Privacy-Preserving Moving Object Detection

Computer Science Tripos - Part II

Candidate Number

Department of Computer Science and Technology
University of Cambridge

Friday 13 May, 2022

Declaration of Originality

. . . DECLARATION OF ORIGINALITY . . .

Proforma

. . . PROFORMA . . .

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the modern world, computers have improved almost every aspect of our lives. Recently, home security has become the latest target of the technology revolution. Companies like Ring [9] and Eufy [4] offer IoT devices like doorbells and cameras to allow their customers to monitor their property 24/7. On top of traditional surveillance, these companies also provide software solutions to monitor the footage recorded by their devices and interpret it. For example, a doorbell may recognise who is at the front door and allow them to enter, or alert the user to the presence of a stranger if it doesn't. However, the computational intensity of these inferences means footage must be transferred from the devices to more powerful servers.

In order to preserve privacy, video is encrypted before it is sent to the server. However, the footage must be decrypted when the inference algorithms are executing. This is an immediate privacy concern. Having the ability to decrypt the footage exposes the opportunity for employees of these companies to access constant surveillance of peoples' homes. The possibilities for exploitation are endless. Malicious actors could use this information to monitor people's location, appraise their belongings, or use the contents of footage for extortion, to name a few. Homomorphic Encryption may provide a solution to this.

Homomorphic Encryption (henceforth HE) is a cryptographic method of encrypting data such that mathematical operations can be performed on encrypted data, or *ciphertext*, itself, rather than on the raw data, or *plaintext*. For example, consider the operation 3×5. In a traditional encryption scheme, the plain values 3 and 5 would be multiplied before encrypting the result. Using a homomorphic scheme, the 3 and 5 can be encrypted, and the ciphertexts multiplied so that when the ciphertext is decrypted, the plaintext is 15. An open question is, can this technique be scaled to more complex algorithms, like those required for surveillance?

More specifically, is it possible to extract the moving objects from a frame of HE video data? Moving object detection, also known as *foreground extraction* or *background subtraction*, is fundamental to modern surveillance systems. Detecting when, for example, somebody enters a property, allows the security systems to alert their owners, possibly pre-empting a break-in. To perform this analysis, the contents of a video must be modelled using a, usually probabilistic, function that allows significant changes in a pixels' value to be discerned. The difficulty of this arises when accounting for environmental changes that cause numerical variation, such as light levels when moving from day to night or different weather conditions causing objects to distort.

1.2 Related Work

The lack of privacy caused by constant surveillance is not a new concern. There have been many attempts at solving video inference in the encrypted domain, but none are without flaws. For example, in 2013, Chu et al. [3] proposed an encryption scheme that supports real-time moving object detection, but this was quickly shown to suffer from information leakage, leaving it vulnerable to chosen-plaintext attacks[1]. Similarly, in 2017, Lin et al. [7] proposed a different encryption scheme to achieve the same goal by only encrypting some of the bits in each pixel, but this is unprotected against steganographic[2] attacks. Therefore, while research has been able to solve the weaknesses in privacy, it is yet to offer a solution that also preserves security against adversaries directly attacking the encryption, making them useless to real-world applications.

Likewise, researchers have been investigating inference using HE for many years. In 2012, Graepel et al. [5] introduced machine learning in the HE domain. Dowlin et al. [5] built upon this when they developed the CryptoNets model for deep learning with HE in 2016. However, deep learning neural networks are considered overly complex for moving-object detection. Instead, GMMs are the most widely used technique for background modelling. There is much less research into this area of unsupervised learning within the HE domain. The best example appears to be when, in 2013, Pathak and Raj [8] proposed a HE implementation of a GMM for audio inference. But there does not seem to be any investigations linking HE and GMMs to video analysis.

It appears that the most prevailing explanation for this lack of research is HE's inapplicability to real-time applications, due to its high computational complexity. While this may be true now, it is important to acknowledge that advances in computing capability will reduce the relative difficulty of HE operations. Consequently, more insight into its applicability will become increasingly valuable, as suggested by the trend in the growing popularity of HE research. This dissertation attempts to offer some beginnings to this insight as it attempts to find the constraining limitations of current HE implementations with respect to surveillance.

1.3 Aims and Contributions

This dissertation documents the design and implementation of a potential solution to the questions posed in §1.1, while attempting to follow the constraints re-

[1] A *chosen-plaintext attack* is a scenario in which an adversary can encrypt plaintexts of their choosing, and analyse the corresponding ciphertext in an attempt to break the encryption.

