BIOMETRIC ANONYMIZER SYSTEM AND METHOD

CROSS REFERENCE TO RELATED APPLICATIONS

[0000] This application claims priority to U.S. Provisional Application No. 63/568,690 filed on March 22, 2024, the entire contents of which are herein incorporated by reference.

FIELD

[0001] The present disclosure relates generally to anonymizing data, and more particularly to anonymizing biometric data to remove personally identifiable information, and even more particularly to anonymizing biometric data to remove personally identifiable information whilst maintaining a homomorphic relationship to the original data.

BACKGROUND

[0002] Biometric information about an individual may be subject to strict terms of use. Privacy regulation defines that use to be for an authorized purpose. As well biometric information is often retained without direct consent from the individual concerned. And whether consent is given or not, obligations pertaining to the retention and use of biometric information are increasing and breaches of obligation, depending on the jurisdiction involved, are often subject to significant and usually punitive consequences.

[0003] Managing the balance between custodial obligations and respecting the boundaries of authorized purpose, is problematic in many cases, especially with privacy regulation and case history continuing to evolve. Organizations handling personally identifiable information (PII) face an ongoing challenge of correctly asserting: when their right to retain PII data ends and what data pursuits are within the bounds of an authorized purpose. Alternatively, organizations can take an overly conservative position, that of not using available data to support their endeavors, particularly if alignment with approved purpose, or right to retain, is in doubt.

[0004] It remains desirable therefore, to develop further improvements and advancements in relation to collecting, maintaining, sharing, and/or anonymizing biometric information and/or PII to overcome shortcomings of known techniques, and to provide additional advantages thereto.

[0005] This section is intended to introduce various aspects of the art, which may be associated with the present disclosure. This discussion is believed to assist in providing a framework to facilitate a better understanding of particular aspects of the present disclosure. Accordingly, it should be understood that this section should be read in this light, and not necessarily as admissions of prior art.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Embodiments of the present disclosure will now be described, by way of example only, with reference to the attached Figures.

[0007] FIG. 1 illustrates a system and method for encoding and matching biometric information in accordance with an embodiment of the present disclosure.

[0008] FIG. 2 illustrates a system and method for encoding, anonymizing, and matching biometric information in accordance with an embodiment of the present disclosure.

[0009] FIG. 3 illustrates an example of applying Binning to a normal distribution in accordance with an embodiment of the present disclosure.

[0010] FIG. 4 illustrates a system and method for applying a new permutation key in accordance with an embodiment of the present disclosure.

[0011] FIG. 5 is a block diagram of an example computing device or system for implementing one or more systems, aspects, embodiments, methods, or operations of the present disclosure.

[0012] Throughout the drawings, sometimes only one or fewer than all of the instances of an element visible in the view are designated by a lead line and reference character, for the sake only of simplicity and to avoid clutter. It will be understood, however, that in such cases, in accordance with the corresponding description, that all other instances are likewise designated and encompasses by the corresponding description.

DETAILED DESCRIPTION

[0013] In an aspect, the present disclosure provides for a computer implemented method for anonymizing data, including receiving biometric data having personally identifiable information associated with a person, encoding a record with a plurality of biometric identifiers of the biometric data, each of the plurality of biometric identifiers comprising a numeric value, and generating an anonymized record based on transforming the numeric value of each of the

plurality of biometric identifiers, the anonymized record being disassociated from the personally identifiable information.

[0014] As a further example in accordance with the present disclosure, transforming the numeric value includes normalizing the numeric value of each of the plurality of biometric identifiers.

[0015] As a further example in accordance with the present disclosure, normalizing includes scaling the numeric value to a normal distribution curve of a plurality of representative samples, each of the plurality of representative samples having a plurality of corresponding biometric identifiers.

[0016] As a further example in accordance with the present disclosure, normalizing includes scaling and shifting the numeric value to a normal distribution curve of a plurality of representative samples, each of the plurality of representative samples having a plurality of corresponding biometric identifiers.

[0017] As a further example in accordance with the present disclosure, the normal distribution curve is based on an average mean value of a mean of the plurality of representative samples and an average mean value of a standard deviation of the plurality of representative samples.

[0018] As a further example in accordance with the present disclosure, transforming the numeric value includes applying a many-to-one mapping to the numerical value.

[0019] As a further example in accordance with the present disclosure, transforming the numeric value includes assigning the numeric value of each of the plurality of biometric identifiers to a representative value of a bin of a plurality of bins.

