Faisal Harbi CSCI 4243 Final SBIR Grant

Overview

The purpose of this project is to preserve data privacy while still allowing key aspects about the information to be verified. The project aims to accomplish this by utilizing homomorphic encryption technology. By applying basic arithmetic operations to the encrypted data, key factors such as the minimum value can be revealed. The ultimate goal is to integrate this technology into a software as a service (SaaS) platform in which individuals can register for and use.

The objectives of this project are to develop the encryption algorithm for the outlined application, provide a secure platform that allows a user to apply the encryption on their private data and send it, and to build the interface for access to this technology. These objectives will involve extensive research and development within the fields of security and cloud-computing in order to deliver a service that offers data privacy on demand.

Intellectual Merit

This Small Business Innovation Research Phase 1 project will provide research on the use of homomorphic encryption as a data privacy preservation algorithm for instances where crucial information needs to be verified, but not entirely revealed. Development of this project will involve researching and analyzing the encryption algorithm in order to utilize it for data specific scenarios. The exploration of this algorithm has the potential of discovering many data privacy preserving applications for homomorphic encryption.

There are several technical challenges in further developing and applying such an encryption method for practical use. Although the encryption technique is malleable and can be used for a wide variety of cases, designing it as a general-purpose algorithm that can be applied to several data scenarios at once is difficult. This challenge in generalizing the encryption algorithm could potentially limit the variety of uses for it.

Another hurdle in developing this project is securing the platform in which the algorithm itself is stored and accessed by the public. As cybersecurity attacks become more prevalent and sophisticated, ensuring the safety and integrity of a user’s data within the platform will become an ever-evolving routine.

Broader/Commercial Impact

The economic impact of this project is immense. Many individuals that are reluctant to share personal information with businesses for fear of data misuse or other nefarious activities will have the ability to keep their data private while still providing businesses the key verifying facts that they need. This new security platform has the potential to alleviate many data privacy concerns among individuals during business transactions and ultimately facilitate economic activity.

The benefits of this technology for both individuals and businesses could span across many industries. One sector that could benefit from this project is the financial industry. Banks routinely request verification of assets for loan applications in order to verify that a borrower holds at least a certain level of funds in their account. While the bank needs to verify that the balance meets their requirement, they do not necessarily need to know the exact amount. This technology would allow the bank to verify that the balance meets a minimum requirement without exposing the borrower’s actual balance.

This project could also have a positive impact on the real estate rental industry. Landlords and management companies usually need to verify a tenant's income before they can approve a lease application. Utilizing this technology, a tenant would be able to prove that their income meets the minimum requirement without exposing their actual income.

Elevator Pitch

This unique technology is directly addressing the needs of customers who apprehensively and unnecessarily share their private data with institutions that only need to verify that it meets a certain requirement. The proposed project will provide a direct solution for those that wish to preserve the privacy of their data, but still allow another party to verify certain aspects of it. In addition to these customers, this product is also designed for those that entirely refuse to conduct business with institutions that require verification of personal information because they simply do not want to reveal their private data. This technology will allow these users to engage with such businesses without compromising their private information.

The proposed project will provide an incredibly valuable security feature for customers who electronically share their private data. Rather than revealing their personal information, customers will have the power to keep their sensitive data private while still providing proof that it meets a particular requirement. This powerful tool will also allow those that avoid business transactions with companies that require private data verification the ability to engage with these businesses. This will create a wave of economic opportunities for these individuals by allowing them to confidently do business with a variety of institutions.

The banking sector is one area where customers can benefit from this technology. Customers that either refused to do business with banks because they required verification of their assets or apprehensively revealed their bank statements will now have the opportunity to complete the qualification process without revealing all their personal information.

The most important benefit that this technology offers to its customers is privacy. The ability to provide verification of private data without actually revealing that personal information is a revolutionary tool that has yet to be commercialized.

