УДК 004.9

МЕТОДИ АНОНІМІЗАЦІЇ МЕДИЧНИХ БАЗ ДАНИХ

Б. М. Гавриш[1], О. В. Тимченко[2,3], Н. О. Кустра[1]

[1]Національний університет «Львівська політехніка», вул. Степана Бандери, 12, Львів, 79013, Україна

[2]Українська академія друкарства, вул. Під Голоском, 19, Львів, 79020, Україна

[3]Uniwersytet Warmińsko-Mazurski w Olsztynie, ul. Michała Oczapowskiego, 2, Olsztyn, 10-719, Polska

Розвиток електронних форм зберігання, обробки та передачі медичних даних вплинув не лише на підвищення якості медичної допомоги пацієнтів, а й на розвиток нових методів отримання знань із медичних баз неуповноваженими особами. Щоб запобігти розголошенню конфіденційних даних, медичні інформаційні системи потрібно постійно перевіряти на безпеку в усіх структурах системи. Технічні заходи безпеки мають бути доповнені заходами фізичної, особистої та організаційної безпеки. Методи, описані у статті, є основними методами анонімізації медичних даних, на основі яких були розроблені інші методи анонімізації, такі як l-диверсифікація, (X, Y)-з'єднуваність, (X, Y)-конфіденційність, LKC-закритість конфіденційності, обмежена довіра та персоналізована конфіденційність. Завдяки методам анонімізації у медичних базах даних можна звести до мінімуму ефективність атак на дані пацієнтів.

Ключові слова: *анонімізація, псевдоідентифікатор, атака, захист систем, безпека.*

Постановка проблеми. Медична документація пацієнта є ключовим елементом під час його лікування, оскільки містить всю інформацію про стан здоров'я, проведені аналізи, перебування в стаціонарі та процедури, які проводилися протягом багатьох років. Ще кілька років тому медична документація була переважно у паперовій формі. Сьогодні його повільно витісняють електронні форми. Один факт не змінився з роками — найчастіше саме пацієнт відповідає за доставку його до іншого лікувального закладу. Тому в екстрених випадках під час лікування в новому закладі рівень знань про цього пацієнта дорівнює нулю. Цю проблему вирішує запровадження електронної медичної картки ЕМК (англ. Electronic Medical Record — EMR) [1]. ЕМК — це віртуальний документ, який складається з усіх медичних записів у цифровій формі, що належать одному пацієнту. Завдяки цьому рішенню інформацію про пацієнта можна створювати, зберігати та використовувати в багатьох різних медичних установах і робити доступною для пацієнта в одному документі у вебдодатку.

Аналіз останніх досліджень та публікацій. Запровадження електронної системи медичної документації має переваги, але породжує й нові проблеми. Перевагами є підвищення якості медичної допомоги пацієнтам, більш ефективне та набагато ефективніше управління (система електронного рецепту дає змогу контролювати небажані взаємодії між препаратами, призначеними та прийнятими одночасно), підтримку рішень лікарів та зменшення медичних помилок до 55 % [1], дистанційне лікування, яке добре у великих містах, міжміське, міжконтинентальне. Найсерйознішим наслідком передачі ресурсів з лікарняних баз у мережу є проблеми з контролем та захистом інформації, що міститься в медичних документах. Цілісність цифрових об'єктів також є проблемою у випадку багатомодульних систем EHR.

Мета статті — представити основні загрози безпеці медичних інформаційних систем та методи захисту, звернувши увагу на способи приховування знань у медичних базах даних.

Виклад основного матеріалу дослідження. *Безпека інформаційних систем.* Найсерйознішою проблемою, з якою стикаються сучасні інформаційні системи, є забезпечення безпеки конфіденційних даних. Закон України № 2297–VI про захист персональних даних від 2010 р. регулює відносини, пов'язані із захистом персональних даних під час їх обробки. Через лояльність до пацієнтів, які довірили свої дані медичному закладу, персональні дані наражаються на низку небезпек. Медичні БД піддаються різним видам атак, незалежно від того, чи зберігаються вони стаціонарно в комп'ютерах медичних установ, чи доступні іншим медичним установам через інтернет-додатки. В обох випадках дані з баз даних (БД) мають бути належним чином захищені від несанкціонованого доступу. Далі описані основні категорії загроз медичним системам, приділяючи особливу увагу безпеці медичних інтернет-додатків та методу аналізу цих загроз.

