УДК 621.96 (043.2)

А.І. Мужик, науковий співробітник
І.М. Мужик, старший викладач

ПІДХОДИ ДО ОБРОБКИ ПЕРСОНАЛЬНИХ ДАНИХ

Національний авіаційний університет, e-mail: myzhuk@mail.ru

У роботі розглянуто алгоритм визначення з переліку даних таких, які відносяться до категорії персональних даних. Приведено класифікацію категорій персональних даних.

***Ключові слова**: персональні дані, захист інформації, інформаційна безпека, загрози*

Вступ і постановка задачі . Широке впровадження інформаційних технологій приводить до виникнення проблеми захисту інформації в цілому і, зокрема, до проблеми неконтрольованого поширення даних, що підлягають обробці. Стосовно неконтрольованого поширення інформації особливо чутлива така категорія інформації, як персональні дані. Опинившись за межами захищеної інформаційної системи, персональні дані можуть бути доступними практично необмеженому колу користувачів. За таких умов, інформація може бути знищена чи спотворена, або може бути використана з метою нанесення шкоди особі, якої вона стосується. При цьому шкода може бути не тільки моральною, але й матеріальною. Тому питанням захисту персональних даних від неконтрольованого їх поширення приділяється особлива увага зі сторони як міжнародного співтовариства, так і урядів багатьох держав світу. Питання захисту персональних даних є актуальними для операторів інформаційних систем та фахівців з інформаційної безпеки. У зв'язку з цим виникає проблема автоматичного визначення з потоків даних, що підлягають обробці у інформаційних системах таких даних, які відносяться до категорії персональних.

Аналіз останніх досліджень та публікацій. Результати опублікованих досліджень в галузі інформаційної безпеки (безпеки інформаційних систем) [1-4] показують, що питанням внутрішньої безпеки інформаційних систем приділяється значна увага у багатьох державах світу. Це стосується і питань, які пов'язані з неконтрольованим поширенням даних. Зростання уваги до безпеки інформаційних систем, згідно опублікованих даних, обумовлено стабільно зростаючою кількістю зафіксованих випадків витоку інформації у всіх країнах світу. Аналіз отриманої інформації показав, що 70-90% випадків витоку даних, які втрачаються в цілому, складають персональні дані (ПД). При цьому третина цих даних втрачається мережевим шляхом. В той же час, приблизно однакові частки втрати персональних даних спостерігаються за рахунок не тільки навмисних дій співробітників компаній, але й через їх необережність. Слід також відмітити, що загрози в інформаційній безпеці за своєю актуальністю посідають друге місце серед основних загроз бізнесу [2], таких як економічна нестабільність, промисловий шпіонаж, викрадення інтелектуальної власності, нанесення шкоди репутації, тощо.

Серед загроз в інформаційній безпеці виділяють дві групи загроз: внутрішні та зовнішні [5]. До зовнішніх загроз відносять загрози, що виникають та якими керують за межами інформаційних систем (ІС). Такі загрози направлені або відносяться до ресурсів ІС. Внутрішні загрози виникають безпосередньо в границях ІС, що використовуються в межах організацій. Загрози такого типу можуть надходити від технічного обладнання, недосконалих програмних засобів, навмисних або не навмисних дій персоналу, який є користувачем ІС. Слід відмітити, що проблема внутрішніх загроз інформаційній безпеці викликана незахищеністю ІС організацій і установ та відсутністю ефективного рішення протидії таким загрозам.

На сьогоднішній день, практично на всіх підприємствах використовуються програмні і/або апаратні засоби захисту інформації і ІС, які призначені для боротьби із зовнішніми загрозами. Такі засоби досить ефективно протистоять зовнішнім загрозам. Стосовно засобів захисту ІС від внутрішніх загроз, можна відмітити, що тільки незначна частина компаній та організацій займається їх використанням, хоча необхідність у цих засобах об'єктивно існує [1].

Аналіз літературних джерел показує, що наявні алгоритми та запропоновані на ринку системи виявлення конфіденційної інформації у мережевому потоці не є ефективними для попередження несанкціонованого витоку персональних даних. Це потребує удосконалення існуючих технології контентної фільтрації, розробки нових методів виявлення потрібних даних у інформаційному потоці, концептуально змінюючи підходи до їх розпізнавання.

