Project Summary

Proposal Title

PI Names
More PI Names

High-performance computing applications such as medical imaging, financial data processing, scientific data visualization and simulations are increasingly performed in distributed computing platforms like peer-to-peer or cloud networks. Many of these applications require the use of sensitive information and/or proprietary software that cannot be shared between different data owners. Encrypted-domain secure multiparty computation (SMC) techniques based on homomorphic encryption and garbled circuits are increasingly studied for such applications. Despite significant improvements over the past decade, these protocols require long security parameters, which render them unusable in most practical applications. Information-theoretic (IT) SMC like Shamirs secret-sharing can provide a more computationally-efficient platform as their security does not rely on the hardness of any computational problems. Our preliminary results indicate that Shamirs secret sharing is 100-10,000 times faster than garble circuit on performing basic arithmetic operations over encrypted data. On the other hand, there are a number of challenges in applying IT-SMC, including vulnerability to collusion attacks, handling of floating point operations, bandwidth expansion, and lack of software support. We propose to tackle these challenges through a wide range of computational tools that can make IT-SMC scalable to the most demanding high-performance computing applications over realistic distributed networks.

The **Intellectual Merits** of the project include (1) a novel evolutionary game-theoretic framework that deters collusion attacks using a retaliation mechanism and undercover computing agents; (2) low-complexity and low-latency cryptographic protocols to prevent subliminal communication among colluding agents; (3) provably-secure floating-point operations that are optimized for numerical accuracy, and (4) novel ways of compressing encrypted data for IT-SMC via side information.

The **Broader Impact** include (1) development of realistic high-performance computing applications including medical image rendering and privacy-protected data mining; (2) development of a new cyber-security undergraduate curriculum; (3) international collaboration that supports exchange of graduate students; (4) outreach program to rural and disadvantage communities in New York.

SaTC: Medium: Securing Sensitive Information in Cloud Computing - Project Description Pradeep K. Atrey (patrey@albany.edu)

1 Introduction

A huge amount of data (approximately 2000 terabytes every minute) is generated in many areas such as social media (e.g. Twitter, Facebook and YouTube), electronic surveillance (e.g. CCTV videos and phone call records), bioinformatics (e.g. DNA sequences), e-health (e.g. electronic health records including medical images), environment sensing (e.g. weather data) and business (e.g. stock market and bank transactions) every day. The storage and processing of such large-scale data poses a constant challenge to these organizations.

It is a common practice to outsource the large-scale data storage and analysis tasks to distributed high-performance computing servers such as cloud data centers (CDCs). Such solutions deliver highly scalable and virtualized resources to efficiently perform the required services. However, since these third-party service providers can often be untrustworthy, their use brings in serious security and privacy concerns [Nanavati et al. (2014)].

1.1 Motivating Scenarios

The following examples emphasize the need to consider security and privacy issues with CDCs:

- An adversary having access to medical data of patients can misuse it in a number ways. First, for economical benefits, the adversary may illegally sell the disease information of patients to other interested parties such as insurance companies. Second, for publicity, both the health information and the name of the admitting hospital of a prominent person may be leaked to the public and media [Taitsman et al. (2013)]. Third, medical image can be purposefully modified to provide misleading information to doctors [Anderson (1994)].

- Video surveillance footage stored on CDCs may consist of sensitive information and therefore may need to be protected. With the invention of modern video processing tools and techniques, it is easy to interpret the information contained in a video and mine it using computer algorithms to gain the usable information such as an individual's identity, current location, movements and time-stamps related to various events [Saini et al. (2014)], affecting the individual's privacy. Furthermore, an adversary can utilize, track and infer the behavioral patterns based on an individual's activities.

- Call centers record several hours of customer calls (audio data), most of which contain confidential information such as individuals' date of birth, address and credit card. In order to save cost, call centers often store data on CDCs [BBH Solutions (2014)]. A rogue or malicious employee within the call center or CDC may use this

confidential information to their own benefit. Therefore, the security of such sensitive audio records on CDCs is of utmost importance.

Due to these potential threats, there exists a number of federal laws, such as the *HIPPA* act by USA, the *PIPED* act by Canada, and the *Data Protection Act* by European countries, to protect the privacy of citizens.

To overcome the security and privacy issue, one can encrypt the data using traditional encryption methods such as Advanced Encryption Standard (AES) or Chaos-based Encryption before sending it to CDCs. There are two major limitations with the traditional key-based encryption methods: 1) they often suffer from single-point vulnerability meaning that their security depends upon the secrecy of the encryption key, and 2) they can provide only secured data storage but present the challenge of processing the data in the encrypted form.

1.2 Research Challenges and Goals

Encrypted-domain (ED) secure multiparty computation (SMC) techniques based on homomorphic encryption and garbled circuits are being increasingly studied for such purposes. Despite significant improvements over the past decade, these protocols require long security parameters, which render them unusable in most practical applications. Information-theoretic (IT) SMC like Shamir's secret-sharing (SSS) can provide a perfectly secure and more computationally-efficient platform as its security does not rely on the difficulty of any computational problems. In a (k, n)-SSS technique, secret data is broken into multiple arbitrary shares (n), out of which only a few shares ($k \leq n$) are required to reconstruct the secret. Our preliminary results indicate that SSS is 100-10,000 times faster than garble circuit when performing basic arithmetic operations over encrypted data [*Samson, could you please provide a reference here*]. However, there are a number of challenges in applying SSS, such as:

1. Vulnerability to collusion attacks: The (k, n)-SSS technique works under the assumption that k or more share holders do not collude with each other. However, it is difficult to guarantee the validity of such an assumption in many practical scenarios. Therefore, it is very important to devise a method to deal with such collusion while using the SSS-technique for ED multimedia data analytics.

