Summary Page

An astute American privilege in our society is the ability to accumulate good standing credit, and with that comes the options of requesting credit loans of all sorts for any needed purpose. Citizens may acquire loans through many means, the most widely used through their own banking institutions or from other standing banks. However, sometimes loan seekers utilize third party creditor/debit institutions like www.quickezloans.com, which provide the service of high risk loans that have higher interest rates than more secure and trustworthy banking institutions. In situations like these, loanees risk a lot of their personal information, including banking/billing information to services that may not be trustworthy. This is where a service like eVerify may become helpful.

This Small Business Innovation Research Phase I project has broad impacts on the way citizens of our country acquire loans for any personal means. By providing a service that bridges the gap between client confidentiality, and third party authenticity, we propose a compromise that upholds the standards of the CIA triad-confidentiality integrity and availability. In a situation where a client needs to prove his net worth, through means of physical financial assets such as the accumulated sum of financial account balances across multiple banking institutions, privacy can be easily negated. Where a client may have businesses across multiple banks, it can be difficult to prove to creditors their total financial assets without disclosing information such as how much money they have in each prospective bank. eVerify seeks to provide a solution to this problem by allowing a client to upload multiple bank statements from their respective banks, extract sensitive information like account balances, and sum them via means of the additive homomorphic properties of the Paillier cryptosystem. We utilize this algorithmic component to provide a sense of security to mask each account balance from the bank they were extracted from. In addition to this service, we provide the authenticity means creditors desire to verify that our clients do in fact have businesses with each of these banks, and that these financial numbers are not made up, or have been altered in any way shape or form.

A second algorithmic component utilizes optical character recognition to extract sensitive account balances from client uploaded bank statements, in addition to blurring the account balance and retaining the original bank statement pdf file's originality. This proves to the verifier; whether be a trusted bank providing the loan, or any other third party creditor organization that our client can prove that their financial assets exceed a given threshold that can be used to verify that the client has the ability to pay back the loan in question. This is the first step in establishing a sense of trust between loanee and loaner, so that the institution providing the loan can ensure the client isn't fabricating their own financial assets. In a world where trustworthy sources are few and far inbetween, we provide a service to bridge this gap and offer our clients the ability to secure their privacy when it comes to sensitive banking information, while also verifying that these provided values are in fact owned by our clients.

This is only the first step in establishing trust in a potential loan request, where there may exist other information required to determine if a requested loan is acceptable to give to our clients. We expect to add additional qualifiers to help this verification process, while still providing the confidentiality our clients so desperately desire in this world of information theft.

Elevator Pitch

Eventually, many American citizens in their life come to a point where they require some type of short term loan, whether be an open or close-credit loan for personal or non-personal means. With a good credit score, acquiring a loan from your own banking institution might seem like a breeze, especially if you have previous experiencing paying back loans with good faith. However, for people with low or no credit score, acquiring a loan can be a stressful and difficult process, especially if they get rejected from their own bank. Sometimes customers turn toward insecure third party creditor organizations, who make massive profits off of high interest rates on relatively low amount loans, and request a lot of personal information from the requesting loanee. In a world full of potential adversaries waiting to jump on the opportunity to steal banking information, such as routing numbers, billing information or even identity theft, it can be difficult to decide who to trust when applying for a loan and what information is safe to provide to the prospective loaning organization.

eVerify seeks to provide a compromise between client confidentiality and third party creditor authenticity. To elaborate, we understand how sensitive our clients banking information is, especially when it pertains to information regarding our loved ones and family assets. On the contrary, creditors do not willingly provide loans to requesting loanees without fully assuring they can repay the loan back in good faith. eVerify considers both of these requirements and does not take sides in regards to preferring complete confidentiality and anonymity, or complete authentication and verification. This makes the eVerify innovation unique in the sense that we try and find a middle ground between loaners and loanees, for acquiring loans of any sort for a quick, painless and secure process for both parties involved. This applies specifically for clients who have financial assets across multiple banking institutions, of which they would rather not disclose what percentage of their assets belong to which banking institution.

