SECURE VIDEO PROCESSING: PROBLEMS AND CHALLENGES

Wenjun Lu, Avinash Varna, and Min Wu

Department of Electrical and Computer Engineering
University of Maryland, College Park, USA

ABSTRACT

Secure signal processing is an emerging technology to enable signal processing tasks in a secure and privacy-preserving fashion. It has attracted a great amount of research attention due to the increasing demand to enable rich functionalities for private data stored online. Desirable functionalities may include search, analysis, clustering, etc. In this paper, we discuss the research issues and challenges in secure video processing with focus on the application of secure online video management. Video is different from text due to its large data volume and rich content diversity. To be practical, secure video processing requires efficient solutions that may involve a trade-off between security and complexity. We look at three representative video processing tasks and review existing techniques that can be applied. Many of the tasks do not have efficient solutions yet, and we discuss the challenges and research questions that need to be addressed.

Index Terms— Secure signal processing, Privacy protection.

1. INTRODUCTION

The Internet is now embracing the era of cloud computing, where web-based services are gradually replacing traditional desktops for tasks of storing and managing information and performing computation. Information of different scope and importance, from personal data to commercial and government documents, are being managed by various online service providers. To fully achieve cloud computing's ultimate objective and potential, security and privacy protection is one of the most important issues that have to be addressed.

In this paper, we focus on video data and discuss the problems and challenges in securely managing private video online. Video can be considered as a sequence of images possibly accompanied by audio information. Video processing techniques typically can be adapted from image and audio processing techniques, but the rich information and large data volume of video brings unique challenges and also attracts active research. In recent years, due to the prevalence of smartphones and portable camcorders, video information has been generated at an unprecedented pace. The large size of video data and high computational requirement for processing videos makes it desirable to store and process videos in the cloud. As such, privacy protection and secure processing techniques for video becomes a pressing need.

Traditional privacy protection for online information typically relies on password-based access control. The private information might be encrypted during transfer, but will be decrypted and typically stored in plaintext on the server. Given that cloud services may attract more attacks and are vulnerable to untrustworthy system administrators, storing private information in plaintext brings serious privacy concerns. To address this problem, researchers have been studying the possibility of signal processing in the encrypted domain. Encrypting the private information brings privacy protection to a higher level, while the capability of performing signal processing in the encrypted domain can leverage the cloud's computation power to provide functionalities over the encrypted data. Research in secure signal processing can help address the security and privacy concerns about cloud computing and make cloud computing reach its full potential.

In the area of secure signal processing, Erkin et al. [1] provided an overview of related cryptographic primitives and some applications of secure signal processing in data analysis and content protection. Recently, there has been research work on secure computation targeting different application s [2–6]. Yiu et al. [2] considered privacy preserving range query over geospatial coordinates using a kd-tree. Wong et al. [3] proposed secure k-NN computation that preserves the original dot-product between two vectors while keeping the actual distances secret from the server. Erkin et al. [4] and Sadeghi et al. [5] studied the problem of privacy preserving face recognition, where one party wants to verify the existence of a given face in another party's database. The two parties want to keep their own data secret from each other. Homomorphic encryption is used to allow similarity computation in the encrypted domain. Similar techniques are also used by Jiang et al. [6] to identify the existence of similar text documents between two parties. Ye et al. [7] proposed k-anonymous quantization to achieve some trade-off between computational complexity and privacy protection in secure biometric matching. For searching over encrypted image database, Lu et al. [8, 9] proposed to encrypt visual features and search indexes of images and perform similarity comparison directly in the encrypted domain.

The above mentioned work have focused on text and images, and the processing task is either matching an exact copy or search for similar items. Many of the primitives used in above works are still valuable building blocks for secure video processing, and many of the techniques developed there have the potential to be extended for videos. However, the rich information contained in video and its temporal nature bring unique challenges and opportunities to secure video processing. In this paper, we provide an overview of such

Email: {wenjunlu, varna, minwu}@umd.edu

opportunities and challenges. We study several representative video processing tasks, including video summarization, classification, and search. For each of these tasks, we review related primitives and techniques that can be helpful for performing the task in a privacy-preserving fashion, and analyze challenges and possible future research directions. For some tasks, there is a straightforward extension from existing work, while for most others, it is still an open problem and calls for more research effort. In addition to specific processing tasks, the fundamental question of how to define security for multimedia data has great importance but has yet to be carefully studied. A correct understanding of the necessary level of protection for multimedia data is important in designing an efficient practical scheme for real-world applications. Overall, we believe it is beneficial to provide such an overview of research issues and challenges in secure video processing to attract more interest from the community.