This innovative technology utilizes homomorphic encryption to keep data private while allowing basic arithmetic operations to be performed on it. By applying mathematical procedures to this encrypted data, key aspects can be determined, such as whether the value is greater or less than a certain amount. The encryption algorithm will be available as an internet based service, allowing customers to sign up and pay whenever they need to share their private data with another party.

While homomorphic encryption is not new technology, the concept of applying it to a variety of private data verification transactions and making it available as a SaaS platform is a new and unique application of its powerful facility. Commercializing and bundling this encryption technology into a secure verification tool is not only unique, but it is also revolutionary to the private data verification domain.

Commercial Opportunity and Social Impact

The market for our product encompasses many business sectors and has far reaching implications. The primary goal is to address all instances in which private numerical data needs to be shared with another party for verification purposes. This objective is specifically aimed at scenarios in which the data doesn't need to be revealed, but rather an aspect of it needs to be verified. For example, another party might need to confirm that a value is at least or no greater than a certain amount. In these cases, the actual value is not as important as whether it meets a particular requirement or not.

The primary markets that rely on this type of verification process includes the financial and real estate industries. These industries usually need to verify that a customer's income or assets is at least a certain amount for qualification purposes. Therefore, our product aims to address these specific industries by providing a solution for the application process of their business transactions. These transactions include applications for mortgages, residential/commercial leases, business loans, and any other application that requires personal information such as income or assets to be verified for guideline requirements.

The economic activity within these industries is immense and, for the most part, collectively stable. In the real estate industry, over 5 million homes will be sold this year alone and most of those homeowners will require a mortgage (*Total home sales in the United States from 2011 to 2016 (in millions), n.d.*). Almost every mortgage application will require that the borrower provide income/asset documents in order to prove that they are financially qualified for the loan amount that they are requesting and that they have the means to pay their monthly mortgage payment.

Another massive sector within the real estate industry is the residential rental market. Over 2,600 new lease applications are signed every day (*Rental Statistics (Rental Clock), n.d.*). Nearly every one of those lease applications will require paystubs, tax returns or other income documents from the lessee. The landlord requests proof of income in order to verify that the renter makes enough to afford the monthly rent.

The customers for our innovative product are individuals who need to complete the types of applications that are outlined above and are at the same time very uneasy about sharing private documents such as tax returns and bank statements with other parties. Millions of consumers across the U.S. are very conscious about their personal data and yet most will have no choice but to share that private data with businesses during crucial transactions such as when applying for a mortgage or lease. Our product aims to address the privacy concerns of these consumers during these types of transactions by giving them the opportunity to use our application on a per verification basis. A verification will consist of any number of documents being uploaded to our database and shared with one other party for a single transaction. The application will allow the consumer to hide numerical data such as their income or account balance by encrypting it. This encrypted balance will be hidden from the party that needs to verify that it meets their minimum requirement, but the application will allow them to perform basic mathematical operations on the value in order to answer basic verification questions such as whether or not the value is at least a certain amount.

This unique product has no competition. There are currently no technologies available that offer anything remotely similar. Consumers have no choice but to simply reveal their personal information when submitting applications that require income or asset verification. Therefore, this innovative product will revolutionize the way applications are submitted within the financial and real estate industries.

The major risk in introducing our product to the market is acceptance by the institutions that require personal information verification. Businesses such as banks and mortgage companies set the guidelines on their application process, not the consumer. Acceptance and demand from the consumer alone will not be enough to successfully commercialize this product. It will take confidence in our technology and an openness to a new verification procedure from institutions that have been processing applications the same way for decades to make our product completely successful.

Our plan in successfully commercializing this product begins by making this product available online for personal use. Transactions may include verifying your income with a private landlord willing to use the technology for a lease agreement or sharing your bank statements with some unconventional party that is open to the idea of this technology. After receiving positive reviews from customers, we will start approaching large financial and real estate institutions and attempt to convince them to accept our verification process as a legitimate alternative to their standard application procedures. This would allow their customers to have the option of using our technology when submitting personal documents to them. By allowing their customers to use a verification process that keeps private data hidden, financial institutions will have a convincing selling point for apprehensive consumers who place a high value on data privacy.