With the idea in mind that a potential loanee has funds across several different banks or even funds in money market accounts or stocks, concatenating these funds and proving their net worth exceeds a given threshold while also providing privacy in regards to those funds can be a challenge. This is where the additive homomorphic properties of the Paillier cryptosystem come into work. By encrypting all financial account balances, summing them on an insecure network, and then providing the corresponding decryption key to the verifier in question, alongside the user provided encrypted threshold, the verifier can confirm whether or not the loanees net worth is above or below a given agreed upon threshold. To verify that these account balances are in fact from the client, we utilize optical character recognition to scan over client provided bank statements, extract all sensitive financial balance information, and blur their

location respectfully while maintaining the originality of the bank statement. This allows the creditor to view an original bank statement provided from the respective banking institution without knowledge of the account balances associated with that bank. We allow the creditor to verify this threshold request, including all blurred provided bank statements alongside the client provided threshold amount. If the request appears legitimate and authentic, they may accept the verification and review the result to determine if the client requesting the amount is suitable to obtain the loan, in combination with the terms initially provided in the original loan request.

Societal & Broader Commercial Impact

The notion of privacy in this age of the ever increasing interconnectedness of the Internet of Things is a far too frequently ignored. When a company becomes more concerned with generating profits than protecting its clients' sensitive information, privacy becomes a luxury of the past; a commodity too costly and irrelevant to devote any effort and resources in. This is where eVerify selective data privacy comes in.

We target an audience of prospective loan seeking citizens, for either personal loans or those backed by a physical asset such as a car or a house. The number of home sales is forecast to reach 5.8 million dollars in the end of 2016. Over 2,600 new lease application are signed every day, on top of new car sales exceeding 17 million sales in 2015, of which 85% were financed. We live in a world where individuals must build credit, must take responsibility to pay off their debt, and must authenticate to creditors that they are able to and willing to pay back all debts owed. The potential market opportunity of lending money with interest, and assuming responsibility for loans unpaid is both an American affluence as much as it is an American burden. Therefore eVerify was created to authenticate its clients' banking information while still protecting the sensitive nature of that information. The application satisfies the needs of both the client, and the third party loaner/debitor requirements.

Our customers range from individuals seeking closed-end credit loans, such as mortgages, car loans, student or personal loans, to open-end credit loans also known as revolving loans. The eVerify business model revolves around the concept of selective data privacy, while also supplying adequate verified data use to properly authenticate clients with bank accounts, pay stubs or transactional history requests. The information our client deems proprietary or sensitive to unauthorized entities, is extracted by our use of optical character recognition and then blurred for 100% confidentiality.

In terms of competition, not many companies are taking customer privacy as seriously as they should be, especially in a market involving sensitive banking information of their clients. With credit and identity fraud a much compelling reason to hide personal information not necessary for proper authentication with third party verifiers, it would make sense other companies come up with means of their own version of selective data privacy. However, our ability to uphold the core concepts of the CIA triad-confidentiality, integrity and availability, protected with the difficulty of breaking the Paillier cryptosystem to protect sensitive data,

competition should not be such a problem at first. Imperative objectives include properly creating our authentication documents to be as legitimate and compelling as possible to our third party creditors, to approve and validate that our clients due in fact hold *n* accounts across *n* banks, and proving that it is indeed our clients money in those accounts. I believe however, that the use of our optical character recognition software will pinpoint the necessary confidential information to be kept secret, and our use of Paillier homomorphisms to sum encrypted monetary amounts into our encrypted overall sum to provide the end goal of our application.

Depending on how far consumers will pay for the protection of their privacy including information regarding family members and loved ones, eVerify should be a commodity held in a great light in today's ever growing world of the IoT, where security should become everyone's most important risk factor. With that being said, if eVerify selective data privacy can cheaply provide this sense of safety and protection of our clients, we will have a large clientele willing to use our application when validating monetary bank amounts with prospective third party creditor organizations with a 100% trust factor maintained on both sides of the deal. Requirements of such an application include a database full of potential bank statement templates that would have to be accumulated to help speed up our optical character recognition software in acquiring the specific areas posed to be confidential and blurred. On top of a database of prospective users, certain qualifying information would need to be gathered from our clients before they register an account with this application. And lastly, a trustful and business-like relationship with third party credit agencies must be bridged to create complete authenticity and trust between creditor and loanee.