Безпека медичних інтернет-додатків. Медичні дані можуть бути доступні пацієнту у вебдодатках через веббраузери. Така форма надання інформації пацієнтам та лікарям стає дедалі популярнішою завдяки швидкому та легкому доступу до інтернету [2, 3].

Останні медичні інтернет-додатки мають тип клієнт-сервер [2]. Вони характеризуються тим, що встановлюються не на локальний комп'ютер користувача, а запускаються через веббраузер. Користувач керує програмою за допомогою списків вибору та полів редагування. Програми мають багатошарову архітектуру і завдяки розподілу на незалежні модулі управління БД реалізація бізнес-логіки та підтримка інтерфейсу користувача здійснюються окремо. Окремі модулі можуть працювати на різних машинах, нижнім рівням не потрібно знати про вищі. Кожен шар може бути реалізований різними мовами різними командами [2, 3].

Так багато зручностей відкривають ще більше можливостей для атаки на конфіденційні дані, тому ключову роль у створенні нових медичних ІТ-систем відіграє належний захист кожного прикладного рівня та постійний моніторинг ефективності безпеки.

Метою механізмів безпеки є мінімізація ризику перехоплення конфіденційних даних неуповноваженими особами. Щоб мінімізувати ризики, пов'язані з порушенням безпеки додатків, розроблені спеціальні стратегії управління ризиками. Однією з широко використовуваних моделей ризику є модель Microsoft, яка складається з п'яти етапів [2]:

1. Визначення цілей стратегії безпеки.

2. Уточнення характеристик програми, які будуть корисні для виявлення загроз на четвертому етапі.

3. Декомпозиція програми з метою відокремлення модулів, де безпека має вирішальне значення.

4. Ідентифікація індивідуальних небезпек на основі інформації з другого та третього пункту.

5. Огляд усіх прикладних рівнів, визначення загроз у кожному рівні та оцінка ступеня загрози.

Для класифікації загроз використовується методологія STRIDE (Spoofing Identify, Tampering, Resudiability, Information Disclosure, Denial of Service, Elevation of Privilege) [2]. У межах цієї методології виділено шість груп загроз та запропоновано методи зменшення цих загроз. Модуль, вилучений під час декомпозиції, повинен пройти перевірку з точки зору зазначених загроз.

Основні критерії загроз. Існує шість основних категорій ризиків у медичних інформаційних системах. Кожна категорії має свої способи боротьби з цими загрозами [2].

Видавання себе за іншу особу. Спуфінг є основним ризиком безпеки програми, коли особа або програма маскується під іншу за допомогою фальсифікації даних, включає в себе ідентичність іншого користувача, щоб отримати доступ до даних, поки програма працює з однаковими правами для всіх користувачів. Адміністратори (з найширшими повноваженнями), які можуть змінювати медичні БД, зазнають найбільшого ризику. Для зменшення цієї загрози використовується протокол SSL (Secure Sockets Layer). SSL-з'єднання необхідно використовувати в чутливих до безпеки даних, при зборі або збереженні даних пацієнтів, при передачі між комп'ютерами в мережі файлів cookie (і. Cookies), які містять ідентифікатори сеансів, паролі, логіни. Для захисту від спуфінгу також рекомендується використовувати надійні механізми аутентифікації, шифрувати паролі, шифрувати облікові дані, що надсилаються по мережі, та контролювати доступ до адміністративних профілів [2].

Маніпулювання даними. Загрози цієї категорії містять загрозу, що виникає внаслідок надмірної довіри до перевірки даних, що здійснюється на стороні клієнта. Змінюючи методи GET і POST протоколу HTTP, можна за запитами отримувати конфіденційні дані. Ризик цієї категорії також охоплює значення змінних середовища вебсервера, щоб отримати контроль над додатком на стороні сервера. Загрози цієї категорії можна зменшити шляхом захисту рівнів даних протоколами, що забезпечують цілісність (наприклад, IPSec), використанням протоколів, стійких до маніпуляцій, або використанням електронного підпису [2].

Відмова в дії. Цей ризик полягає у приховуванні модифікації даних, коли такі зміни були внесені. У медичних програмах необхідно використовувати механізми відстеження та контролю користувачів, а також перевіряти процес зміни даних. Для зменшення цієї загрози використовують електронний підпис, механізми генерації єдиного пароля для підтвердження зміни, реєстрація всіх переміщень даних користувачів [2].

