Ежегодная международная научно-практическая конференция
«РусКрипто'2022»

конференция РусКрипто

Безопасность персональных данных: новый взгляд на старую проблему

Профессор кафедры БИТ д.т.н. Минзов Анатолий Степанович
Заведующий кафедрой БИТ к.т.н. Невский Александр Юрьевич
Доцент кафедры БИТ НИУ к.т.н. Баронов Олег Рюрикович

МЭИ

Содержание

1. Почему проблемы защиты персональных данных все еще остаются актуальными ?
2. Современная трактовка содержания термина «персональные данные» в стандарте GPRS. Что за ней стоит ?
3. Новый взгляд на защиту персональных данных субъекта. Как защитить субъект ПДН от угроз при их разглашении ?
4. Некоторые выводы из этого исследования.

1.Почему проблемы защиты персональных данных все еще остаются актуальными ?

История защиты персональных данных

- Под ***privacy*** в общем понимается неприкосновенность частной жизни.
- В 1890 году два американских юриста – Сэмюэль Уоррен и Луи Брэндайс – определили это понятие как «*право быть оставленным в покое*».
- В 1948 году во 2-й статьи Всеобщей декларации прав человека, утвержденной ООН определялось, что никто, иначе как по законному решению суда или уполномоченных органов, не может посягать на тайну личной жизни и переписки человека.
- В 1981 году Конвенция, принятая Советом Европы, посвящена защите прав и свобод граждан при автоматической обработке их персональных данных.
- В 2006 появился первый вариант ФЗ РФ №152 «О персональных данных».
- В 2018 году появился новый регламент защиты персональных данных GDPR (General Data Protection Regulation).
- Частично эти изменения были уточнены в ФЗ РФ №152

Модель защиты информации

(Постановление Правит.РФ №1119 от 01.11.2012 N 1119)

Исходные данные:

T – актуальные угрозы, $T = \{t_1, t_2, t_3\}$, где t_1, t_2, t_3 типы угроз $t_1 > t_2 > t_3$.

K – категория ПДН, $K = \{k_0, k_1, k_2, k_3, k_4\}$, (k_0 – обезличенные данные; k_1 – общедоступные данные; k_2 – специальные категории ПДН; k_3 – биометрические данные; k_4 – иные ПДН).

n^0 – количество штатных сотрудников в организации;

n – количество записей ПДН не являющихся сотрудниками оператора;

P – персональные данные для различных категорий ПДН.

C – уровень защищенности информационной системы, $C = \{c_1, c_2, c_3, c_4\}$,

Условие классификации информационной системы по 1 уровню защищенности:

$$\forall p_i, \{(t = t_1) \cap [(k = k_2) \cup (k = k_3) \cup (k = k_4)]\} \cup [(t = t_2) \cap (k = k_3) \cap (n > 100000)] \Rightarrow c_1$$

Условие классификации информационной системы по 2 уровню защищенности:

$$\forall p_i, [(t = t_1) \cap [(k = k_1)] \cup \{(t = t_2) \cap (k = k_2^0) \cup [(k = k_2) \cap (n < 100000)]\} \cup [(t = t_2) \cap (k = k_3)] \cup [(t = t_2) \cap (k = k_1) \cap (n > 100000)] \cup [(t = t_2) \cap (k = k_4) \cap (n > 100000)] \cup [(t = t_3) \cap (k = k_2) \cap (n > 100000)] \Rightarrow c_2$$

Условие классификации информационной системы по 4 уровню защищенности:

$$\forall p_i, [(t = t_3) \cap (k = k_1)] \cup [(t = t_3) \cap (k = k_4^o)] \cup [(t = t_3) \cap (k = k_4) \cap (n < 100000)] \Rightarrow c_4$$

Современные проблемы защиты персональных данных

1. Глубокое противоречие между целями защиты ПДН и практическими результатами обеспечения их информационной безопасности.
2. До настоящего времени не определен круг задач, решаемых в организациях с использованием ПДН. Отсюда у них возникает неуемное желание узнать о субъекте ПДН как можно больше.
3. Расширение понятия ПДН в область «любой» информации о субъекте ПДН не нашло отражения в системе защиты информации.
4. Отсутствие механизмов оценки достоверности ПДН может привести к существенным ошибкам в оценке достоверности действий субъекта ПДН в процессах правосудной деятельности.

2. Современная трактовка содержания термина «персональные данные» в стандарте GDPR. Что за ней стоит ?

