Data integrity is supported for the transaction and media data delivery process. Enhanced Traitor Tracing Protocol is applied to verify the data release operations. Location and time verification procedures are integrated with

the system. Traitor tracing carried out with integrity verification process. Misbehave proxy identification is carried out with Four party anonymous communication protocol.

6. CONCLUSION

Legal distribution of multimedia contents is an important factor in multimedia data market environment. Traitor Tracing Protocol is used to identify the illegal distribution multimedia content. The Traitor Tracing Protocol is enhanced with misbehave proxy management mechanism. Spatio-temporal factors and security on communication and verification tasks are upgraded in the system. Context aware illegal data transmission discovery model is adapted in the multimedia data security process. Security and privacy are provided for anonymous data communication under the Peer to Peer (P2P) environment. Misbehave proxy identification is supported by the multimedia data security system. Graph search overhead is reduced by the system. The system protects the data privacy in the piracy protection process. Request based piracy verification process is carried out on the media data. The system reduces the communication and computational overhead in the verification process.

REFERENCES

[1] N. Khodabakhshi and M. Hefeeda, "Spider: A system for finding 3D video copies," in ACM Trans. Multimedia Comput., Commun., Appl. (TOMM), Feb. 2013, vol. 9, no. 1, pp. 7:1–7:20.
[2] A. Stupar, S. Michel, and R. Schenkel, "Rankreduce – processing k-nearest neighbor queries on top of mapreduce," in Proc. Workshop Large-Scale Distrib. Syst. Inf. Retrieval (LSDS-IR'10), Geneva, Switzerland, Jul. 2010, pp. 13–18.
[3] M. Aly, M. Munich, and P. Perona, "Distributed Kd-Trees for retrieval from very large image collections," in Proc. Brit. Mach. Vis. Conf. (BMVC), Dundee, U.K., Aug. 2011.
[4] Z. Liu, T. Liu, D. Gibbon, and B. Shahraray, "Effective, and scalable video copy detection," in Proc. ACM Conf. Multimedia Inf. Retrieval (MIR'10), Philadelphia, PA, USA, Mar. 2010, pp. 119–128.
[5] H. Liao, J. Han, and J. Fang, "Multi-dimensional index on hadoop distributed file system," in Proc. IEEE Conf. Netw., Archit. Storage (NAS'10), Macau, China, Jul. 2010, pp. 240–249.
[6] W. Lu, Y. Shen, S. Chen, and B. Ooi, "Efficient processing of k nearest neighbor joins using MapReduce," in Proc. VLDB Endowment (PVLDB), Jun. 2012, vol. 5, no. 10, pp. 1016–1027.
[7] E. Metois, M. Shull, and J. Wolosewicz, "Detecting online abuse in images. Markmonitor Inc.," U.S. Patent 7925044, Apr. 12, 2011.
[8] P. Ram and A. Gray, "Which space partitioning tree to use for search," in Proc. Adv. Neural Inf. Process. Syst. (NIPS'13), Lake Tahoe, NV, USA, Dec. 2013, pp. 656–664.
[9] S. Ioffe, "Full-length video fingerprinting. Google Inc.," U.S. Patent 8229219, Jul. 24, 2012.
[10] Mohamed Hefeeda , Tarek ElGamal, Kiana Calagari and Ahmed Abdelsadek, "Cloud-Based Multimedia Content Protection System", IEEE Transactions On Multimedia, Vol. 17, No. 3, March 2015