ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ
ԻՆՍՏԻՏՈՒՏ

Սանթրոսյան Ջեմմա Սպարտակի

ՏՎՅԱԼՆԵՐԻ ՀԵՆՔԵՐՈՒՄ ԻՆՖՈՐՄԱՑԻԱՅԻ ԳԱՂՏՆԻՈՒԹՅՈՒՆ-ՕԳՏԱԿԱՐՈՒԹՅՈՒՆ ՓՈԽՋԻՋՄԱՆ ԽՆԴՐԻ ԻՆՖՈՐՄԱՑԻՈՆ-ՏԵՍԱԿԱՆ ՄՈՏԵՑՈՒՄԸ

Ե.13.05 - «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ ֆիզիկամաթեմատիկական գիտությունների թեկնածուի գիտական աստիճանի համար

ՍԵՂՄԱԳԻՐ

Երևան 2026

INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS

Santrosyan Jemma

AN INFORMATION-THEORETIC APPROACH TO THE PRIVACY–UTILITY TRADEOFF PROBLEM IN DATABASES

ABSTRACT

of the dissertation for obtaining a Ph.D. degree in **Physical and Mathematical Sciences** on specialty 05.13.05 “Mathematical modelling, numerical methods and program complexes”

Yerevan 2026

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝ ֆիզ. մաթ. գիտ. դոկտ. Մ. Ե. Հարությունյան

Պաշտոնական ընդդիմախոսներ՝ տեխ. գիտ. դոկտ. Դ.Գ.Ասատրյան

ֆիզ. մաթ. գիտ. թեկն. Ա.Օ. Եսայան

Առաջատար կազմակերպություն՝ Երևանի պետական համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2026թ. դեկտեմբերի 24-ին ժամը 14:00-ին Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԻԱՊԻ գրադարանում:

Սեղմագիրն առաքված է 2026 թ.-ի նոյեմբերի 24-ին:

Մասնագիտական խորհրդի գիտական

քարտուղար ֆիզ. մաթ. գիտ. դոկտոր՝ Մ. Ե. Հարությունյան

The topic of the dissertation was approved at the Institute of Informatics and Automation Problems.

Scientific supervisor: M. E. Harutyunyan, D. Ph. M. S.
Official opponents: D. G. Asatryan, D. T. S.
A. O. Yesayan, Ph.D.

Leading organization: Yerevan State University

The Defense will take place on 24 December 2026; at 14:00, at the Specialized Council 037 “Informatics” at the Institute of Informatics and Automation Problems.
Address: Yerevan, 0014, P. Sevak 1.

The Dissertation is available at the library of IIAP.

The abstract is delivered on 2 4 November, 2026.

Scientific Secretary of the Specialized Council, D.Ph.M.S. M. E. Haroutunian

Աշխատանքի ընդհանուր նկարագիրը

Հետազոտության թեմայի արդիականությունը:

Ժամանակակից տվյալների համակարգերն ավելի ու ավելի են հիմնվում մեծածավալ տվյալների հենքերի հավաքագրման, մշակման և հրապարակման վրա: Այս տվյալների հենքերը արժեքավոր են, քանի որ ապահովում են վերլուծության, կանխատեսման, կառավարման, անհատականացման և որոշումների կայացման հնարավորություններ: Միևնույն ժամանակ, օգտակար տվյալների հրապարակումը կարող է հանգեցնել զգայուն տեղեկատվության արտահոսքի՝ հանրային և մասնավոր հատկանիշների միջև առկա վիճակագրական կախվածության պատճառով:

Կոմունալ ծառայությունների մատակարարը, առողջապահական տվյալների վերլուծաբանը, ամպային համակարգի կառավարիչը կամ մեքենայական ուսուցման ծառայությունը կարող է կարիք ունենալ մաքրված (անանունացված) չափումների հասանելիության՝ կանխատեսման, մշտադիտարկման կամ կառավարման նպատակներով: Միևնույն ժամանակ, այդ չափումները կարող են բացահայտել անձնական բնույթի տեղեկություններ, ինչպիսիք են՝ բնակության վայրում մարդկանց ներկայության օրինաչափությունները, առողջական վիճակը, սարքավորումների օգտագործումը, տեղաշարժի հետագծերը կամ տնային տնտեսության վարքագիծը:

Հավաքագրվող տվյալների վերլուծությունները պահանջում են դրանց հրապարակումը առանց գաղտնագրման, ինչը կարող է հանգեցնել զգայուն ինֆորմացիայի արտահոսքին, այսինքն անձնական գաղտնիության խափանմանը: Ի տարբերություն գաղտնիության ապահովման խնդրի (secrecy problem), անձնական տվյալների գաղտնիության պաշտպանության (privacy problem) դեպքում տվյալների բացահայտումը հնարավորություն է տալիս ստանալ տեղեկատվական օգտակարություն, միաժամանակ ստեղծելով անձնական գաղտնիության խախտման հնարավոր ռիսկ: Հրապարակային և գաղտնի տվյալների միջև վիճակագրական կախվածությունների պատճառով օգտակար տվյալների հաղորդումը կարող է բացահայտել նաև գաղտնի բաղադրիչին վերաբերող տեղեկություններ:

Այս պրոբլեմի վերաբերյալ գրականությունում առկա են մի շարք մոտեցումներ, որոնք ոչ բոլոր կիրառություններում են ընդունելի և չեն լուծում ոլորտի բոլոր խնդիրները: Տվյալների հենքերում ինֆորմացիայի օգտակարության և գաղտնիության համատեղ ապահովումը կապված է հրապարակվող տվյալների վերականգնման ճշգրտության, դրանց հաղորդման համար անհրաժեշտ արագության և գաղտնի տեղեկատվության պահպանման պահանջների հետ:

Ինֆորմացիոն տեսական գաղտնիությունն այս խնդրի նկարագրման համար տրամադրում է բնական և հարմար միջոց: Հետևելով տվյալների հենքերի համար առաջարկված **օգտակարություն–գաղտնիություն (utility–privacy)** շրջանակին[1] տվյալների հենքի յուրաքանչյուր գրառում կարելի է բաժանել երկու բաղադրիչի՝ հանրային բաղադրիչ **X**, որը օգտակար է և պետք է վերականգնվի որոշակի ճշգրտությամբ, և մասնավոր բաղադրիչ