2. Handling of floating point operations: The SSS technique works under a finite field, meaning that it deals with integer numbers only. However, in most multimedia data analytics tasks we deal with floating point numbers. Integrating floating point operations in SSS is quite challenging as it usually results data expansion and loss in accuracy.

3. Bandwidth expansion: SSS-based secure processing of data comes at the expense of data overhead, which needs to be minimized.

4. Loss in accuracy: Performing multimedia data analytics task operations in ED can result in some loss in accuracy compared to when these operations are undertaken in the plaintext domain (PD). The challenge is to minimize this loss.

5. Suitability for different analysis tasks in different domains: Although SSS-based ED data processing techniques are able to offer support for fundamental operations (i.e. addition, subtraction, multiplication and division) only, it remains a question whether a particular data analysis task can be suitably represented using these four fundamental operations.

The **Intellectual Merits** of the project include (1) a novel evolutionary game-theoretic framework that deters collusion attacks using a retaliation mechanism and undercover computing agents; (2) low-complexity and low-latency cryptographic protocols to prevent subliminal communication among colluding agents; (3) provably-secure floating-point operations that are optimized for numerical accuracy, (4) novel ways of compressing encrypted data for IT-SMC via side information, and (5) adaptability of the proposed IT-SMC in various large-scale data analytics applications such as multimedia surveillance, social media and medical imaging..

The **Broader Impacts** includes the following: (1) (from a technical perspective) this work will design and develop a realistic secure cloud-based high-performance computing framework applicable to different scenarios such as medical image rendering, privacy-protected social data mining, and privacy-aware multimedia surveillance; (2) (from a social perspective) the proposed project will provide people with a sense of security and privacy while using high computing platforms such as CDCs for storage and processing of their data collected via different sources in various application scenarios including surveillance, social media and medical imaging; (3) (from an educational perspective) it will help to develop a new cyber security and privacy curriculum at the undergraduate as well as the graduate level; (4) (from the perspective of workforce creation) through the proposed project, the PIs highly anticipate that the students graduating under this research project will possess the technical skills and knowledge needed by the cyber security and privacy industry in the US.

2 Background and Related Work

In the past decade, secure multimedia processing has been an emerging area of research. In December 2006, a three year-long project dedicated to this topic, called SPEED (Signal Processing in the Encrypted Domain or s.p.e.d.) [Erkin et al. (2006)] started in Europe. This was a major stepping stone in the field. The following year, there was a special issue published in EURASIP Journal on Information Security on SPEED [Piva and Katzenbeisser (2008)]. Thereafter, many special sessions on SPEED have been organised at various workshops [SPEED-2007 (2007), SPEED-2009 (2009) and conferences Barni and Piva (2008), Rane and Barni (2011)]. Some works [Lu et al. (2011), Puech et al. (2012)], including a keynote speech, were also presented at the various platforms [Lagendijk (2009)]. There are some surveys [Piva and Katzenbeisser (2008), Erkin et al. (2007), Prins et al. (2006)] that mainly summarize the details of applying cryptographic primitives to signal processing operations in ED. They consider various signal processing tasks as independent piece-wise operations and present an overview of the homomorphic mathematical details making s.p.e.d. plausible. There are some works [Fontaine and Galand (2007), Gentry

(2009), Lagendijk et al. (2013), Aguilar et al. (2013)] that present a review for the availability of the latest trends and techniques in homomorphic cryptosystems. Such works emphasize the provable security and privacy of the discussed cryptographic primitives, along with their ability to be utilized for s.p.e.d. and other cloud based secure processing frameworks.

The ability to manipulate multimedia data in ED is largely based on the following two assumptions:

1. *Homomorphic encryption:* Homomorphism is a transformation from one type of algebraic structure into another such that the structure is preserved. This means that for every manipulation of the original data, there is a corresponding manipulation of the transformed data. Furthermore, these manipulations can be classified as additive and multiplicative homomorphism. For example, $D(E(m_1) + E(m_2) mod O) = (m_1 + m_2) mod O$ and $D(E(m_1) \times E(m_2) mod O) = (m_1 \times m_2) mod O$ are examples of additive and multiplicative homomorphism, respectively.

   Many public key cryptosystems (PKCs) utilize particular homomorphic properties to carry out the processing of the signals in ED. The PKCs are mostly based on the difficulty of solving some computationally hard problems, for example, the ElGamal cryptosystem [Elgamal (1985)] - a discrete logarithm in finite field with large (prime) numbers, the RSA cryptosystem [Rivest et al. (1978)] - a factorization of large composite numbers, Paillier cryptosystem [Paillier (1999)] - deciding if a number is a n^{th} power of $\mathbb{Z}_N$ for a large enough composite N. A table stating homomorphism in several cryptosystems can be found in [Erkin et al. (2007)].

2. *Independent sample-wise processing:* Encryption is applied on individual signal samples independently. Erkin et al. (2007) state that although there are no measures applied to hide the temporal or spatial characteristics of the signal, the use of sophisticated encryption schemes that are semantically secure achieves this property automatically.

The authors in [Bianchi et al. (2009), Erkin et al. (2007)] and [Bianchi et al. (2008)], have realized that PKCs operate on very large algebraic structures. This further facilitated the use of a *probabilistic cryptosystem* [Goldwasser and Micali (1984)] instead of a deterministic cryptosystem, to ensure that for any two encrypted signals it is almost impossible to decide if they hide the same sample signal value. The importance of using a probabilistic cryptosystem for encryption lies in the fact that the encryption function is no longer one to one, but one to many and the decryption function is many to one. There is a huge expansion in the ciphertext space with a space determined by a random blinding factor (unknown to the decryption function). The probabilistic cryptosystems also retain their homomorphic property within the original ciphertext space. The chosen plaintext attack involves listing all possible plaintexts and their corresponding ciphertexts, which becomes computationally difficult for a sufficiently large size of the blinding space.

