Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 04, April-2016

Security in Cloud based Multimedia System: A Review

Madhuram Kabra
Student
Dept. of Computer Engineering
VJTI, Matunga, Mumbai

Prajakta Patil
Student
Dept. of computer Engineering
VJTI, Matunga, Mumbai

Atoshi Mondal
Student
Dept. of Computer Engineering
VJTI, Matunga, Mumbai

Prof. Jijnasa Patil
Dept.
Computer Engineering
VJTI, Matunga, Mumbai

***Abstract*— With the emergence of various multimedia services, Cloud Based Multimedia System (CMS) is gaining more and more popularity. The Internet traffic is expected to be dominated by the multimedia content by 72% by 2019. With such popularity, CMS needs to be secured, should provide an efficient quality of service, and should be balanced. This paper provides a comprehensive review of these characteristics, aiming on the fundamental design considerations, such as, robustness, scalability, availability, overall performance.**

***Index Terms*— Cloud Computing, Multimedia System, Load Balancing, Content Protection. Introduction**

I. INTRODUCTION

Cloud Computing is a fast, growing and emerging technology that could provide elasticity, scalability, universal availability, and cost-effectiveness. On the other hand Multimedia Systems is integrated, digitally represented, computer controlled and interactive. Together they form a Cloud-Based Multimedia System (CMS). With the recent advances in the technology the cloud-based multimedia system is emerging as a novel computing paradigm processing multimedia applications along with catering to user's demands. Having said that, need to provide this technology to the users efficiently is not as easily achievable as it looks.

This paper will be focusing on the review of various works done by the researchers on the one of the main characteristics of the CMS. Namely, security (content protection).

In recent years, the world has seen a major advances in the processing and recording equipment for the multimedia content. Also a lot of free hosting sites have been made available. This has made the task of duplicating copyrighted materials for e.g., videos, audios, graphical images etc. relatively easy. This may result into substantial loss of revenues for the creators. Finding these illegally-made copies over the internet is complex and infeasible operation, simply because of the sheer volume of the multimedia content available over the internet.

General architecture of a cloud-based multimedia system is centralized hierarchical in nature [3]. It has a resource manager, cluster heads, and server clusters. Resource manager deals with the clients negotiating the resources required by them as well as its billing. It assigns the clients' request to a server cluster. Server cluster is decided based on the clients' requirements and the task to be performed for the client.

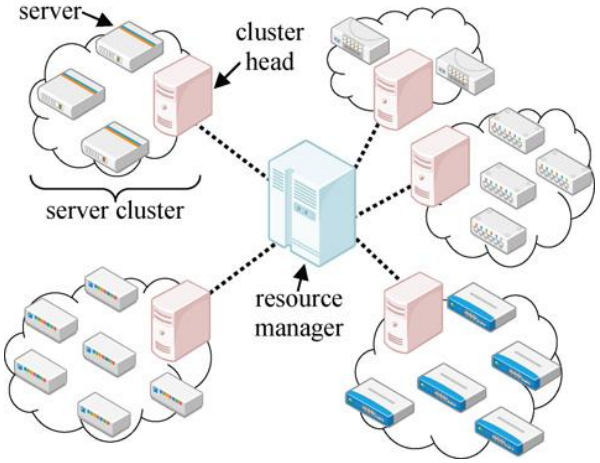

Fig. 1. Illustration of a centralized hierarchical cloud-based multimedia system.

Each server cluster has its own task characteristics. Then the cluster head distributes the task among the servers within its cluster. This is shown in Fig. 1. The architecture, on paper, is easy to implement but practically it is severely complicated. This is more crucial considering that a cloud service provider doesn't provide a single service; rather multiple services to generate more income.

In recent years, various multimedia applications, services and devices have emerged. This has led the multimedia content to be the major traffic in the Internet which keeps increasing rapidly. According to [2] every second a million minutes of video content will cross over the Internet traffic. Globally, 72% of all Internet multimedia traffic will cross content delivery networks by 2019, up from 57% in 2014. Service providers like YouTube stores their video assets in the cloud, delivering the multimedia content to the consumers

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 04, April-2016

cross cloud. With the increase in the service providers, it becomes a huge challenge to determine which content is authenticated and which is fabricated or downright copied violating copyrights of the content owners.

The rest of the paper is organized as follows. In Section II, we will see various proposed protection system for CMS.

II. PROTECTION IN CLOUD-BASED MULTIMEDIA SYSTEMS

Security or protection has always been major issue in the digital world. CMS is also no different. As mentioned earlier, with the recent advances in the technologies one can easily duplicate the original multimedia content of the content owners causing a significant revenue loss for the owners. Various methods have been devised and proposed for protection. We will now see some of them one by one.