[0020] As a further example in accordance with the present disclosure, the plurality of bins includes at least 80 bins.

[0021] As a further example in accordance with the present disclosure, transforming the numeric value includes re-arranging an ordering of the plurality of biometric identifiers of the encoded record to a first order different from an original order.

[0022] As a further example in accordance with the present disclosure, re-arranging the ordering includes applying a first permutation key.

[0023] A further example in accordance with the present disclosure includes, restoring the original order of the plurality of biometric identifiers of the encoded record based on applying an inverse of the first permutation key, and re-arranging the ordering of the plurality

of biometric identifiers of the encoded record to a second order different from the first order based on applying a second permutation key.

[0024] A further example in accordance with the present disclosure includes, maintaining a record of the plurality of biometric identifiers of the encoded record in the first order until completing re-arranging of the plurality of biometric identifiers of the encoded record to the second order.

[0025] As a further example in accordance with the present disclosure, the anonymized record includes a first anonymized record having a first plurality of biometric identifiers, each of the first plurality of biometric identifiers comprising a first numeric value, the method further including receiving a second anonymized record comprising a second plurality of biometric identifiers, each of the second plurality of biometric identifiers comprising a second numeric value, and determining a match between the first anonymized record and the second anonymized record based on comparing the first and the second numeric values of each of the first and second plurality of biometric identifiers, wherein the match comprises a confidence interval and/or a confidence level.

[0026] As a further example in accordance with the present disclosure, determining the match includes evaluating a closeness between the first and the second numeric values.

[0027] In an aspect, the present disclosure provides for a system for anonymizing data, the system including a biometric device for obtaining biometric information associated with a person, and a processor configured to execute a method in accordance with the present disclosure.

[0028] In an aspect, the present disclosure provides for a computer readable medium having instructions stored thereon for anonymizing data that, when executed by a processor, cause the processor to perform a method in accordance with the present disclosure.

[0029] For the purpose of promoting an understanding of the principles of the disclosure, reference may be made to the features illustrated in the drawings and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the disclosure is thereby intended. Any alterations and further modifications, and any further applications of the principles of the disclosure as described herein are contemplated as would normally occur to one skilled in the art to which the disclosure relates. It will be apparent to those skilled in the relevant art that some features that are not relevant to the present disclosure may not be shown in the drawings for the sake of clarity.

[0030] At the outset, for ease of reference, certain terms used in this application and their meaning as used in this context are set forth below. To the extent a term used herein is not defined below, it should be given the broadest definition persons in the pertinent art have given that term as reflected in at least one printed publication or issued patent. Further, the present processes are not limited by the usage of the terms shown below, as all equivalents, synonyms, new developments and terms or processes that serve the same or a similar purpose are considered to be within the scope of the present disclosure.

[0031] Personal Identifiable Information (PII) includes representations of information that permit the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means. Regulations may further define a company's responsibilities for handling, protecting, and sharing PII.

[0032] Biometric identifiers include distinctive and/or measurable characteristics for labelling, describing, and identifying individuals, and may include biometric information such as body measurements and calculations related to human characteristics in quantity. Biometric identifiers may be categorized as physiological, for example, relating to the shape, motion such as gait, derivatives of motion such as signature, sound such as voice, chemistry such as biological sample testing, biological signatures for example iris scans, DNA, finger or palm prints, structure such as medical imagery, geometric or the measurement of body shape or body parts such as hands or ears, and behavioral characteristics, that result in the identification or partial identification of an individual. Biometric identifiers containing PII may be referred to as, for example, PII Biometric Data.

[0033] In order to address questions such as, what percentage of a sample population exhibits a biometric characteristic, members of a sample population need to be tested for the biometric characteristic, with matches separated from non-matches. When working with alphanumeric PII data, such as name, address, gender, age, or date of birth, matching and non-matching may be established through simple binary operators, such as equal-to, less-than, greater-than, etc. For example, Age = X or Year of Birth is between A and B. Even when applying advanced pre-processing techniques to such alphanumeric PII data, such as data correction, standardization, or bucketing, (for example postal addresses usually require data correction and or standardization, or matching age by comparing members of a sample population with an age range) the end result remains the same: matching and non-matching of data may be accomplished through simple operators. Biometric information may however

require matching through non simple operators, such as for example by measuring a "closeness" or determining a confidence level. Consider the example of two facial images of an individual: a first image taken at a first time and a second image taken at a second time. The second image likely does not match the first image pixel-for-pixel, limiting the utility of a simple, binary operator to determine a match. Reasons for non-matching pixels can include, but are not limited to: changes to lighting conditions, changes to the angle of photography, differences in facial expressions, or positional shift of objects and faces with the image dimensions. Despite these differences, the two images can be "matched" if the faces found in the two images are "close enough" in biometric terms.