Recently, the SSS-based technique has become very popular among researchers working in the area of secure domain multimedia processing. The SSS technique was proposed by Shamir (1979)]and Blakley (1979). Many works thereafter emphasize the importance

of the additive and multiplicative homomorphic properties of SSS for sharing and reconstructing secret images [Islam et al. (2009), Chang et al. (2008)]. There are very few works [Upmanyu et al. (2009)] using SSS that involve direct processing of the encrypted secret data, along with the usual application of sharing and reconstructing the secret.

Upmanyu et al. (2009) proposed a Chinese remainder theorem based secret sharing method for change detection in surveillance videos. Their system shattered each original video frame (secret) into multiple shares by using a shatter function, involving the scaling of each video frame's pixels by a factor and adding a random noise to the resultant pixels under the modulo prime domain. However, this work has two shortcomings. First, the proposal is made only for carrying out integer addition and subtraction operations. Second, the authors themselves admit that their system is inefficient in performing division operations, as it may lead to choosing a prime number in such a way that the size of the modulo domain is increased more than required. However, an alternative approach is proposed by using an additional computation server where the merge function is applied to respective residues obtained from other independent servers and the division/comparison is performed in the real domain. Hence, there is a suggested requirement to secure the intermediate information against attacks. Furthermore, there is an additional communication overhead in sending the residues to the additional server and receiving the processed data back for this purpose.

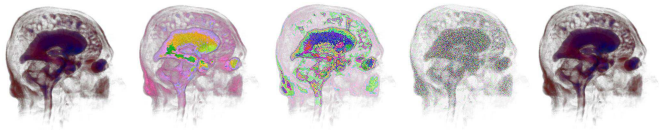
In addition to the above works, the Principle Investigators (PIs) have used the SSS-based technique for ED processing of images, as detailed in next section.

3 Recent Work of PIs Pertaining to the Proposal

The PIs have been contributing significantly to the area of security and privacy. They advanced the concept of encrypted domain processing of images by demonstrating the use of the SSS method. The SSS-based technique possesses the homomorphism property and allows ED processing of data over cloud. The PIs have applied this method in three different applications as described in the following subsections.

3.1 Secure cloud-based 3D medical data visualization [Mohanty et al. (2012), Mohanty et al. (2013b)]

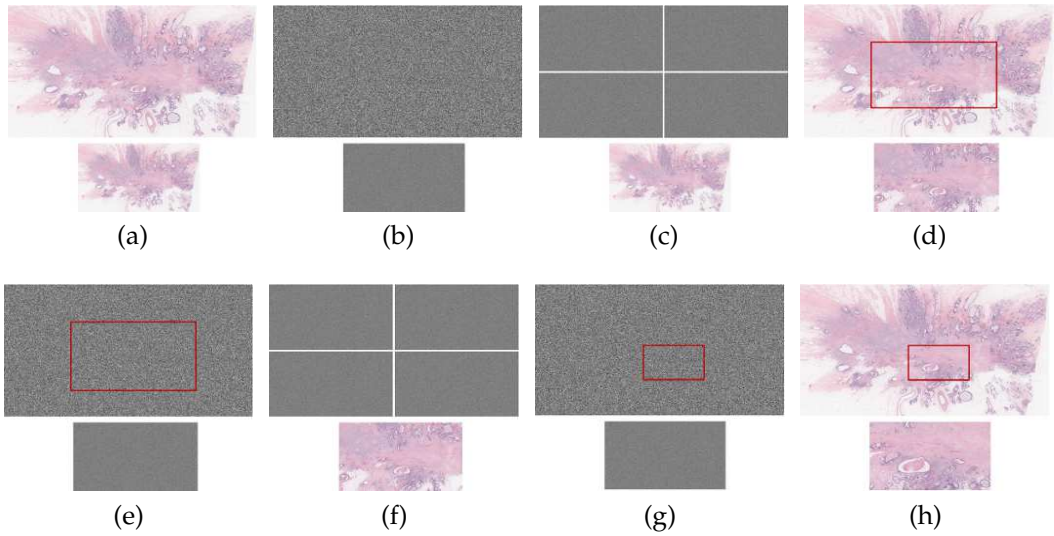
Using server-side rendering for remote data visualization may introduce significant latency, preventing interactive, real-time medical image analysis. To reduce interaction latency, cloud-based visualization services can be engaged to render data on CDCs that are geographically closer to the client. For visualization of medical data, however, such cloud-based architecture presents new security and privacy challenges. To overcome this security issue, the PIs proposed a new technique for cloud-based remote medical data visualization that protects the security of medical data in CDCs. To achieve this, the PIs integrated SSS with pre-classification volume ray-casting and proposed a novel secured volume ray-casting technique that hides the color coded information of the original medical data during the data rendering step at the data centers. Experiments and analyses



(a) (b) (c) (d) (e)

Figure 1: Secured volume ray-casting on Head volume data: (a) secret, (b, c, d) 1^{st} , 2^{nd} and 5^{th} shares, and (e) reconstructed.

show that our proposed method is highly secure, incurs insignificant computational overhead, and can be optimized to provide low visualization latency in the presence of high network bandwidth. Figure 1 shows the result of the pre-classification secured volume ray-casting on head MRI data.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2: Scaling and cropping of shadow images by datacenters: (a) is the required scaled image, (b) is a scaled shadow image, (c) is the recovered scaled secret image; (d) is the required cropped image, (e) is a cropped shadow image, (f) is the recovered cropped secret image; (g) is a zoomed shadow image (prepared by first scaling and then cropping) and (h) is the recovered zoomed secret image.

