РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2538913

СПОСОБ ДЕПЕРСОНАЛИЗАЦИИ ПЕРСОНАЛЬНЫХ ДАННЫХ

Патентообладатель(ли): ***федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики" (Университет ИТМО) (RU)***

Автор(ы): ***Куракин Александр Сергеевич (RU)***

Заявка № **2012144274**

Приоритет изобретения **16 октября 2012 г.**

Зарегистрировано в Государственном реестре изобретений Российской Федерации ***24 ноября 2014 г.***

Срок действия патента истекает **16 октября 2032 г.**

Врио руководителя Федеральной службы по интеллектуальной собственности

Л.Л. Кирий

РОССИЙСКАЯ ФЕДЕРАЦИЯ

(19) **RU** (11) **2 538 913** (13) **C2**

(51) МПК
G06F 21/00 (2013.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2012144274/08, 16.10.2012**

(24) Дата начала отсчета срока действия патента:
16.10.2012

Приоритет(ы):
(22) Дата подачи заявки: **16.10.2012**

(43) Дата публикации заявки: **27.04.2014** Бюл. № 12

(45) Опубликовано: **10.01.2015** Бюл. № 1

(56) Список документов, цитированных в отчете о поиске: **RU 2439693 C1, 10.01.2012. RU 2309450 C1, 27.10.2007. RU 2263950 C2, 10.11.2005. US 7975149 B2, 05.07.2011. US 8074283 B2, 06.12.2011**

Адрес для переписки:
197101, Санкт-Петербург, Кронверкский пр., 49, Университет ИТМО, ОИС и НТИ

(72) Автор(ы):
Куракин Александр Сергеевич (RU)

(73) Патентообладатель(и):
федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики" (Университет ИТМО) (RU)

(54) СПОСОБ ДЕПЕРСОНАЛИЗАЦИИ ПЕРСОНАЛЬНЫХ ДАННЫХ

(57) Реферат:

Изобретение относится к области защиты информации, хранимой в информационных системах персональных данных (ИСПДн), от несанкционированного доступа (НСД) и может быть использовано на стадиях разработки и оптимизации ИСПДн в защищенном исполнении. Техническим результатом является повышение уровня безопасности ИСПДн. Способ деперсонализации персональных данных обеспечивает защиту ИСПДн от НСД на стадиях разработки и оптимизации, оперирует персональными данными субъектов, хранящимися и обрабатываемыми в ИСПДН, и осуществляет двухэтапное перемешивание данных, относящихся к разным субъектам, используя перестановки первого и второго уровней, при этом на первом этапе исходное множество данных $D(d_1, d_2, \ldots, d_N)$, где N - число атрибутов, разбивается на непересекающиеся подмножества данных A_i, относящихся к одному атрибуту d_i, а на втором этапе происходит непосредственно перестановка данных сначала внутри подмножеств A_i и затем элементами перестановки являются сами подмножества. При росте количества субъектов ПДн уменьшается вероятность подбора параметров деперсонализации, соответственно повышается защищенность ИСПДн. Разбиение исходного множества данных на подмножества позволяет сократить размерность задачи и упростить ее практическую реализацию. 1 з.п. ф-лы, 2 табл.

RU 2538913 C2

RU 2538913 C2

RUSSIAN FEDERATION

(19) **RU** (11) **2 538 913** (13) **C2**

(51) Int. Cl.
G06F 21/00 (2013.01)

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) ABSTRACT OF INVENTION

(21)(22) Application: **2012144274/08, 16.10.2012**

(24) Effective date for property rights:
16.10.2012

Priority:
(22) Date of filing: **16.10.2012**

(43) Application published: **27.04.2014** Bull. № **12**

(45) Date of publication: **10.01.2015** Bull. № **1**

Mail address:
197101, Sankt-Peterburg, Kronverkskij pr., 49, Universitet ITMO, OIS i NTI

(72) Inventor(s):
Kurakin Aleksandr Sergeevich (RU)

(73) Proprietor(s):
federal'noe gosudarstvennoe avtonomnoe obrazovatel'noe uchrezhdenie vysshego obrazovanija "Sankt-Peterburgskij natsional'nyj issledovatel'skij universitet informatsionnykh tekhnologij, mekhaniki i optiki" (Universitet ITMO) (RU)

(54) METHOD FOR DEPERSONALISATION OF PERSONAL DATA

(57) Abstract:

FIELD: physics, computer engineering.