A. Watermarking

Kahng [4] proposed an approach to this problem using watermarking. In his paper, Kahng establishes a principle of watermarking the intellectual property. Watermarking is a mechanism which can be used for identification of the multimedia content. Watermarking is some distinctive information which is integrated in the content itself and remains so permanently. It is nearly invisible to the human or machine inspection and thus, is difficult to remove. Before we look into the approach we need to see is non-intrusive watermarking.

Non-intrusive Watermarking: The watermarking methods that can be transparently integrated within existing design flows via pre- or post-processing.

The following ingredients form the *context* for a non-intrusive watermarking procedure:

- An *optimization problem* with known difficult complexity, i.e., either achieving an acceptable solution, or enumerating enough acceptable solutions, is prohibitively costly. The solution space of the optimization problem should be large enough to accommodate a digital watermark.
- A well-defined *interpretation* of the solutions of the optimization problem as intellectual property.
- Existing *algorithms* and/or *off-the-shelf software* that solve the *optimization problem*, likely without any kind of watermarking involved.
- *Protection requirements.* For e.g. (i) removing or forging a watermark must be as hard as re-creating the design; and (ii) tampering with a watermark must be provable in court.

Their general strategy is as follows:

- Map the content owner's signature into a set of constraints decided by the owner. These constraints can be independently held for a particular solution.
- Now, if disproportionately, many of these constraints are satisfied, then the presence of the signature is indicated.

The key or integral part of the watermarking are the constraints that a content owner chooses, and the method which is implemented for them to be satisfied (e.g., pre- or post-processing). Because these choices dramatically affects the strength of the watermark and the degradation of solution quality caused by watermarking.

Watermarking by far is mathematically sound and practical. However, it is not an optimal approach for the content protection. Watermarking needs the content owner to embed the watermark into the multimedia content before it can be released. Although, these contents would be secured from tampering, it leaves the contents already released in the Internet dangling. It does not provide any sort of protection to them.

B. Audio Fingerprinting

Fingerprints or signatures are a piece of identifiers that are extracted from a multimedia content. Fingerprints can be extracted both audio and video. In research literature, the process of fingerprint or signature identification is also known as content-based copy detection (CBCD). We will discuss the approaches proposed for audio fingerprinting in this subsection and for video in further subsections.

An audio fingerprint is a content-based compact signature that summarizes an audio recording. Audio Fingerprinting technologies enables us to monitor audio independent of its format and without the need of meta-data or watermark embedding. It can link unlabeled audio to corresponding metadata. The different approaches proposed for audio fingerprinting are usually described with different rationales and terminology depending on the background: Pattern matching, Multimedia (Music) Information Retrieval or Cryptography (Robust Hashing).

General working of audio fingerprinting or Content-based audio identification (CBID) systems are as follows. Firstly, they extract a perceptual digest of a piece of audio content, i.e. the fingerprint and it is stored in a database. When the system is presented with an unlabeled audio, the system calculates its fingerprint and matches against those stored in the database. Using fingerprints and matching algorithms, distorted versions of a recording can be identified as the same audio content.

Thus, there are two fundamental processes in audio fingerprinting, i.e., fingerprint extraction and matching algorithms.

In spite of the different rationales behind the identification task, methods share certain aspects. As shown in Fig. 2 the framework consists of two processes: fingerprint extraction and matching. The fingerprint extraction consists of a front-end, and a fingerprint modeling block. The front-end is responsible for computing a set of measurements from the signal. The fingerprint model block defines the final fingerprint representation for eg: a vector, or a tree.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 04, April-2016

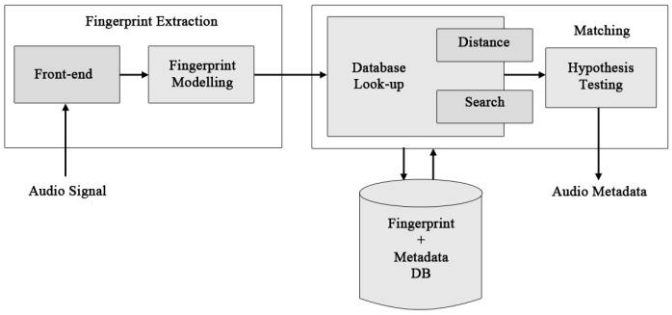

Fig. 2. Content-based Audio Identification Framework

The matching algorithm searches the database of fingerprints for the best match. Distance and Search compares fingerprints such that the results are calculated as fast and as reliable as possible. Finally, the last block of the system – the hypothesis testing computes a reliability measure indicating how sure the system is about an identification.