Задача попередження витоку ПД мережевим шляхом за межі захищених ІС ускладнюється дуже загальним їх визначенням, тобто визначенням поняття ПД. На сьогоднішній день не існує жодного вичерпного переліку даних, які однозначно відносяться до категорії ПД. Виходячи з положень Закону України "Про захист персональних даних" [6] та чинних нормативних актів можна

стверджувати, що основним критерієм, який визначає приналежність певних даних до категорії ПД, є характерна таким даним властивість ідентифікувати за ними особу, до якої вони відносяться. Теоретично така ідентифікація можлива за умови, якщо всі особи, в рамках наявної інформації, мають унікальні дані і є хоча б одна особа, якій відповідають наведені (визначені) дані. Очевидно, що ймовірносна оцінка відповідності даних певній особі могла би бути визначена шляхом визначення переліку даних, які можна ідентифікувати як персональні. Але така процедура не передбачена чинними нормативними актами.

Оскільки склад і структура ПД у кожному конкретному випадку можуть бути різними, то, з точки зору попередження витоку ПД, першочерговою задачею є задача ідентифікації з потоку даних, що передаються мережевими каналами та обробляються в інформаційних системах, як персональних. Ця задача особливо актуальна при автоматизованій ідентифікації ПД.

Задачі дослідження. В роботі буде розглянуто структуру персональних даних, визначені класи та категорії персональних даних, досліджені класи ризику. Буде розглянуто алгоритм для автоматизованого виявлення у масивах інформації даних, які відносяться до категорії персональних.

Результати досліджень. За властивістю ідентифікувати деяку особу ПД можна поділити на дві групи (рис.1). До першої групи ПД входять ідентифікаційні дані, які однозначно визначають особу. Однак такі дані не несуть ніякої додаткової інформації про особу, що ідентифікується. Наприклад, прізвище, ім’я, по батькові; серія та номер паспорту; ідентифікаційний код, серія та номер водійського посвідчення тощо. Такі ідентифікаційні дані розподіляються на дві підгрупи: унікальні ідентифікаційні дані, які однозначно відносяться до конкретної особи та дані, що можуть відноситись до обмеженого кола осіб. Наприклад, серія та номер відомчого посвідчення, перепустки, адреса проживання тощо. Дані, що відносяться до обмеженого кола осіб, з метою ідентифікації за ними конкретної особи потребують певного уточнення.

До другої групи ПД входять особисті відомості про особу. Вони несуть певну інформацію про особу. Наприклад, вік, стать, майнові відомості, відомості про сімейний стан, освіту, політичні погляди, релігійні переконання тощо. За певним набором таких даних теоретично можливо визначити особу, до якої вони відносяться. У даному випадку перелік особистих даних повинен бути досить широким, і тим більшим, чим ширше коло осіб, до яких такі дані можуть відноситись. При цьому, як правило, з метою звуження кола осіб, до яких такі дані можуть відноситись, виникає необхідність звертатися до додаткових джерел інформації – інших баз ПД. У відповідності з нормами чинного законодавства такі джерела інформації (бази персональних даних) повинні бути закритими. Доступ до них може мати обмежене коло користувачів. При цьому не допускається вільне поширення даних, що зберігаються у базах ПД. За таких умов, передбачається, що отримати з них уточнюючу інформацію не є можливим. Тому можна припустити, що за особистими відомостями, які є знеособленими, ідентифікувати особу, до якої такі дані можуть відноситись, неможливо.

Особисті відомості, у відповідності до вимог Закону України “Про захист персональних даних” поділяться на дві підгрупи. Стаття 7 цього закону передбачає особливі вимоги до обробки ПД, які несуть чутливу інформацію. До таких даних відносяться відомості про расове або етнічне походження, політичні, релігійні або світоглядні переконання, членство в політичних партіях та професійних спілках, звинувачення у скоєнні злочину або засудження до кримінального покарання, а також даних, що стосуються здоров'я чи статевого життя.