[0034] One example of determining closeness for matching biometric data includes determining an accumulated variance across a number of measurements. After processing the measurements of two biometric samples, a biometric match algorithm will, instead of returning a boolean "yes this is a match" or "no this is not a match", return a confidence level (derived from the extent of overall 'variance' between the two biometric samples). In the case of Facial Recognition, a confidence level may correspond to how likely the two facial images are the same individual.

[0035] In an embodiment, the confidence level may map to a confidence interval. For example, a plurality of confidence intervals may be defined, such as "match", "partial match", and "not matched", each interval being associated with a range of confidence levels. For example, a first confidence interval may correspond to a first range of confidence levels, a second confidence interval may correspond to a second range of confidence levels, a third confidence interval may correspond to a third range of confidence levels, and so forth. In an embodiment, the first confidence interval is "match" and the first range of confidence levels is greater than or equal to 91%; the second confidence interval is "partial match" and the second range of confidence levels is less than 91% and greater than or equal to 85%; and, the third confidence interval is "not matched" and the third range of confidence levels is less than 85%. Confidence levels and confidence intervals, and their interpretation, typically reflect the underlying biometric sample processing system (BSPS).

[0036] FIG. 1 illustrates an embodiment of a system and steps for matching biometric information, in particular, for facial recognition.

[0037] Step 1 may include capturing a biometric sample 12 (e.g. face image, voice recording, fingerprint scan, etc.) using a biometric capturing device 10, such as, for example,

a camera for capturing images of a person's face. Step 1 may further include sending the captured biometric sample 12 (e.g. jpeg, wav, data file etc.) to an identification and selection engine 20 for integrity and sensibility checks and to isolate relevant parts of the biometric sample 12. In an embodiment, step 1 may further apply an enhancement to the captured biometric sample 12, to for example, improve clarity for subsequent steps.

[0038] Step 2 may include sending an enhanced biometric sample 12' to an encoder 30.

[0039] Step 3 may include creating a first *k*-dimensional (k-d) vector encoding (E1) representing the biometric sample, the first vector having k measurements corresponding to each feature of the biometric sample.

[0040] Step 4 may include retrieving or looking up a previously stored vector encoding (E2) and further forwarding E1 and/or E2 to a matcher 50. In an embodiment, the matcher 50 may also look up and/or retrieve one or more encoding vectors. In an embodiment, step 4 may further include adding E1 to a database 40 of previously stored vector encodings of known biometric samples.

[0041] Step 5 may include determining a closeness of match between E1 and E2. In an embodiment, determining a closeness of match comprises measuring a Euclidean distance between E1 and E2. Step 5 may further include providing an output comprising a confidence level and/or confidence interval.

[0042] A potential draw back of a system according to FIG. 1 includes instances wherein custody over E1 and E2 belongs to different entities. Legal and privacy regulations may limit the entity having custody over E1 from sharing PII with the entity having custody over E2, and vice-versa, thereby precluding the possibility to determine a match between E1 and E2 without contravening legal and/or privacy regulations. Thus, for this reason, and to relieve obligations of the entity having custody over E1, there is a need to remove or otherwise anonymize PII to allow for and permit sharing of biometric information, creating a technical problem associated with balancing both anonymizing PII while maintaining sufficient fidelity with the anonymized PII to carry out further functions, such as matching a closeness of the anonymized PII data.

[0043] Anonymizing biometric information comprising PII may require addressing one or more dimensions of the data. For example, anonymized data may benefit from normalizing each feature of the encoded k-dimension vector to a normal distribution curve, to marginalize

or eliminate distinguishing differences between the distribution of each encoded feature. As another example, applying a many-to-one mapping to each feature of the encoded k-dimension can marginalize or eliminate the ability to reverse the mapping of the feature to its original value. As yet another example, applying a permutation to the encoded vector to re-arrange the order of features in the encoded k-dimension vector may offer a further degree of anonymity to the data. Generally, each of the foregoing examples removes the ability to recover PII by increasing the obscurity or anonymity of the underlying data, whilst also maintaining a homomorphic relationship with the original data and allowing for matching with other similarly anonymized data at a level of accuracy commensurate to matching undertaken with the original data.