2. PROBLEM FORMULATION

The application considered in this paper is secure online video management, where users store their private videos in encrypted form on remote servers and the server performs processing tasks over the encrypted videos. This is different from the traditional two-party computation problem, where both parties hold private data and want to compute some common function of their joint data, while keeping their private information secret from each other. In our problem, the users owns the private data and the server merely stores the information and performs processing on behalf of the user. We illustrate the overall system in Fig. 1.

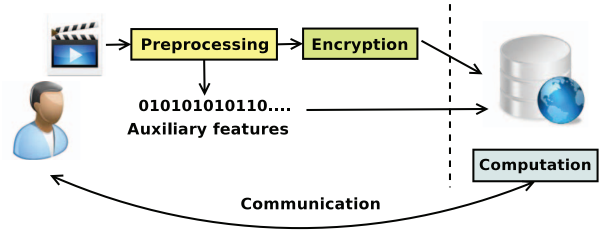

Fig. 1. System diagram for secure video processing

There are two parties in the system: the user who owns the private data and the server who stores the encrypted video and performs processing tasks. There are three key components in the system: preprocessing the video to obtain auxiliary information, encryption of the video, and computation in the encrypted domain by the server. Obtaining auxiliary information is often necessary to assist secure computation by the server, which would otherwise be highly computationally intensive or incur large communication overhead. The auxiliary features need to be properly secured to prevent information leak. Video encryption is an area with rich literature. Videos can be encrypted before or after compression [10], with different computational complexity and different level of protection from full encryption to partial encryption [11]. Since online servers typically have large computational power, a secure processing scheme should be designed to outsource as much work as possible to the server while minimizing the preprocessing task at the user-side. However, depending on the specific task, it may still require additional communication between the user and server during the execution of the task. Another design goal then would be to minimize any additional communication cost.

Overall, there are three aspects when evaluating a secure video processing technique:

- *Security*: this captures the protection level for the private video data, i.e., how much information is revealed from the encrypted video and its auxiliary features.
- *Performance*: this evaluates the accuracy of the processing task performed on the encrypted data as opposed to plaintext data.
- *Complexity*: this includes the computational complexity of preprocessing on the user-side and the communication cost during the processing task.

Good designs of secure video processing should achieve a good and flexible trade-off among the three aspects above.

3. VIDEO PROCESSING TASKS

In this section, we first discuss the problem of defining security for multimedia data, which is important for understanding the proper level of protection needed for online video management. Then we study some representative video processing tasks and discuss helpful existing tools and further research challenges for performing these tasks in a privacy-preserving fashion. It should be noted that video contains multiple modalities such as visual, audio, and text. Due to space limits, this paper focuses on utilizing visual data for different processing tasks.

3.1. Security Definition for Multimedia

To design a secure video processing technique, it is necessary to have a clear security objective and definition. By treating multimedia as ordinary data and applying cryptographic ciphers such as RSA and AES, information leak is minimized but such an approach is inefficient for practical video processing applications. For example, it can affect the compression efficiency and format compliance of video data. As such, there have been work on format compliant encryption [12] and partial encryption [11] for video, which are more efficient but may not provide the highest level of protection. For example, in partial encryption, a low resolution version of the video might be available.

As evidenced above, efficient and secure video processing techniques will inevitably have trade-off between efficiency and security. It is therefore important to quantify what is the proper protection level required for each specific application. Multimedia data is different from text data in the sense that it contains very rich information and almost infinite possibilities of signal representation. Video also contains both visual part and audio data, both of which require proper protection and can potentially be used to enable functionality in the encrypted domain. Furthermore, it is often the semantic content and visual appearance of the data rather than the exact signal values and statistics that we want to protect. In the system diagram shown in Fig. 1, we need to properly evaluate how much content information will be revealed from the encrypted data and auxiliary features. A quantitative study of the proper protection level and amount of content information for multimedia data can be a beneficial research direction for the secure signal processing community.

3.2. Video Search and Retrieval

We consider here the application where the user wants to search his/her private database using video queries while keeping the query and database content secret from the server. [8, 9] considered a similar problem of content-based search over encrypted image database. The idea is to encrypt visual features or search indexes from images in a distance preserving fashion, which allows the server to compare the similarity of encrypted images directly in the encrypted database without additional communication with the user. This is critical for large databases beyond a few hundreds entries, for which the computational cost and communication bandwidth involved using traditional cryptography techniques such as homomorphic encryption will be formidably expensive. The above mentioned techniques thus provide a practical solution for applications that do not require the highest level of security and cannot afford the resources, and the performance is very close to retrieval over plaintext database.