The average consumer seeking a loan or mortgage credit does not know what to expect when it comes to what personal information of theirs gets shared with third party credit organizations. eVerify provides an alternative to this consumer gap in understanding, and provides a bridge of compromise between confidentiality and authentication. The potential broader impacts of such a compromise implies that the consumer has control of what they deem sensitive information on their behalf, while in addition providing adequate authentication means to verify bank account holdings and current bank standings. The ability to provide both these criteria satisfy what eVerify, selective data privacy is seeking from its initial construction, to its final goal of providing complete confidentiality on the consumer's behalf. This notion of consumer privacy in the delicate process of loanee authentication and verification is a step in the right direction of general consumer privacy and confidentiality. This transformative concept steps outside the typical box of well defined criteria for obtaining a loan or potential mortgage, and defines a guideline or “compromise” between the two parties involved. When the third party creditor recognizes a validated and authenticated client with proper standing bank credentials, we allow some leeway to then provide semi confidentiality for our client in the form of blurring client specific sensitive information.

eVerify’s mechanism for success lies upon two initial assumptions. The first, that our third party creditor/loaner can validate and agree that our client does in fact hold accounts across *n* potentially different banks. And the second, that after this validation step occurs, our optical character recognition software can identify specific sensitive information and blur them

respectfully in a fashion that does not alter the corresponding criteria to our first step of success. When both steps are met, our transaction is considered verified and approved, and the corresponding output designates that agreement/disagreement between our client and third party creditor. As CS majors concentrating in cyber security, and big data analytics, we offer the skills required to create a backend application to store user data, in addition to the creation of a secure cryptosystem responsible for encrypting sensitive client information into the resulting final equation of our product. Which in turn, determines the outcome of the loaning request.

Adequate resources on our behalf to achieve our end goal for our product are not that specific nor complicated. A backend SQL database, a front end web app, optical character recognition API call and a cryptographic encryption/decryption function provide the backbone algorithmic components of eVerify. To diminish the harmful results of ubiquitous user error, we will limit what factors the client has control over. An example of such is to ensure no overlap of required validation on the third party creditors behalf gets blurred and remains confidential on the client / users behalf on due part of their own action. Third party validation and authentication prioritizes client confidentiality and legitimacy, as the integrity of the verifiers comes first. Our second goal in mind is then to provide the means of confidentiality on the client's behalf, by allowing them to select which fields may remain private after designated authenticated fields have been visually identified and solidified in place.

I believe additional broader impacts may be revealed in areas of individual well-being in society, increased partnership between academia and industry, improved national security and enhanced infrastructure for both research and education.By providing a service that prides itself around the CIA Triad, consumer privacy becomes an index of individual well-being in today's every connected world with consequences of the likes of targeted marketing, and malicious identity and credit fraud. Consumers no longer need to hide in fear that their banking information or pay stub history are being sold and abused on the black market. Furthermore, promoting the importance of consumer privacy further reinforces the general trend of maintaining the importance of users personal information and how third parties treat that respective information when provided by the consumer.

Coining the term “compromise” between third party creditors and our clients means we are bridging a gap formed through academia, to support ongoing relationships that exist in the credit/loan transactional industry. Where both parties utilize the same platform to achieve 100% efficiency of their end goal, with the enhanced benefits of consumer privacy, in turn generates a positive and trustworthy relationship between these two parties using our application. If both parties are happy and we are providing a unique service to our clients, there should be no reason why any third party loaning agency would not use our application.

The integrity of our application relies on the computational complexity of breaking the Paillier cryptosystem. If eVerify can standardize a secure means of creating validating bank statements and formulating loan and credit requests, we subsequently increase the overall assurance of national security. Knowing that no data gets altered between the initial input of

plaintext through the encryption process and into our algorithm to compute proper financial coverage, we have set up a means to securely determine if one is eligible to apply for any given potential loan or mortgage request.

The use of the Paillier cryptosystem and its intrinsic homomorphic properties may pave way for additional research of using homomorphisms in selective data privacy. An example of such would be a secure electronic voting system, or perhaps a means of evaluating website content data and its relevance to prospective web users. Homomorphisms are a rather unique property of certain cryptographic systems, in which not much research has been conducted pertaining its usefulness in general purpose application.

eVerify clearly displays potential for both commercial and broader societal impacts of the loaning process that exists in this country. We believe that the general trend towards the protection of consumer privacy will prompt other industries to follow suit. Whether utilizing the effective properties of homomorphic cryptosystems, or some other means of selective data privacy, this trend will hopefully continue on to the later half of the century to define the default criteria for what encompasses consumer's rights over their own personal information.