[0044] Advantages achieved by aspects of the present disclosure include, but are not limited to: anonymizing PII to a degree whereby the anonymized data does not permit recovery, discovery, or regeneration of the original PII; providing real-time and/or near real-time processing equivalent to matching a closeness of the original PII; achieving measurable match accuracies with low false positives and low false negatives commensurate to the accuracy and matching performance when operating on the original PII; and, sharing data between different entities without disclosing PII.

[0045] Systems and methods in accordance with the present disclosure may derive an anonymized form of PII that may match against similarly anonymized PII, for biometric applications including, but not limited to: facial recognition, fingerprint recognition, voice recognition, retina recognition, DNA matching, and writing recognition. Aspects herein thus provide anonymized data void of PII (from which PII data cannot be recovered), for use in matching biometric identifiers without needing to retain PII. In an embodiment, anonymizing PII occurs at the time of capturing PII, or shortly thereafter. Example embodiments in accordance with the present disclosure may be applied wherein: each biometric sample comprises a set of k values; matching the closeness of the anonymized samples is associative with respect to the k values; and, the number of dimensions of the k-dimension encoding renders brute force attempts, of k factorial, computationally infeasible. In an embodiment, an equal weighting is applied to each of the k values when matching a closeness of anonymized data.

[0046] Aspects of the present disclosure provide systems and methods for converting biometric information associated with a person to a derivative form without PII. The PII filtered

or anonymized form of biometric information obfuscates identifying persons associated with the original biometric information. However, the anonymized biometric information can allow for determining, for example, what percentage of a sample population exhibited a biometric characteristic, or what percentage of a sample population had the same biometric identifier e.g. was a facial image present in two different population samples.

[0047] Aspects of the present disclosure provide systems and methods which may allow organizations having custody over PII to do one or more of: using biometric information for additional purposes (by and large globally, PII is restricted to the purpose for which it was collected); sharing PII anonymized biometric information with affiliated agencies (once biometric data is void of PII it is no longer bound under the same terms as PII is); retaining the PII anonymized biometric information past the point-in-time when the original purpose has been satisfied (and past when PII must be removed); removing PII before the original purpose has been satisfied, for example, as soon as PII is captured and converted to PII anonymized biometric information; limiting retention of PII to specific warranted cases (and then subjecting those cases to stricter access management), and reducing burden to adhere to responsible custody of PII (which may also invite a heightened expectation of acting more responsibly).

[0048] Anonymizing PII in accordance with the present disclosure may include: anonymizing PII data wherein PII cannot be recovered from the anonymized form; maintaining characteristics of the biometric information allowing for matching a closeness in the anonymized form without degradation of accuracy as compared to matching with the original PII data; and, matching anonymized data does not result in a prohibitive increase in computational time or resources. In an embodiment, the systems and methods disclosed herein may comprise an expected complexity of O(N).

[0049] FIG. 2 illustrates an example of anonymizing first data captured by a first party (A) for matching with second data obtained by a second party (B). As illustrated in FIG. 2, a plurality of steps permit sharing of data between two different parties without exposing any underlying PII to one another. For example, the first party (A) may be a first company and the second party (B) may be a database 40' external to the first party (A), such as a third party cloud service.

[0050] Step 1 may include capturing, by the first party (A), a biometric sample 12 (e.g. face image, voice recording, fingerprint scan, etc.) using a biometric capturing device 10, such as, for example, a camera for capturing images of a person's face. Step 1 may further include

sending the captured biometric sample 12 (e.g. jpeg, wav, data file etc.) to an identification and selection engine 20 for integrity and sensibility checks and to isolate relevant parts of the biometric sample 12. In an embodiment, step 1 may further apply an enhancement to the captured biometric sample, to for example, improve clarity for subsequent steps.

[0051] Step 2 may include sending an enhanced biometric sample 12' to an encoder 30.

[0052] Step 3 may include creating a first *k*-dimensional (k-d) vector encoding (E1) representing the biometric sample 12, the first vector having k measurements for each of the biometric sample's features.

[0053] Step 3.5 may include using an anonymizer 35 in accordance with an embodiment of the present disclosure, to anonymize E1.

[0054] Step 4 may include retrieving or looking up a previously stored and anonymized vector encoding (E2) and further forwarding the anonymized E1 and/or E2 to a matcher 50'. In an embodiment, the matcher 50' may look up and/or retrieve one or more anonymized encoding vectors. In an embodiment, step 4 may further include adding the anonymized E1 to a database 40' of previously stored anonymized vector encodings of known biometric samples.