Розкриття інформації. Ця загроза може бути спричинена неправильною поведінкою веббраузера. Секретну інформацію можуть розкривати програми, клієнти яких працюють із спільними БД. Щоб запобігти небажаному розголошенню інформації, медичний вебдодаток має запобігати запам'ятовуванню важливих даних на стороні клієнта, використовувати надійні механізми авторизації та шифрування, подбати про відокремлення рівня даних від інтерфейсу користувача та захистити рівні зв'язку з конфіденційністю (протоколи (SSL/TLS, IPSec) [2].

Блокування доступу до сервісу. Ця загроза полягає в блокуванні доступу до програми, яка може здійснюватися на кількох рівнях програми. Цей тип нападу не приносить користі зловмиснику, але може зашкодити пацієнту, чия медична документація втрачена. Атака DOS (відмова в обслуговуванні) полягає в наповненні сервера багатьма запитами від різних користувачів, над якими отримано контроль, і блокуванні сервера. Найголовніше — це можливість розрізняти підвищений інтерес користувачів до програми на сервері. Для захисту від цієї загрози необхідно використовувати механізми на багатьох рівнях, але найчастіше для контролю ресурсів системи, мережевого трафіку використовується IDS (Intrusion Detection System) — системи виявлення вторгнень та IPS (Intrusion Prevention System) — система профілактики вторгненням [2].

Несанкціоноване отримання більших привілеїв. Ця загроза полягає в тому, що стороння особа отримує більше прав. З метою зниження ризику необхідно контролювати рівень дозволів, керуючись принципом надання найменших необхідних дозволів для роботи програми, використовувати механізми поділу процесів та віртуалізацію серверів. Дуже часто програми не захищені від атак XSS (Cross-Site Scripting). Атака полягає у виконанні коду, який надає користувачеві адміністративні права без відома адміністратора [2].

Аналіз загроз. Для забезпечення належного контролю за безпекою медичних інформаційних систем необхідно постійно контролювати рівень загроз. Ризики, спричинені внутрішньою структурою програми, та ризики, пов'язані з внутрішнім і зовнішнім потоком даних, потрібно розглядати окремо. Аналіз загроз, які можуть виникнути на кожному прикладному рівні, використовується для виявлення прогалин у системі. Кожну небезпеку потрібно оцінювати за шкалою від 0 до 10, причому 0 — це найнижчий ризик, а 10 — значний рівень ризику. Для оцінки ступеня ризику можна використовувати метод DREAD (Damage, Reproductibility, Exploitability, Affected Users, Discoverability), завдяки якому визначається рівень п'яти загроз щодо програми [2]. Типи та оцінка цих загроз наведені в табл. 1.

Таблиця 1

Типи та ступені загроз за методом DREAD

Тип загрози	Ступінь небезпеки
Рівень пошкодження (у разі успішної атаки)	0 — без пошкоджень 5 — розголошення конфіденційних даних користувача 10 — повне руйнування системи та втрата даних
Труднощі відновлення стану системи до атаки	0 — стан неможливо або важко відтворити 5 — відновний стан за певних умов 10 — легко відтворити стан
Легкість використання уразливості	0 — потребує розширених знань з мережі та програмування, а також передових інструментів 5 — можна використовувати з наявними інструментами 10 — напад можливий з боку особи без спеціальних компетенцій
Кількість користувачів у групі ризику	0 — близький до нуля 5 — деякі користувачі, не всі 10 — всі користувачі
Рівень труднощів локалізації люку (trapdoor)	0 — люк, який дуже важко знайти 5 — локалізується під час моніторингу мережі 10 — легко знайти навіть користувачеві без спеціальних знань

Методи захисту медичних інформаційних систем. У зв'язку з тим, що медичні дані містять багато конфіденційних даних, їх потрібно захищати на багатьох рівнях одночасно. Захист медичних систем має містити фізичну, технічну, особисту та організаційну безпеку.