[1] Sankar L., Rajagopalan S. R., Poor H. V., "Utility-privacy tradeoffs in databases: An information-theoretic approach", IEEE Transactions on Information Forensics and Security, vol. 8, no. 6, pp. 838–852, 2013. doi: 10.1109/TIFS.2013.2253320

Y, որը տվյալների հրապարակումից հետո պետք է հնարավորինս անորոշ մնա: Այսպիսով, տվյալների մաքրման (sanitization) գործընթացը կարելի է դիտարկել որպես **կորուստներով աղբյուրի կոդավորում՝ գաղտնիության սահմանափակմամբ**[2]:

Շենոնի արագություն–շեղում հայտնի տեսությունը բնութագրում է վերականգնման տրված ճշգրտությունն ապահովող փոքրագույն կոդավորման արագությունը: Գաղտնի բաղադրիչով աղբյուրի կոդավորման վերաբերյալ Յամամոտոյի արդյունքներում վերականգնման շեղման սահմանափակմանը միանում է գաղտնի բաղադրիչի անորոշության պահպանման պահանջը: Տվյալների հենքերում օգտակարության և գաղտնիության փոխառնչությունների ինֆորմացիոն-տեսական ձևակերպմամբ այդ մոտեցումը կիրառվում է տվյալների հրապարակման խնդրում. հրապարակային բաղադրիչը պետք է վերականգնվի պահանջվող ճշգրտությամբ, իսկ գաղտնի բաղադրիչը հրապարակումից հետո պետք է մնա հնարավորինս անորոշ:

Վերականգնման միջին շեղման նկատմամբ սահմանափակումը, սակայն, ինքնին չի երաշխավորում, որ շեղման թույլատրելի մակարդակի գերազանցման հավանականությունը բավականաչափ փոքր կլինի: Ուստի օգտակարության և գաղտնիության պահանջների հետ մեկտեղ անհրաժեշտ է դիտարկել նաև հուսալիության չափանիշը: Հուսալիության պահանջը սահմանվում է բլոկի երկարության մեծացմանը զուգընթաց շեղման շեմի գերազանցման հավանականության ցուցչային նվազմամբ: Այդ պահանջի հաշվառումը հնարավորություն է տալիս գնահատել, թե ինչ կոդավորման արագություն է անհրաժեշտ վերականգնման ճշգրտությունը և գաղտնիության սահմանված մակարդակը տրված հուսալիությամբ ապահովելու համար: Սա հանգեցնում է **հուսալիություն-շեղում–արագություն (Rate-Reliability-Distortion) ֆունկցիայի** հասկացությանը, որը նկարագրում է կոդավորման արագությունը՝ որպես աղավաղման մակարդակի և սխալի նախանշված ցուցչի կամ հուսալիության պարամետրի ֆունկցիա: Այս ֆունկցիան ուսումնասիրվել է աղբյուրի տարբեր մոդելների համար [3]:

Անձնական ինֆորմացիայի պաշտպանության խնդիրներում հարկ է հաշվի առնել նաև կողմնակի ինֆորմացիայի հնարավոր առկայությունը, որը ազդում է կոդավորման հնարավորությունների և գաղտնիության մակարդակի վրա: Հետևաբար, անհրաժեշտ է ուսումնասիրել ինչպես գաղտնի բաղադրիչով աղբյուրի կոդավորումը, այնպես էլ կողմնակի ինֆորմացիայով մոդելները և այդ երկու պայմանների համատեղ առկայությունը: Ատենախոսությունը նվիրված է այս խնդիրների ուսումնասիրությանը և ստացված արդյունքների կիրառությանը տվյալների հենքերում օգտակարության ու գաղտնիության փոխզիջման խնդրում՝ հուսալիության պահանջի հաշվառմամբ:

---

[2] Yamamoto H., "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers", IEEE Transactions on Information Theory, vol. 29, no. 6, pp. 918-923, 1983. doi: 10.1109/TIT.1983.1056749

[3] Haroutunian E. A., Haroutunian M. E., Harutyunyan A. N., "Reliability criteria in information theory and in statistical hypothesis testing", Foundations and Trends in Communications and Information Theory, vol. 4, nos. 2–3, pp. 97–263, 2007. doi: 10.1561/0100000008.

Աշխատանքի նպատակը և հիմնական խնդիրները:

Ատենախոսության նպատակն է ուսումնասիրել տվյալների հենքերում ինֆորմացիայի գաղտնիություն-օգտակարություն փոխզիջման խնդիրը հուսալիության բաղադրիչի ներմուծմամբ և լուծել այդ խնդիրը ինֆորմացիայի տեսության գործիքակազմի ներդրմամբ:

Այդ նպատակին հասնելու համար դիտարկվել են հետևյալ խնդիրները:

1. Հետազոտել աղբյուրի կոդավորման խնդիրը փոխկապակցված ելքերով միակողմանի աղբյուրի համար:
2. Դիտարկել աղբյուրի մոդելը կողմնակի ինֆորմացիայի առկայության դեպքում, որը հայտնի է կոդավորիչում և ապակոդավորիչում:
3. Միավորել երկու մոդելները, այսինքն դիտարկել աղբյուրի կոդավորման խնդիրը կոռելացված ելքերով, միաժամանակ ունենալով կողմնակի ինֆորմացիա՝ հասանելի կոդավորիչին և ապակոդավորիչին:
4. Կիրառել ստացված արդյունքները տվյալների հենքերում օգտակարության ու գաղտնիության համատեղ ապահովման խնդրում:

Հետազոտության օբյեկտը և առարկան:

Հետազոտության օբյեկտը վերջավոր այբուբեններով, առանց հիշողության դիսկրետ աղբյուրների կոդավորման և տվյալների հենքերի հրապարակման մոդելներն են: Հետազոտության առարկան կոդավորման արագության, վերականգնման շեղման, հուսալիության ցուցչի և գաղտնի բաղադրիչի անորոշության համատեղ սահմանափակումներով հասանելի տիրույթներն ու դրանք բնութագրող ֆունկցիաներն են:

Հետազոտության մեթոդները:

Աշխատանքում կիրառվում են ինֆորմացիայի տեսության հավանականային և կոմբինատոր մեթոդները: Հիմնական արդյունքների ապացույցները կատարվում են կազմերի մեթոդի, կազմային դասերի հզորությունների և հավանականությունների գնահատականների, ինչպես նաև ծածկման լեմմաների համապատասխան ձևափոխությունների միջոցով: Հասանելիության ապացույցների համար կառուցվում են սահմանված պահանջները բավարարող կոդեր, իսկ հակադարձ մասերում ստացվում են կոդավորման արագության և անորոշության անհրաժեշտ գնահատականները:

Գիտական նորույթը:

Աշխատանքի հիմնական նորույթը գաղտնիություն-օգտակարություն խնդրում հուսալիության E ցուցչի ներմուծումն է, որը քանակապես բնութագրում է ապակոդավորման սխալի հավանականության ցուցչային նվազման արագությունը: Ստացված արդյունքները բնութագրում են կոդավորման արագության, հուսալիության, վերականգնման շեղման և գաղտնիության միջև փոխզիջումները: Առաջարկված մոտեցումը հնարավորություն է տալիս ուսումնասիրելու այդ փոխզիջումները հուսալիության տարբեր պահանջների դեպքում՝ դասական խնդիրը ներառելով որպես սահմանային դեպք:

Աշխատանքի տեսական և գործնական նշանակությունը:

Ստացված արդյունքները բնութագրում են աղբյուրի կոդավորման այն հնարավորությունները, որոնց դեպքում վերականգնման ճշգրտությունը, հուսալիությունը և գաղտնիության մակարդակը կարող են ապահովվել միաժամանակ: Համապատասխան ֆունկցիաները որոշում են տրված սահմանափակումները բավարարող փոքրագույն կոդավորման արագությունը և գաղտնի բաղադրիչի անորոշության առավելագույն հասանելի մակարդակը: Արդյունքները տրամադրում են միասնական ինֆորմացիոն-տեսական հիմք, որն արտացոլում է սեղմման արդյունավետության, վերականգնման ճշգրտության, հուսալիության և գաղտնիության փոխազդեցությունը:

Տվյալների հենքի հրապարակումը կորուստներով աղբյուրի կոդավորման միջոցով ներկայացնելը հնարավորություն է տալիս հաղորդման արագությունը, օգտակարությունը, գաղտնիությունը և հուսալիությունը ուսումնասիրել միասնական մոդելում: Այս մոտեցումը տեսական հիմք է ապահովում տարբեր կիրառություններում տվյալների հուսալի և գաղտնիությունը պահպանող հրապարակման մեխանիզմների վերլուծության համար: Խելացի էլեկտրացանցի թվային օրինակը, իրականացված R միջավայրում, ցույց է տալիս հուսալիության երաշխիքների ապահովման հետ կապված գաղտնիության և հաղորդակցման արագության փոփոխությունները:

Ստացված արդյունքների քննարկումը և փորձարկումը:

Ատենախոսության հիմնական արդյունքները զեկուցվել և քննարկվել են՝

- «Computer Science and Information Technologies» 15-րդ միջազգային գիտաժողովում (CSIT 2025, Երևան, 2025 թ. սեպտեմբերի 22–26),
- «Collaborative Technologies and Data Science in Smart City Applications» միջազգային գիտաժողովում (CODASSCA 2026, Երևան, 2026 թ. սեպտեմբերի 7–9):
- Վանաձորի պետական համալսարանի (ՎՊՀ) գիտական սեմինարներում,
- Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի (ԻԱՊԻ) գիտական սեմինարում:

Հրապարակումներ:

Ատենախոսության հիմնական արդյունքները հրապարակվել են չորս գիտական աշխատանքներում, որոնց ցանկը բերված է սեղմագրի վերջում:

Աշխատանքի բովանդակությունը

Ներածություն բաժնում հիմնավորվում է ատենախոսության արդիականությունը, ձևակերպված է աշխատանքի նպատակը, դիտարկված խնդիրները, գիտական նորույթը, կիրառական նշանակությունը և պաշտպանության ներկայացված հիմնական դրույթները:

Առաջին գլուխն ունի ներածական բնույթ: Վերլուծվել են տվյալների հենքերում մասնավոր ինֆորմացիայի պաշտպանության հիմնական մոտեցումները: Ձևակերպվել է օգտակարություն–գաղտնիություն փոխզիջման խնդիրը և հիմնավորվել է դրա ուսումնասիրության անհրաժեշտությունը՝ վերականգնման հուսալիության հաշվառմամբ:

Բերվել են ինֆորմացիայի տեսության հիմնական հասկացությունները՝ X դիսկրետ

պատահական մեծության էնտրոպիան՝

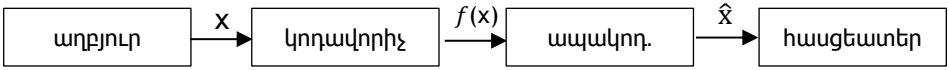
$$H_P(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x),$$

X, Y պատահական մեծությունների փոխադարձ ինֆորմացիան՝

$$I_P(X;Y) = H_P(X) - H_P(X \mid Y).$$

Ներկայացվել են նաև պայմանական փոխադարձ ինֆորմացիան, Մարկովի շղթան և Կուլբակի–Լեյբլերի $D(P \parallel Q)$ ինֆորմացիոն տարամիտությունը: Բոլոր լոգարիթմների հիմքը 2 է:

Ներկայացվել են կազմերի մեթոդի հիմնական գնահատականներն ու լեմմաները, որոնք կիրառվել են հետագա ապացույցներում: Ներկայացվել են աղբյուրի կոդավորման Շենոնի դասական խնդիրը և արագություն–շեղում ֆունկցիան, որը որոշում է տրված ճշգրտությամբ վերականգնման փոքրագույն արագությունը:



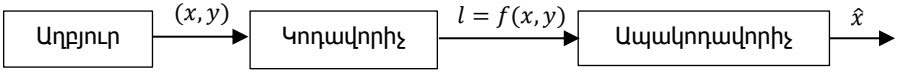
Նկար 1: Անաղմուկ կապի համակարգի գծապատկերը:

Առանձին ներկայացվել է արագություն–հուսալիություն–շեղում $R(E,\Delta,P^*)$ ֆունկցիան: Հուսալիության $E > 0$ ցուցչով պահանջը սահմանվել է թույլատրելի շեղման գերազանցման հավանականության ցուցչային նվազմամբ՝ կամայական $\delta > 0$-ի և բավականաչափ մեծ N-երի համար՝

$$e(f_N, g_N, P^*, \Delta) \le exp\{-N(E-\delta)\}.$$

Այստեղ N-ը բլոկի երկարությունն է: Ներկայացվել է հուսալիության տրված պահանջն ապահովող փոքրագույն արագության արտահայտությունը:

Դիտարկվել է նաև գաղտնի բաղադրիչով աղբյուրի կոդավորման Յամամոտոյի մոդելը, որտեղ X-ը վերականգնվող օգտակար բաղադրիչն է, իսկ Y-ը՝ գաղտնի բաղադրիչը:



Նկար 2: Գաղտնի բաղադրիչով աղբյուրի մոդելը

Գաղտնիությունը գնահատվում է հաղորդված ինդեքսի պայմանում պահպանվող անորոշությամբ: Ներկայացվել են արագության, շեղման և անորոշության համատեղ հասանելիության դասական սահմանները, որոնք հիմք են ծառայում հետագա գլուխների ուսումնասիրությունների համար:

Երկրորդ գլուխը նվիրված է գաղտնի բաղադրիչով աղբյուրի կոդավորման խնդրին՝ հուսալիության պահանջի հաշվառմամբ [1]: Դիտարկվել է վերջավոր այբուբեններով, P^*

համատեղ բաշխմամբ, առանց հիշողության դիսկրետ աղբյուր՝ փոխկապակցված (X,Y) ելքերով: Օգտակար X բաղադրիչը պետք է վերականգնվի շեղման տրված մակարդակով, իսկ Y բաղադրիչը՝ հնարավորինս գաղտնի մնա: N երկարությամբ բլոկը կոդավորվում է $J = f_N(\boldsymbol{x}, \mathbf{y})$ ինդեքսով, որի հիման վրա կառուցվում է $\hat{\mathbf{x}} = g_N(J)$ վերականգնումը:

Մեկ աղբյուրային ազդանշանին բաժին ընկնող կոդավորման արագությունը հավասար է

$$R_N = \frac{1}{N}\log L(N):$$

Տրված $E > 0$-ի դեպքում (R, Δ_d, Δ_e) եռյակը սահմանվել է որպես *E-հասանելի*, եթե կամայական $\varepsilon > 0$-ի և $\delta > 0$-ի համար բավականաչափ մեծ N-երի դեպքում գոյություն ունեն հետևյալ պահանջները բավարարող կոդեր՝

$$\frac{1}{N}\log L(N) \le R + \varepsilon,$$
$$e(f_N, g_N, P^*, \Delta_d) \le exp\{-N(E-\delta)\},$$
$$\frac{1}{N}H(\mathbf{Y} \mid L(N)) \ge \Delta_e - \varepsilon:$$

Այստեղ $L(N)$-ը կոդային ինդեքսների քանակն է, Δ_d-ն՝ թույլատրելի շեղման շեմը, իսկ Δ_e-ն՝ պահպանվող անորոշության պահանջվող նվազագույն մակարդակը: Սխալի հավանականությունը սահմանվել է որպես շեղման շեմի գերազանցման հավանականություն:

Բոլոր E-հասանելի եռյակների բազմությունը նշանակվել է $\mathcal{R}^*(E)$-ով: Ներմուծվել են արագություն–հուսալիություն–շեղում–անորոշություն $R^*(E, \Delta_d, \Delta_e)$ (RRDE) և անորոշություն–հուսալիություն–շեղում $\Gamma^*(E, \Delta_d)$ (ERD) ֆունկցիաները: Դրանք որոշում են համապատասխանաբար տրված պահանջներն ապահովող փոքրագույն արագությունը և հասանելի առավելագույն անորոշությունը:

Արդյունքների ձևակերպման համար դիտարկվել է բաշխումների հետևյալ բազմությունը՝

$$\alpha(E, P^*) = \{P : D(P \parallel P^*) \le E\}:$$

$\mathcal{Q}(P, \Delta_d, \Delta_e)$-ով նշանակվել է $Q_p(\hat{x} \mid x, y)$ պայմանական բաշխումների բազմությունը, որոնց համար՝

$$\mathbb{E}_{P,Q_P} d(X, \hat{X}) \le \Delta_d, \quad H_{P,Q_P}(Y \mid \hat{X}) \ge \Delta_e:$$

$\mathcal{Q}(P, \Delta_d)$-ով նշանակվել է միայն $\mathbb{E}_{P,Q_p} d(X, \hat{X}) \le \Delta_d$ սահմանափակումը բավարարող պայմանական բաշխումների բազմությունը:

Թեորեմ 2.1: *Տրված P^*-ի դեպքում, յուրաքանչյուր $E > 0$-ի համար՝*

$$\mathcal{R}^*(E) = \left\{ \begin{array}{c} (R, \Delta_d, \Delta_e) : \Delta_d \ge 0, \Delta_e \ge 0 \\ 0 \le R_e \le \min\limits_{P \in \alpha(E,P^*)} \max\limits_{Q_P \in \mathcal{Q}(\mathcal{P}, \Delta_d)} H_{P,Q_P}(Y \mid \hat{X}), \\ R \ge \max\limits_{P \in \alpha(E,P^*)} \min\limits_{Q_P \in \mathcal{Q}(\mathcal{P}, \Delta_d, \Delta_e)} I_{P,Q_P}(X, Y; \hat{X}) \end{array} \right\}:$$

Հետևանք 2.1: *Անորոշություն–հուսալիություն–շեղում (*ERD*) ֆունկցիան հավասար է՝*

$$\Gamma^*(E,\Delta_d)=\min_{P\in\alpha(E,P^*)}\max_{Q_P\in\mathcal{Q}(\mathcal{P},\Delta_d)} H_{P,Q_P}\big(Y\mid\hat{X}\big):$$

Հետևանք 2.2: *Շեղում–անորոշություն* E*-հասանելիության տիրույթն (*DE*))ունի հետևյալ տեսքը՝*

$$\mathcal{R}^*_{\Delta_d,\Delta_e}(E)=\begin{Bmatrix}R(E,\Delta_d,\Delta_e):\Delta_d\geq 0,\\ 0\leq\Delta_e\leq\Gamma^*(E,\Delta_d)\end{Bmatrix}:$$