3.2 Scaling and cropping of encrypted images over cloud [Mohanty et al. (2013a)]

In this work, the PI Atrey's team proposed an image sharing scheme that allows the user to retrieve a scaled or cropped version of the secret image by operating directly on the shadow images, therefore reducing the amount of data sent from the data stores to the user. Results and analyses show that our scheme is highly secure, requires low computational cost, and supports a large number of scale factors with arbitrary crop. As shown in Figure 2, the CDC can, however, scale and crop its shadow image to produce a noise-like

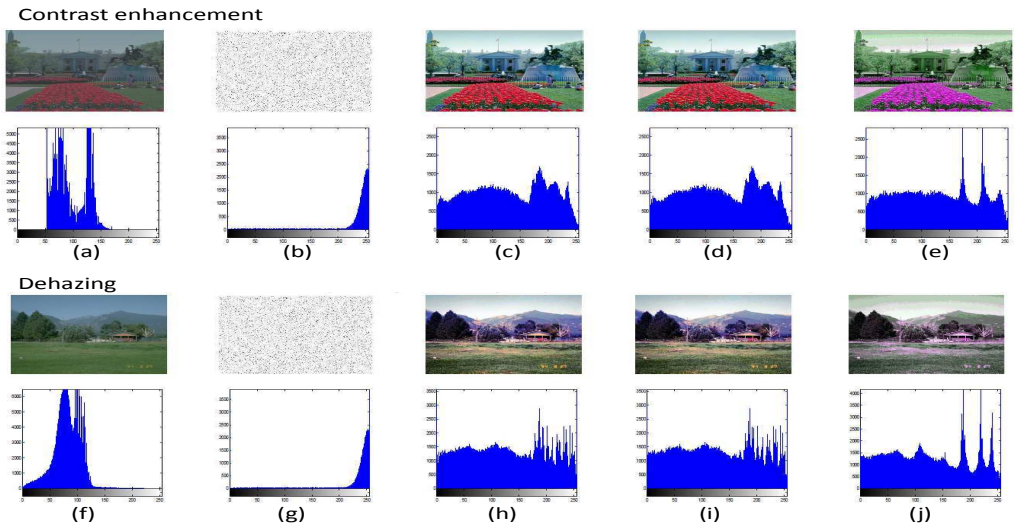


Figure 3: An illustration of the functioning of the proposed method for image enhancement in ED over cloud: (a-f) Secret image (b-g) Share image (c-h) Results in PD (d-i) Results in ED using Scheme I and (e-j) Results in ED using Scheme II

scaled/cropped image that is required by the user to recover the scaled/cropped secret image.

3.3 Cloud-based enhancement of quality of degraded images in ED [Lathey et al. (2013), Lathey and Atrey (2014)]

In this work, PI Atrey's team presented an improved, efficient and secure method (as compared to Mohanty et al. (2012), Mohanty et al. (2013a), Upmanyu et al. (2009)) for enhancing images in ED. We emphasized the feasibility of performing the division operation (including non-terminating decimal quotients) in ED. We showed the application of the proposed method for various image enhancement operations including noise removal, anti-aliasing, edge sharpening, contrast enhancement, and dehazing. The online demos of the work can also be viewed at https://sites.google.com/site/ankitaresearchdemos/home. The challenge of making the division operation compatible in both the real and modulo domains is addressed by adapting preprocessing schemes suitable for the image data. One of the schemes (Scheme I) led to an error-free information theoretically secure solution with some data overhead, whereas the other scheme (Scheme II) represented an error bound information theoretically secure solution, with no or constant data overhead. In Figure 3, we show the results of the two schemes for performing the contrast enhancement and dehazing operations in ED.

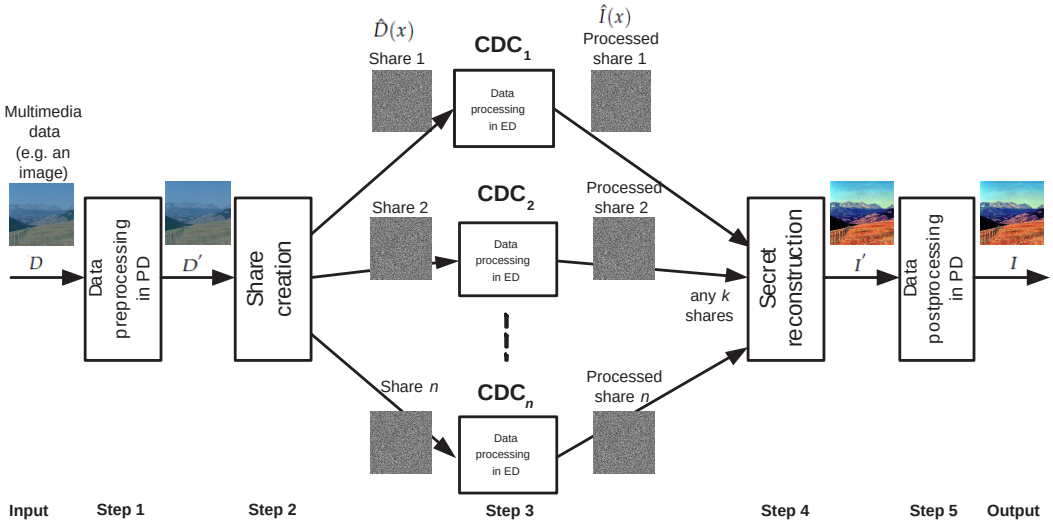


Figure 4: An illustration of the functioning of the proposed framework for image data processing in ED over cloud

4 Research Plan

4.1 Objective 1: A secure cloud-based multimedia processing framework

In this section, we briefly describe the general architecture of the proposed framework. As illustrated in Figure 4, it allows an authorized user to reconstruct the processed data from CDCs. The core idea is to use the (k, n)-SSS technique that divides the original raw data into N obfuscated shares. The n shares are outsourced to n different CDCs for storage and processing. In order to obtain the processed data or an inference, an authorized user first obtains the processed shares from any $(k \leq n)$ CDCs and then reconstructs the processed version of the data. In this scheme, the user cannot reconstruct the secret even if any $k-1$ shares are available.

