Type of manuscript

Multimedia Content Protection System

#1Dhanashree Parmar, #2Pooja Talekar, #3Priyanka Marathe
1dhanu07parmar@gmail.com,
2pooja03talekar@gmail.com,
3priyankamarathe0108@gmail.com
Nutan Maharashtra Institute of Engineering and Technology,
Talegaon Dabhade, Pune

ABSTRACT

A new approach to protect our videos and large text documents is discussed here. Our design aims to provide a security for your content (here content refers to videos, multimedia and large text documents) in the online world so that no other person would copy the same and claim it as there content. The system is cloud based all the computing of the content is done in the cloud. This system has three main components (i)Signature Generation of the video , large text files or any document (ii) Comparison of the signature of your reference content with the other content which might possible been copied from the original content , (iii)Possible outcome of success or failure is known and for any altered content level of copy can be detected. The comparison is fast and is in the cloud .Hence our model not only can detect any pirated content but also gives level of copy for any altered content in an online process.

ARTICLE INFO

Keywords— ***Copy detection video depth Signatures, images, cloud applications***

INTRODUCTION

Latest development technologies as well as the availability of online free hosting sites have made it very easy to duplicate copyrighted materials such as videos, images, important documents and music clips files. Illegally redistributing these contents over the Internet will result in great loss in terms of revenues for content owners. Finding these illegally made copies over the Internet is a very complex and expensive task as the volume of data available over the internet is huge and comparing content to match and identify copies is very complex. Here we present a novel system for any file content protection on cloud infrastructures. The system can be used to protect various multimedia content types like videos, images, music or any large text files. The system can run on private clouds, public clouds, or any combination of public private clouds. The design is cost effective because it uses the computing resources on demand. The design can be scaled up and down to support varying amounts of file's content being protected. The contributions of this paper are as follows:-

- Complete multi-cloud system for multimedia content protection. The system can support different types of file content and can effectively use varying computing resources.
- A novel method for generating signature for any multimedia or large text files which is simple and effective.

The focus is on the approach for protecting multimedia content, which is content based copy detection (CBCD) . In this approach, signatures (or fingerprints) are extracted from original objects. Signatures are also created from query (suspected) 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 (particularly the block based) are the most widely used. However, 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 sig-natures are computationally intensive and not widely used in practice. For more details, see surveys for audio fingerprinting and 2-D video fingerprinting .

II. Related work

The problem of protecting various types of multimedia con-tent has attracted significant attention from academia and industry. One approach to this problem is using watermarking, in which 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.

Distributed Matching Engine:

Unlike many of the previous works, 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 is on distributed techniques that can scale to large datasets which are geographically separated.

III. Existing System

The problem in existing system protecting various types of multimedia content has attracted significant attention from academia and industry. One approach to this problem is using watermarking , in which 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.

Signatures or fingerprints extraction technique is another important one. Signatures are also created from query (suspected) 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 (particularly the blockbased) are the most widely used. However, 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.