[0055] Step 5 may include determining a closeness of match between the anonymized E1 and E2. In an embodiment, determining a closeness of match comprises measuring a Euclidean distance between E1 and E2. Step 5 may further include providing an output comprising a confidence level and/or confidence interval.

[0056] In accordance with the present disclosure, anonymizing data to prevent recovery of PII includes applying a series of transformation to the biometric information containing PII. Embodiments in accordance with the present disclosure both anonymize data in a manner which prevents recovery of PII while also preserving a homomorphic relationship with the original data and allowing for matching a closeness with other anonymized data, at or near real time. For example, using anonymized data does not result in a meaningful increase in a closeness calculation computational effort when compared to a closeness calculation of the same original data before anonymization.

[0057] Consider an example of matching anonymized biometric information relating to facial recognition having 128 features (i.e. $k = 128$).

[0058] As a precursor, embodiments in accordance with the present disclosure may first collect a sample set of facial images of a population to determine statistical characteristics

of the sample set, such as the mean and standard deviation of various facial features used to perform facial recognition for further use in matching anonymized biometric information. For each sample image, generate a 128-d encoding representing 128 biometric identifiers or features f_i. For each of the 128 features f_i, determine an approximation to the mean μ_i and the standard deviation σ_i as calculated for the feature f_i across the sample set of facial images; and, calculate the average mean value μ_{avr} of all 128 μ_i and the average mean value σ_{avr} of all 128 σ_i.

[0059] In a first step, a method in accordance with the present disclosure may acquire, capture, obtain, or retrieve biometric information containing PII of a person, such as by using a camera to capture an image of a person's face or otherwise receiving and/or retrieving an image of a person's face. The first step may further include defining a bounding box about a face captured in the image and returning the *n* box coordinates for the image file.

[0060] In a second step, a method in accordance with the present disclosure may encode the image as a *k* dimension vector (E1), for example, a 128-dimension vector, each dimension corresponding to a biometric identifier used for facial recognition.

[0061] In a third step, a method in accordance with the present disclosure may anonymize the *k* dimension vector (E1) based on applying a series of transformations to E1. Generally, each transformation will increase the difficulty in recovering PII while also maintaining sufficient fidelity to calculate a closeness of match with other similarly anonymized data.

[0062] In an embodiment, a first transformation may comprise normalizing each feature f_i of E1. For example, because the characteristic distribution (e.g. the mean and standard deviation of the normal distribution) for each feature could be used to help identify which feature each position represents, reducing the possibility of recovering PII may further including transforming and/or translating the distribution for each feature so that the standard normal distribution has a different value, such as a mean of 0 and a standard deviation of 1. For example, a transformation can be applied to the location of each feature f_i on the standard normal distribution. In an embodiment, the transformation comprises:

$$\textit{normalized } f_i \textit{ value} = (\textit{original } f_i \textit{ value} - \mu_i) \,/\, \sigma_i$$

[0063] The *normalized f_i value* of this standard normal distribution may be further converted or scaled to a normal distribution with mean μ_{avr} and standard deviation σ_{avr}, by applying a further transform, such as for example:

scaled normalized f_i value = (*normalized f_i value* - μ_{avr}) / σ_{avr}

[0064] After normalizing and converting the feature f_i, even looking at a large sample of representations, each feature will have, in consideration of distributions, the same distribution as all the other features, thereby reducing and/or eliminating the possibility of distinguishing each feature based on its distribution as each will have the same normalized normal distribution. Advantageously, determining a Euclidean distance between two such normalized vectors maintains a lossless match function with respect to the original baseline.

[0065] In an embodiment, a second transformation may comprise applying a many-to-one mapping to the k-dimensioned vector, such as by binning, which may function to obfuscate PII similar to blurring or pixelating an image to remove identifiable features. In a first step, Binning may comprise selecting a number of bins, β. The lower the number of bins, the greater the anonymizing or "blurring" effect. However, too few bins may negatively impact the accuracy of matching a closeness of anonymized data.

[0066] In a second step, Binning may comprise dividing the range of the scaled normal distribution into β regions. In an embodiment, each of the β regions comprises an equal probability, in other words, an equal area under a given normal distribution curve. As such, regions closer to the mean/middle of the curve will have relatively smaller widths than regions closer to the tails of the curve. FIG. 3 illustrates an embodiment of Binning in accordance with the present disclosure.