Video search and similarity comparison also relies on comparing visual features or summarizations, which can be extracted from key frames or group of frames. Therefore, the above techniques for approximate distance comparison between encrypted features can be effectively applied for video retrieval. In the preprocessing step on the user-side, the user performs visual feature extraction from the video. Then these visual features are properly encrypted in a distance-preserving fashion as in [8, 9] and stored on the server as auxiliary information. During search, visual features from the query image or video are extracted and encrypted before being sent to the server. The server compares the encrypted query features to encrypted features in the database and returns the encrypted videos with the highest similarity. For visual features extracted from key frames of the videos, similarity comparison can be performed over different subsequences of two videos and accumulated to get an overall similarity score.

3.3. Video Classification

Another interesting processing task is to automatically assign tags, such as beaches, portrait, indoor, to videos and classify the collection into different categories. Privacy-preserving video classification and annotation is desirable because it can better organize and present the private video collection for the users. Since there is a huge amount of public videos available online, one approach would be to let the server compare private videos with already labeled public videos and assign tags based on the most similar public videos. In contrast to video search and retrieval considered above, where the comparison is done between the encrypted features to obtain similarity ranking of videos, video classification here requires comparing private video with public data in a secure way. Depending on the security objective, the classification results can be made available or oblivious to the server.

One of the cryptography tools that can be used to compare encrypted information with public ones is homomorphic encryption. Homomorphic encryption allows the computation of addition or multiplication between encrypted values. Using the Pallier additive homomorphic encryption [13] as an example, the summation of two values can be computed by the product of their encrypted values and the result is also homomorphicly encrypted.

$$\mathrm{E}(x+y) = \mathrm{E}(x) \cdot \mathrm{E}(y). \tag{1}$$

Given this property, the inner-product between an encrypted vector and a public vector can be computed by the server as

$$\mathrm{E}(\mathbf{f} \cdot \mathbf{g}) = \Pi_{i=1}^{n} \mathrm{E}(f_i)^{g_i}. \tag{2}$$

Since the decryption key is kept secret from the server, the encrypted distances between $\mathbf{f}$ and all the public features $\mathbf{g}$ have to be sent back to the user for decryption. If the classification results should be oblivious to the server, the category information of all the public videos also need to be sent to the user, and the final correct category will be encrypted before being sent back to the server. Given the potentially huge size of public video data and large number of visual features, such a cryptography approach brings serious communication cost and makes it highly inefficient for practical use.

In order to reduce communication complexity, only a small set of the encrypted distances should be sent back to the user. Therefore, the server needs to be able to determine which public videos are more likely to be similar to the private video. Since storing the visual features without any protection reveals content information, while using a cryptography only approach brings high communication cost, one possible approach that can trade-off some security for improved efficiency would be to store a coarse representation of the visual features and allow the server to quickly ignore candidates that are unlikely to be similar. The locality sensitive hashing (LSH) technique [14] can be used here to give a coarse representation of the features. LSH essentially performs random projection of the feature onto random vectors as shown below.

$$h(\mathbf{f}) = \lfloor \frac{\mathbf{r} \cdot \mathbf{f}}{w} \rfloor. \tag{3}$$

Here, $\mathbf{r}$ is a Gaussian random vector and w is the quantization width. The feature points that are close-by in the original space will be similar after projection with a high probability. By using multiple hash functions, the coarse representation of a feature $\mathbf{f}$ will be $[h_1(\mathbf{f}), \cdots, h_k(\mathbf{f})]$. The server can compare only those public videos whose coarse features fall into the same quantization bucket as the private feature. By controlling the dimension k, different trade-off between security and communication complexity can be achieved. A larger k means more information about the private video is revealed, but also less public videos will fall into the same bucket, thus less communication cost. A smaller k will imply more videos in the same bucket, thus more uncertainty for the server to guess the content of the private video. An important research issue will be to quantify the information revealed from these coarse features, so that we can use the right level of protection to achieve the best possible efficiency.

3.4. Video Summarization

Video summarization is the task of extracting a set of salient images called key frames to represent the video content [15]. The extracted key frames can be used for fast browsing of the video and creating index for retrieval. Since private videos are stored in encrypted form on the server, browsing the video content requires real-time decryption. By restricting the decryption to only key frames, the real-time browsing of encrypted video can be made possible. There are different ways to extract key frames from videos. Sampling based method obtains key frames by randomly or uniformly sampling from the video. To apply sampling based key frame extraction, the video has to be encrypted in a format compliant way, so that the server can identify frame structures without decryption.