SUBSTANCE: invention relates to protecting information stored in personal data information systems from unauthorised access and can be used at steps of developing and optimising secure personal data information systems. The technical result is a higher level of security of personal data information systems. The method for depersonalisation of personal data protects personal data information systems from unauthorised access at development and optimisation steps, handles personal data of subjects stored and processed in the personal data information systems and performs two-step movement of data relating to different subjects using first and second layer rearrangement, wherein at the first step a source data set $D(d_1, d_2, \ldots, d_N)$, where N is the number of attributes, is broken down into non-overlapping data subsets A_i relating to one attribute d_i, and at the second step the data are directly rearranged first within the subsets A_i and the rearrangement elements are then the subsets themselves. As the number of personal data subjects increases, the probability of selecting depersonalisation parameters decreases, thereby improving security of personal data information systems.

EFFECT: breaking down a source data set into subsets enables to reduce the dimension of the problem and simplify practical implementation thereof.

2 cl, 2 tbl

RU 2538913 C2

RU 2538913 C2

Изобретение относится к области защиты информации, хранимой в информационных системах персональных данных (ИСПДн), от несанкционированного доступа (НСД) и может быть использовано на стадиях разработки и оптимизации ИСПДн в защищенном исполнении.

Известен способ защиты от несанкционированного доступа к информации пользователя в системе обработки информации (патент RU №2309450, МПК G06F 12/14, дата приоритета 26.04.2006, дата публикации 27.10.2007) [1], основанный на том, что формирование сервисных служб системы обработки информации производится из доступного пользователю набора функциональных блоков, расположенных на различных серверах системы. Рабочая информация пользователя подвергается преобразованию, уникальному для каждого обращения пользователя к системе обработки информации, сведения о хранении учетной записи пользователя также подвергаются уникальному для данного случая преобразованию и сохраняются в других местах системы обработки информации. Недостатком известного технического решения является то, что выполнение указанных в способе требований влечет за собой значительные материальные затраты на внедрение дополнительных функциональных блоков.

Наиболее близким к предлагаемому изобретению по совокупности существенных признаков и принятым в качестве прототипа является способ защиты текстовой информации от несанкционированного доступа (патент RU №2439693, МПК G06F 21/24, дата приоритета 04.06.2010, дата публикации 10.01.2012) [2], использующий искажение в системах передачи данных без использования секретных ключей и пин-кодов. Способ включает: шифрование текстового сообщения A, его передачу, дешифрование принятого текстового сообщения A, предоставление восстановленного сообщения пользователю, при этом перед шифрованием на передающей стороне искажают исходное текстовое сообщение A с помощью известного пин-кода P путем отображения -ого слова, где $i = \overline{1, I}$, a - количество слов в исходном сообщении, в соответствующий код D_i по таблице возможных значений, замены кода D_i на код D_{K_i} путем сдвига кода D_i по кольцу на заданное в пин-коде P количество строк k в прямом направлении, а на приемной стороне после дешифрации восстанавливают код D_i принятого -ого слова путем сдвига кода D_{K_i} по кольцу на заданное в пин-коде P количество строк k в обратном направлении, осуществляют отображение кода D_i в соответствующее i-е слово по таблице возможных значений. Под таблицей возможных значений понимается таблица соответствия слов русского языка, находящихся в простой форме, расположенных в алфавитном порядке, которым поставлен в соответствие код W, равный логарифму по основанию 2 порядкового номера слова. При использовании данного способа из-за ошибочного приема злоумышленником хотя бы одного бита информации в силу свойств применяемого алгоритма шифрования, несмотря на то что при шифровании используется открытый ключ, при дешифровании перехваченных сообщений происходит лавинообразное размножение ошибок. После восстановления дешифрованных сообщений с ошибками в силу свойств применяемого алгоритма искажения текстовой информации будет получено множество текстов, слова в которых связаны грамматически, но автоматизированный анализ полученных текстов невозможен, то есть анализ может выполняться только экспертным путем.