Հետևանք 2.3: *Արագություն–հուսալիություն–շեղում–անորոշություն ֆունկցիան (*RRDE*) հավասար է՝*

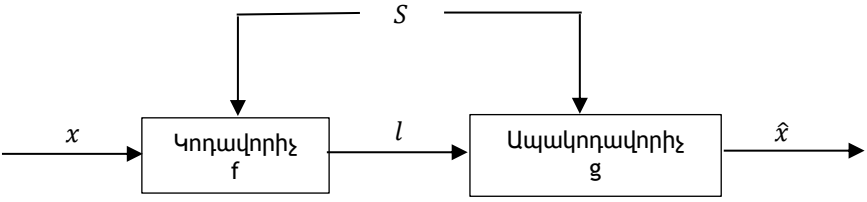
$$R^*(E,\Delta_d,\Delta_e)=\max_{P\in\alpha(E,P^*)}\min_{Q_P\in\mathcal{Q}(\mathcal{P},\Delta_d,\Delta_e)} I_{P,Q_P}\big(X,Y;\hat{X}\big):$$

Հետևանք 2.4: *Երբ E-ն ձգտում է 0-ի,* RRDE *և* ERD *ֆունկցիաների սահմանները համընկնում են Յամամոտոյի դասական խնդրում ներկայացված RDR և ED ֆունկցիաների հետ՝*

$$\lim_{E\to 0}R^*(E,\Delta_d,\Delta_e)=R^*(\Delta_d,\Delta_e)=\min_{Q^*_P\in\mathcal{Q}(\mathcal{P}^*,\Delta_d,\Delta_e)} I_{P^*,Q_{P^*}}\big(X,Y;\hat{X}\big)$$
$$\lim_{E\to 0}\Gamma^*(E,\Delta_d)=\Gamma^*(\Delta_d)=\max_{Q_{P^*}\in\mathcal{Q}(\mathcal{P}^*,\Delta_d)} H_{P^*,Q_{P^*}}\big(Y\mid\hat{X}\big):$$

Թեորեմի ապացույցը կատարվել է կազմերի մեթոդի և ծածկման լեմմայի ձևափոխության միջոցով: Ուղիղ մասում կառուցվել են արագության, շեղման, անորոշության և հուսալիության պահանջները բավարարող կոդեր: Հակադարձ մասում ստացվել են արագության և պահպանվող անորոշության անհրաժեշտ գնահատականները:

Երրորդ գլուխը նվիրված է կողմնակի ինֆորմացիայով աղբյուրի արագություն–հուսալիություն–շեղում ֆունկցիայի ուսումնասիրությանը [2]: Դիտարկվել է վերջավոր այբուբեններով, առանց հիշողության դիսկրետ աղբյուր՝ $P^*(x,s)$ համատեղ բաշխմամբ: Կողմնակի ինֆորմացիայի **s** հաջորդականությունը հայտնի է և՛ կոդավորիչին, և՛ ապակոդավորիչին ու օգտագործվում է կոդավորման և վերականգնման ընթացքում:



Նկար 3: Առանց հիշողության դիսկրետ աղբյուրի մոդել՝
կոդավորիչին և ապակոդավորիչին հասանելի կողմնակի ինֆորմացիայով:

N երկարությամբ բլոկների համար կոդը սահմանվել է հետևյալ արտապատկերումներով՝

$$f_N:\mathcal{X}^N\times\mathcal{S}^N\to\{1,2,\dots,L(N)\},$$
$$g_N:\{1,2,\dots,L(N)\}\times\mathcal{S}^N\to\hat{\mathcal{X}}^N:$$

Այստեղ $L(N)$-ը կոդի ծավալն է: Նպատակն է որոշել աղբյուրի հաղորդագրությունների վերականգնման տրված շեղումն ու հուսալիությունն ապահովող կոդավորման փոքրագույն արագությունը:

$Q(P,\Delta)$-ով նշանակվել է $Q_P(\hat{x} \mid x,s)$ պայմանական բաշխումների բազմությունը, որոնց համար $P(x,s)Q_P(\hat{x} \mid x,s)$ համատեղ բաշխման նկատմամբ տեղի ունի միջին շեղման սահմանափակումը՝

$$\mathbb{E}_{P,Q_P} d(X,\hat{X}) \le \Delta:$$

Թեորեմ 3.1: *Տրված P^* բաշխման, յուրաքանչյուր $E > 0$ և $\Delta \ge 0$ արժեքների համար*՝

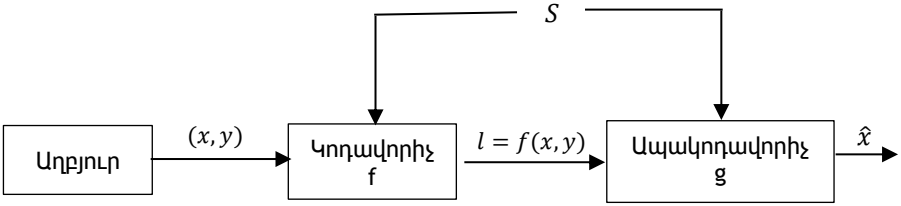
$$R(E,\Delta,P^*) = \max_{P \in \alpha(E,P^*)} \min_{Q_P \in Q(P,\Delta)} I_{P,Q_P}(X;\hat{X} \mid S):$$

Այս արտահայտությունում կողմնակի ինֆորմացիայի ազդեցությունը նկարագրվում է $I_{P,Q_P}(X;\hat{X} \mid S)$ պայմանական փոխադարձ ինֆորմացիայով, իսկ հուսալիության պահանջը՝ աղբյուրի P^* բաշխման շուրջ ինֆորմացիոն տարամիտությամբ որոշվող բաշխումների բազմությամբ:

Թեորեմի ապացույցում կիրառվել են կազմերի մեթոդը և ծածկման լեմմայի համապատասխան ձևափոխությունը: Յուրաքանչյուր ֆիքսված վիճակների հաջորդականության համար կառուցվել են պայմանական կազմային դասերը ծածկող կոդեր: Ուղիղ մասում ստացվել է արագության վերին գնահատականը, իսկ հակադարձ մասում՝ ստորինը: Գնահատականների համընկնումով որոշվել է կոդավորման նվազագույն անհրաժեշտ արագությունը:

Ստացված արդյունքը բնութագրում է արագության, հուսալիության և շեղման փոխառնչությունը երկու կողմերին հայտնի կողմնակի ինֆորմացիայի առկայությամբ :

Չորրորդ գլուխը նվիրված է գաղտնի բաղադրիչով և կողմնակի ինֆորմացիայով աղբյուրի կոդավորման խնդրին՝ արագության, հուսալիության, շեղման և անորոշության համատեղ սահմանափակումներով: Դիտարկվել է վերջավոր այբուբեններով, առանց հիշողության դիսկրետ աղբյուր՝ (X,Y,S) ելքերով և P^* համատեղ բաշխմամբ: Օգտակար X բաղադրիչը ենթակա է վերականգնման, Y-ը՝ պաշտպանության, իսկ կողմնակի ինֆորմացիայի S հաջորդականությունը հայտնի է և՛ կոդավորիչին, և՛ ապակոդավորիչին:



Նկար 4: Միակողմանի հաղորդում փոխկապակցված աղբյուրներով ևընդհանուր կողմնակի ինֆորմացիայով

N երկարությամբ բլոկների համար կոդը սահմանվել է արտապատկերումներով՝

$$f_N: \mathcal{X}^N \times \mathcal{Y}^N \times \mathcal{S}^N \to \{1,2,\dots,L(N)\},$$

$$g_N: \{1,2,\dots,L(N)\} \times \mathcal{S}^N \to \hat{\mathcal{X}}^N:$$

Հրապարակվող ինդեքսը նշանակվել է $L = f_N(\boldsymbol{x}, \mathbf{y}, \mathbf{s})$-ով, իսկ վերականգնումը՝ $\hat{\mathbf{x}} = g_N(\mathbf{y} \mid L(N), \mathbf{s})$-ով։ Գաղտնիության գնահատման մեջ հաշվի են առնվել և՛ հաղորդված ինդեքսը, և՛ ստացողին հայտնի կողմնակի ինֆորմացիան՝

$$R_e = \frac{1}{N} H(\mathbf{Y} \mid L(N), \mathbf{S}):$$

Տրված $E > 0$-ի դեպքում (R, Δ_d, Δ_e) եռյակը սահմանվել է որպես E-հասանելի, եթե կամայական $\varepsilon > 0$-ի և $\delta > 0$-ի համար բավականաչափ մեծ N-երի դեպքում գոյություն ունեն հետևյալ պահանջները բավարարող կոդեր՝

$$\frac{1}{N} \log L(N) \le R + \varepsilon,$$
$$e(f_N, g_N, P^*, \Delta_d) \le exp\{-N(E - \delta)\},$$
$$R_e \ge \Delta_e - \varepsilon:$$

Այստեղ Δ_d-ն թույլատրելի շեղման շեմն է, իսկ Δ_e-ն՝ պահպանվող անորոշության պահանջվող նվազագույն մակարդակը։ Ներմուծվել են $R^*(E)$ հասանելիության տիրույթը, արագություն–հուսալիություն–շեղում–անորոշություն $R^*(E, \Delta_d, \Delta_e)$ (RRDE) և անորոշություն–հուսալիություն–շեղում $\Gamma^*(E, \Delta_d)$ (ERD) ֆունկցիաները։

$\mathcal{Q}(P, \Delta_d, \Delta_e)$-ով նշանակվել է $Q_P(\hat{x} \mid x, y, s)$ պայմանական բաշխումների բազմությունը, որոնց համար $P(x, y, s)Q_P(\hat{x} \mid x, y, s)$ համատեղ բաշխման նկատմամբ՝

$$\mathbb{E}_{P,Q_P} d(X, \hat{X}) \le \Delta_d,$$
$$H_{P,Q_P}(Y \mid \hat{X}, S) \ge \Delta_e:$$

$\mathcal{Q}(P, \Delta_d)$-ը սահմանվել է հետևյալ միավորմամբ՝

$$\mathcal{Q}(P, \Delta_d) = \bigcup_{H_{P,Q_P}(Y|X,S) \le \Delta_e \le H_{P,Q_P}(Y|S)} \mathcal{Q}(P, \Delta_d, \Delta_e):$$

Թեորեմ 4.1: *Աղբյուրի տրված P^* բաշխման և յուրաքանչյուր $E > 0$ համար*՝

$$R^*(E) = \left\{ \begin{array}{c} (R, \Delta_d, \Delta_e): \Delta_d \ge 0, \Delta_e \ge 0, \\ 0 \le R_e \le \min\limits_{P \in \alpha(E,P^*)} \max\limits_{Q_P \in \mathcal{Q}(P,\Delta_d)} H_{P,Q_P}(Y, \mid \hat{X}, S), \\ R \ge \max\limits_{P \in \alpha(E,P^*)} \min\limits_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X, Y; \hat{X} \mid S) \end{array} \right\}:$$

Այս արդյունքը բնութագրում է կողմնակի ինֆորմացիայի առկայության դեպքում արագության, շեղման, հուսալիության և գաղտնիության հիմնական փոխառնչությունը։

Հետևանք 4.1: *Անորոշություն–հուսալիություն–շեղում ֆունկցիան հավասար է*՝

$$\Gamma^*(E, \Delta_d) = \min_{P \in \alpha(E,P^*)} \max_{Q_P \in \mathcal{Q}(P,\Delta_d)} H_{P,Q_P}(Y \mid \hat{X}, S):$$

Հետևանք 4.2: *Շեղում–անորոշություն* E*-հասանելիության տիրույթն ունի հետևյալ տեսքը*՝

$$\mathcal{R}^*_{\Delta_d,\Delta_e}(E) = \left\{ \begin{array}{c} R(E, \Delta_d, \Delta_e): \Delta_d \ge 0, \\ 0 \le \Delta_e \le \Gamma^*(E, \Delta_d) \end{array} \right\}:$$