[0067] In an embodiment, the boundaries of each of the β regions may be set using *Z critical values* equal to the probabilities: 1/β, 2/β, … (β-1)/β. Given Z(pr) as the function to calculate the *Z critical values* for probability pr, then the β regions (each with equal probability) will be:

$$\{[-\infty < x \leq Z(1/\beta)],\ [Z(1/\beta) < x \leq Z(2/\beta)],\ \ldots\ [Z((\beta-1)/\beta) < x \leq +\infty]\}$$

[0068] In a third step, Binning may comprise calculating a representative value for each bin $[Val_{left} < x \leq Val_{right}]$. In an embodiment, the representative value is a midpoint of the bin's

range. In an embodiment, the representative value is set to $Z((Prob_{left} + Prob_{right})/2)$ where $Prob_n$ is a probability of a value being less than Val_n.

[0069] In a fourth step, Binning may comprise sorting each of the *k*-dimension features into a corresponding bin and replacing the value of the feature f_i with the representative value of the corresponding bin. In this regard, Binning functions to provide a many-to-one mapping that is not reversible: each feature f_i having a value within the range of a given bin will be mapped to the same representative value, thus, the representative value can represent a range of values. Thus, the *k*-dimension vector E1 where k = 128 may comprise a set of values $\{b_{v1}, b_{v2}, \ldots b_{v128}\}$ reflecting the representative value of their associated bins which can be used to determine a closeness of matching with similarly binned data, wherein an accuracy of the match is dependent on a number of bins as discussed. As the number of bins β increases, the delta between a value of the feature f_i and the representative value of a bin may decrease as a result of each bin having an increasingly smaller width as the number of bins increases. Accordingly, as the number of bins increases, the representative value of a given bin converges closer to a given value of a feature f_i. In an embodiment, the number of bins β is greater than or equal to 80.

[0070] In an embodiment, a third transformation may comprise anonymizing the representation of values, for example, shuffling the original, normalized and/or binned values of E1. In an embodiment anonymizing the representation of values comprises applying a permutation to the normalized and binned values, resulting in a re-ordering of the values. For example, the 128-d vector E1 comprises 128! permutations (more than 3.8×10^{125} possibilities), rendering the use of brute force methods to unlock the correct permutation computationally infeasible. In combination with normalizing, even if a second party had access to the large sampling of the permuted anonymized representations from the first party, the distribution values of each of the 128 features will be exactly the same, and will fail to grant the second party insight into which of the actual 128 features each anonymized feature actually represents. In an embodiment, the permutation is selected and known to a first party and not shared with a second party. In an embodiment, the permutation comprises a permutation key.

[0071] FIG. 4 illustrates an example of changing a permutation key. A first party (A) captures biometric information of a person comprising PII. In accordance with the present disclosure, the first party further anonymizes the biometric information before sending to a second party (B). In this regard, the first party selects a permutation key to apply to all biometric

samples during the step of anonymizing. However, at a future point in time, the first party may wish to apply a different permutation key.

[0072] In a first step, the first party may submit a request 60 to a key update engine 70 to change from a first permutation key to a second permutation key.

[0073] In a second step, the key update engine 70 may request all records anonymized with a given permutation key from parties external to the first party, such as for example requesting all records anonymized with a first permutation key that are stored in a database 40' by the second party (B).

[0074] In a third step, the anonymizer 35 receives the records anonymized with the first permutation key, for use in mapping the records to their representative order (e.g. the first value is f_7, the second value is f_{89}, etc). The anonymizer 35 may further run an inverse of the permutation key for each record, undoing the re-ordering which resulted from the original application of the first permutation key, thus retrieving the original feature ordering ($\{f_1, f_2, \ldots, f_{128}\}$) of each encoding. The anonymizer 35 may then further apply a second permutation key to permute a new feature order, and further supply the newly re-ordered records to the key update engine. The anonymizer 35 may further store the second permutation key, for use in anonymizing future biometric samples.

[0075] In a fourth step, the key update engine may supply the newly re-ordered records to the second party (B), which have been permuted based on the second permutation key. The two sets of records ($Set_{1,}$, the original encoded with the original first permutation key, and Set_2 encoded with the new second permutation key) can co-exist for an indefinite grace period. This allows a zero-outage transfer period when new search request (base on either key) will work. Once the transfer period is complete (i.e. all users of the first key have been transferred over to using the new permutation key), then Set_1 can be deleted.