R&D Plan

Of most key importance for eVerify, is the notion of what are the trusted/un-trusted entities of the application. Currently, eVerify is a trusted entity where our clients can trustfully upload their bank statements straight from their online banking website. This notion of trust is imperative as eVerify will have access to the plaintext financial information our client is providing. Therefore, our clients must trust that our application keeps this sensitive information private and protected against any outside adversary. Because our current model of optical character recognition utilizes an outside source API call to extract financial account balances from the bank statements, it is imperative that this information is kept private from any outside entities. After the API call is made, the information is extracted and placed in a text file that the Paillier java class takes as input for encryption. Once Paillier is finished with this input file, it is immediately deleted and furthermore no plaintexts are stored on the backend database. As for the bank statements, the exact location of the sensitive financial information are immediately extracted and replaced with arbitrary X characters, then applied a CSS burring element for confidentiality. These blurred documents are then stored on our database and sent to the verifier for validation. No unblurred documents are stored on the database as well. With this established notion of security surrounding the original bank statements, eVerify has assured the clients that their uploaded documents are safe from any untrusted entity seeking to reverse the blurring element and extract the plaintext values that the optical character recognition API call has found. Yet the originality of the document stays intact, in terms of the respective bank logo and the rest of the trademarks left behind from the online bank statement pdf to ensure to our verifiers that the document is in fact authentic and only the account balances have been blurred for client privacy.

eVerify utilizes the Paillier cryptosystem's additive homomorphic property to sum any number of encrypted ciphertexts into a concatenated ciphertext sum amount. The following equation represents how additive homomorphism is used in the context of eVerify:

$$D(\prod_{i=1}^{Nv} C_i \bmod n^2) = \sum_{i=1}^{Nv} V_i \bmod n$$

This equation effectively states that the summation of initial plaintext elements modulo n, equates to the decryption of the multiplication of said ciphertext elements module n^2. This allows eVerify to take in many plaintext financial account balances, encrypt them, then sum them to create a concatenated ciphertextSum value. This ciphertextSum when decrypted, equates to the corresponding summation of initial plaintext values used. eVerify utilizes this encryption and summation process to ensure no values have been altered and more importantly, remain confidential in their encrypted state. This also allows everify to store encrypted values in our backend database, ensuring even a data breach in that aspect provides no information to a potential adversary about the integer value amounts from the inputted plaintext elements. This allows our clients to upload any number of bank statements from any number of bank accounts, and retain the integrity of their summed amount through a protected encryption process.

The plaintext elements are extracted through an optical character recognition API call, which effectively converts any uploaded bank statement pdf into a very long html string. This string is then parsed for specific tokens, such as "On Deposit:", "Available Now" and most importantly "$". The string parser then removes any ubiquitous characters such as "," in addition to removing any decimal value, which effectively rounds the plaintext element down. This is necessary because the Paillier java class utilizes BigInteger class type to perform operations on large numbers starting at 256 bits, instead of a BigDecimal object type. Once the account information is extracted, it is placed in a plaintext.txt file and fed as input to the Paillier class. Afterwards, this sensitive plaintext.txt file is immediately deleted from our server as all that remains are the encrypted values, and eventually their respective summed total. In place of the extracted financial account balances, arbitrary X elements are inserted into the html string and then a CSS blurring element is applied for client confidentiality of the initial bank statement. This is an added security element incase any third party attempts to reverse engineer the CSS blurring attribute to reveal the plaintext elements behind, which will just be simple X's.

Each document's ciphertextSum is then stored in the database in a pending verification table, denoted with a pending verification ID, an encrypted threshold amount created by the client, alongside a foreign key corresponding to the actual blurred document that resides in a documents column in the database. Restating again, no plaintext sensitive information are stored on the database for obvious security reasons. When a pending verification becomes approved by a third party verifier, it becomes an approved verification, and the Paillier class is

called once more for the decryption of all ciphertextSum's related to all client uploaded documents associated with that given approved verification. Paillier is fed the encrypted threshold amount, and looks for a ciphertext.txt file that the SQL database has created once a verification becomes approved. Paillier then takes this ciphertext.txt file as input, alongside an encrypted threshold value, and performs the proper decryption of both elements. Finally, it subtracts the decrypted ciphertextSum value from the decrypted threshold and either a positive or negative value is found. This gives the result of the verification; whether or not the client's provided financial assets across all banks provided exceeds the given client generated threshold.