Front End is a process which converts an audio signal into a sequence of relevant features to feed the fingerprint model block. It is a five step process:

- *Pre-processing:* the audio is digitalized and converted into a general format. Usually, it is converted into raw format, or to a certain sampling rate (5 to 44.1 KHz).
- *Framing & Overlap:* the signal is divided into frames and overlapped to assure robustness. Here also, when choosing the values, there is a trade-off between rate of change in the spectrum and system complexity.
- *Transformation:* The idea is to transform the set of measurements to a new set of features. These transforms are suitably chosen to reduce the redundancy. Usually standard time to frequency transforms like Fast Fourier Transform (FFT) are used. Although, various others transforms have been proposed: Walsh-Hadamard Transform [5]. Mihak *et al.* [6] and Burges *et al.* [7] uses Modulated Complex Transform (MCLT).
- *Feature Extraction:* Once on a time-frequency representation, we apply additional transformations to generate the final acoustic vectors so that the dimensionality is reduced and at the same time invariance to distortions is increased. There are many techniques developed and applied to extract perceptually meaningful parameters. Cano *et al.* [8] and Blum *et al.* [9] used Mel-frequency Cepstrum Coefficients (MFCC) while Allamanche *et al.* [10] chose Spectral Flatness Measure (SFM).
- *Post-processing:* further refinements if required.

Fingerprint Modelling Block usually receives a sequence of feature vectors calculated on a frame by frame basis. It exploits redundancies in the frame time vicinity, inside a recording and across the whole database, to reduce the fingerprint size. The type of model chosen conditions the distance metric and also the design of indexing algorithms for fast retrieval. Fingerprint can be achieved in the form of a vector like in [11] and [12]. Also it could be sequences of features like binary vector sequences in [13] and [9].

Distances and Searching Methods are very much depended to the type of model chosen. For e.g., if the model chosen is vectors, then correlation is used. Also an issue of how to efficiently do the comparisons is answered, which is usually depended upon fingerprint representation. [9] Uses a slightly modified version of Euclidean distance. In [13], Manhattan distance is used and in a special where quantization is binary, Hamming distance is used.

And finally, *Hypothesis Testing* aims to answer whether the query is present or not in the repository. It provides scores to the fingerprints extracted from the query to decide correct identification.

C. 2-D Video Fingerprinting

Content-based copy detection (CBCD) or fingerprinting of 2-D videos has been a research problem for a long time. A video fingerprint uniquely differentiates one video from other. The general strategy of CBCD for videos in CMS is as follows:

- First, get the unknown video segment and extract a fingerprint.
- Once, the fingerprint is extracted, discard the video, it is no longer required.
- Next, perform the search query with the database for finding the match with the query fingerprint.
- Give a score to each fingerprint matched.
- The metadata of the candidate closest to the query is declared as the result.

Lee *et al.* [14] work on this can be considered as academic state-of-the art. Their approach is similar to the general strategy.

The methods for creating and matching signatures are classified into four categories:

- *Spatial:* Spatial signatures deals with the luminance pattern of the video. A class of spatial signatures is designed to characterize luminance patterns in a video frame. In such designs, a video image is first converted to the YUV color space; the luminance (Y) component is kept, and the chrominance components (U, V) are discarded. The luminance image is further subdivided into a fixed-sized grid of blocks. The subdivision of a video frame serves two purposes. First, it leads up to block-based signatures that are robust to changes in pixel values; second, it produces a compact and fixed-sized frame fingerprint consisting of fixed number of block signatures. Some of the works in Spatial are [15], [16], [17], [18] and [19]
- *Temporal:* In temporal we use the durations for signature. First, a video sequence is segmented into shots. Then, the duration of each shot is concatenated to form the fingerprint of the video. Shivakumar *et al.* [20] and [21] has done major work in temporal signatures. Other works in temporal signatures are Cheung *et al.* [22], Chen *et al.* [23], and Law-To *et al.* [24]
- *Color:* In this approach, a level-quantized histogram is computed for Y, U, and V components for each video frame. To reduce the resulting signature data, a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 04, April-2016

polynomial approximation is used to model the pixel counts in each bin of the histograms along the temporal direction. A special distance metric based on histogram intersection is used as a similarity measure. Hampapur *et al.* [17] and Li *et al.* [18] have used color signatures for content protection.

- *Transform-domain:* This is the most practically implemented approach. There are some designs that compute video signatures in a transform domain. Various transformations used are polar Fourier Transform [26], Radon Transform [27], Discrete Cosine Transform [28] etc.