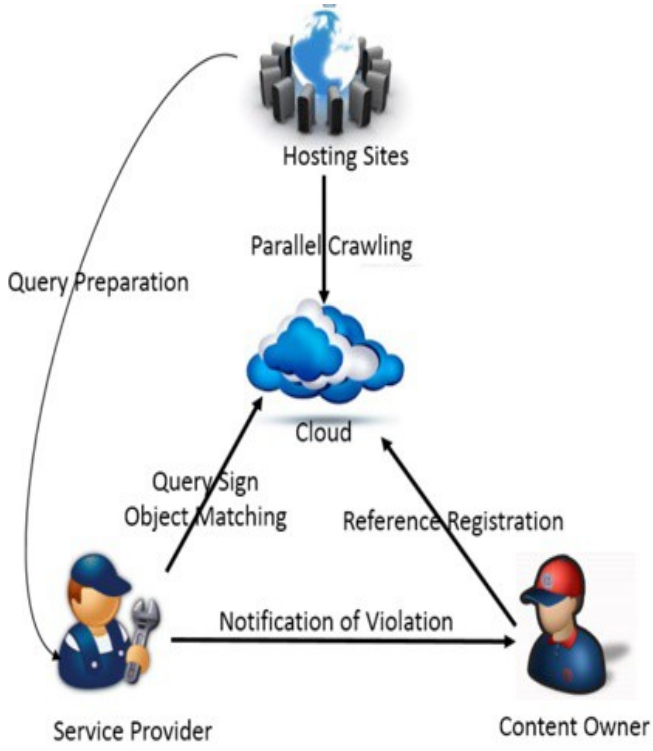
IV. Proposed System

The proposed system uses spatial signature techniques there are two main components, the cloud server and the clients. The client users are given provision to store their files in the cloud. But how safe will their data or copy write contents be from any third party? Chances are that the files can be viewed and copied. To prevent this copying and duplication of data we have come up with a simple content protection method that can assure the user that no matter who views the file, they will not be able to replicate the data in the system. Example, YouTube videos can be seen by everyone but our system will not allow any two users to have the same videos. The credit always goes to the owners and not anyone else. We developed new system for multimedia content protection on cloud framework. This novel system can be used to protect different multimedia content types. In our novel system we evolved fully multicloud system for multimedia content protection.

The proposed system supports various types of multimedia content and can actively occupy varying computing resources. New approach can be used generating signatures for videos. This novel approach generates signatures that occupy the depth in stereo content without computing the depth signal itself, which is a computationally expensive process. The novel design for capture high dimensional multimedia objects for distributed matching engine. This design provides the primitive function of finding nearest neighbors for large scale datasets. The design also offers an auxiliary function for further processing of the neighbors. This two stage design used in novel system to easily support various types of multimedia content. The focus of this paper is on the other approach for protecting multimedia content, which is content-based copy detection. In this proposed system from original objects, signatures are extracted. Signatures are also generated from query and objects downloaded from web. Then, the complimentary is measure

between original and suspected objects to find potential copies.

V. Proposed System Architecture



The architecture is as shown in the figure above. In which user firstly upload the video and also create the signature for that video. Whenever user upload their video system will check the signature for that video. If the signature is matched with in our database system will send the notification to the data owner.

Admin will check the details for each video and its signature.

Query Preparation:

A content protection system has three main parties: (i) content owners (e.g., Disney), (ii) hosting sites The first party is interested in protecting the copyright of some of its multimedia objects, by finding whether these objects or parts of them are posted on hosting sites (the second party).

Accuracy:

The system should have high accuracy in terms of finding all copies (high recall) while not reporting false copies (high precision). Achieving high accuracy is challenging, because copied multimedia .

Reference Registration:

Creates signatures from objects that content owners are interested in protecting, and inserts them in the distributed index.

Query Preparation:

Creates signatures from objects downloaded from online sites, which are called query signatures. It then uploads these signatures to a common storage.

Object Matching:

Compares query signatures versus reference signatures in the distributed index to find potential copies. It also sends notifications to content owners if copies are found.

Parallel Crawling:

Downloads multimedia objects from various online hosting sites.

Content Owner:

The one who owns or upload data on the cloud, hosting sites is the content owner. he will be notified if the contents are copied or misused.

Service Provider:

Administrator of the hosting sites is the Service provider.

Hosting Sites:

The sites like YouTube, instagram, twitter, facebook

Parameters	Existing system	Proposed system
Algorithm	SIFT	CBIR MD5 Multipart
Techniques	Watermarking	Digital signature & watermarking
Disadvantages	1.Less secure. 2.Watermarking is not effectively use. 3.High cost.	Overcome all disadvan-tages.

VI. Algorithms

A. CBIR:

"Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image.

The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself.

CBIR is desirable because searches that rely purely on metadata are dependent on annotation quality and completeness. Having humans manually annotate images by entering keywords or metadata in a large database can be time consuming and may not capture the keywords desired to describe the image. The evaluation of the effectiveness of keyword image search is subjective and has not been well-defined. In the same regard, CBIR systems have similar challenges in defining success.

B. MD5:

The **MD5 algorithm** is a widely used hash function producing a 128-bit hash value. It is used as a checksum to verify data integrity, but only against unintentional corruption.

1. Step 1 – append padded bits: A single “1” bit is appended to the message, and then “0” bits are appended so that the length in bits equals 448 modulo 512.

2. Step 2- append length: A 64 bit representation of b is appended to the result of the previous step. The resulting message has a length that is an exact multiple of 512 bits.

3. Step 3 – Initialize MD Buffer :A four-word buffer (A,B,C,D) is used to compute the message digest. Here each of A,B,C,D, is a 32 bit register

4. Step 4 – Process message in 16-word blocks: Four auxiliary functions that take as input three 32-bit words and produce as output one 32-bit word.

5. Step 5 – output: The message digest produced as output is A, B, C, D. That is,

output begins with the low-order byte of A, and end with the high-order byte of D.