[2] *Steganography* describes the technique of information hiding. Like cryptography, steganography attempts to prevent adversaries from reading messages. Unlike cryptography, where the existence of a message is known but its contents are not, steganography attempts to hide the message's existence.

stricting the aforementioned real-world systems. In particular, the contribution of the work is:

- The creation of a client-server application simulating the device-server stack utilised by existing products, allowing secure transmission of video data from client to server and back again after performing inference.
- The use of Microsoft's Secure Encrypted Arithmetic Library (SEAL) [6] to integrate the CKKS HE scheme [2] for encrypting videos while they are away from the client.
- The implementation of a series of algorithms for enabling private and plain inference of video data to extract moving objects by producing a mask that can be applied to videos in the clear by the client.
- An investigation of Gaussian Mixture Models (GMMs) for HE encrypted background subtraction, beginning with the work by Stauffer and Grimson [10] then moving into more general Expectation-Maximisation GMM algorithms [SOURCE?].
- As an extension, the fabrication of a CKKS implementation from scratch, called MeKKS, based on the Homomorphic Encryption for Arithmetic of Approximate Numbers paper by Cheon et al. [2, 1] to improve understanding of HE.
- A demonstration of the efficacy of the above solutions using timing, accuracy, and *(hopefully)* energy usage data to compare inference of CKKS and MeKKS solutions to plain videos, highlighting the advantages of the MeKKS implementation being targeted to this application over the more generic CKKS.

Chapter 2

Preparation

Could have sub sectons for dofferent theories, methods, etc.

2.1 Preliminaries

Processes related to . . .

