УДК 621.383

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ БОЛЬШИХ ДАННЫХ НА ОСНОВЕ ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ В СООТВЕТСТВИИ С ТРЕБОВАНИЯМИ ЗАКОНОДАТЕЛЬСТВА РБ

А.М. Тимофеев
Доцент кафедры защиты информации БГУИР, кандидат технических наук, доцент
TAMvks@mail.ru

А.Н. Шишпаренок
Учащийся учреждения образования «Национальный детский технопарк» по направлению «Информационная безопасность»
sashapistol22102008@gmail.com

В.Е. Юруть
Учащийся учреждения образования «Национальный детский технопарк» по направлению «Информационная безопасность»
vadimurut6@gmail.com

А.М. Тимофеев
Окончил Учреждение образования «Высший государственный колледж связи». Область научных интересов связана с фотоэлектронными процессами в фотоприемниках при одноквантовой регистрации, методами регистрации сверхслабых оптических потоков, а также комплексным обеспечением защиты информации в системах и сетях связи.

А.Н. Шишпаренок
Учащийся ГУО «Средняя школа №20 г. Борисова» (9 класс), учащийся учреждения образования «Национальный детский технопарк» в рамках индивидуальной учебной программы дополнительного образования одаренных детей и молодежи по направлению «Информационная безопасность» (22.11.2023 г. – 22.02.2024 гг.), научный руководитель доцент А.М. Тимофеев.

В.Е. Юруть
Учащийся ГУО «Средняя школа №3 г. Волковыска» (11 класс), учащийся учреждения образования «Национальный детский технопарк» в рамках индивидуальной учебной программы дополнительного образования одаренных детей и молодежи по направлению «Информационная безопасность» (22.11.2023 г. – 22.02.2024 гг.), научный руководитель доцент А.М. Тимофеев.

Аннотация. Применительно к системам обработки больших данных, содержащих персональные данные физических лиц, получена реализация метода введения идентификатора с учетом требований законодательства Республики Беларусь в сфере защиты информации.

Предложенная реализация позволяет выполнить обезличивание персональных данных, в результате чего нелегитимному пользователю системы становится невозможным без использования дополнительной информации определить принадлежность этих данных конкретному физическому лицу.

Ключевые слова: обезличивание, персональные данные, обработка персональных данных, субъект персональных данных, идентификатор.

Введение. В настоящее время существует потребность создания и обработки огромных массивов данных, которые содержат информацию различного назначения. Зачастую в составе таких массивов присутствуют персональные данные [1-4].

Под персональными данными будем понимать любую информацию, относящуюся к идентифицированному физическому лицу или физическому лицу, которое может быть идентифицировано [5].

Одной из основных задач при этом является обеспечение информационной безопасности персональных данных, поскольку собственник (владелец) информационной системы (или систем) обязан выполнять работы по защите информации этих данных, если они не являются общедоступными персональными данными [6].

Общедоступными персональными данными называют персональные данные, распространенные самим субъектом персональных данных либо с его согласия, либо распространенные в соответствии с требованиями законодательных актов [6].

Субъект персональных данных – это физическое лицо, в отношении которого осуществляется обработка персональных данных [6].

Обработка персональных данных – это действие или совокупность действий, совершаемых с персональными данными, включающие сбор, систематизацию, хранение, изменение, использование, обезличивание, блокирование, распространение, предоставление, удаление персональных данных [6].

Под обезличиванием будем понимать действия, в результате которых становится невозможным без использования дополнительной информации
определить принадлежность персональных данных к конкретному субъекту персональных данных [7, 8].

В соответствии с [6] для защиты персональных данных необходимо использовать методы их обезличивания. Одним из таких методов является метод введения идентификатора.

Метод введения идентификатора реализуется путем замены персональных данных, или части персональных данных, позволяющих идентифицировать субъект персональных данных, их идентификатором и создания таблицы соответствия с последующим раздельным хранением идентификаторов и таблиц [1]. В связи с тем, что до настоящего времени отсутствуют технические решения, которые позволяют выполнить обезличивание персональных данных, содержащихся в базах данных, относящихся к категории больших, целью данной работы являлась разработка технического решения по обеспечению информационной безопасности таких данных.

Объект исследования: структурированные и неструктурированные данные огромных объемов (большие данные).

Предмет исследования: применение метода введения идентификаторов для обеспечения информационной безопасности больших данных, содержащих персональные данные.

Реализация метода введения идентификатора. На рисунке 1 представлена программная реализация метода введения идентификатора.

При реализации метода введения идентификаторов важно, чтобы обезличенные данные можно было восстановить, применив процедуру деобезличивания [8].

Обезличенные данные – это действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных к конкретному субъекту персональных данных [8].

```rust
use bitvec::prelude::*;
use des::{Block, FromHexStr, MainKey, ToHexString};
use miette::{IntoDiagnostic, Result};
use std::io;

fn main() -> Result<()> {
    let main_key = MainKey::from_hex_str("AAAABBBBCCCCDDDD").into_diagnostic()?;
    let iv = Block::from_hex_str("FFFFFFFFFFFFFFFF").into_diagnostic()?;
    let iv = iv.encode(&main_key).into_diagnostic()?;

    println!("Input data in hex format (as one hex string): ");
    let input = get_input_string().into_diagnostic()?;
    let input = input.to_lowercase();
    let input = input.trim_start_matches("0x");
    let input = BitVec::from_hex_str(input).into_diagnostic()?;

    let output = encode(input, iv).to_hex_string();
    let (output, output_end) = output.split_at(output.len() - 16);
    let output_end = output_end.trim_start_matches('0');
    println!("\n{}{}", output, output_end);

    Ok(())
}

fn encode(input: BitVec, encoded_iv: Block) -> BitVec {
    let iv = encoded_iv.into_bitvec();
    let mut output: BitVec<usize, bitvec::order::LocalBits> = BitVec::with_capacity(input.len());
    let input = input.chunks(64);
    for chunk in input {
        let mut new_chunk = chunk.to_bitvec();
        new_chunk ^= iv.clone();
        output.extend(new_chunk);
    }
    output
}

fn get_input_string() -> io::Result<String> {
    let mut input = String::new();
    io::stdin().read_line(&mut input)?;
    Ok(input.trim().to_string())
}
```

исходные коды на языке программирования *Rust*

Рисунок 1. Реализация метода введения идентификаторов для обеспечения информационной безопасности больших данных, содержащих в своем составе персональные данные

Под деобезличиванием будем понимать действия, в результате которых обезличенные данные принимают вид, позволяющий определить их принадлежность к конкретному субъекту персональных данных, то есть, становятся персональными данными [8].

В этой связи применение методов, не обладающих свойством обратимости, нецелесообразно. Например, в случае реализации метода введения идентификатора с использованием односторонних функций хеширования (или односторонних функций, или дайджест-функций, или хеш-функций) отсутствует возможность выполнить процедуру деобезличивания [9-11].

В настоящей работе предложена и выполнена реализация метода введения идентификаторов, а также определены базовые принципы выполнения процедур обезличивания и деобезличивания персональных данных. Разработана компьютерная программа, которая позволяет выполнять процедуры обезличивания и деобезличивания персональных данных. В качестве языка программирования выбран язык *Rust* (рисунок 1). Отметим, что данная программная реализация может быть выполнена на любом языке программирования высокого уровня [12-17], поскольку не требует сложных математических операций.

Сущность технического решения, реализованного в компьютерной программе, заключается в следующем. Вначале персональные данные, подлежащие обезличиванию, кодируют с применением стандартных кодировочных таблиц, например, *UTF*-16. Важно отметить, что в качестве кодировочного формата может быть использована любая другая стандартная система, отличная от *UTF*-16, либо собственная («внутрифирменная») кодировочная таблица.

Затем формируют первый блок идентификатора ID_1. Для этого вначале зашифровывают начальный вектор, который также называют вектором инициализации, или синхропосылкой [9, 10]. Далее применяют стандартный блочный шифр. В качестве такого шифра в работе использован стандарт DES, поскольку он прост в реализации, имеет достаточно высокий уровень информационной безопасности (с учетом схемных решений и режим его использования [9, 10]) и характеризуется высокой скоростью работы, что является весьма важным при обработке больших данных [1, 2]. Важно отметить, что вместо стандарта *DES* допустимо использовать стандарты *AES*, ГОСТ 28147-89, СТБ 34.101.31-2020 либо любой другой блочный шифр, имеющий криптостойкость, достаточную с учетом требований криптографической защиты информации, предъявляемых при эксплуатации информационной системы [9-11]. Результат зашифрования суммируют по модулю 2 с первым блоком закодированных персональных данных, получая первый блок идентификатора ID_1.

Важно отметить, что разрядность блока закодированных персональных данных целесообразно выбирать равной разрядности двоичного числа, полученного на выходе используемого блочного шифра. Например, в случае применения СТБ 34.101.31-2020, разрядность такого блока может быть в диапазоне от 1 до 128 бит.

Аналогичным образом формируют остальные блоки идентификатора $ID_2 \div ID_N$ (где N – общее число блоков, определяемых объемом персональных данных), за исключением следующего. Перед зашифрованием очередного блока вначале обновляют входное значение, удалив старшие биты в количестве, равном выбранной разрядности блока. После этого оставшиеся блоки сдвигают влево на такое число двоичных разрядов, а в освободившееся битовое пространство записывают предыдущий блок идентификатора. Например, при выработке второго блока ID_2 предыдущим блоком идентификатора является ID_1.

Заключение. По результатам выполненной работы предложена реализация метода введения идентификаторов применительно к обезличиванию персональных данных, содержащихся в больших данных.

Определены принципы обеспечения информационной безопасности на основе криптографических и криптоподобных преобразований информации, полученная программная реализация метода введения идентификаторов для обеспечения информационной безопасности больших данных, содержащих в своем составе персональные данные.

Установлено, что полученные результаты могут быть использованы для обеспечения информационной безопасности персональных данных как при их обезличивании, так и

при деобезличивания, включая все основные этапы по обработке таких данных – сбор, систематизацию, хранение, изменение, использование и предоставление.

Работа выполнена при поддержке Учреждения образования «Национальный детский технопарк» (индивидуальная учебная программа дополнительного образования одаренных детей и молодежи по направлению «Информационная безопасность»).

Список литературы

[1] Big Data = Большие данные : учеб. пособие / И. Б. Тесленко [и др.] ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир: ВлГУ, 2021. – 123 с

[2] Коллинз, М. Защита сетей. Подход на основе анализа данных / М. Коллинз. – Москва : ДМК Пресс, 2020. – 308 с.

[3] Диогенес, Ю. Кибербезопасность: стратегии атак и обороны / Ю. Диогенес, Э. Озкайя. – Москва : ДМК Пресс, 2020. – 326 с.

[4] Семкин, С. Н. Основы правового обеспечения защиты информации : учеб. пособие для вузов / С. Н. Семкин, А. Н. Семкин. – М. : Горячая линия-Телеком, 2008. – 238 с.

[5] Национальный реестр. Регистрационный номер 2/2819. Закон Республики Беларусь от 7 мая 2021 г. № 99-З «О защите персональных данных» [Электронный ресурс]. – Режим доступа : https://pravo.by/document/?guid=12551&p0=H12100099. – Дата доступа: 4.02.2024.

[6] Национальный реестр. Регистрационный номер 7/4470. Приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 20 февраля 2020 г. № 66 «О мерах по реализации Указа Президента Республики Беларусь от 9 декабря 2019 г. № 449» [Электронный ресурс]. – Режим доступа : https://pravo.by/document/?guid=12551&p0=T62004470. – Дата доступа: 4.02.2024.