Sampling based extraction is simple but may not capture the dynamic content of the video. More advanced techniques use low level visual features to identify important frames. To perform content-based key frame extraction, we envision there can be two general approaches. One is to apply partial encryption of the video and make certain information available after encryption. For example, a partial encryption can scramble all the static content but leaves some selected motion vectors unencrypted. A limited number of motion vectors may not be sufficient to reconstruct the video content, but can be used by the server to extract key frames that contain large motion or abrupt transition. Another approach is to preprocess the video and identify key frames on the user side. The user then applies format compliant encryption on the video and sends the video along with the key frame numbers to the server. In this case, the key frame selection algorithm has to be very efficient to reduce computational cost on the user side. If the user wants more functionality, he can also extract auxiliary features, such as color statistics, salient point features, and then sends a coarse representation or distance-preserving encrypted version of these features to the server. The server can use the features to perform key frame extraction and even tasks such as face, object detection and tracking. Similar to the task of video classification, it is important to evaluate how much information is revealed through the partial encryption of video and storage of auxiliary features.

4. CONCLUSIONS

In this paper, we considered the application scenario of providing functionality over encrypted private videos stored online. We studied three representative processing tasks, namely, video search, classification, and summarization, and discussed related techniques and remaining challenges. Given the large size and rich information of video data, it is important to design highly efficient yet privacy-aware processing techniques. Success in this direction will depend on future research efforts to address the question of how to properly define security and the proper level of protection for multimedia data. We believe better understanding of the above question along with future advancement in cryptography can open up possibilities for a complete range of secure processing and editing operations for private videos stored online.

5. REFERENCES

[1] Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, and M. Barni, "Protection and retrieval of encrypted multimedia content: when cryptography meets signal processing," *EURASIP Journal on Information Security*, vol. 7, no. 2, pp. 1–20, 2007.

[2] M. Yiu, G. Ghinita, C. S. Jensen, and P. Kalnis, "Outsourcing search services on private spatial data," in *Proceedings of IEEE Int. Conf. on Data Engineering*, 2009, pp. 1140–1143.

[3] W. Wong, D. Cheung, B. Kao, and N. Mamoulis, "Secure kNN computation on encrypted databases," in *Proceedings of SIGMOD Intl. Conf. on Management of Data*, 2009, pp. 139–152.

[4] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, "Privacy-preserving face recognition," *Privacy Preserving Technologies, LNCS*, vol. 5672, pp. 235–253, 2009.

[5] A. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *Intl. Conf. on Information Security and Cryptology*, 2009.

[6] W. Jiang, M. Murugesan, C. Clifton, and L. Si, "Similar document detection with limited information disclosure," in *IEEE Intl. Conf. on Data Engineering*, 2008.

[7] Shuiming Ye, Ying Luo, Jian Zhao, and Sen-Ching S. Cheung, "Anonymous biometric access control," *EURASIP J. Inf. Secur.*, vol. 2009, January 2009.

[8] W. Lu, A. L. Varna, A. Swaminathan, and M. Wu, "Secure image retrieval through feature protection," in *IEEE Conf. on Acoustics, Speech and Signal Processing*, April 2009.

[9] W. Lu, A. Swaminathan, A. L. Varna, and M. Wu, "Enabling search over encrypted multimedia databases," in *SPIE/IS&T Media Forensics and Security*, Jan. 2009, pp. 7254–18.

[10] F. Liu and H. Koenig, "A survey of video encryption algorithms," *Computers & Security*, vol. 29, no. 1, pp. 3–15, 2010.

[11] A. Massoudi, F. Lefebvre, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater, "Overview on selective encryption of image and video: challenges and perspectives," *EURASIP Journal Information Security*, vol. 2008, pp. 1–18, 2008.

[12] Y. Mao and M. Wu, "A joint signal processing and cryptographic approach to multimedia encryption," *IEEE Trans. on Image Processing*, vol. 15, no. 7, pp. 2061–2075, 2006.

[13] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *EUROCRYPT'99: Proceedings of the Intl. Conf. on Theory and Application of Cryptographic Techniques*, 1999, pp. 223–238.

[14] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proceedings of the 25th Very Large Database (VLDB) Conference*, 1999, pp. 518–529.

[15] Ying Li, Shih-Hung Lee, Chia-Hung Yeh, and C.-C.J. Kuo, "Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 79–89, 2006.