К недостаткам данного способа можно отнести то, что при большой размерности задачи приходится хранить большую таблицу возможных значений для кодирования,

что понижает защищенность информации. При небольшом количестве слов в текстовой информации высока вероятность успешного применения метода полного перебора и получения исходного сообщения.

Задача, на решение которой направлено предлагаемое изобретение, заключается в разработке надежного способа деперсонализации персональных данных, позволяющего повысить уровень безопасности ИСПДн на стадиях разработки и оптимизации путем перемешивания персональных данных, относящихся к различным субъектам и снизить требования к уровню защищенности данных, сократив, тем самым, соответствующие расходы.

Сущность изобретения заключается в перемешивании персональных данных, хранящихся в ИСПДн, относящихся к различным субъектам. Данный способ обладает следующими преимуществами: персональные данные хранятся в одной информационной системе и значительно снижается вероятность успеха контекстного анализа.

В качестве исходных данных рассматривается таблица персональных данных $D(d_1, d_2, \ldots, d_N)$, где N - число атрибутов, а M - число строк таблицы, множество данных A_i, относящееся к одному атрибуту - $d_i(i=1, 2, \ldots, N)$. Все элементы каждого множества пронумерованы.

Способ обеспечивает перемешивание данных каждого множества атрибутов исходной таблицы пошагово. На каждом шаге используется принцип циклических перестановок.

На первом шаге множество данных A_i, относящееся к одному атрибуту, разбивается на K_i $(M>K_i>1)$ непересекающихся подмножеств, где число элементов подмножества A_{ij} равно $M_{ij}(M>M_{ij}>1)$, $j=1, 2, \ldots, K_i$. Разбиение каждого множества должно обладать следующими свойствами:

1) подмножества разбиения включают все элементы множества данных одного атрибута;

2) каждое подмножество не пусто, а пересечение любых двух подмножеств пусто;

3) все элементы в подмножествах упорядочены как по внутренним номерам (номера элементов внутри подмножества), там и по внешней нумерации самих подмножеств в разбиении;

4) суммарное число элементов всех подмножеств множества данных одного атрибута равно общему числу элементов этого множества.

Для каждого подмножества из разбиения определяется циклическая перестановка (подстановка) $p_{ij}(r_{ij})$, в которой производится циклический сдвиг всех элементов подмножества на некоторое число, называемое параметром перестановки. Таким образом, перестановки для всех подмножеств множества данных одного атрибута можно задать набором (вектором) параметров этих перестановок. Данный вектор задает первый уровень способа перемешивания, т.е. перестановки первого уровня.

На втором шаге способа рассматривается циклическая перестановка второго уровня $p_{0i}(r_{0i})$, элементами которой выступают подмножества, состоящие из K_i элементов, из описанного ранее разбиения. В результате применения данной перестановки производится циклический сдвиг элементов на некоторую величину - параметр перестановки второго уровня.

В результате последовательного проведения перестановок первого и второго уровней (или одной результирующей перестановки $p_i(r_{0i}, r_i,)$) получается перемешивание элементов множества данных одного атрибута так, что меняется нумерация этих элементов по отношению к исходной нумерации.

Доступность персональных данных (получение достоверных персональных сведений при легитимном обращении к ним) обеспечивается посредством решения обратного способа деперсонализации. Решением обратного способа деперсонализации является формирование исходной таблицы.

Для оценки защищенности предложенного способа деперсонализации используют такую характеристику, как число вариантов деперсонализации, получаемых при применении данного способа. При большом количестве записей число вариантов получается очень большим, что обеспечивает очень малую вероятность подбора параметров и соответственно хорошую защиту обезличенных данных.