This product is aimed at addressing consumer confidence when sharing personal information and data security. Many customers across the country are becoming increasingly aware of security issues that are arising with the advance of technology. Sharing personal information with a business is a security risk, regardless of how reputable that company is. There is always the possibility of unauthorized data access. Keeping a customer’s private data hidden will mitigate the effects of a security breach and give more confidence to consumers. In addition, giving consumers the option to hide personal information such as income or assets will make the private data verification process much more comfortable for customers.

This newfound confidence in sharing personal information with businesses will facilitate economic activity across the entire spectrum. A consumer’s confidence in sharing income or asset documents for a business loan will spawn new commercial activity and have downstream effects that will impact all industries. Apprehensive individuals that refuse to share personal documents and are forced to fund transactions such as home and auto purchases will have the confidence they need to provide their personal documents and obtain financing, maximizing leverage. Security breaches will have reduced effects, as information stored in databases will have less private information.

Developed properly and secured appropriately, this technology will only have a positive impact on the economy. However, without the proper security measures in place, this product could have adverse effects that would impact every industry. Modifying the data being verified by another party is the biggest concern. If the data being verified was fabricated, applications or transactions would be processed based on false data. The impact on the financial and real estate industries would be as grave as the 2008 financial crisis if this technology becomes the cornerstone of the application process within those sectors. Since the wellbeing of our economy relies on the security of this technology, fortifying this product is of paramount concern.

The most vulnerable area of our technology is the algorithm that encrypts the data intended to be kept private, as access and modification of this would allow the alteration of the data itself. We will take extensive measures to specifically ensure that this encryption process is done on a secure server. We will also have a mechanism that detects unauthorized access to secure areas of the application, which will allow us to shut down the entire platform in cases where a security breach has occurred. This will prevent transactions from being processed if there is a chance that the encryption algorithm was changed in any way. These measures will ensure a safe application that can be used confidently for any transaction.

Technical Discussion and R&D Plan

This project involves many highly technical features that will present many challenges over the course of its development. Security will be a major concern, as the primary goal of this product is to preserve data privacy. Although sensitive data will be stored in a secured database, ensuring that data integrity is maintained at all times is still a priority. In order to address this security concern, the highly sensitive data, such as passwords and bank balances, must be processed with additional security measures before being stored. This involves encrypting the data before it is stored in the database. This will mitigate the effects of a data breach.

Another technical challenge will involve the image processing API of the application. Image processing is still a developing technology that does not always provide 100% accuracy. Since the success of the product depends on its ability to accurately read and encrypt vales from documents on a consistent basis, the image processing API must provide absolute precision. Misinterpreted values from a bank statement would provide erroneous data that could mistakenly inform banks that a consumer has either enough funds or insufficient assets. In order to address this issue, there will be a substantial amount of effort invested in researching the most advanced image processing web services and testing their ability to provide accurate and reliable results.

Technical challenges are not the only obstacles that this projects needs to overcome. Both the banks and the consumers will need to trust and understand the benefits of using the proposed technology. Consumers that value data privacy and have the desire to reduce the amount of personal information that is shared with other parties will be the primary target for this product. However, they need to have confidence that the technology is trustworthy and reliable. Banks will need to trust the system as well in order to integrate this technology into their business processes. Providing quality assurance, rigorous testing results and educating both the consumer and bank on the technology will be key in the ultimate success of this project.

The focus of the project during the initial phase will involve addressing the security and image processing technologies of the application. Upon successfully implementing solutions to these technological challenges, thorough testing will be performed in order to verify accuracy, reliability and security. Gaining customer and bank confidence will be addressed upon the completion of the final phase of the project.