Հետևանք 4.3: *Արագություն–հուսալիություն–շեղում–անորոշություն ֆունկցիան հավասար է*՝

$$R^*(E, \Delta_d, \Delta_e) = \max_{P \in \alpha(E,P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X, Y; \hat{X} \mid S):$$

Հետևանք 4.4: *Երբ E -ն ձգտում է զրոյի, արագություն–հուսալիություն–շեղում–անորոշություն և անորոշություն–հուսալիություն–շեղում ֆունկցիաների սահմանները համընկնում են Յամամոտոյի աշխատանքում ներկայացված համապատասխան ֆունկցիաների հետ՝*

$$\lim_{E\to 0} R^*(E,\Delta_d,\Delta_e) = R^*(\Delta_d,\Delta_e) = \min_{Q_P^*\in\mathcal{Q}(P^*,\Delta_d,\Delta_e)} I_{P^*,Q_P^*}(X,Y;\hat{X}\mid S),$$
$$\lim_{E\to 0} \Gamma^*(E,\Delta_d) = \Gamma^*(\Delta_d) = \max_{Q_{P^*}\in\mathcal{Q}(P^*,\Delta_d)} H_{P^*,Q_{P^*}}(Y\mid \hat{X},S):$$

Հետևանք 4.5: *Գաղտնի բաղադրիչի բացակայության դեպքում հասանելի տիրույթը համընկնում է* [2] *աշխատանքում ներկայացված տիրույթի հետ՝*

$$R(E,\Delta,P^*) = \max_{P\in\alpha(E,P^*)} \min_{Q_P\in\mathcal{Q}(P,\Delta)} I_{P,Q_P}(X;\hat{X}\mid S):$$

Հետևանք 4.6: *Ապակողավորիչին հայտնի կողմնակի ինֆորմացիայի բացակայության դեպքում արդյունքը հանգում է* [1] *աշխատանքում ներկայացված արդյունքին՝*

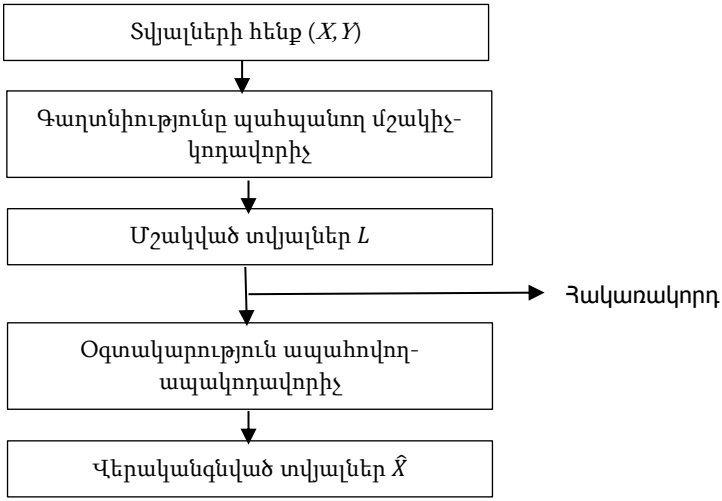
$$\mathcal{R}^*(E) = \left\{ \begin{array}{c} (R,\Delta_d,\Delta_e)\colon \Delta_d\geq 0,\ \Delta_e\geq 0, \\ 0\leq R_e\leq \min\limits_{P\in\alpha(E,P^*)} \max\limits_{Q_P\in\mathcal{Q}(P,\Delta_d)} H_{P,Q_P}(Y\mid\hat{X}), \\ R\geq \max\limits_{P\in\alpha(E,P^*)} \min\limits_{Q_P\in\mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X,Y;\hat{X}) \end{array} \right\}:$$

Այս հետևանքները բացահայտում են ընդհանուր մոդելի և նախորդ գլուխների արդյունքների կապը: Գաղտնի բաղադրիչի բացակայության դեպքում արդյունքը հանգում է երրորդ գլխում ուսումնասիրված՝ կողմնակի ինֆորմացիայով աղբյուրի արագություն–հուսալիություն–շեղում ֆունկցիային: Կողմնակի ինֆորմացիայի բացակայության դեպքում ստացվում է երկրորդ գլխի գաղտնի բաղադրիչով մոդելը:

Ապացույցը հիմնված է կազմերի մեթոդի և ծածկման լեմմայի համապատասխան ձևափոխության վրա: Յուրաքանչյուր վիճակների հաջորդականության համար կառուցվել են պայմանական կազմային դասերը ծածկող կոդեր: Ուղիղ մասում ստուգվել են շեղման, հուսալիության և անորոշության պահանջները և ստացվել է արագության վերին գնահատականը: Հակադարձ մասում ապացուցվել են անհրաժեշտ ստորին գնահատականն ու անորոշության վերին սահմանը:

Հինգերորդ գլուխը նվիրված է նախորդ գլուխների արդյունքների կիրառությանը տվյալների հենքերում գաղտնիություն–օգտակարություն փոխզիջման խնդրում: Տվյալների հենքը մոդելավորվել է որպես P^* համատեղ բաշխմամբ առանց հիշողության աղբյուր, որտեղ X-ը հրապարակային օգտակար բաղադրիչն է, իսկ Y-ը՝ գաղտնի բաղադրիչը: Հրապարակման մեխանիզմը ձևավորում է $L = f_N(\mathbf{x},\mathbf{y})$ ինդեքսը և $\hat{\mathbf{x}} = g_N(L)$ վերականգնումը:

Դիտարկվել են ընդհանուր մեխանիզմը, որն օգտագործում է երկու բաղադրիչները, և միայն հրապարակային բաղադրիչից կախված մեխանիզմը: Վերջին դեպքում ներմուծվում է $Y\to X\to\hat{X}$ Մարկովի շղթայի սահմանափակումը:



Նկար 5: Տվյալների համացանցի ճարտարապետություն տվյալների բազայի մաքրման շրջանակի համար

Իրերի ինտերնետի (IoT) համակարգում $\mathbf{X} = (X_1, \dots, X_N)$ կարող է ներկայացնել տվիչների օգտակար չափումները կամ բնութագրերը, որոնք նախատեսվում է հրապարակել տվյալների վերլուծության համար, մինչդեռ $\mathbf{Y} = (Y_1, \dots, Y_N)$ ներկայացնում է այդ չափումների հետ փոխկապակցված զգայուն փոփոխականները: Այս մոդելի իրական կիրառությունների բնորոշ օրինակներն են՝