In the following, we describe the specific steps that are undertaken in the proposed framework.

Step 1: Preprocessing the original secret data In order to enable the processing of data in ED, we may need to preprocess it. The preprocessing task usually depends on the type of operations to be performed on the share data. Mathematically, we represent it as:

$$D^{'} = \mathcal{P}_{pre}(D) \tag{1}$$

where D and $D^{'}$ are the original and preprocessed data, respectively, and $\mathcal{P}_{pre}$ is the suitable preprocessing operation.

Step 2: Creating obfuscated data shares using SSS Theoretically, the (k, n)-SSS technique involves sharing a secret (an integer value) among a set of n participants in such a

way that any $k \leq n$ participants can compute the secret, but a group of $k-1$ participant(s) cannot do so [Shamir (1979)]. Individual shares are completely arbitrary and do not reveal any information about the secret. The obfuscated data shares are created using the following equation, which is evaluated under the modulo prime, p:

$$\hat{D}(x) = (D^{'} + \sum_{j=1}^{k-1} a_j x^j) mod p \tag{2}$$

where a_j coefficients are randomly chosen from $\mathcal{Z}_p$. The first coefficient $D^{'}$ is the secret that is divided into shares $(x, \hat{D}(x))$ for different values of x. The shares $(x, \hat{D}(x)), 1 \leq x \leq n$ are sent to n different CDCs.

Step 3: Data processing operations on obfuscated shares Performing data processing operations on a data shares, $\hat{D}$(x), is represented as:

$$\hat{I}(x) = \mathcal{G}(\hat{D}(x)) \tag{3}$$

where $\mathcal{G}$ is the given data processing operation (for example, a quality enhancement operation).

Step 4: Obtaining the CDC-processed data In order to obtain the processed data, an authorized user accesses the k number of $\hat{I}$ data shares from any k CDCs and uses the Lagrange Interpolation formula to reconstruct the data $I^{'}$. Mathematically, we represent it as:

$$I^{'} = \mathcal{L}(\hat{I}(j), 1 \leq j \leq k) \tag{4}$$

where $\mathcal{L}$ is the Lagrange Interpolation function that provides the CDC-processed data $I^{'}$ in PD.

Step 5: Postprocessing the CDC-processed data In some cases we may need to perform some postprocessing in order to obtain the desired results. This is represented as:

$$I = \mathcal{P}_{post}(I^{'}) \tag{5}$$

where I is the final output data, and $\mathcal{P}_{post}$ is the postprocessing operation.

Our research objective [RO.1] is to design and develop this framework. Note that in our framework, **we intend to perform the majority of the data processing tasks in ED on CDCs** (in Step 3) and **aim to minimize the preprocessing and postprocessing operations** (in Step 1 and Step 5), which will mainly depend upon the task in Step 3.

4.2 Objective 2: A game-theoretic method to resist collusion attack

A novel game theoretic method will be proposed ... (Samson: please add your thoughts here).

4.3 Objective 3: Handling of floating point operations

The SSS technique works under a finite field, meaning that it deals with integer numbers only. However, in most multimedia data analytics tasks we deal with floating point numbers. Integrating floating point operations in SSS is quite challenging as it usually results data expansion and loss of security as well as accuracy of the processing task.

There are **two possible ways to solve this problem**. **First**, we can exclude the modular prime operation from SSS. However, due to the exclusion of the modular prime operation, the SSS technique no longer works in a finite field. Therefore, the modified SSS can lose IT security, and can even be unusable in certain situations. Moreover, due to the change from the finite to the infinite field, there could be data expansion. Security and data expansion analysis of such a modified SSS-based data processing framework needs further investigation. Hence, **our research objective [RO.3a] is to investigate suitable frameworks to have a tradeoff between security and data overhead**.

Second, we can address the incompatibility issue of a cryptosystem with a floating point number by converting the floating point numbers to fixed point numbers before applying SSS. For instance, a number D can be converted into an integer D' using the following equation:

$$D' = round(D, d) \times 10^d \tag{6}$$

where d is the number of decimal places up to which we want to round D. Clearly, this preprocessing operation (as mentioned in Step 1 of the proposed framework) could result in a rounding off error and/or data overhead. How much impact the rounding off error and the data overhead would have on the performance of the proposed framework needs further introspection. Therefore, **our research objective [RO.3b] is to investigate proper frameworks that can offer a tradeoff between loss in accuracy (of the processing task due to rounding off error) and data overhead**.

4.4 Objective 4: Bandwidth expansion

In the SSS-based framework, extra bandwidth is required to transmit n data shares to n CDCs and to transmit k data shares to the user. In many real time applications such as CDC-based medical image rendering where a user (e.g. a doctor) needs to see the rendered image immediately upon request, the bandwidth expansion could be a major concern.

Although the increase in bandwidth requirement is clearly a limitation of the SSS-based framework, **there are a number of ways to minimize this increase**. **First**, we can use a variant of SSS, called (l, k, n)-Ramp Secret Sharing (RSS), in which the size of the share can be reduced by a factor of l. In a (l, k, n)-RSS scheme, $l \leq k$ is the number of polynomial coefficients chosen from the secret data. It is well known that, unlike the SSS scheme, the RSS scheme is not IT secure, but it is still considered computationally secure. Hence, in many practical scenarios, RSS becomes a favorable choice. However, it remains to be examined whether the RSS scheme can support the homomorphic operations for a given multimedia processing task. **Our research objective [RO.4a] is to investigate proper frameworks to have a tradeoff between security and data expansion**.