The project will consist of a web application for both users and financial institutions to access. Users will have the option to upload financial asset summary documents, which will ultimately be shared with the financial institution that needs to verify that their asset balance meets their requirements. Documents that are uploaded by the user will be processed using an OCR (optical character recognition) API. The API will scan the uploaded documents and return an HTML string containing all the read characters within the document. This string will be parsed using a regular expression algorithm, which will extract balance values and store them in a text file. The extracted values within the string will then be replaced with arbitrary characters in order to protect the sensitive data on the consumer's documents. The parsed string will ultimately be stored in a database for both the consumer and bank to access. The text file containing all the extracted values will be processed using an encryption algorithm, which will encrypt all the values and sum them. The encrypted sum will ultimately be stored in the database.

Once a document has been uploaded and processed. The consumer will enter the threshold, which is the amount that needs to be verified by the bank, and the institution that the documents are being shared with. This information is all that is needed in order to submit a complete verification transaction, which will subsequently need action from the financial institution. The new verification transaction will be saved in the database so that the bank can log in to their account and view the new submission.

The bank will need to respond to the verification submitted by the consumer by viewing the financial documents uploaded and deciding whether they are acceptable or not. If the documents are satisfactory, the bank will accept the verification. Upon accepting the verification, the encrypted value will then be sent to the encryption algorithm. The encryption algorithm will subtract the threshold value from the summed encrypted value extracted from the documents and return the difference. This difference will determine whether the customer has sufficient funds in their account based on its positive or negative value. The final result of the verification procedure will be saved in the database and can be accessed by both the bank and consumer. This process allows a financial institution to verify a consumer's minimum asset balance without revealing the actual balance of their accounts.

The most important technical factors of this project that need to be addressed during the initial phase of this project are the image processing algorithms and the security of the information stored. This will involve research into all the existing image processing technologies available in order to determine which solution will provide our project with the accuracy and consistently it depends on. Each possible solution will be tested against many different documents from various financial institutions so that a realistic assessment can be made.

Once the technical challenges are addressed, the project's target market and commercial opportunity will be examined carefully. This order is important. Investing significant time and resources into developing reliable and effective technology will provide a product that can be accurately reviewed for commercial practicality. From this step, surveys can be conducted to explore the opinions of the proposed project. Consumer responses on the technology and whether they would consider using it will be gathered and reviewed. In addition, financial institutions will be surveyed as well. It is important that both the consumers and organizations that they do business with are equally receptive to the technology. Customers must want to use the product and businesses must be willing to incorporate the technology into their current commercial processes.

This project must implement an efficient and reliable image processing algorithm in order to be able to provide a viable product to the market. As stated previously, the success of this product relies on its ability to accurately process documents. Therefore, finding and implementing an effective OCR API will be a major achievement. Security is also a paramount concern for this project. Securing the database by deploying a robust encryption technique will ensure one of the most critical aspects of this project. Accomplishing these two technical objectives will be the most important landmark of this project.