Фізична безпека стосується захисту комп'ютерів і приміщень, в яких ці комп'ютери розташовані, від сторонніх осіб. Технічні заходи безпеки містять такі питання, як резервне копіювання даних, використання антивірусних програм і віртуальних приватних мереж (VPN), використання систем безпеки електронної пошти та аутентифікації користувачів [3]. Особиста та організаційна безпека стосується захисту медичних даних від необережності людей, які використовують системи на законних підставах. Системи ідентифікації користувачів не будуть корисними, якщо люди, які використовують систему, залишать свої ідентифікатори в місцях, доступних для третіх осіб. Від системи архівування даних буде мало користі, якщо персонал не буде дбати про резервне копіювання даних. Тому важливо навчити тих, хто має право використовувати медичні дані. Організаційні заходи безпеки передбачають розробку та передачу сценаріїв надзвичайних ситуацій (хакерська атака, проникнення в систему комп'ютерного вірусу) особам, уповноваженим на використання інформаційної системи, та навчання цих осіб. Кожен із описаних заходів безпеки нічого не означає, коли функціонує самостійно. Для того щоб медична інформаційна система була безпечною, необхідно інтегрувати технічні засоби захисту, відповідний рівень підготовки працівників та відповідні організаційні рішення [2, 4].

Розглянемо основні методи захисту конфіденційності в медичних БД.

Методи анонімізації медичних даних

У літературі згадуються чотири основні методи атак на таблиці медичних БД [5]:

1) об'єднання записів даних;

2) посилання на атрибути;

3) приєднання до таблиць;

4) ймовірнісні методи.

Метод прив'язки до записів і атрибутів передбачає, що значення псевдоідентифікатора (PID) жертви відоме і виконує пошук конфіденційних даних. У разі приєднання до таблиць зловмисник визначає, чи є в опублікованій таблиці запис жертви. Імовірнісні методи використовуються для отримання більш широких знань про жертву на основі анонімних даних [5−8]. У відповідь на такі атаки використовуються методи забезпечення анонімності даних, що зберігаються в таблицях. Деякі з цих методів описані далі.

Анонімність методом k-анонімізації

Метод k-анонімізації — це метод, який запобігає порушенню конфіденційності пацієнта шляхом додавання зовнішніх записів до таблиці даних. У таблиці з даними пацієнта можна вибрати атрибути, які становлять псевдоідентифікатор пацієнта PID. Усі записи БД групуються за окремими значеннями PID [5−8]. Зазвичай такий поділ визначає групи, які не дуже численні. Невелика кількість записів у групі призводить до невеликої кількості значень, які можна призначити пацієнту, чиї дані ви хочете отримати. Отже, шахрай, маючи доступ до інших знань пацієнта, може виділити запис пацієнта з групи, зазначеної перед PID. Щоб захистити медичні дані від розкриття, потрібно забезпечити мінімальну кількість k-записів у групах, ідентифікованих псевдо-PID [5−7, 9]. На практиці цей метод потребує корекції даних для атрибутів, які складають PID. У табл. 2 наведено приклади даних пацієнтів із БД захворювань щитовидної залози. PID складається з трьох атрибутів: стать, місто та значення тиреотропного гормону (ТТГ – TSH, thyroid stimulating hormone) у пацієнта.

Таблиця 2

Витяг з БД пацієнтів з діагнозом захворювання щитовидної залози, що містить записи пацієнтів, ідентифікованих за трьома PID

Псевдоідентифікатор	**Стать** (атрибут a)	**Місто** (атрибут b)	**TSH** (атрибут c)	**Захворювання щитовидної залози** (атрибут d)
PID_1	a_3	b_1	c_2	d_1
PID_2	a_2	b_1	c_3	d_2
PID_3	a_1	b_2	c_1	d_1
PID_1	a_3	b_1	c_2	d_1
PID_1	a_3	b_1	c_2	d_2
PID_3	a_1	b_2	c_1	d_2
PID_3	a_1	b_2	c_1	d_2
PID_1	a_3	b_1	c_2	d_2

На основі табл. 2. може бути виділено три групи, що визначаються значенням псевдоідентифікатора. Серед цих груп є одна з назвою PID_2, яка становить один запис. Шахрай, маючи знання з інших джерел, наприклад табл. 3, може витягувати конфіденційні дані пацієнтів.

Таблиця 3

Додатковий фрагмент БД пацієнтів із захворюваннями щитовидної залози

Пацієнт	**Стать**	**Місто**	**TSH**
(об'єкт)	(атрибут a)	(атрибут b)	(атрибут c)
x_1	a_3	b_1	c_2
x_2	a_2	b_1	c_3
x_3	a_1	b_2	c_1

На основі табл. 2 та 3 шахрай може отримати конфіденційні дані пацієнта x_2 у вигляді діагностованого захворювання щитовидної залози d_2. Метод k-анонімізації не захищає медичну БД від розкриття конфіденційних даних за допомогою статистичних методів, наприклад, інформації про те, що пацієнт x_3 з ймовірністю приблизно 0,67 може мати хворобу d_2.

