

Privacy Threats to Medical and Nutritional Data in Digital Healthcare Systems and a Conceptual Model for Their Protection

Matyoqubov Og'abek Bahrombek o'g'li

Urganch State University

Yusupov Bahodir Karomatovich

Institute of Information and Communication Technologies and Military

Communications

Abstract

The paper examines personal medical and nutritional data streams on digital healthcare platforms and the cyber threats targeting them. Re-identification, attribute inference and membership inference attacks caused by hidden correlations between quasi-identifiers and clinical and dietary attributes are classified. A multilevel aggregative security model and a method for estimating attribute criticality weights using Shannon mutual information are proposed.

Keywords

digital healthcare, sensitive personal data, diet, quasi-identifiers, attribute inference attack, Shannon entropy, aggregative security model.

Raqamli tibbiyot tizimlarida tibbiy va ovqatlanish ma'lumotlarining maxfiylik tahdidlari hamda ularni himoyalashning konseptual modeli

Matyoqubov Og'abek Bahrombek o'g'li

Urganch davlat universiteti

Yusupov Bahodir Karomatovich

Axborot-kommunikatsiya texnologiyalari va harbiy aloqa instituti

Annotatsiya. Tezisdan raqamli tibbiyot platformalaridagi shaxsiy tibbiy va ovqatlanish ma'lumotlari oqimi hamda unga qaratilgan kiberxavfsizlik tahdidlari tadqiq etilgan. Kva-identifikatorlar bilan klinik va parhez atributlari o'rtasidagi yashirin

korrelyatsiya tufayli yuzaga keladigan qayta aniqlash, atribut va a'zolik xulosasi hujumlari tasniflangan. Ko'p sathli agregativ xavfsizlik modeli va Shennon o'zaro axborot miqdori asosida atributlarning kritiklik vaznlarini aniqlash usuli taklif etilgan.

Kalit so'zlar: raqamli tibbiyot, shaxsga doir nozik ma'lumotlar, ovqatlanish ratsioni, kva-identifikatorlar, atribut xulosasi hujumi, Shennon entropiyasi, agregativ xavfsizlik modeli.

1. Kirish va muammoning dolzarbligi

Taqiladigan datchiklar va mobil parhez ilovalari natijasida foydalanuvchilarning fiziologik holati va ratsioniga oid ma'lumotlar real vaqtda bulutli bazalarga uzatilmoqda. IBM hisobotiga ko'ra, 2024-yilda sog'liqni saqlashda ma'lumotlar sizishining o'rtacha zarari 9,77 mln AQSH dollari bo'lib, tarmoqlar orasida eng yuqori bo'lgan [3].

O'RQ-547-son Qonunning 19-moddasida jismoniy va ruhiy salomatlikka oid ma'lumotlar maxsus toifaga kiritilgan [1]. Vazirlar Mahkamasining 570-son qaroriga ko'ra, bunday tizimlar birinchi darajali himoyaga ega bo'lishi shart [2].

2. Ma'lumotlar xususiyatlari va maxfiylik tahdidlari tahlili

Foydalanuvchi profili geterogen va vaqtga bog'liq bo'lib, to'rtta atributdan iborat kortej sifatida ifodalanadi:

$$P_u = \langle DI_u, QI_u, SA_m(t), SA_n(t) \rangle \quad (1)$$

bunda DI_u – to'g'ridan-to'g'ri identifikatorlar, QI_u – kva-identifikatorlar, $SA_m(t)$ – klinik, $SA_n(t)$ – ovqatlanish atributlari oqimi. Sweeney tadqiqotiga ko'ra, jins, tug'ilgan sana va pochta indeksi AQSh aholisining 87 foizini yagona aniqlaydi [4]. Tahdidlar 1-jadvalda keltirilgan.

1-jadval. Ma'lumot toifalarining maxfiylik darajasi va tahdidlar matritsasi

Ma'lumot toifasi	Atributlar	Maxfiylik darajasi	Asosiy tahdidlar
To'g'ridan-to'g'ri identifikatorlar (DI)	F.I.Sh., JShShIR, pasport, telefon	O'ta yuqori	Identifikatorlarni o'g'irlash, kamalak jadvallari hujumi
Kva-identifikatorlar (QI)	Yosh, jins, tana massasi indeksi, hudud	Yuqori	Tashqi bazalar bilan bog'lash va qayta aniqlash
Ovqatlanish atributlari (SAn)	Kaloriya, uglevod, natriy, taom vaqti, allergenlar	O'rta-yuqori	Atribut xulosasi, ijtimoiy va e'tiqodiy profil yaratish
Klinik ko'rsatkichlar (SAm)	Glyukoza, qon bosimi, puls, tashxislar	O'ta yuqori	A'zolik xulosasi, rekonstruksiya, kanalni tinglash

Sintaktik modellar (k -anonimlik [4], l -xilma-xillik [5], t -yaqinlik [6]) ko'p o'lchovli oqimlarda o'lchamlilik muammosiga uchrab, tahliliy foydalilikni sezilarli yo'qotadi [7].