Table 1: Secure data processing over cloud from the perspective of application areas, analysis tasks and type of media and features to be processed

Application area	Analysis tasks	Type of media and features used
Social media and networks	Hate posts detection	Text, audio, images, video and social features
	Personality assessment	Text, images, video, demographic and social features
	Cyber bullying detection	Text and social features
Multimedia surveillance	Data quality improvement	Image, video, audio
	Face detection and recognition	Image, video
	Suspicious event detection	Image, video, audio
	Data search and retrieval	Text, image, video, audio
Medical imaging	3D medical data rendering and visualization	3D data, image
	Medical data surface rendering	2D images
	Affine transformations	Image, video

Another option to reduce the bandwidth expansion can be choosing the values of k and n intelligently. Note that small values of k and n lead to less bandwidth requirement; however, it can affect the fault tolerance property (data integrity and availability) of the SSS-based framework. The (k, n)-SSS based framework ensures data integrity and provides protection against tampering of data that can occur at any of the CDCs. The $k < n$ condition results in $\binom{n}{k}$ different ways of reconstructing the secret data. Therefore, if any adversary tampers with one or more data shares, then the reconstructed data from the tampered shares will differ from each other. Hence, by comparing two or more reconstructed data, a user can detect any tampering with data shares. In addition, with a certain probability, a user can also infer the uncorrupted data out of all available reconstructed data (i.e. higher availability). If data shares of n_1 ($n_1 < n$ and $(n - n_1) > k$) CDCs have been corrupted, then, there exist $\binom{n-n_1}{k}$ number of other ways to reconstruct the uncorrupted data. Therefore, for $\binom{n-n_1}{k} > 1$, a user can infer the reconstructed data with a probability of $1 - \frac{1}{\binom{n-n_1}{k}}$. To this end, **our research objective [RO.4b] is to investigate a tradeoff between fault-tolerance capability and data expansion**.

4.5 Objective 5: Novel applications of our secure cloud-based data processing framework

In this section, we outline the novel applications that we propose to take on using our secure cloud-based data processing framework. Our goal is to target the following three application areas: 1) social media, 2) multimedia surveillance, and 3) medical imaging. Our goal is to find out how particular data analysis tasks in these areas can be represented in the four fundamental operations (i.e. addition, subtraction, multiplication and division) that are supported by the SSS-based framework. The tasks that we aim to perform in ED are provided in Table 1.

4.5.1 Secure analysis of social media and network data

The use of social media and networks such as Facebook, Youtube and Twitter has grown exponentially in recent years, and these social platforms continue to host a large amount of people's personal data ranging from their demographic information (age, location, gender, etc.) to their behavioral information (preferences, habits, companions, activities, etc.). This increasing amount of data poses storage and processing challenges. To overcome these challenges, such data is being outsourced to third party CDCs, where various analysis tasks such as personality assessment, epidemic spread detection, hate posts detection, spending behavior analysis, identification of important events and trends over social media, and cyber bullying detection are performed. Although third party CDCs usually offer high-end storage and computing facilities, they often lack users' trust and therefore security and privacy become the primary concerns. In this part of the research project, the PIs aim to investigate the following research questions:

- **[RO.5.1a]** Can we detect the instances of hate posts using the encrypted social media and network data over cloud?
- **[RO.5.1b]** Can we find a user's personality by processing his/her social media and network data in ED over cloud?
- **[RO.5.1c]** Can we detect cyber bullying instances by processing the social media and network data in ED over cloud?

The PI, Atrey, has already been working with his colleagues Prof. Alex 'Sandy' Pentland from Media Lab., Massachusetts Institute of Technology, Prof. Vivek K. Singh from Rutgers University and Prof. Abdulmotaleb El Saddik from University of Ottawa on personality assessment and cyber bullying projects. In this project, our focus will not be to invent any new algorithms for these tasks, but to discover the solutions that can be used to perform these tasks in ED, keeping security of data and privacy of people intact.

4.5.2 Secure processing of multimedia surveillance data

Due to various terrorist threats, multimedia surveillance systems have seen a world-wide growth in recent years. It is noteworthy that these systems capture a huge amount of data that must be: 1) stored so that it can be accessed in case of a post-incident investigation, and 2) processed for identification of suspicious people and their unlawful activities. In the era of cloud computing, such space demanding and high-end computing tasks are usually outsourced to *CDC*s. However, the involvement of the third-party servers raises privacy concerns, particularly in the context of surveillance that involves people. In this part of the research project, the PIs aim to investigate the following research questions:

- **[RO.5.2a]** Can we improve the quality of video and audio data in ED over cloud?
- **[RO.5.2b]** Can we perform face detection and recognition by processing video data in ED over cloud?

- **[RO.5.2c]** Can we perform suspicious event detection using image, video and audio data in ED over cloud?
- **[RO.5.2d]** Can we perform search and retrieval from the image, video and audio data in ED over cloud?

The PIs have been contributing in the surveillance area and they now aim to investigate the feasibility of existing automated surveillance methods in ED over cloud. Recently the PI, Atrey, has developed methods for performing image enhancement in ED over cloud, [Lathey et al. (2013), Lathey and Atrey (2014)] (as mentioned in Section 3.3) and through this project, one of the objectives of the PIs is to extend these methods to perform quality enhancement in ED on other types of data such as video and audio.