Анонімність за допомогою методу анонімізації (X, Y)

Узагальненням методу k-анонімізації є (X, Y)-метод анонімізації. Цей метод полягає в поділі атрибутів на два набори X і Y. X — це набір атрибутів, які утворюють псевдоідентифікатор, а набір Y — набір значень псевдоідентифікатора. Необхідно, щоб для кожного значення атрибута X було k різних значень атрибута Y [5, 7]. У табл. 4 наведено поділ даних за допомогою (X, Y) методу анонімізації.

Таблиця 4

Фрагмент БД, що містить записи, вибрані методом анонімізації (X, Y)

Псевдоідентифікатор	**Стать (атрибут a)**	**Місто (атрибут b)**	**TSH (атрибут c)**
Y	X		
PID_1	a_2	b_1	c_3
PID_2	a_2	b_1	c_3
PID_3	a_2	b_1	c_3
PID_4	a_2	b_2	c_2
PID_5	a_2	b_2	c_2
PID_6	a_2	b_2	c_2
PID_7	a_2	b_2	c_2
PID_8	a_2	b_2	c_2

Анонімність за допомогою методу (α, k)-анонімізації

Цей метод поєднує методи k-анонімізації та α-дисоціації. Згідно з методом k-анонімізації, мінімальна кількість записів у групі для цього PID не може бути меншою за k. Крім того, для заданого значення w, ймовірність появи не перевищує α у всіх класах [5, 9]. Табл. 5 показує умови (0.67, 3)-анонімізації.

Таблиця 5

Фрагмент БД, що містить записи, виділені методом (0.67, 3)-анонімізації

Псевдоіден-	**Стать**	**Місто**	**TSH**	**Захворювання щитовидної залози**
тифікатор	(атрибут a)	(атрибут b)	(атрибут c)	(атрибут d)
PID_1	a_3	b_1	c_2	d_1
PID_2	a_2	b_1	c_3	d_2
PID_3	a_1	b_2	c_1	d_1
PID_1	a_3	b_1	c_2	d_2
PID_2	a_2	b_1	c_3	d_3
PID_3	a_1	b_2	c_1	d_2
PID_3	a_1	b_2	c_1	d_3
PID_1	a_3	b_1	c_2	d_3
PID_2	a_2	b_1	c_3	d_1

Анонімність за допомогою методу (k, e)-анонімізації

Цей метод використовується для захисту медичних даних у числовому вигляді. У методі (k, e)-анонімізації записи мають бути розділені на групи, що містять принаймні k різних чутливих значень, а максимальна різниця між цими чутливими значеннями має бути e [5, 10]. У табл. 6 показано поділ даних за допомогою методу (5, 2.3)-анонімізації.

Таблиця 6

Фрагмент БД, що містить записи, виділені методом (5, 2.3)-анонімізації

Псевдоідентифікатор	**Стать**	**Місто**	**TSH**
	(атрибут a)	(атрибут b)	(атрибут c)
PID_1	a_2	b_1	1,26
PID_2	a_2	b_1	3,12
PID_3	a_2	b_1	0,59
PID_4	a_2	b_1	2,01
PID_5	a_2	b_1	2,88