C. Multipart

This algorithm is used for file uploading on the cloud storage. A HTTP multipart request is a HTTP request that HTTP clients construct to send files and data over to a HTTP Server. It is commonly used by browsers and HTTP clients to upload files to the server. Under the hood of a HTTP request with multipart form data

VII. Advantages

- Accuracy in matching the contents.
- Computational Efficiency of signature generation is faster.
- Scalability and Reliability.
- Cost Efficiency.
- The system can run on private clouds, public clouds, or any combination of public-private clouds.
- Our design achieves rapid deployment of content protection systems, because it is based on cloud infrastructures that can quickly provide computing hardware and software resources.
- The design is cost effective because it uses the computing resources on demand.
- The design can be scaled up and down to support varying amounts of multimedia content being protected.

VIII. Conclusion

This system is used to detect and prevent the possibilities of any other third party trying to violate any copyrighted material. The Material here refers to any type of multimedia content like a video, images, music etc. The system intends to protect the contents of any multimedia content file irrespective of the type of file for files are available on a cloud platform. The contents of the multimedia video cannot be replicated, If the contents are copied then it should not be allowed to be uploaded into the cloud facility. Unique technique is employed to protect the contents of the file.

Signatures are generated for every video which is unique to that video. Signature generation and comparison uses well known algorithms effectively. The algorithms used are simple and effective. The duplicate copies of the video contents are determined by analyzing these signatures. The level or the percentage of copy is determined from which the contents were copied.

For example, our current system is optimized for batch processing. Thus, it may not be suitable for online detection of illegally distributed multimedia streams of live events such as soccer games. In live events, only small segments of the video are available and immediate detection of copyright infringement is crucial to minimize financial losses. To support online detection, the matching engine of our system needs to be implemented using a distributed programming framework that supports online processing, such as Spark. In addition, composite signature schemes that combine multiple modalities may be needed to quickly identify short video segments. Furthermore, the crawler component needs to be customized to find online sites that offer pirated video streams and obtain

segments of these streams for checking against reference streams, for which the signatures would also need to be generated online. Another future direction for the work in this paper is to design signatures for recent and complex formats of 3-D videos such as multiview plus depth. A multiview plus depth video has multiple texture and depth components, which allow users to view a scene from different angles. Signatures for such videos would need to capture this complexity, while being efficient to compute, compare, and store.

VIII. ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to our guide **Prof. Pramod Patil** and our HOD **Prof. Nitin Wankhede** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of (Organization Name)for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

REFERENCES

[1] A. Abdelsadek, "Distributed index for matching multimedia objects," M.S. thesis, School of Comput. Sci., Simon Fraser Univ., Burnaby, BC, Canada, 2014.

[2] A. Abdelsadek and M. Hefeeda, "Dimo: Distributed index for matching multimedia objects using MapReduce," in *Proc. ACMMultimedia Syst. Conf. (MMSys'14)*, Singapore, Mar. 2014, pp. 115–125.

[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] J. Bentley, "Multidimensional binary search trees used for associative searching," in *Commun. ACM*, Sep. 1975, vol. 18, no. 9, pp. 509–517.

[5] P. Cano, E. Batle, T. Kalker, and J. Haitsma, "A review of algorithms for audio fingerprinting," in *Proc. IEEE Workshop Multimedia Signal Process.*, Dec. 2002, pp. 169–173.

[6] J. Dean and S. Ghemawat, “MapReduce: Simplified data processing on large clusters,” in *Proc. Symp. Oper. Syst. Design Implementation (OSDI’04)*, San Francisco, CA, USA, Dec. 2004, pp. 137–150.

[7] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR’09)*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[8] A. Hampapur, K. Hyun, and R. Bolle, “Comparison of sequence matching techniques for video copy detection,” in *Proc. SPIE Conf. Storage Retrieval Media Databases (SPIE’02)*, San Jose, CA, USA, Jan. 2002, pp. 194–201.

[9] S. Ioffe, “Full-length video fingerprinting. Google Inc.,” U.S. Patent 8229219, Jul. 24, 2012.

[10] A. Kahng, J. Lach, W. Mangione-Smith, S. Mantik, I. Markov, M. Potkonjak, P. Tucker, H. Wang, and G. Wolfe, “Watermarking techniques for intellectual property protection,” in *Proc. 35th Annu. Design Autom. Conf. (DAC’98)*, San Francisco, CA, USA, Jun. 1998, pp. 776–781.

[11] 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.

[12] S. Lee and C. Yoo, “Robust video fingerprinting for content-based video identification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 983–988, Jul. 2008.