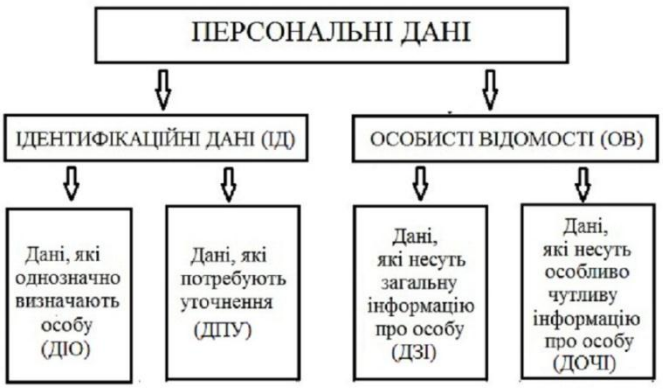

Рисунок 1 - Класифікація персональних даних

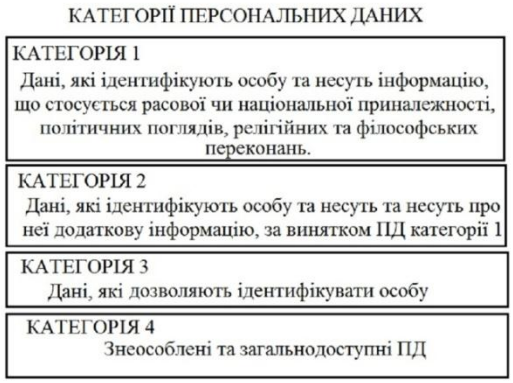

Рисунок 2 - Категорії персональних даних

Щодо можливого характеру загрози для збереження цілісності та конфіденційності ПД, які обробляються у відповідних базах та ІС, а також необхідності впровадження відповідних заходів безпеки даних, ПД діляться на чотири класи ризику:

Клас ризику 4: ризик відсутній. ПД, що обробляються, вже знаходяться у вільному доступі. При цьому вважається, що використання таких ПД не містить ризиків для суб'єктів ПД. Для їх захисту не потрібні жодні спеціальні заходи безпеки;

*Клас ризику 3***:** незначний рівень ризику. В цьому класі, у випадку втрати або несанкціонованого чи неналежного доступу до ПД особи, наслідки для особи є такими, що для їх запобігання буде достатньо використовувати звичайні (стандартні) заходи захисту інформації. До цієї групи відносяться бази даних бухгалтерії та відділу кадрів невеликих підприємств, бібліотек, комунальних організацій, а також клієнтські бази торгівельних та сервісних організацій (із певними виключеннями);

Клас ризику 2: середній рівень ризику. У цьому класі втрата або неавторизоване чи неналежне використання ПД суб'єкта може спричинити додаткові негативні наслідки. До баз ПД цього класу відносяться бази, що містять дані про особисте життя громадян, расове або етнічне походження, політичні, релігійні або світоглядні переконання, членство в політичних партіях та професійних спілках, бази даних, що містять або можуть містити опосередковану інформацію про світоглядне переконання, статеве життя чи здоров'я. Наприклад, бази абонентів телекомунікаційних компаній, інтернет-сервіс провайдерів тощо. Для таких баз даних може бути необхідним проведення незалежної оцінки вжитих заходів щодо захисту ПД.

Клас ризику 1: високий рівень ризику. У випадку, якщо несанкціоновані дії із ПД можуть мати серйозні наслідки для суб'єкта персональних даних, для їх захисту повинні бути впроваджені належні засоби захисту, а також обов'язково проводитися незалежна оцінка таких заходів.

На основі наведених класифікації та класів ризику ПД можна виділити чотири категорії ПД, які умовно відповідають даним, включеним до певних класів ризику (рис.2). До четвертої категорії відносяться знеособлені та загальнодоступні дані. До третьої категорії відносяться ідентифікаційні дані, вони ніякої інформації про особу не несуть. Очевидно, що поширення таких даних третьої та четвертої категорій не можуть завдати відчутної шкоди особі, до якої вони відносяться. З проведеної класифікації випливає, що захисту від несанкціонованого поширення підлягають дані першої та другої категорії.

Виходячи з класифікації ПД (рис.3) можна виділити три групи ПД:

- дані, що дозволяють однозначно ідентифікувати громадянина (ІД);
- дані, що розкривають загальну інформацію, але не дозволяють однозначно ідентифікувати громадянина (ДЗІ);
- дані, що розкривають особливо чутливу інформацію, але не дозволяють однозначно ідентифікувати громадянина (ДОЧ).

На основі такого розмежування ПД можна представити у вигляді наступних множин:

ІД – множини даних, що дозволяють однозначно ідентифікувати громадянина: ІД (1-6).

ДЗІ – множини даних, які розкривають загальну інформацію, але не дозволяють однозначно ідентифікувати громадянина: ДЗІ (1-16).

ДОЧІ – множини даних, які розкривають інформацію про релігійні, расові, національні, політичні погляди, але не дозволяють однозначно ідентифікувати громадянина: ДОЧІ (1-6).

Елементи розглянутих множин представлені в табл.1.

Таблиця 1.

Можливі види персональних даних

Тип даних	Елемент	Найменування елемента
ІД	1	Прізвище, ім'я, по батькові
	2	Дані, що характеризують фізіологічні особливості, та за якими можна встановити особу
	3	Паспортні дані
	4	Дані свідоцтва про народження
	5	Дані водійського посвідчення
	6	Адреса місця проживання
ДЗІ	1	Дані про освіту
	2	Дані про трудову діяльність
	3	Дані про трудову книжку
	6	Дані з трудового договору
	4	Дані про заробітню плату
	5	Дані з наказів по особовому складу

	6	Дані про атестацію
	7	Дані з матеріалів службових розслідувань
	8	Дані про підвищення кваліфікації
	9	Дані про проходження професійної перепідготовки
	10	Дані про військовий облік
	11	Дані про доходи
	12	Дані про майно
	13	Дані про індивідуальні номери платника податків
	14	Дані страхового свідоцтва
	15	Дані про сімейний стан та склад сім'ї
	16	Дані про пільги
ДОЧІ	1	Дані про здоров'я
	2	Дані про расову приналежність
	3	Дані про національну приналежність
	4	Дані про політичні погляди
	5	Дані про релігійні переконання
	6	Дані про інтимне життя

Для однозначного визначення належності даних до категорії ПД, використовуючи дані табл.1, можна сформувати множини персональних даних, представлені на рис.3.

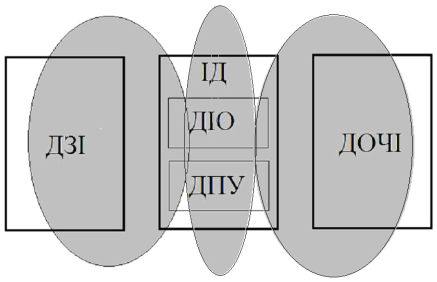

Рисунок 3 - Групи персональних даних

У результаті аналізу можливих варіантів об'єднання груп даних (рис.3) видно, що до категорії ПД будуть відноситись дані, до складу яких входять елементи з груп ІД та ДОЧІ і ІД та ДЗІ. Такі дані будуть відповідно відноситись до 1 та 2 класу ризику ПД. Дані до яких входять елементи з підгруп ДІО та ДПУ також будуть персональними та будуть відноситись до 3 класу ризику. Дані, до яких входять елементи груп ДЗІ та ДОЧІ можуть бути віднесеними до 4 класу ризику, тому що вони є знеособленими. Такі дані не потребують захисту.

Отримані результати показали, що з метою визначення відсутності персональних даних у масиві інформації, що підлягає обробці, достатньо будувати алгоритм її обробки з визначенням відсутності ПД групи ІД у цьому масиві. При виявлені у масиві інформації, що підлягає обробці, групи ІД необхідно проводити подальший аналіз даних з метою виявлення інших ПД, що стосуються особи, до якої відносяться виявлені ПД групи ІД. При виявлені додаткових ПД необхідно встановлювати до якої групи вони відносяться, з наступним визначенням категорії виявлених ПД.

Висновки: Розроблено алгоритм встановлення відповідності даних категорії персональних даних. В основі роботи алгоритму лежить аналіз комбінацій груп персональних даних. Персональними даними 4 категорії є множини даних, що несуть загальну інформацію про особу та особливо чутливих даних, так як вони не дають можливості ідентифікувати особу. Персональними даними 3 категорії є множина даних, що ідентифікують особу. Персональними даними 2 категорії є множини даних, що ідентифікують особу та даних з загальною інформацією, так як до неї входять дані, що ідентифікують громадянина, а також несуть додаткову інформацію про нього. Персональними даними 1 категорії є множини даних, що ідентифікують особу та дані з особливо чутливою інформацією що стосується расової, національної приналежності, політичних поглядів, релігійних і філософських переконань, стану здоров'я, інтимного життя. Отриманий алгоритм дозволяє визначати у масиві інформації дані, які можна віднести до категорії персональних.

Список літературних джерел

1. http://gatton.uky.edu/FACULTY/PAYNE/ACC324/CSISurvey2010.pdf
2. http://www.kaspersky.ru/other/custom-html/brfwn/Bezopasnost_Screen.pdf
3. Ерыгин А.В. Анализ эффективности систем предотвращения утечек конфиденциальной информации из локальных сетей. // Журнал «Вестник СибАДИ», №2 (20) 2011 год. - с.47-52.
4. Волчинская, Е.К. Персональные данные в России 2010 [Текст] / Е.К. Волчинская // Защита персональных данных. Опыт правового регулирования. - 2010. - N6. - С. 5-7.
5. Марков, А.П. Проблемы и решения по защите персональных данных в информационных системах персональных данных [Текст] / А.П. Марков, Б. И. Сухинин // Компьютерная безопасность. - Улан-Уде: ВСГТУ. - 2009. - №5. – с. 20-27.
6. Закон України "Про захист персональних даних"