2.1.1 Threat Model

2.1.2 Homomorphic Encryption

2.1.3 Moving Object Detection

Table 2.1: tablecaption

c1	c2	c3	c4	c5	c6	c7

2.2 Project Strategy

2.2.1 Requirements Analysis

2.2.2 Methodology

2.3 Starting Point

2.3.1 Knowledge and Experience

2.3.2 Use of Libraries and Datasets

2.3.3 Pre-Implementation Decisions

2.3.4 Computer Resources

2.4 Tools Used

2.4.1 Software

2.4.2 Licensing

2.5 Testing

Chapter 3

Implementation

What you actually did

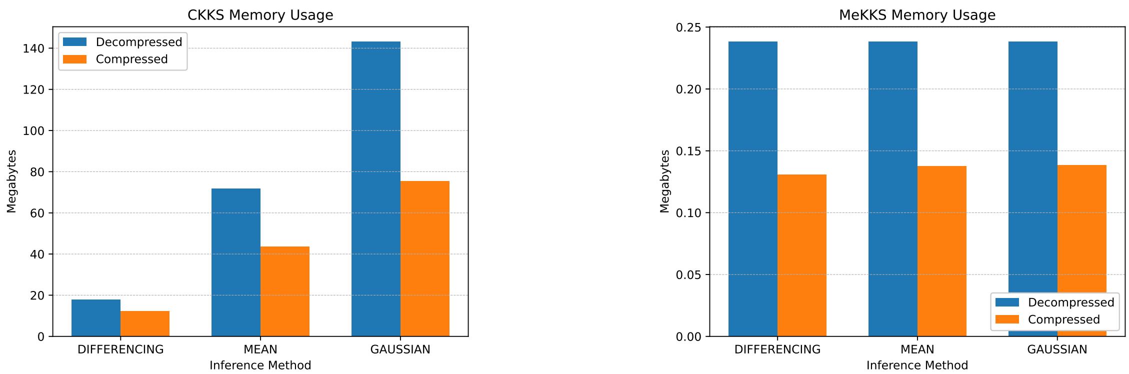
3.1 Implementation

3.2 Evaluation

Chapter 4

Evaluation

4.1 Results and Analyses



(a) CKKS Encryption Scheme *(b) MeKKS Encryption Scheme*

Figure 4.1: Memory Usage of Encrypted Frames

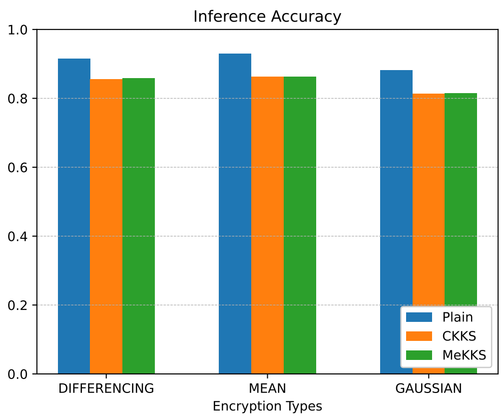
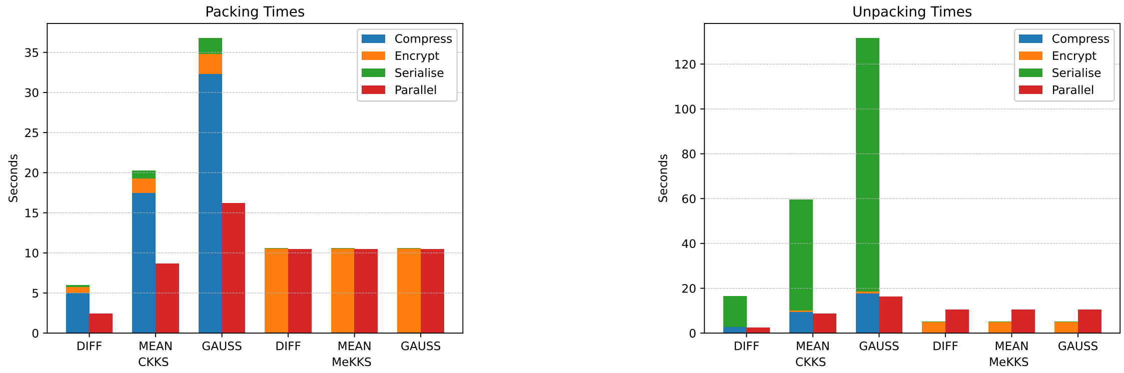
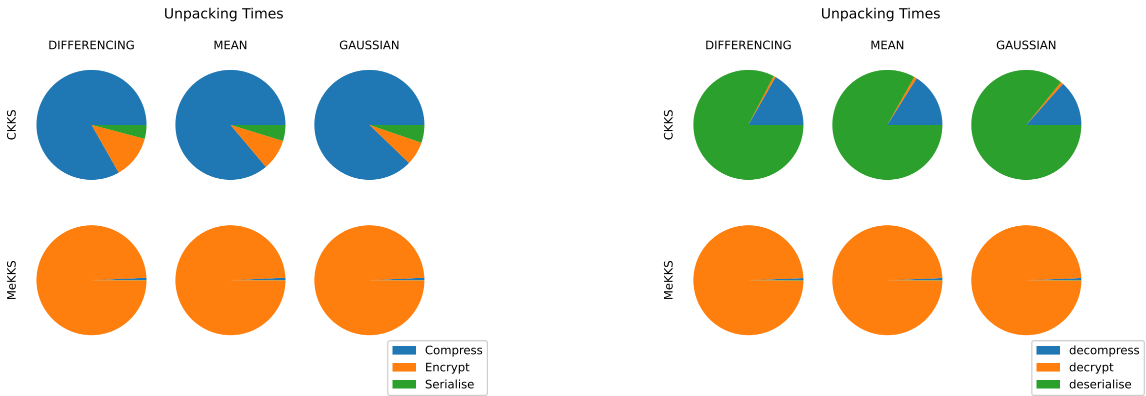


Figure 4.2: Inference Accuracy by Encryption Type



(a) Packing *(b) Unpacking*

Figure 4.3: Packing and Unpacking Times



(a) Packing *(b) Unpacking*

Figure 4.4: Operations Performed During Packing and Unpacking

4.2 Discussion

Did your findings support your hypothesis? Why? Why not?

Chapter 5

Conclusions

This Chapter concludes the thesis by summarizing the findings from the study, the contributions and possible limitations of the approach. It can also identify issues that were not solved, or new problems that came up during the work, and suggests possible directions going forward.

Bibliography

[1] Jung Hee Cheon, Kyoohyung Han, Andrey Kim, Miran Kim, and Yongsoo Song. “Bootstrapping for Approximate Homomorphic Encryption”. In: *Advances in Cryptology* (2018).

[2] Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. “Homomorphic Encryption for Arithmetic of Approximate Numbers”. In: *Lecture Notes in Computer Science* (2016).

[3] Kuan-Yu Chu, Yin-Hsi Kuo, and Winston H. Hsu. “Real-time privacy-preserving moving object detection in the cloud”. In: *Proceedings of the 21st ACM international conference on Multimedia* (2013).

[4] Eufy. *Eufy Homepage*. https://uk.eufylife.com/. est. 2016.

[5] Thore Graepel, Kristin Lauter, and Michael Naehrig. “ML Confidential: Machine Learning on Encrypted Data”. In: *The International Conference on Information Security and Cryptology* (2012).

[6] Kim Laine. “Simple Encrypted Arithmetic Library 2.3.1”. In: *Microsoft Research TechReport* (2017).

[7] Chih-Yang Lin, Kahlil Muchtar, Jia-Ying Lin, Yu-Hsien Sung, and Chia-Hung Yeh. “Moving object detection in the encrypted domain”. In: *Multimedia Tools and Applications* (2017).

[8] Manas A. Pathak and Bhiksha Raj. “Privacy-Preserving Speaker Verification and Identification Using Gaussian Mixture Models”. In: *IEEE* (2013).

[9] Ring. *Ring Homepage*. https://en-uk.ring.com/. est. 2012.

[10] Chris Stauffer and W.E.L Grimson. “Adaptive background mixture models for real-time tracking”. In: *IEEE* (1999).

Appendix A

Project Proposal