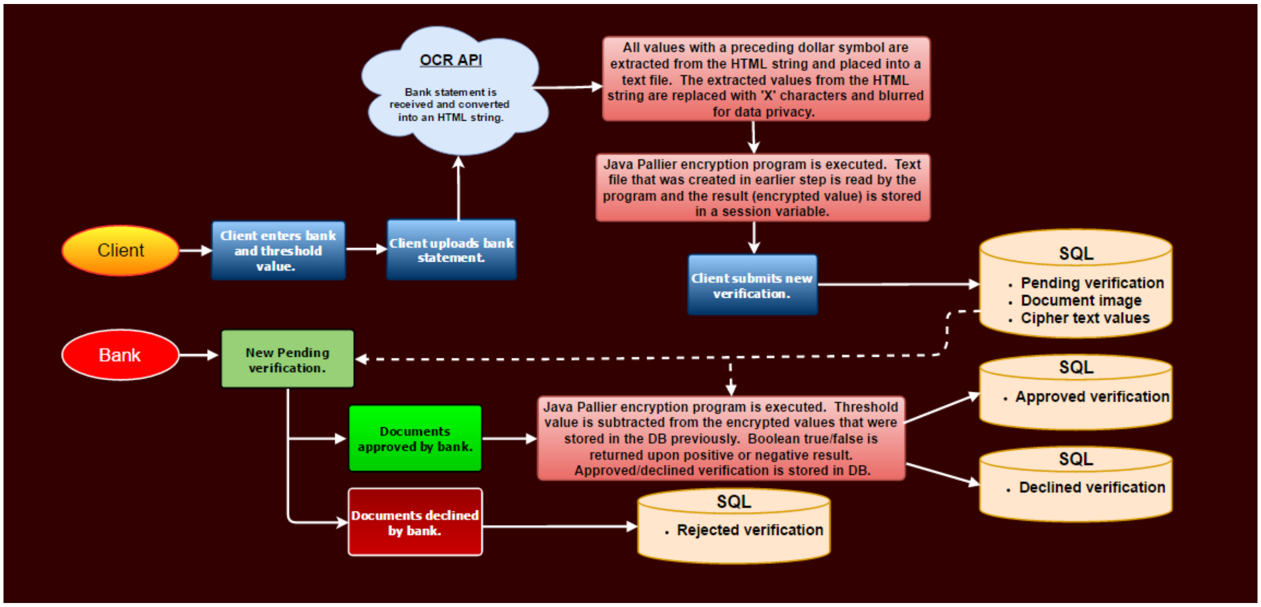
The first objective will be to build the basic structure of the UI. Once, the foundation of the interface has been built, finding and implementing an accurate and reliable image processing API will be the next step. The UI will be tested by loading the application and uploading documents to the interface. Documents should be processed accurately through the image processing API and displayed on the screen. These two objectives will be completed within the first two months.

Next, a regular expression algorithm will be developed to extract the balance values from uploaded documents. The algorithm will be tested by uploading different bank statements from various institutions and confirming that the values are correct. During the development of this algorithm, the encryption program will be built. This program should encrypt all values passed to it and sum them. The sum should be returned as a cipher text value. Both the regular expression algorithm and encryption program will then be integrated into the application. This entire process should take 2 months to complete.

The next goal will involve building the database and securing it. Tables for users, institutions, transactions, documents and cipher text values will need to be created. Once the database and relevant tables are built, this entire system will need to be integrated into the application. Queries to the database will be written and built into the application so that its integration and functionality can be tested. Users and companies should be able to log into the UI at this point. In addition, transactions stored in the database should be readily accessible through interface. Development of this part of the application should take about 1 month.

At this point, the basic structure and systems have been built and integrated. The remaining objectives include: further developing the UI into a commercial grade product, maximizing security throughout the entire system, and optimizing application performance. The UI should reflect quality, reliability, and efficiency. Users of the interface should be able to easily log in and execute their desired actions quickly. Data in the system should be stored efficiently and secured. The entire application should be very quick and reliable across all the integrated components. The final product will be tested thoroughly for accuracy and consistency. This final step in the development of the project will take approximately 2 months.

Project Diagram


Client
Client enters bank and threshold value.
Client uploads bank statement.
OCR API
Bank statement is received and converted into an HTML string.
Client submits new verification.
SQL
Pending verification
Document image
Cipher text values
Bank
New Pending verification.
Documents approved by bank.
Documents declined by bank.
SQL
Approved verification
SQL
Declined verification
SQL
Rejected verification

All values with a preceding dollar symbol are extracted from the HTML string and placed into a text file. The extracted values from the HTML string are replaced with 'X' characters and blurred for data privacy.

Java Pallier encryption program is executed. Text file that was created in earlier step is read by the program and the result (encrypted value) is stored in a session variable.

Java Pallier encryption program is executed. Threshold value is subtracted from the encrypted values that were stored in the DB previously. Boolean true/false is returned upon positive or negative result. Approved/declined verification is stored in DB.

Sources

Total home sales in the United States from 2011 to 2016 (in millions). Retrieved from https://www.statista.com/statistics/275156/total-home-sales-in-the-united-states-from-2009/

Rental Statistics (Rental Clock). Retrieved from http://www.rentalprotectionagency.com/rental-statistics.php