Висновки. Тільки інтеграція правильної технічної безпеки, належного рівня навчання співробітників та організаційних рішень може мінімізувати ймовірність виявлення знань, захищених медичною системою. Серед перерахованих вище категорій безпеки найбільш динамічно розвивається технічна безпека, зокрема резервне копіювання даних, антивірусні програми, програмно-апаратні системи аутентифікації користувачів. З метою підвищення захисту БД від розкриття даних пацієнтів використовуються різноманітні алгоритми поділу даних на менші, специфічні таблиці. Найпоширенішими методами є k-анонімізація, (k, e)-анонімізація, (X, Y)-анонімізація та (α, k)-анонімізація, які поділяють записи БД на таблиці з урахуванням кількості записів для певного псевдоідентифікатора, а найчастіше також з урахуванням інших факторів, наприклад статистичного (α) або диференціального (e). Завдяки методам приховування знань у медичних БД можна звести до мінімуму ефективність різноманітних форм атак на дані пацієнтів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Hassan N. H., Maarop N., Ismail Z., Abidin W. Z. Information security culture in health informatics environment: A qualitative approach. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Langkawi, Malaysia, 2017. Pp. 1–6. DOI: 10.1109/ICRIIS.2017.8002450.
2. Kester Q.-A., Nana L., Pascu A. C., Gire S., Eghan J. M., Quaynor N. N. A Security Technique for Authentication and Security of Medical Images in Health Information Systems, 2015 15th International Conference on Computational Science and Its Applications, Banff, AB, Canada, 2015. Pp. 8–13. DOI: 10.1109/ICCSA.2015.8.
3. Wang Y., Gong L., Zhang M. Remote Disaster Recovery and Backup of Rehabilitation Medical Archives Information System Construction under the Background of Big Data, 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022. Pp. 575–578. DOI: 10.1109/ICSCDS53736.2022.9760774.
4. Chiuchisan I., Balan D.-G., Geman O., Chiuchisan I., Gordin I. A security approach for health care information systems, 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 2017. Pp. 721–724. DOI: 10.1109/EHB.2017.7995525.
5. Zhu J., Chen Z. Exploration of Application Security for Medical Electronic Health Card, 2022 International Conference on Artificial Intelligence in Everything (AIE), Lefkosa, Cyprus, 2022. Pp. 451–454. DOI: 10.1109/AIE57029.2022.00092.
6. Özarar M., Akansu A., Hasbay B. Impact of Cyber Maturity Level on Health Sector, 2021 International Conference on Information Security and Cryptology (ISCTURKEY), Ankara, Turkey, 2021. Pp. 127–131. DOI: 10.1109/ISCTURKEY53027.2021.9654395.
7. Humayed A., Lin J., Li F., Luo B. Cyber-Physical Systems Security—A Survey. *IEEE Internet of Things Journal*. Dec. 2017. Vol. 4. № 6. Pp. 1802–1831. DOI: 10.1109/JIOT.2017.2703172.
8. Zhang M., Chen Y., Lin J. A Privacy-Preserving Optimization of Neighborhood-Based Recommendation for Medical-Aided Diagnosis and Treatment. *IEEE Internet of Things Journal*. 1 July, 2021. Vol. 8. № 13. Pp. 10830–10842. DOI: 10.1109/JIOT.2021.3051060.

9. Mohsen Nia A., Sur-Kolay S., Raghunathan A., Jha N. K. Physiological Information Leakage: A New Frontier in Health Information Security. *IEEE Transactions on Emerging Topics in Computing*. July-Sept. 2016. Vol. 4. № 3. Pp. 321–334. DOI: 10.1109/TETC.2015.2478003.
10. Indumathi J. et al. Block Chain Based Internet of Medical Things for Uninterrupted, Ubiquitous, User-Friendly, Unflappable, Unblemished, Unlimited Health Care Services (BC IoMT U6 HCS). *IEEE Access*. 2020. Vol. 8. Pp. 216856–216872. DOI: 10.1109/ACCESS.2020.3040240.