[0076] The outcome of applying a series of transformations to biometric PII as disclosed herein results in anonymized data wherein PII cannot be recovered, and wherein the anonymized data shares a homomorphic relationship with the original data, and permits matching between other similarly anonymized data without drastically increasing computational load.

[0077] FIG. 5 is a block diagram of an example computerized device or system 500 that may be used in implementing one or more aspects, components, sub-components,

operations, and so forth, of an embodiment of a biometric anonymizer in accordance with the present disclosure, including for embodiments of a method of anonymizing biometric data.

[0078] Computerized system 500 may include one or more of a processor 502, memory 504, a mass storage device 510, an input/output (I/O) interface 506, and a communications subsystem 508. Further, system 500 may comprise multiples, for example multiple processors 502, and/or multiple memories 504, etc. Processor 502 may comprise one or more of a digital processor, an analog processor, a digital circuit designed to process information, an analog circuit designed to process information, a state machine, and/or other mechanisms for electronically processing information. These processing units may be physically located within the same device, or the processor 502 may represent processing functionality of a plurality of devices operating in coordination. The processor 502 may be configured to execute modules by software; hardware; firmware; some combination of software, hardware, and/or firmware; and/or other mechanisms for configuring processing capabilities on the processor 502, or to otherwise perform the functionality attributed to the module and may include one or more physical processors during execution of processor readable instructions, the processor readable instructions, circuitry, hardware, storage media, or any other components.

[0079] One or more of the components or subsystems of computerized system 500 may be interconnected by way of one or more buses 512 or in any other suitable manner.

[0080] The bus 512 may be one or more of any type of several bus architectures including a memory bus, storage bus, memory controller bus, peripheral bus, or the like. The CPU 502 may comprise any type of electronic data processor. The memory 504 may comprise any type of system memory such as dynamic random access memory (DRAM), static random access memory (SRAM), synchronous DRAM (SDRAM), read-only memory (ROM), a combination thereof, or the like. In an embodiment, the memory may include ROM for use at boot-up, and DRAM for program and data storage for use while executing programs.

[0081] The mass storage device 510 may comprise any type of storage device configured to store data, programs, and other information and to make the data, programs, and other information accessible via the bus 512. The mass storage device 510 may comprise one or more of a solid state drive, hard disk drive, a magnetic disk drive, an optical disk drive, or the like. In some embodiments, data, programs, or other information may be stored remotely, for example in the cloud. Computerized system 500 may send or receive information

to the remote storage in any suitable way, including via communications subsystem 508 over a network or other data communication medium.

[0082] The I/O interface 506 may provide interfaces for enabling wired and/or wireless communications between computerized system 500 and one or more other devices or systems. For instance, I/O interface 506 may be used to communicatively couple with sensors, such as cameras or video cameras. Furthermore, additional or fewer interfaces may be utilized. For example, one or more serial interfaces such as Universal Serial Bus (USB) (not shown) may be provided.

[0083] Computerized system 500 may be used to configure, operate, control, monitor, sense, and/or adjust devices, systems, and/or methods according to the present disclosure.

[0084] A communications subsystem 508 may be provided for one or both of transmitting and receiving signals over any form or medium of digital data communication, including a communication network. Examples of communication networks include a local area network (LAN), a wide area network (WAN), an inter-network such as the Internet, and peer-to-peer networks such as ad hoc peer-to-peer networks. Communications subsystem 508 may include any component or collection of components for enabling communications over one or more wired and wireless interfaces. These interfaces may include but are not limited to USB, Ethernet (e.g. IEEE 802.3), high-definition multimedia interface (HDMI), Firewire™ (e.g. IEEE 1394), Thunderbolt™, WiFi™ (e.g. IEEE 802.11), WiMAX (e.g. IEEE 802.16), Bluetooth™, or Near-field communications (NFC), as well as GPRS, UMTS, LTE, LTE-A, and dedicated short range communication (DSRC). Communication subsystem 508 may include one or more ports or other components (not shown) for one or more wired connections. Additionally or alternatively, communication subsystem 508 may include one or more transmitters, receivers, and/or antenna elements (none of which are shown).

[0085] Computerized system 500 of FIG. 5 is merely an example and is not meant to be limiting. Various embodiments may utilize some or all of the components shown or described. Some embodiments may use other components not shown or described but known to persons skilled in the art.