GDPR - General Data Protection Regulation

(https://gdpr-info.eu)

Несколько важных терминов

- **Идентификация** – это определение пользователя в автоматизированной системе по его уникальному признаку – идентификатору. По существу, это техническая процедура.
- **Аутентификация** – это проверка подлинности лица, которое хочет получить доступ к системе. В том случае, если используется несколько способов подтверждения – такая аутентификация называется многофакторной.

Сущность понятия «персональные данные»

(Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 02.07.2021) "О персональных данных»)

Персональные данные - любая информация, относящаяся к прямо или косвенно (directly or indirectly) **определенному** или **определяемому** (identified or identifiable) физическому лицу (субъекту персональных данных) (by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person) *(ФЗ № 152, GDPR)*

Остаются неопределёнными следующие понятия:

1. *Любая ?*
2. *Прямо или косвенно определенная, как это представить ?*
3. *Определенному кем, когда, зачем ? Почему не применен термин «идентифицированному» ?*
4. *Определяемому с какой целью, кем ? идентифицируемому*

Другая интерпретация определения ПДН

$$N_k \rightarrow F(a_k,\ x_1, x_2, \dots, x_n) \rightarrow Z_k \mid (p \geq p_k) \qquad (1)$$

$$F^{-1}(x_1, x_2, \dots, x_n) \rightarrow a_k \rightarrow N_k \mid (p \geq p_k), \qquad (2)$$

где

N_k – вектор параметров *определенного* физического лица (ФЛ). Этот параметр устанавливается в зависимости от решаемых социальных, экономических, политических и других задач, в которых требуется информация о владельце ПДН .

$F(a_k, x_1, \dots, x_n)$ – функция преобразования вектора параметров ФЛ во внутреннюю форму их хранения в базе данных и назначение внутреннего идентификатора a_k для задачи Z_k, которая решается с применением этого набора персональных данных.

$F^{-1}(x_1, \dots, x_n)$ – обратная функция определения ФЛ по предъявленным параметрам. Эта функция решает 2 задачи: определение наличия в БД предъявленных параметров *(идентификация)* и выдача заключения о результатах *аутентификации* определяемого ФЛ.

p –достоверность аутентификации. p_k –требуемая достоверность решения задачи Z_k.

3. Новый взгляд на защиту персональных данных субъекта

Уточним параметры решаемых задач

Переменные $\{ x_1, x_2, \dots , x_n\}$ представляют собой ***измеряемые, определяемые или диагностируемые*** характеристики персональных данных. Каждая характеристика описывается кортежом: *<наименование признака, его значение, частота в БД>.*

Z_k – задача (k) , для решения которой требуются определенные наборы параметров ПДН.

p – достоверность аутентификации по группе параметров вектора N_k . Значения вероятности p могут рассматриваться как независимые события, поэтому общая достоверность по n признакам может быть определена по формуле сложения вероятностей для независимых событий.

p_k – требуемая вероятность решения задачи обработки ПДН. В настоящее время этот параметр практически не используется. Он должен устанавливаться нормативными требованиями, исходя из допустимой погрешности решения задач обработки ПДН.

Идентифицирующие признаки $\{x_i\}$

1. ***Отраслевые идентификаторы*** (СНИЛС, ИНН, номер и серия паспорта, номер личного автомобиля, и др.).
2. ***Характеристики электронных идентификаторов личности*** (карты, RFID-метки, электронный паспорт, электронная подпись, электронные идентификаторы, вживляемые в тело владельца ПДН).
3. ***Признаки, сохраняющие свойства личности на длительное время, в том числе и после жизни человека*** (медицинская карта с историями заболеваний, личные дела, ДНК, социальные сети, поисковые системы, страницы сайтов и блогов с участием владельца ПДН, архивы электронной почты и мессенджеров.
4. ***Неустойчивые признаки владельца ПДН, сохраняемые от нескольких до десятков лет*** (психофизиологические параметры личности человека и его фотографии, профили в различных информационных системах, репутация и т.д.).
5. ***Признаки, которые можно определить, измерить или сопоставить только в период жизни взрослого человека:*** (паспорт с биометрическими показателями личности, отпечатки пальцев, рисунок сосудов рук, радужная оболочка глаза, группа крови, рентгеновские снимки, описательные словесные портреты, телосложение, рост, размер обуви).