The next phase of eVerify includes extending our application to allow the uploading of multiple pdf files, as currently only one pdf may be used, where only a checking and savings account are summed. Once several documents can be uploaded, our database may be populated with more ciphertextSum values that require to be summed again for the final subtraction operation with the client provided threshold. Then, once a verifier approves a given pending verification, our database must create a ciphertext.txt file with all associated ciphertext values from each client provided bank statement. Once this has been created, Paillier must be called again with an additional input value of the threshold to be used in the final subtraction operation. But first, Paillier must sum all ciphertextSum values from each document to create a final ciphertextSum, as a document may contain multiple account balances in themselves and an approved verification may consist of multiple bank documents. Once Paillier generates the final ciphertextSum, it can decrypt the value alongside the inputted encrypted threshold value, then finally subtract both values to generate the final result. The result will either be a positive or negative integer, proving that our client's combined financial assets across all uploaded banks exceeds the client generated threshold value.

There must be more abstraction of trusted/untrusted entities in the eVerify application. Currently, eVerify is considered a trusted entity, which might change in the future to provide more usefulness of the Paillier encryption scheme. To provide an example, it would be in eVerify's best interest to be considered an untrusted entity where our client would not trust us with their uploaded bank statements and corresponding financial information. Instead, our clients banking institutions would encrypt their financial balances utilizing the same Paillier encryption scheme alongside the client generated threshold value, all using the same encryption keys . eVerify would then be provided these encrypted values from all banks associated with a given verification request, alongside the corresponding modulus n, necessary for the summation of all ciphertext values. eVerify would therefore only be able to see all encrypted financial account balances, encrypted threshold, and modulus n. With only the responsibility of utilizing the additive homomorphic properties of the Paillier cryptosystem, everify would sum all encrypted values, and send the corresponding ciphertextSum and encrypted threshold value to the verifier for approval. The verifier; another untrusted entity, would then need to be supplied the decryption keys used for the final decryption of all encrypted values to perform the final subtraction operation on their own behalf, to see the final positive/negative result. The verifier would have to request the decryption keys from the same

banking institutions that encrypted the values in the first place. These are two important assumptions that would have to be made.

eVerify must be abstracted as an untrusted entity, to prove the usefulness of the Paillier scheme is valid. This includes the experimentation of an additional masking element that must be applied to the encrypted ciphertextSum value. In it's current condition, subtracting the decrypted ciphertextSum from the decrypted threshold indicates exactly how much funds the client has away from that threshold, which voids the privacy eVerify is currently attempting to provide our clients. A masking element, when applied to a decrypted ciphertextSum, would hide the true value difference from the threshold without altering the true positive/negative result the verifier is looking for. This experimentation will be done over winter break, alongside the addition of the capability to upload multiple bank statements, if our current use of the optical character recognition API call is still used. However, if eVerify is relying on the banks to encrypt our clients respective financial account balances, there would be no need for the optical character recognition API call or the blurring element.

The presumed assumption of different banks utilizing the same Paillier scheme to share the same encryption/decryption keys will have to be another project in its own. A separate trusted Paillier scheme will have to be created and allowed to be shared amongst all of the client's banks, and a secure means of transporting the encrypted values to eVerify would need to be established, alongside the decryption keys to the untrusted verifiers. However, if this scheme can be accomplished, eVerify need not know any sensitive client information at all, as the only trusted entities can be the client's corresponding banking institutions. This also allows our third party verifier to be any type of verifier; an online loan organization, a local untrusted loan shark, etc. Demoing this new concept would be quite difficult, as the abstraction of the bank's behalf has now presumed to be something handled on their end, not inside of the eVerify application. However, it would still be possible to achieve the same task with an increased confidence in client trust with our application, as client need not supply eVerify with any sensitive information and an optical character recognition API call would no longer be necessary to validate the authenticity of client provided financial information. This depends on the legitimacy of the banking institutions ability to properly encrypt all clients financial balances, which would have to be experimented post winter break.

The following image depicts the current systems design diagram of the eVerify application: