

**УМЕНЬШЕНИЕ ОБЪЕМА ХРАНИМЫХ ПЕРСОНАЛЬНЫХ
ДАННЫХ ПРИ ИХ ОБЕЗЛИЧИВАНИИ НА ОСНОВЕ МЕТОДА
ИЗМЕНЕНИЯ СОСТАВА**

А.М. Тимофеев¹, Д.В. Шляхтун²

¹Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», г. Минск, Республика Беларусь

²Учреждение образования «Национальный детский технопарк», г. Минск, Республика Беларусь

Аннотация. В работе представлена процедура обезличивания персональных данных на основе метода изменения состава с акцентом на оптимизацию ресурсов хранения. Научная новизна заключается в использовании 9-битного кодирования алфавитно-цифровых символов, что позволяет исключить избыточное представление данных, характерное для UTF-16. Предложенный алгоритм не только обеспечивает высокий уровень информационной безопасности за счет изменения статистических свойств

персональных данных, но и уменьшает объем хранимых данных. Это делает процедуру эффективной для применения в крупномасштабных информационных системах с ограниченными дисковыми ресурсами.

Ключевые слова: информационные системы; персональные данные; защита информации; обезличивание персональных данных; методы обезличивания персональных данных; метод изменения состава.

REDUCING THE VOLUME OF STORED PERSONAL DATA BY DEPERSONALIZING IT BASED ON THE METHOD OF CHANGING THE COMPOSITION

Timofeev A.¹, Shliakhtun D.²

¹*Educational Institution "Belarusian State University of Informatics and Radioelectronics", Minsk, Republic of Belarus*

²*Educational Institution "National Children's Technopark", Minsk, Republic of Belarus*

Abstract. This article presents a personal data anonymization procedure based on a compositional modification method with an emphasis on storage resource optimization. The scientific innovation lies in the use of 9-bit alphanumeric encoding, which eliminates the redundant data representation typical of UTF-16. The proposed algorithm not only ensures a high level of information security by modifying the statistical properties of personal data but also reduces the volume of stored data. This makes the procedure effective for use in large-scale information systems with limited disk resources.

Keywords: information systems; personal data; information protection; depersonalization of personal data; methods of depersonalization of personal data; method of changing the composition.

Введение

Одной из ключевых задач при создании современных информационных систем является обезличивание персональных данных [1 – 3]. Применяемые в настоящее время методы позволяют скрыть сведения от посторонних, сохраняя доступ к ним только для легитимных пользователей с помощью секретных параметров деобезличивания. Законодательство Республики Беларусь допускает использование метода изменения состава, который предполагает обобщение, модификацию или удаление идентифицирующей информации, для обезличивания персональных данных.

Однако традиционные подходы к реализации этого метода имеют недостатки: уязвимость перед статистическим анализом и необходимость выделения памяти для хранения измененных фрагментов. Более того, использование стандартных кодировок, например, UTF-16, создает избыточность, увеличивая нагрузку на инфраструктуру хранения и каналы передачи данных.

В данной работе представлена процедура обезличивания персональных данных на основе метода изменения состава, лишенная указанных недостатков. Предложенное решение ориентировано не только на модификацию статистических свойств персональных данных для повышения уровня их информационной безопасности, но и на значительное сокращение их объема.

Целью исследования являлась разработка процедуры обезличивания персональных данных, повышающей уровень их информационной безопасности и уменьшающей объем хранимых данных за счет оптимизации их бинарного представления. Объектом работы являлся метод изменения состава. Предметом исследования являлось повышение уровня информационной безопасности персональных данных с уменьшением используемого для их хранения объема баз данных.

Кодирование персональных данных

Предложенная процедура обезличивания персональных данных основана на переходе от стандартного шестнадцатибитного кодирования к оптимизированному девятибитному представлению символов. Это позволяет устранить структурную избыточность данных и уменьшить их объем для последующего хранения. Процесс подготовки информации заключается в преобразовании каждого символа исходного текста в компактный девятибитный блок. На основе анализа старшего байта исходного кода формируется бит-селектор: значение 0 соответствует кириллическому диапазону с префиксом 04, а значение 1 используется для латиницы и служебных знаков с префиксом 00. К сформированному флагу добавляется восьмибитное значение младшего байта символа. Переход к использованию девяти бит вместо шестнадцати обеспечивает сокращение объема хранимых данных на 43,7% в сравнении с исходным представлением в формате UTF-16.

Сформированная бинарная последовательность выравнивается до кратности шестидесяти четырем битам и подвергается криптографическому преобразованию методом наложения гаммы. Применение функции формирования ключа на основе пароля и добавления случайного шестнадцатибитного числа обеспечивает изменение статистических свойств данных, надежно защищая их от несанкционированного семантического анализа. Восстановление информации легитимным пользователем происходит путем расшифрования и декомпозиции потока блоками по девять бит. Алгоритм считывает бит-селектор, автоматически реконструирует соответствующий префикс и объединяет его с последующими восемью битами для получения исходного кода символа. Данный механизм исключает необходимость хранения дополнительных таблиц соответствия

или исходных фрагментов данных, так как вся информация для деобезличивания извлекается непосредственно из структуры девятибитных блоков. В результате предложенный метод позволяет существенно снизить требования к ресурсам систем хранения при обеспечении высокого уровня информационной безопасности.

Заключение

Разработана процедура обезличивания персональных данных на основе метода изменения состава, направленная на повышение уровня их информационной безопасности и оптимизацию ресурсов хранения.

В работе представлен способ 9-битного кодирования символов, позволяющий отказаться от избыточного представления данных. Предложенный подход обеспечивает снижение физического объема хранимых данных на 43,7% и сокращение длины соответствующих битовых потоков.

Список использованных источников

1. Ворона, В. А. (2023) *Биометрическая идентификация личности*. Москва, Горячая линия-Телеком.
2. Коллинз, М. (2020) *Защита сетей. Подход на основе анализа данных*. Москва, ДМК Пресс.
3. Остапенко, Г. А. (2020) *Информационные операции и атаки в социотехнических системах: организационно-правовые аспекты противодействия*. Москва, Горячая линия-Телеком.

References

1. Vorona V. A. (2023) *Biometric Identification of Personality*. Moscow, Goryachaya Liniya-Telecom (in Russian).
2. Collins M. (2020) *A Data-Based Approach*. Moscow, DMK Press (in Russian).
3. Ostapenko G. A. (2020) *Information Operations and Attacks in Socio-Technical Systems: Organizational and Legal Aspects of Counteraction*. Moscow, DMK Press (in Russian).

Сведения об авторах

Тимофеев А.М., канд. техн. наук, доц., доц. каф. защиты информации, Учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», TAMvks@mail.ru

Шляхтун Д.В., учащаяся Учреждения образования «Национальный детский технопарк» по направлению «Информационная безопасность», rd2859002@gmail.com

Information about the authors

Timofeev A., Cand. Sci. (Tech.), Associate Professor, Associate Professor of the Department of Information Security, Educational Institution "Belarusian State University of Informatics and Radioelectronics", TAMvks@mail.ru

Shliakhtun D., Student of the Educational Institution "National Children's Technopark" in "Information Security", rd2859002@gmail.com