4.5.3 Secure medical data rendering and visualization

Remote data visualization allows medical experts located at a distance from the server to analyze images that are captured by host hospitals. It typically uses a *client-server* architecture, where the host hospital that captures the data acts as the server and the remote display device acts as the client. In this architecture, the 3D medical data rendering is usually performed on the server since it produces better quality images. Recently, it has been proposed that CDCs can be used as *dummy* servers for rendering [Dorn et al. (2011)]. Although cloud-based rendering has many advantages over conventional server-side rendering, security is a major issue when medical data rendering is carried out by the third party cloud providers. To address the security and privacy concerns, the PIs aim to find out the following:

- **[RO.5.3a]** Can we perform surface rendering (3D volume reconstruction from a set of 2D medical images) in ED over cloud?
- **[RO.5.3b]** Can we perform affine transformations (scaling, rotation, cropping and translation) on 2D compressed images and 3D compressed volume medical data in ED over cloud?
- **[RO.5.3c]** Can we detect abnormalities in the encrypted medical images over cloud?

In this area, the PI, Atrey and his colleagues have designed and developed techniques for secure cloud-based volume ray casting [Mohanty et al. (2012)] (as mentioned in Section 3.1) and secure scaling and cropping of large uncompressed images [Mohanty et al. (2013b)] (as mentioned in Section 3.2). Furthermore, the PIs aim to extend their work, among other things, for compressed medical data.

5 Integrated Research, Education and Outreach Activities

What are your potential contributions to developing human resources in science and engineering at postdoc, graduate, and undergrad levels? — NEED TO WRITE —

Table 2: Research project timeline

PHASES and TASKS	Human Resources	2015-16	2016-17	2017-18	2018-19
PHASE 1 - General framework					
RO.1: Design and development of secure cloud-based multimedia processing framework	PIs	X			
RO.2: Integration of game theoretic approach into the framework	PIs, 0.33 Postdoc, 1 PhD	X			
RO.3a & b: Identification of a suitable strategy for handling with floating point operations	PIs, 0.33 Postdoc, 1 MS	X	X		
RO.4a & b: Identification of a suitable approach for minimize the bandwidth expansion	PIs, 0.33 Postdoc, 1 MS	X	X		
PHASE 2 - Secure processing of social media and networks data					
RO.5.1a: Hate posts detection	PIs, 0.33 Postdoc, 1 PhD		X	X	
RO.5.1b: Personality assessment	PIs, 0.33 Postdoc, 1 PhD		X	X	
RO.5.1c: Cyber bullying detection	PIs, 0.33 Postdoc, 1 PhD		X	X	
PHASE 3 - Secure processing of multimedia surveillance data					
RO.5.2a: Video and audio data quality improvement	PIs, 2 MS			X	X
RO.5.2b: Face detection and recognition	PIs, 0.33 Postdoc, 1 PhD			X	X
RO.5.2c: Suspicious event detection	PIs, 0.33 Postdoc, 1 PhD			X	X
RO.5.2d: Data search and retrieval	PIs, 0.33 Postdoc, 1 PhD			X	X
PHASE 4 - Secure medical data rendering and visualization					
RO.5.3a: 3D medical data rendering and visualization	PIs, 0.33 Postdoc, 1 MS				X
RO.5.3b: Medical data surface rendering	PIs, 0.33 Postdoc, 0.5 PhD				X
RO.5.3b: Affine transformations	PIs, 0.33 Postdoc, 0.5 PhD				X

6 Research Schedule and Project Milestones

The timeline for the project in terms of phases and tasks is shown in Table 2. It indicates the anticipated major research milestones and the personnel involved for each year during the four-year planned duration for this project.

7 Project Management and Roles of the Participating Researchers

The management of the project will be coordinated among the PIs, drawing from each otter's expertise and experience. While the PI, Atrey, has extensive experience in designing and developing novel methods for performing multimedia (text, image, video and audio) analysis tasks applicable to surveillance and social media areas, the PI, Cheung, has largely contributed in security and privacy areas. Recently both PIs have gained interest in secure and privacy issues in processing multimedia data over cloud and have found encouraging results from their preliminary works. This has given rise to a natural tie-up between the two PIs for this project.

The design and development of the general framework as well as well as secure data analysis tasks would be shared between the two universities. Besides regular email communication, a monthly meeting will serve to bring the teams from UAlbany and UKentucky together for communication and coordination. This meeting will take place via videoconferencing and will enable basic co-ordination for development and integration,

joint learning for continued education in the domain, discussion of completed work, planning for future work and critical feedback for improvement. In addition, face-to-face meetings between PIs and other personnel will be planned twice a year. Further, one-on-one conversations via phone, e-mail and web video conferencing will help coordination of specific tasks across the two sites. Co-located team members at each site will meet every week for better coordination.

We will setup a password protected web server for sharing information and resources among the team. This would include collected data, documents produced by the team, team calendar, relevant articles, and source code for data analysis and developed software. Such mechanisms are already in place and have proven useful in the past.

8 Results from Prior Research Projects

The PI, Pradeep K. Atrey, has had no NSF support in the past five years.

References

Aguilar, M., C. Xlim, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, 2013: Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain. *IEEE Signal Processing Magazine*, **30(2)**, 108–117.

Anderson, C., 1994: Easy-to-alter digital images raise fears of tampering. *Science*, **263**, 317–318.

Barni, M. and A. Piva, 2008: Special issue on signal processing in the encrypted domain. In *EURASIP European Signal Processing Conference*, Lausanne, Switzerland.

BBH Solutions, 2014: Call center cloud integration. Http://www.bbhinc.com/clients/detail/call-center-cloud-integration/.

Bianchi, T., A. Piva, and M. Barni, 2008: Efficient pointwise and blockwise encrypted operations. In *Proceedings of* 10^{th} *ACM Workshop on Multimedia and Security*, Oxford, UK, pp. 85–90.

Bianchi, T., T. Veugen, A. Piva, and M. Barni, 2009: Processing in the encrypted domain using a composite signal representation: PROS and CONS. In *Proceedings of* 1^{st} *IEEE International Workshop on Information Forensics and Security*, London, UK, pp. 176–180.