D. 3-D Video Fingerprinting

Three dimensional videos although existing since 1915, didn't became worldwide popular until 2000s. Thus, it wasn't until recently that a need to provide the security for 3-D video was realized. As such it doesn't have much works done for it. There are 3 major works that we were able to find in our research [29] [30] [1]. In [29], they first create signatures for depth and visual for 3-D videos. Then the signature is compared with the database of the reference video's signatures. It is done using temporal and spatial characteristics of the videos. The system generates a score indicating whether the video matches any referenced video or not and in case of matching what portion is matching with the query video. We found the process to be efficient and precise, however it works on the assumption that the depth maps are either given or estimated. This method is suitable for 3-D videos which are encoded in video plus depth format. But not for stereotypic videos, where estimating depth maps are expensive.

Second work [30], focus on two of the major attacks possible on the 3-D video. First, *view interpolation/deletion attack* and second, *display plane/change of focus attack. View interpolation* attacks are those in which the user or end user software records a personal video for the arbitrarily selected viewpoints and illegally distribute them. This attack is based on anti-stereoscopic concept. *Display plane/change of focus* attacks the user can misuse the free-view setup and change the display plane or the chose the region of focus to record the video with these changes and wrongfully use and distribute it. These attacks are counter-measured by a scale invariant feature descriptor (SIFT) based fingerprinting mechanism. They create a SIFT descriptor using multiscale Gaussian pyramid derived from given image. Then they find the localized maxima and minima for all the feature points. Then these feature points are compared with the centralized database. The major drawback it faces is the storage overhead and search complexity as comparing all SIFT points in each frame is not practical in large database.

The work in [1], computes the signatures for the stereo format 3-D videos. The video is divided into two views left and right for left and right eyes respectively. And then the signature is computed for both the views respectively. Instead of using entire video to compute the signature only a part of the video is taken using sampling. This reduces computation and storage costs. This sampling could be random or periodic, for e.g. every fifteenth frame. Once, the signature is computed it is then matched against the database. The database instead of being centralized like in previous works is distributed. A distributed matching engine is used to search through the database using crawlers. Crawlers downloads video from internet in parallel and checks with the queried video to find the match.

III. SUMMARY

We have presented a review of research in the recent area of content protection. We saw how the work evolved from watermarking to fingerprinting. We also saw various works from past to recent in audio fingerprinting. The different tasks involved in an audio fingerprinting system have been described. The purpose of each block has been commented along with some hints of the proposed solutions. We also saw the research in video fingerprinting. Video fingerprinting has come a long way since it began almost two decades ago and developed into a technology that is adopted by the industry. We saw how two-dimensional videos and three-dimensional videos are protected. Key areas of research include designs of video signatures, fingerprinting and fingerprint matching algorithms. Among the large number of video signature techniques, they can be classified into spatial, temporal, color, and transform-domain signatures. Although none is perfect, spatial signatures are found to be the overall winner in terms of robustness, discriminability, compactness, and computational complexity. Temporal and color signatures can provide enhanced discriminability when used together with spatial signatures. Fingerprint matching by exhaustive search has a linear time complexity with regard to the size of reference database. Fortunately, effective approximation techniques have been developed that provide a dramatic reduction in computational complexity, speeding up fingerprint queries by several orders of magnitude over an exhaustive search with a negligible loss in accuracy. This made it possible to build practical fingerprint matching systems that are scalable. Moving forward, researchers and practitioners are also exploring and experimenting other applications of video fingerprinting, including contextual advertising, video asset management, and content-based video search.

REFERENCES

[1] M. Hefeeda, T. ElGamal, K. Calagari, and A. Abdelsadek, "Cloud-based Multimedia Content Protection System" *IEEE Transactions on Multimedia*, Vol. 17, no. 3, pp. 420-433, March 2015

[2] Cisco Visual Networking Index: Forecast and Methodology, 2014-2019 White Paper [Online].

[3] C. Lin, H. Chin, and D. Deng, "Dynamic Multiservice Load Balancing in Cloud-Based Multimedia System" *IEEE Systems Journal*, Vol. 8, no. 1, pp 225-234, March 2014

[4] 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.

[5] S. Subramanya, R.Simha, B. Narahari, and A. Youssef, "Transform based indexing of audio data for multimedia databases," in *Proc. of Int. Conf. on Computational Intelligence and Multimedia Applications*, New Delhi, India, Sept. 1999.

[6] M. Mihak and R. Venkatesan, "A perceptual audio hashing algorithm: a tool for robust audio identification and information hiding," in *4th Workshop on Information Hiding*, 2001.