REFERENCES

1. Hassan, N. H., Maarop, N., Ismail, Z., & Abidin, W. Z. (2017). Information security culture in health informatics environment: A qualitative approach. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), Langkawi, Malaysia, 1–6. DOI: 10.1109/ICRIIS.2017.8002450 (in English).
2. Kester, Q.-A., Nana, L., Pascu, A. C., Gire, S., Eghan, J. M., & Quaynor, N. N. (2015). A Security Technique for Authentication and Security of Medical Images in Health Information Systems, 2015 15th International Conference on Computational Science and Its Applications, Banff, AB, Canada, 8–13. DOI: 10.1109/ICCSA.2015.8 (in English).
3. Wang, Y., Gong, L., & Zhang, M. (2022). Remote Disaster Recovery and Backup of Rehabilitation Medical Archives Information System Construction under the Background of Big Data, 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 575–578. DOI: 10.1109/ICSCDS53736.2022.9760774 (in English).
4. Chiuchisan, I., Balan, D.-G., Geman, O., Chiuchisan, I., & Gordin, I. (2017). A security approach for health care information systems, 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 721–724. DOI: 10.1109/EHB.2017.7995525 (in English).
5. Zhu, J., & Chen, Z. (2022). Exploration of Application Security for Medical Electronic Health Card, 2022 International Conference on Artificial Intelligence in Everything (AIE), Lefkosa, Cyprus, 451–454. DOI: 10.1109/AIE57029.2022.00092 (in English).
6. Özarar, M., Akansu, A., & Hasbay, B. (2021). Impact of Cyber Maturity Level on Health Sector, 2021 International Conference on Information Security and Cryptology (ISCTURKEY), Ankara, Turkey, 127–131. DOI: 10.1109/ISCTURKEY53027.2021.9654395 (in English).
7. Humayed, A., Lin, J., Li, F., & Luo, B. (Dec. 2017). Cyber-Physical Systems Security—A Survey. *IEEE Internet of Things Journal*, 4, 6, 1802–1831. DOI: 10.1109/JIOT.2017.2703172 (in English).
8. Zhang, M., Chen, Y., & Lin, J. (1 July, 2021). A Privacy-Preserving Optimization of Neighborhood-Based Recommendation for Medical-Aided Diagnosis and Treatment. *IEEE Internet of Things Journal*, 8, 13, 10830–10842. DOI: 10.1109/JIOT.2021.3051060 (in English).
9. Mohsen Nia, A., Sur-Kolay, S., Raghunathan, A., & Jha, N. K. (July-Sept. 2016). Physiological Information Leakage: A New Frontier in Health Information Security. *IEEE Transactions on Emerging Topics in Computing*, 4, 3, 321–334. DOI: 10.1109/TETC.2015.2478003 (in English).
10. Indumathi, J. et al. (2020). Block Chain Based Internet of Medical Things for Uninterrupted, Ubiquitous, User-Friendly, Unflappable, Unblemished, Unlimited Health Care Services (BC

IoMT U6 HCS). *IEEE Access*, 8, 216856–216872. DOI: 10.1109/ACCESS.2020.3040240 (in English).

doi: 10.32403/1998-6912-2023-1-66-68-79

METHODS OF ANONYMIZATION OF MEDICAL DATABASES

B. M. Havrysh[1], O. V. Tymchenko[2,3], N. O. Kustra[1]

[1]National University «Lviv Polytechnic»,
12, S. Bandera St., Lviv, 79013, Ukraine

[2]Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine

[3]University of Warmia and Mazury in Olsztyn,
2, Michala Oczapowskiego St., Olsztyn, 10-719, Poland
dana.havrysh@gmail.com

The development of electronic forms of storage, processing and transmission of medical data has influenced not only the improvement of the quality of medical care for patients, but also the development of new methods of obtaining knowledge from medical databases by unauthorized persons. To prevent the disclosure of confidential data, medical information systems must be constantly tested for security in all system structures. Technical security measures should be complemented by physical, personal and organizational security measures. The methods described in the work are the main methods of medical data anonymization, on the basis of which other anonymization methods have been developed, such as: l-diversification, (X, Y)-connectivity, (X, Y)-privacy, LKC-privacy closure, bounded trust, and personalized privacy. Thanks to anonymization methods in medical databases, the effectiveness of attacks on patient data can be minimized.

The patient's medical record is a key element during his treatment, as it contains all the information about the state of health, tests performed, stays in the hospital and procedures performed over the years. A few years ago, medical documentation was mostly in paper form. Currently, it is slowly being replaced by electronic forms. One fact has not changed over the years – most often it is the patient who is responsible for transporting him to another medical institution. Therefore, in emergency cases during treatment in a new institution, the level of knowledge about this patient is zero. This problem is solved by the introduction of the electronic medical record EMR (English Electronic Medical Record - EMR). EMR is a virtual document that consists of all medical records in digital form belonging to one patient. Thanks to this solution, patient information can be created, stored and used in many different medical facilities and made available to the patient in a single document in a web application.

The introduction of an electronic system of medical documentation brings advantages, but also creates new problems. The advantages are improving the quality of medical care for patients, more efficient and much more effective management (the electronic prescription system allows controlling unwanted interactions between drugs prescribed and taken at the same time), supporting the decisions of doctors and reducing medical errors by up to 55%, remote treatment, which is good in big cities, intercity, intercontinental. The most serious consequence of the transfer of resources from hospital databases to the network is problems with control and protection of information contained in medical documents. The integrity of digital objects is also an issue in the case of multi-module EHR systems.

***Keywords**: anonymization, pseudo-identifier, attack, system protection, security.*

Стаття надійшла до редакції 10.02.2023.
Received 10.02.2023.