В совокупности признаков заявленного способа используются следующие терминология и обозначения:

- запись в таблице - совокупность элементов множеств разных атрибутов с одинаковыми номерами, при этом в исходной таблице каждая запись имеет определенный смысл, связанный с конкретным субъектом (физическим лицом), т.е. содержит персональные данные конкретного лица, определенного в этой же записи;

- внешний номер m_{ijk} - номер элемента в подмножестве A_{ij}, имеющего внутренний номер k, $1 \le m_{ijk} \le M$, т.е. m_{ijk} - это порядковый номер элемента во множестве A_i, соответствующий элементу с внутренним номером k;

- циклическая перестановка первого уровня - перестановка, в которой элементы первой строки матрицы, стоящей в правой части равенства, соответствуют внутренним номерам элементов подмножества A_{ij} до перестановки (в исходной таблице), а элементы, стоящие во второй строке, соответствуют внутренним номерами элементов подмножества A_{ij}, стоящим на местах, с номерами, определенными в верхней строке, после перестановки:

$$p_{ij}(r_{ij}) = \begin{pmatrix} 1 & 2 & 3 & \dots & (M_{ij}-1) & M_{ij} \\ (M_{ij}-r_{ij}^{1}+1)(M_{ij}-r_{ij}^{2}+2)(M_{ij}-r_{ij}^{3}+3)\dots(M_{ij}-r_{ij}-1)(M_{ij}-r_{ij}) \end{pmatrix};$$

- параметр перестановки первого уровня r_{ij} - некоторое случайное число, задаваемое генератором случайных чисел (ГСЧ) в интервале $[1; M_{ij}-1]$;

- циклическая перестановка второго уровня - перестановка, в которой элементы верхней строки матрицы перестановки соответствуют исходным номерам подмножеств A_{ij}, а элементы нижней строки матрицы соответствуют номерам подмножеств A_{ij}, стоящим на местах с номерами, определенными в верхней строке, после перестановки:

$$p_{0i}(r_{0i}) = \begin{pmatrix} 1 & 2 & 3 & \dots & (K_i-1) & K_i \\ (K_i-r_{0i}^{1}+1)(K_i-r_{0i}^{2}+2)(K_i-r_{0i}^{3}+3)\dots(K_i-r_{0i}-1)(K_i-r_{0i}) \end{pmatrix};$$

- параметр перестановки второго уровня r_{0i} - некоторое случайное число, задаваемое генератором случайных чисел (ГСЧ) в интервале $[1; K_i-1]$,

- результирующая перестановка - полученная с учетом правил перемножения перестановок первого и второго уровней перестановка, в которой верхняя строка матрицы содержит порядковые номера элементов множества атрибута i, в соответствии с их размещением в столбце после перестановок, а нижняя строка содержит внешние номера элементов множества этого атрибута, соответствующие их размещению в исходной таблице:

$$p_i(r_{0i}, r_i) = \begin{pmatrix} [1 \dots M_{i(K_i - r_{0i}+1)}] & [(M_{i(K_i-r_{0i}+1)} + 1) \dots (M_{i(K_i-r_{0i}+1)} + M_{i(K_i-r_{0i}+2)})] \\ [m_{i(K_i-r_{0i}+1)1} \dots m_{i(K_i-r_{0i}+1)M_{i(K_i-r_{0i}+1)}}] & [m_{i(K_i-r_{0i}+2)1} \dots m_{i(K_i-r_{0i}+2)M_{i(K_i-r_{0i}+2)}}] \\ \dots & [M - M_{i(K_i-r_{0i})} \quad \dots \quad M] \\ \dots & [m_{i(K_i-r_{0i})1} \dots m_{i(K_i-r_{0i})M_{i(K_i-r_{0i})}}] \end{pmatrix}$$

Применение данного способа позволяет обеспечить защиту персональных сведений от несанкционированного доступа, в том числе от компрометации информации при ее утечке по техническим каналам, а также обеспечить гарантированный доступ к персональным данным при легитимном обращении. При этом все персональные сведения хранятся в одной таблице, а их получение посредством контекстного анализа или путем перебора весьма трудоемко, а зачастую практически невозможно. Практическое применение данного способа является аналогом абонентского шифрования. Его реализация подразумевает, что персональные данные хранятся на постоянном запоминающем устройстве (ПЗУ) в деперсонализированном виде. При необходимости работы с персональными данными оператор применяет обратный алгоритм деперсонализации (запускает программу работы с персональными данными, реализующую прямой и обратный алгоритм). Следует отметить, что открытая (персонализированная) информация, с которой работает оператор, как правило, хранится в ОЗУ и только по завершении работы (или команде сохранения/синхронизации) записывается в файл в ПЗУ, где она хранится только в закрытом виде.