[7] C. Burges, J. Platt, and S. Jana, "Extracting noise-robust features from audio data," in *Proc. of the ICASSP*, Florida, USA, May 2002.

[8] P. Cano, E. Batlle, H. Mayer, and H. Neuschmied, "Robust sound modeling for song detection in broadcast audio," in *Proc. AES 112th Int. Conv.*, Munich, Germany, May 2002.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 04, April-2016

[9] T. Blum, D. Keislar, J. Wheaton, and E. Wold, "Method and article of manufacture for content-based analysis, storage, retrieval and segmentation of audio information," U.S. Patent 5,918,223, June, 1999.

[10] E. Allamanche, J. Herre, O. Helmuth, B. Fr¨oba, T. Kasten, and M. Cremer, "Content-based identification of audio material using mpeg-7 low level description," in *Proc. of the Int. Symp. of Music Information Retrieval*, Indiana, USA, Oct. 2002.

[11] (2002) Etantrum. [Online]. Available: http://www.freshmeat.net/projects/songprint

[12] (2002) Musicbrainz trm. musicbrainz-1.1.0.tar.gz. [Online]. Available: ftp://ftp.musicbrainz.org/pub/musicbrainz/

[13] J. Haitsma, T. Kalker, and J. Oostveen, "Robust audio hashing for content identification," in *Proc. of the Content-Based Multimedia Indexing*, Firenze, Italy, Sept. 2001.

[14] 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.

[15] Bhat, D. N. and Nayar, S. K., "Ordinal measures for image correspondence," IEEE Trans. Pattern Ana. Mach. Intell., vol. 20, no. 4, pp. 415-423, Apr. 1998.

[16] Mohan, R., "Video sequence matching," Proc. Int. Conf. Acoust., Speech and Signal Processing (ICASSP), vol. 6, pp. 3697-3700, Jan. 1998.

[17] Hampapur, A., Hyun, K.-H. and Bolle, R. M., "Comparison of sequence matching techniques for video copy detection," Proc. SPIE, Storage and Retrieval for Media Databases, vol. 4676, pp. 194-201, Jan. 2002.

[18] Hua, X.-S., Chen, X. and Zhang, H.-J., "Robust video signature based on ordinal measure," IEEE Int. Conf. Image Proc. (ICIP), vol. 1, pp. 685-688, Oct. 2004.

[19] Kim, C. and Vasudev B., "Spatiotemporal sequence matching for efficient video copy detection," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 1, pp. 127-132, Jan. 2005.

[20] Indyk, P., Iyengar, G. and Shivakumar, N., "Finding pirated video sequences on the Internet," Tech. Rep., Stanford InfoLab, Stanford University, Feb. 1999.

[21] Shivakumar, N., "Detecting digital copyright violations on the Internet," Ph.D. Dissertation, Stanford University, Aug. 1999.

[22] Cheung, S.-C. S and Zakhor, A., "Efficient video similarity measurement with video signature," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 1, pp. 59-74, Jan. 2003.

[23] Chen, L. and Stentiford, F. W. M., "Video sequence matching based on temporal ordinal measurement," Pattern Recognition Letters, vol. 29, no. 13, pp. 1824-1831, Oct. 2008.

[24] Law-To, J., Buisson, O., Gouet-Brunetand, V. and Boujemma, N., "Robust voting algorithms based on labels of behavior for video copy detection," Proc. ACM Int. Conf. on Multimedia, pp. 835-844, 2006.

[25] Li, Y., Jin, J. S. and Zhou, X., "Matching commercial clips from TV streams using a unique, robust, and compact signature," Proc. Digital Imaging Computing: Techniques and Applications, pp. 266-272, Dec. 2005.

[26] Swaminathan, A., Mao, Y. and Wu, M., "Image hashing resilient to geometric and filtering operations," IEEE Workshop on Multimedia Signal Processing (MMSP), pp. 355-358, Sep. 2004.

[27] De Roover, C., De Vleeschouwer, C., Lefèbvre, F. and Macq, B., "Robust video hashing based on radial projections of key frames," IEEE Trans. Signal Proc., vol. 53, no. 10, pp. 4020-4037, Oct. 2005.

[28] Coskun, B. and Sankur, B., "Robust video hash extraction," Proc. European Conf. on Signal Processing (EUSIPCO), pp. 2295-2298, Sep. 2004.

[29] 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.

[30] V. Ramachandra, M. Zwicker, and T. Nguyen, "3D video fingerprinting," in *Proc. 3DTV Conf.: True Vis.—Capture, Transmiss. Display 3D Video (3DTV'08)*, Istanbul, Turkey, May 2008, pp. 81–84.