3. Konseptual va agregativ xavfsizlik modeli

Foydalanuvchi ma'lumotlari oqimi to'rt sathda – identifikatsion dekompozitsiya, korrelyatsion vaznlash, adaptiv transformatsiya va kompozitsion byudjet nazoratida qayta ishlanadi (2-rasm). Diskret vaqt intervali $[t_{k-1}, t_k)$ da tizim yettilik kortej bilan beriladi:

$$M(k) = \{ Z, L, T, F, G, X, Y \}, \quad z_{k+1} = F(L, x_k, z_k), \quad y_{k+1} = G(L, x_k, z_k) \quad (2)$$

Ichki holat $z_k = (B_k, W_k, \Sigma_k)$: B_k – maxfiylik byudjeti taqsimoti, W_k – kritiklik vaznlari, Σ_k – atributlar korrelyatsiya matritsasi. Ovqatlanish atributi X_i orqali klinik tashxis X_j ni aniqlash xavfi Shannon entropiyasi va o'zaro axborot miqdori bilan baholanadi:

$$H(X_i) = - \sum p(x_i) \log_2 p(x_i), \quad I(X_i; X_j) = H(X_i) + H(X_j) - H(X_i, X_j) \quad (3)$$

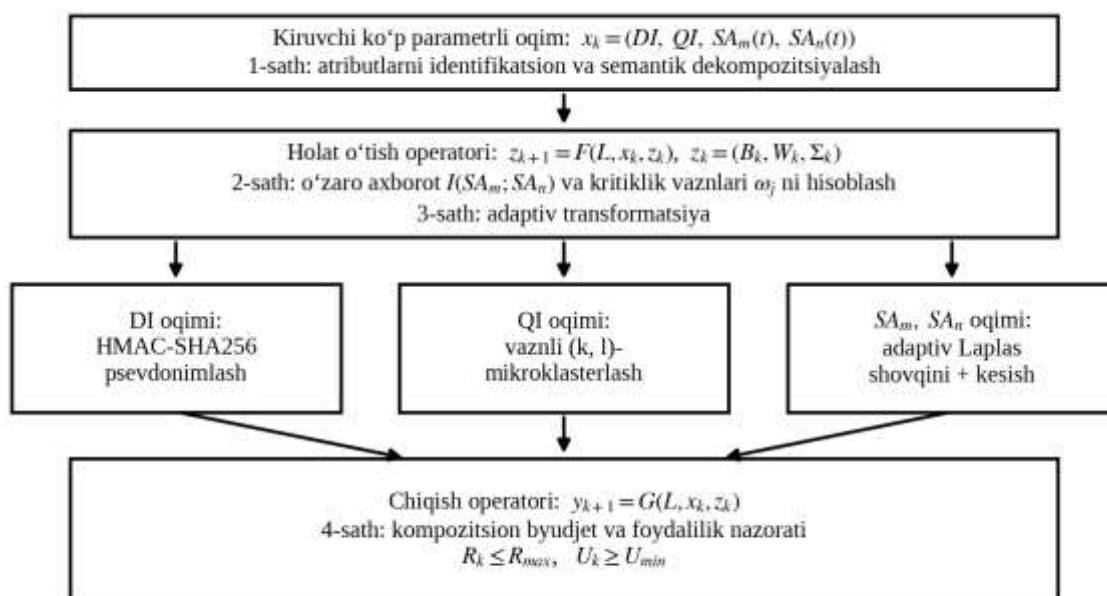
Ularni $[0, 1]$ oraliqda normallashtirib, juftlik koeffitsiyenti ρ_{ij} va kritiklik vazni ω_j topiladi:

$$\rho_{ij} = 2I(X_i; X_j) / [H(X_i) + H(X_j)], \quad \omega_j = \alpha_j(1 + \max \rho_{ij}) / \sum \alpha_r(1 + \max \rho_{ir}) \quad (4)$$

Vaznlar asosida chiqish operatori har bir atributga differensial himoya qo'llaydi [8]:

$$\varepsilon_{k,j} = \varepsilon_k^{(step)} \cdot \omega_j^{-\beta} / \sum \omega_r^{-\beta}, \quad y_{k,j} = \Pi_{[a_j, b_j]}(x_{k,j} + \text{Lap}(\Delta f_j / \varepsilon_{k,j})) \quad (5)$$

Yuqori korrelyatsiyali atributlarga katta ω_j beriladi, ularga kichik $\varepsilon_{k,j}$ ajratiladi.



2-rasm. Ma'lumotlar oqimini himoyalashning konseptual modeli

4. Xulosa

Ma'lumotlar kuchli korrelyatsion bog'liqlikka ega bo'lgani uchun statik nomsizlantirish atribut xulosasi va qayta aniqlash hujumlaridan himoya qila olmaydi. Taklif etilgan to'rt sathli model $M(k)$ va Shennon o'zaro axborotiga asoslangan kritiklik vaznlari yuqori korrelyatsiyali atributlarni aniqlab, ularga adaptiv differensial himoya qo'llash imkonini beradi.

Foydalanilgan adabiyotlar ro'yxati

1. O'zbekiston Respublikasining «Shaxsga doir ma'lumotlar to'g'risida»gi O'RQ-547-son Qonuni, 02.07.2019.
2. O'zbekiston Respublikasi Vazirlar Mahkamasining 570-son qarori, 05.10.2022.
3. IBM Security. Cost of a Data Breach Report 2024. – IBM Corporation, 2024.
4. Sweeney L. k-Anonymity: a model for protecting privacy // International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems. – 2002. – Vol. 10, No. 5. – P. 557–570.
5. Machanavajjhala A., Kifer D., Gehrke J., Venkatasubramanian M. l-Diversity: privacy beyond k-anonymity // ACM Transactions on Knowledge Discovery from Data. – 2007. – Vol. 1, No. 1.
6. Li N., Li T., Venkatasubramanian S. t-Closeness: privacy beyond k-anonymity and l-diversity // IEEE 23rd International Conference on Data Engineering. – 2007. – P. 106–115.
7. Brickell J., Shmatikov V. The cost of privacy: destruction of data-mining utility in anonymized data publishing // Proc. 14th ACM SIGKDD. – 2008. – P. 70–78.
8. Dwork C., Roth A. The algorithmic foundations of differential privacy // Foundations and Trends in Theoretical Computer Science. – 2014. – Vol. 9, No. 3–4. – P. 211–407.