Перечень групп задач, решаемых с использованием ПДН (z_k)

1. Допустимые условия обычного прохода на территорию организации.
2. Условия регистрации при проживании в гостинице.
3. Идентификация проезда в транспорте.
4. Аутентификация в социальной сети и других ресурсах интернет.
5. Вход в информационные системы и АСУ.
6. Проход в системах контроля и управления доступом.
7. Электронная подпись (простая и усиленная).
8. Взаимодействие с органами МВД и другими ведомствами (группа задач).
9. Необходимые сведения при поступление на обучение, работу, службу.
10. Необходимые сведения для оказания медицинских услуг.
11. Необходимые сведения для оказания банковских услуг.
12. Необходимые сведения для оказания образовательных услуг.
13. Необходимые сведения для оказание государственных услуг (нотариальных, ЗАГС и др.).
14. Необходимые сведения для регистрация при страховании.
15. Необходимые сведения для получение визы.
16. Необходимые сведения для проезда через границу.
17. Необходимые сведения для заключение договоров при сделках.
18. Необходимые сведения для электронного и обычного голосования.
19. Признаки и условия подтверждения доказательств участия в различных событиях в правосудной деятельности.

Проблема определения достоверности набора признаков

- Гораздо сложнее определить значение достоверности p_k . Для разных задач оно может быть различным. Например, при решении первой задачи, условие прохождения на территорию организации может быть при $p_k \geq 0,7$.
- При решении 19-й задачи достоверность совершения тяжкого преступления должна быть не ниже 0,99999. Однако, при этом, из совокупности набора признаков F_k должны быть обязательно исключены те из них, распространение которых не контролируется человеком (биоматериал, отпечатки пальцев и другая информация, относящаяся определению ДНК).
- Конкретно граничное значение достоверности совершения преступления должна быть определено на международном уровне обсуждения этой проблемы и может пересматриваться в зависимости от погрешностей измерения параметров признаков ПДН.

Угрозы владельцу ПДН

1. Шантажирование открытыми публикациями компрометирующего характера.
2. Психологическое давление постоянными атаками на владельца ПДН с целью принуждения и манипулирования (троллинг).
3. Сбор сведений о системах предпочтений владельца ПДН.
4. Изменение данных о субъекте ПДН в БД.
5. Манипуляция общественным мнением в отношении владельца ПДН.
6. Раскрытие личной и семейной тайны.
7. Подмена (фальсификация) биологических образцов субъекта ПДН.
8. Фальсификация субъекта ПДН путем использования поддельных документов.
9. Компрометация владельца ПДН путем публикаций аудио и видео записей с личной (частной) информацией.
10. Выполнение несанкционированных действий от имени владельца ПДН.
11. Захват недвижимости владельца ПДН при фальсификации документов.
12. Овладение счетами владельца ПДН.
13. Подлог на месте преступления свидетельств владельца ПДН.

Требования к модели защиты ПДН

Уровень защищенности ПДН в системах их обработки должен определяться:

1. Количеством параметров в принятой организацией модели ПДН (N_k) и зависит от решаемой задачи (Z_k).
2. Характером угроз владельцам ПДН при компрометации данных, содержащихся о них в БД.
3. Требованиями по достоверности аутентификации владельца ПДН (p_k).
4. Полным набором защитных мероприятий, включающих меры по локализации обработки ПДН и контроля за их распространением.
5. Соответствующей моделью ответственности организации за обработку ПДН.

Требования к модели ответственности за обработку ПДН

Уровень ответственности за компрометацию, искажение или передачу информации о ПДН должен определяться:

1. Количеством параметров в принятой организацией модели ПДН (N_k).
2. Количеством владельцев ПДН.
3. Величины (класс) угроз, которые могут быть реализованы для владельцев ПДН.

Выводы

- Современное определение ПДН трактует практически все ИС, имеющие персональные данные, как информационные системы ПДН (ИСПДН).
- Обе задачи имеют вероятностный смысл. Это означает, что достичь определенную вероятность аутентификации возможно большим количеством «слабых» признаков или/и уменьшением множества объектов в базе данных.
- Если первая задача не может быть точно определена, то вторая задача должна иметь конкретные критерии решения задачи аутентификации физ.лица (при устройстве на работу, при допуске к конфиденциальной информации, при определении полноты улик и т.д.).
- Рассмотренная модель защиты ПДН ориентирована на угрозы физическому лицу и может быть использована при выборе модели ПДН для решения конкретной задачи.
- Предложенная модель защиты ПДН стимулирует организации, проводящие обработку ПДН, к уменьшению количества параметров в модели данных и созданию системы контроля за их распространением.

- **Электронная почта:**
  - MinzovAS@mpei.ru
- **Телефон:**
  - +7 926 565-05-70
- **Сайт:**
  - www.mpei.ru