Эти отличительные признаки по сравнению с прототипом позволяют сделать вывод о соответствии заявляемого технического решения критерию «новизна».

Новое свойство совокупности существенных признаков, приводящих к существенному затруднению НСД к персональной информации, хранящейся и обрабатываемой в ИСПДн, путем перемешивания данных, относящихся к различным субъектам, позволяет сделать вывод о соответствии предлагаемого технического решения критерию «изобретательский уровень».

Предлагаемый способ защиты ПДн от НСД опробован в лабораторных условиях. Способ деперсонализации может быть реализован в виде программного обеспечения на языке программирования C#. Исходные данные могут подаваться на вход в виде текстового файла. Также возможна реализация, в которой данные на вход программы поступают непосредственно из информационной системы. Параметры разбиений исходных множеств данных могут задаваться как пользователем, так и программой, используя генератор случайных чисел (ГСЧ).

В результате работы программы пользователь получает деперсонализированные данные в той же форме, в которой они подавались на вход. Кроме того, создается файл, хранящий параметры перестановок и разбиений, который будут необходимы для решения обратного способа деперсонализации.

Для простоты описания работы устройства представим, что алгоритм перестановки, определенный для множества, соответствующего одному атрибуту, применяется ко всем множествам атрибутов исходной таблицы. В этом случае полный алгоритм перестановки задается следующим набором параметров:

1. (K_1, K_2, …, K_N) - множество, определяющее количество подмножеств для множества каждого атрибута, которое определяет подмножества элементов

$(A_{11}, A_{12}, \ldots, A_{1K_1}), (A_{21}, A_{22}, \ldots, A_{2K_2}), \ldots, (A_{N1}, A_{N2}, \ldots, A_{NK_N})$;

2. $((M_{11}, M_{12}, \ldots, M_{1K_1}), (M_{21}, M_{22}, \ldots, M_{2K_2}), \ldots, (M_{N1}, M_{N2}, \ldots, M_{NK_N})$-множество, определяющее число элементов в подмножествах для множества каждого атрибута;

3. $((r_{01}, r_1), (r_{02}, r_2), \ldots, (r_{0N}, r_N))$ - множество параметров перестановок для множества каждого атрибута. Этот набор задает параметры алгоритма деперсонализации для исходной таблицы $D(d_2, d_2, \ldots, d_N)$.

В результате применения процедуры вместо исходной таблицы $D(d_2, d_2, \ldots, d_N)$ получается таблица обезличенных данных $\tilde{D}(d_1, d_2, \ldots, d_N)$.

Набор параметров:

$C(D(d_1, d_2, \ldots, d_N))=\{(K_1, K_2, \ldots, K_N),$

$((M_{11}, M_{12}, \ldots, M_{1K_1}), (M_{21}, M_{22}, \ldots, M_{2K_2}), \ldots, (M_{N1}, M_{N2}, \ldots, M_{NK_N})$,

$((r_{01}, r_1), (r_{02}, r_2), \ldots, (r_{0N}, r_N))\}$

полностью и однозначно задает алгоритм деперсонализации для исходной таблицы $D(d_1, d_2, \ldots, d_N)$.

Пусть исходная таблица $D(d_1, d_2, \ldots, d_N)$ имеет вид (таблица 1):

Таблица 1

Исходная таблица данных

Атрибут d_1	Атрибут d_2	Атрибут d_3	Атрибут d_4	Атрибут d_5	Атрибут d_6
q_1	r_1	s_1	t_1	u_1	v_1
q_2	r_2	s_2	t_2	u_2	v_2
q_3	r_3	s_3	t_3	u_3	v_3
q_4	r_4	s_4	t_4	u_4	v_4
q_5	r_5	s_5	t_5	u_5	v_5
q_6	r_6	s_6	t_6	u_6	v_6
q_7	r_7	s_7	t_7	u_7	v_7
q_8	r_8	s_8	t_8	u_8	v_8
q_9	r_9	s_9	t_9	u_9	v_9
q_{10}	r_{10}	s_{10}	t_{10}	u_{10}	v_{10}