[7] Закон Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации» [Электронный ресурс]. – Режим доступа : https://pravo.by/document/?guid=2012&oldDoc=2008-279/2008-279(014-027).pdf&oldDocPage=1. – Дата доступа: 4.02.2024.

[8] Федеральная служба по надзору в сфере связи, информационных технологий и массовых коммуникаций Российской Федерации. Методические рекомендации по применению Приказа Роскомнадзора от 5 сентября 2013 г. № 996 «Об утверждении требований и методов по обезличиванию персональных данных» [Электронный ресурс]. – Режим доступа : https://rkn.gov.ru/docs/Xerox_Phaser_3200MFP_20131216122746.pdf. – Дата доступа: 4.02.2024.

[9] Тимофеев, А.М. Криптографическая защита информации: учеб.-метод. пособие / А.М. Тимофеев. – Мн.: БГУИР, 2020. – 112 с.

[10] Бутакова, Н.Г. Криптографические методы и средства защиты информации: учебное пособие / Н.Г. Бутакова, Н.В. Федоров. – СПб.: Интермедия, 2020. – 380 с.

[11] Казарин, О.В. Надежность и безопасность программного обеспечения: учебное пособие / О.В. Казарин, И.Б. Шубинский. – М.: Юрайт, 2023. – 342 с.

[12] Programming Rust / J. Blandy, J. Orendorff, L. Tindall. – Sebastopol, Ca: O'Reilly Media, 2021. – 1282 p.

[13] Klabnik, S. The Rust Programming Language / S. Klabnik, C. Nichols. – San Francisco, Ca: No Starch Press, 2021. – 755 p.

[14] Gjengset, J. Rust for rustaceans / J. Gjengset. – San Francisco, Ca: No Starch Press, 2022. – 283 p.

[15] Прайс, М. C# 10 и .NET 6. Современная кросс-платформенная разработка / М. Прайс. – 6-е изд. – СПб.: Питер, 2023. – 848 с.

[16] Керниган, Б. Практика программирования / Б. Керниган, Р. Пайк. – Москва ; Санкт-Петербург : Диалектика, 2019. – 288 с.

[17] Плас, Дж. Вандер. Python для сложных задач : наука о данных и машинное обучение / Плас Дж. Вандер. – Санкт-Петербург : Питер, 2023. – 576 с.

[18]

Авторский вклад

Тимофеев Александр Михайлович – научное руководство исследованием, постановка цели и задач исследования, разработка технического решения по реализации метода обезличивания персональных данных.

Шишпаренок Александр Николаевич – программная реализация метода обезличивания персональных данных, оценка обратимости процедуры обезличивания персональных данных (возможности проведения деобезличивания персональных данных), изучение принципов работы Big Data.

Юруть Вадим Евгеньевич – аналитический обзор литературных источников в рамках исследовательской работы, выполнение расчетной части, оценка свойств обезличенных данных, полученных по результатам предложенной в работе программной реализации метода обезличивания персональных данных, изучение принципов работы Big Data.

BIG DATA INFORMATION SECURITY

A.M. Timofeev
Associate Professor of the Department of Information Security of BSUIR, PhD of Technical Sciences, Associate Professor

A.N. Shyshparonak
Student of the educational institution «National Children's Technopark» in «Information Security»

V.E. Yuruts
Student of the educational institution «National Children's Technopark» in «Information Security»

Abstract. With regard to big data processing systems containing personal data of individuals, the implementation of the identifier input method has been obtained, taking into account the requirements of the legislation of the Republic of Belarus in the area of information protection.

The proposed implementation makes it possible to depersonalize personal data, as a result it becomes impossible for an illegitimate user of the system to determine the belonging of this data to a specific individual without the use of additional information.

Keywords: depersonalization, personal data, processing of personal data, personal data subject, identifier.