Blakley, G. R., 1979: Safeguarding cryptographic keys. In *Proceedings of IEEE National Computer Conference*, New York, USA, p. 313.

Chang, C. C., C. C. Lin, C. H. Lin, and Y. H. Chen, 2008: A novel secret image sharing scheme in color images using small shadow images. *Elsevier Information Sciences*, **178(11)**, 2433–2447.

Dorn, K., V. Ukis, and T. Friese, 2011: A cloud-deployed 3D medical imaging system with dynamically optimized scalability and cloud costs. *Soft. Eng. and Adv. Appl., Euromicro Conf.*, **0**, 155–158.

Elgamal, T., 1985: A public key cryptosystem and a signature scheme based on discrete logarithms. In *Proceedings of* 4^{th} *Annual International Cryptology Conference*, Springer, Santa Barbara, USA, vol. 196 of *Lecture Notes in Computer Science*, pp. 10–18.

Erkin, Z., A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, and M. Barni, 2006: SPEED project. Http://www.speedproject.eu/.

—, 2007: Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing. *EURASIP Journal on Information Security*, **2007(78943)**.

Fontaine, C. and F. Galand, 2007: A survey of homomorphic encryption for nonspecialists. *EURASIP Journal on Information Security*, **1**, 1–15.

Gentry, C., 2009: *A fully homomorphic encryption scheme*. Ph.D. thesis, Stanford University.

Goldwasser, S. and S. Micali, 1984: Probabilistic encryption. *Elsevier Journal of Computer and System Sciences*, **28(2)**, 270–299.

Islam, N., W. Puech, and R. Brouzet, 2009: A homomorphic method for sharing secret images. In 8^{th} *International Workshop on Digital Watermarking*, Springer, Guildford, UK, vol. 5703 of *Lecture Notes in Computer Science*, pp. 121–135.

Lagendijk, R. L., 2009: Secure signal processing : Merging the worlds of signal processing and cryptography. *IEEE Multimedia Signal Processing*, keynote speech.

Lagendijk, R. L., Z. Erkin, and M. Barni, 2013: Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation. *IEEE Signal Processing Magazine*, **30(1)**, 82–105.

Lathey, A. and P. K. Atrey, 2014: Secure image enhancement over cloud. *ACM Trans. on Multimedia Computing, Communications and Applications*, (accepted).

Lathey, A., P. K. Atrey, and N. Joshi, 2013: Homomorphic low pass filtering over cloud. In *Proceedings of* 7^{th} *IEEE International Conference on Semantic Computing*, Irvine, USA, pp. 310–313.

Lu, W., A. Varna, and M. Wu, 2011: Secure video processing: Problems and challenges. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, pp. 5856–5859.

Mohanty, M., P. K. Atrey, and W.-T. Ooi, 2012: Secure medical data visualization over cloud. In *Proceedings of ACM International Conference on Multimedia*, Nara, Japan, pp. 1105–1108.

Mohanty, M., W.-T. Ooi, and P. K. Atrey, 2013a: Scale me, crop me, know me not: Supporting scaling and cropping in secret image sharing. In *Proceedings of IEEE International Conference on Multimedia and Expo*, San Jose, CA, USA, pp. 1–6.

—, 2013b: Secure cloud-based volume ray-casting. In *Proceedings of IEEE International Conference on Cloud Computing Technology and Services*, Bristol, UK, pp. 531–538.

Nanavati, M., P. Colp, B. Aiello, and A. Warfield, 2014: Easy-to-alter digital images raise fears of tampering. *Communications of the ACM*, **57(5)**, 70–79.

Paillier, P., 1999: Public-key cryptosystems based on composite degree residuosity classes. In *Theory and Application of Cryptographic Techniques*, Springer, Prague, Czech Republic, vol. 1592 of *Lecture Notes in Computer Science*, pp. 223–238.

Piva, A. and S. Katzenbeisser, 2008: Special issue on signal processing in the encrypted domain. *EURASIP Journal on Information Security(eds.)*, **2007**.

Prins, J., Z. Erkin, and R. L. Lagendijk, 2006: Literature study: Signal processing in the encrypted domain. Tech. rep., Information and Communication Theory Group, Delft Univerity of Technology.

Puech, W., Z. Erkin, M. Barni, S. Rane, and R. L. Lagendijk, 2012: Emerging cryptographic challenges in image and video processing. In *Proceedings of* 19^{th} *IEEE International Conference on Image Processing*, Orlando, USA, pp. 2629–2632.

Rane, S. and M. Barni, 2011: Special session on secure signal processing. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, pp. 5848–5871.

Rivest, R. L., A. Shamir, and L. Adleman, 1978: A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, **21(2)**, 120–126.

Saini, M., P. K. Atrey, S. Mehrotra, and M. S. Kankanhalli, 2014: W^3-Privacy: Understanding what, when, and where inference channels in multi-camera surveillance video. *Springer International Journal of Multimedia Tools and Applications*, **68(1)**, 135–158.

Shamir, A., 1979: How to share a secret. *Communications of ACM*, **22(11)**, 612–613.

SPEED-2007, 2007: Workshop at the european symposium on research in computer security.

SPEED-2009, 2009: International workshop on signal processing in the encrypted domain.

Taitsman, J. K., C. M. Grimm, and S. Agrawal, 2013: Protecting patient privacy and data security. *The New England Journal of Medicine*, **368**, 977–979.

Upmanyu, M., A. M. Namboodiri, K. Srinathan, and C. V. Jawahar, 2009: Efficient privacy preserving video surveillance. In *Proceedings of* 12^{th} *IEEE International Conference on Computer Vision*, Kyoto, Japan, pp. 1639–1646.

BUDGET JUSTIFICATION