Для этой таблицы заданы следующие параметры алгоритма деперсонализации:

$C(D(d_1, d_2, d_3, d_4, d_5, d_6))$

$=\{(3,2,4,3,3,2), ((3,3,4), (6,4), (2,3,2,3), (3,4,3), (5,2,3), (3,7)),$

$((2, (1,2,3)), (1, (3,1)), (3, (1,2,1/1)), (2, (2,1,2)), (2, (4,1,1)), (1, (1/4)))\}$.

После выполнения алгоритма деперсонализации получаем таблицу 2 - $\tilde{D}(d_1, d_2, \ldots, d_N)$.

Таблица 2

Таблица обезличенных данных

Атрибут d_1	Атрибут d_2	Атрибут d_3	Атрибут d_4	Атрибут d_5	Атрибут d_6
q_{10}	r_8	s_9	t_{10}	u_9	v_8
q_7	r_9	s_{10}	t_8	u_{10}	v_9
q_8	r_{10}	s_8	t_9	u_8	v_{10}
q_9	r_7	s_2	t_3	u_5	v_4

q_2	r_4	s_1	t_1	u_1	v_5
q_3	r_5	s_5	t_2	u_2	v_6

Атрибут d_1	Атрибут d_2	Атрибут d_3	Атрибут d_4	Атрибут d_5	Атрибут d_6
q_1	r_6	s_3	t_5	u_3	v_7
q_6	r_1	s_4	t_6	u_4	v_2
q_4	r_2	s_7	t_7	u_7	v_3
q_5	r_3	s_6	t_4	u_6	v_1

Как видно из примера, в результате применения алгоритма деперсонализации получена преобразованная таблица, в которой записи не соответствуют записям в исходной таблице, что обеспечивает достаточно высокую сложность восстановления исходной таблицы при отсутствии сведений о параметрах алгоритма деперсонализации.

Реализация предлагаемого способа не вызывает затруднений, так как блоки и узлы общеизвестны и широко описаны в технической литературе.

Таким образом, заявляемый способ деперсонализации персональных данных позволяет повысить уровень безопасности ИСПДн на стадиях разработки и оптимизации путем перемешивания персональных данных, относящихся к различным субъектам и снизить требования к обеспечению надлежащего уровня защищенности данных, сократив, тем самым, соответствующие расходы.

Источники информации

1. Патент RU №2309450 «Способ защиты от несанкционированного доступа к информации пользователя в системе обработки информации». G06F 12/14, дата приоритета 26.04.2006, дата публикации 27.10.2007.

2. Патент RU №2439693 «Способ защиты текстовой информации от несанкционированного доступа» МПК G06F 21/24, дата приоритета 04.06.2010, дата публикации 10.01.2012.

3. Куракин А.С. Алгоритм деперсонализации персональных данных // Научно-технический вестник информационных технологий, механики и оптики. СПб НИУ ИТМО, 2012. Выпуск №6.

4. Стенли Р. Перечислительная комбинаторика. М.: Мир, 1990. 440 с.

Формула изобретения

1. Способ деперсонализации персональных данных, заключающийся в производимом при передаче в оперативное запоминающее устройство преобразовании информации, составляющей по совокупности персональные данные, из постоянного запоминающего устройства путем выполнения двухэтапного перемешивания данных, относящихся к разным субъектам, используя перестановки первого и второго уровней, при этом на первом этапе исходное множество данных $D(d_1,d_2,\ldots,d_N)$, где N - число атрибутов, разбивается на непересекающиеся подмножества данных A_i, относящихся к одному атрибуту d_i, и на втором этапе происходит непосредственно перестановка данных сначала внутри подмножеств A_i, а затем элементами перестановки являются сами подмножества; для представления ее пользователю и обратном преобразовании информации при ее записи - передаче из оперативного запоминающего устройства в постоянное запоминающее устройство.

2. Способ по п.1, отличающийся тем, что параметры разбиений исходных множеств данных задаются при помощи генератора случайных чисел.