[0086] In the preceding description, for purposes of explanation, numerous details are set forth in order to provide a thorough understanding of the embodiments. However, it will be apparent to one skilled in the art that these specific details are not required. In other instances, well-known electrical structures and circuits are shown in block diagram form in order not to

obscure the understanding. For example, specific details are not provided as to whether the embodiments described herein are implemented as a software routine, hardware circuit, firmware, or a combination thereof.

[0087] Embodiments of the disclosure can be represented as a computer program product stored in a machine-readable medium (also referred to as a computer-readable medium, a processor-readable medium, or a computer usable medium having a computer-readable program code embodied therein). The machine-readable medium can be any suitable tangible, non-transitory medium, including magnetic, optical, or electrical storage medium including a diskette, compact disk read only memory (CD-ROM), memory device (volatile or non-volatile), or similar storage mechanism. The machine-readable medium can contain various sets of instructions, code sequences, configuration information, or other data, which, when executed, cause a processor to perform steps in a method according to an embodiment of the disclosure. Those of ordinary skill in the art will appreciate that other instructions and operations necessary to implement the described implementations can also be stored on the machine-readable medium. The instructions stored on the machine-readable medium can be executed by a processor or other suitable processing device, and can interface with circuitry to perform the described tasks.

[0088] The above-described embodiments are intended to be examples only. Alterations, modifications and variations can be effected to the particular embodiments by those of skill in the art. The scope of the claims should not be limited by the particular embodiments set forth herein, but should be construed in a manner consistent with the specification as a whole.

WHAT IS CLAIMED IS:

1. A computer implemented method for anonymizing data, comprising:

receiving biometric data having personally identifiable information associated with a person;

encoding a record with a plurality of biometric identifiers of the biometric data, each of the plurality of biometric identifiers comprising a numeric value, and

generating an anonymized record based on transforming the numeric value of each of the plurality of biometric identifiers, the anonymized record being disassociated from the personally identifiable information.

2. The method according to claim 1, wherein transforming the numeric value comprises, normalizing the numeric value of each of the plurality of biometric identifiers.

3. The method according to claim 2, wherein normalizing comprises scaling the numeric value to a normal distribution curve of a plurality of representative samples, each of the plurality of representative samples having a plurality of corresponding biometric identifiers.

4. The method according to claim 2, wherein normalizing comprises scaling and shifting the numeric value to a normal distribution curve of a plurality of representative samples, each of the plurality of representative samples having a plurality of corresponding biometric identifiers.

5. The method according to claim 3 or claim 4, wherein the normal distribution curve is based on an average mean value of a mean of the plurality of representative samples and an average mean value of a standard deviation of the plurality of representative samples.

6. The method according to claim 1, wherein transforming the numeric value comprises applying a many-to-one mapping to the numerical value.

7. The method according to claim 1, wherein transforming the numeric value comprises assigning the numeric value of each of the plurality of biometric identifiers to a representative value of a bin of a plurality of bins.

8. The method according to claim 7, wherein the plurality of bins comprises at least 80 bins.

9. The method according to claim 1, wherein transforming the numeric value comprises re-arranging an ordering of the plurality of biometric identifiers of the encoded record to a first order different from an original order.

10. The method according to claim 9, wherein re-arranging the ordering comprises applying a first permutation key.

11. The method according to claim 10, further comprising:

restoring the original order of the plurality of biometric identifiers of the encoded record based on applying an inverse of the first permutation key, and

re-arranging the ordering of the plurality of biometric identifiers of the encoded record to a second order different from the first order based on applying a second permutation key.

12. The method according to claim 11, further comprising maintaining a record of the plurality of biometric identifiers of the encoded record in the first order until completing re-arranging of the plurality of biometric identifiers of the encoded record to the second order.

13. The method according to claim 1, wherein the anonymized record comprises a first anonymized record having a first plurality of biometric identifiers, each of the first plurality of biometric identifiers comprising a first numeric value, the method further comprising:

receiving a second anonymized record comprising a second plurality of biometric identifiers, each of the second plurality of biometric identifiers comprising a second numeric value, and

determining a match between the first anonymized record and the second anonymized record based on comparing the first and the second numeric values of each of the first and second plurality of biometric identifiers;

wherein the match comprises a confidence interval and/or a confidence level.

14. The method according to claim 13, wherein determining the match comprises evaluating a closeness between the first and the second numeric values.

15. A system, comprising:

a biometric device for obtaining biometric information associated with a person, and

a processor configured to execute the method according to any one of claims 1 to 14.

16. A computer readable medium having instructions stored thereon that when executed by a processor, cause the processor to perform the method according to any one of claims 1 to 14.