- **Խելացի էլեկտրացանց:** Հրապարակվող **X** տվյալները ներառում են էներգասպառման չափումները կամ բեռնվածության բնութագրերը, իսկ **Y** գաղտնի փոփոխականը համապատասխանում է տանը մարդկանց ներկայությանը, առանձին սարքերի օգտագործմանը կամ օգտատիրոջ կենսակերպի առանձնահատկություններին:
- **Կրելի սարքերով առողջության մոնիթորինգ:** Օգտակար **X** տվյալները ներառում են ֆիզիկական ակտիվության գրանցման տվյալները կամ ֆիզիոլոգիական ցուցանիշների ամփոփ տեղեկությունները, իսկ **Y**-ը ներկայացնում է զգայուն տեղեկություններ, օրինակ՝ բժշկական ախտորոշումները, օգտատիրոջ ինքնությունը կամ առողջական վիճակի վերաբերյալ գաղտնի տվյալները:
- **Արդյունաբերական իրերի ինտերնետ (IIoT):** **X** բնութագրերը ներառում են մեքենասարքավորումների հեռաչափական տվյալները, իսկ **Y**-ը ներկայացնում է գաղտնի շահագործման ռեժիմները կամ ընկերության արտադրական ժամանակացույցերը:
- **Խելացի տուն:** Դիտարկվող **X** տվյալները ներառում են շրջակա միջավայրի տվիչների գրանցումները (օրինակ՝ ջերմաստիճանը, խոնավությունը), իսկ **Y**-ը ներկայացնում է բնակիչների ներկայությունը, վարքագծային բնութագրերը կամ առօրյա ռեժիմը:

Հրապարակվող L ինդեքսը կարող է ծայրամասային սարքից կամ ցանցային

անցուղուց փոխանցվել տվյալների վերլուծության ամպային ծառայությանը: Վերականգնված $\hat{\mathbf{X}}$-ն օգտակար է, եթե դրա հիման վրա հնարավոր է կատարել եզրահանգումներ, կանխատեսում կամ կառավարում՝ Δ_d-ը չգերազանցող շեղմամբ:

Գաղտնիության պահանջը

$$\frac{1}{N}H(\mathbf{Y}|L) \geq \Delta_e$$

ապահովում է, որ հրապարակված ինդեքսը չափազանց շատ տեղեկություն չբացահայտի իրերի ինտերնետի համակարգի զգայուն վիճակների մասին:

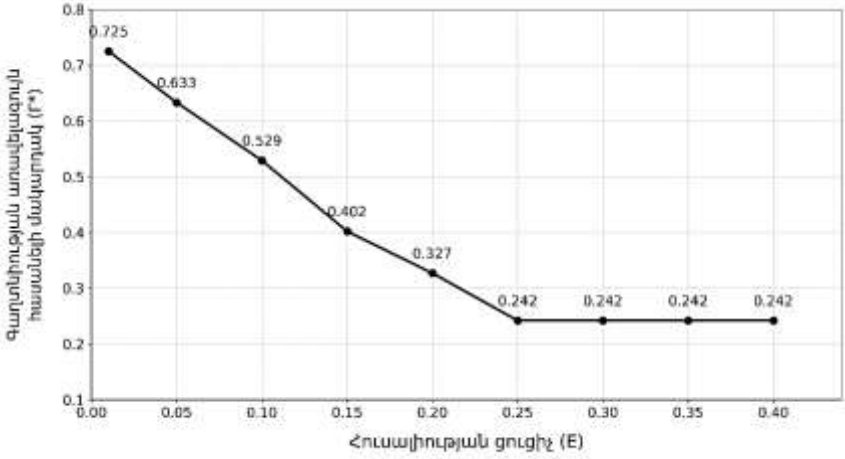
Հուսալիության պահանջը

$$\Pr\{d(\mathbf{x}, \hat{\mathbf{x}}) > \Delta_d\} \leq exp\{-N(E-\delta)\}$$

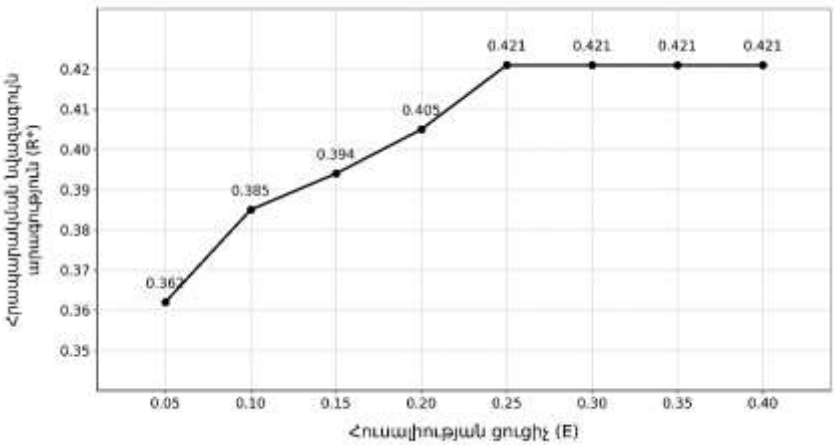
ապահովում է, որ տվիչների չափումների քանակի աճին զուգընթաց օգտակարության անընդունելի խախտումները դառնան ցուցչայնորեն հազվադեպ: Հուսալիության երաշխիքների շնորհիվ դիտարկման ժամանակահատվածի երկարացմանը զուգընթաց վերականգնման լուրջ ձախողումները դառնում են ցուցչայնորեն քիչ հավանական: Այսպիսով, հուսալիության սահմանափակմամբ ձևակերպումը հարմար է իրերի ինտերնետի այն կիրառությունների համար, որոնցում տվյալները պետք է միաժամանակ սեղմվեն և հրապարակվեն՝ պահպանելով օգտակարությունը, գաղտնիությունն ու հուսալիությունը:

Թվային օրինակը դիտարկվել է խելացի էլեկտրացանցի համար. $X, Y \in \{0,1\}$ փոփոխականները ներկայացնում են համապատասխանաբար էներգիայի սպառումը և տնային տնտեսությունում մարդկանց ներկայությունը: Աղբյուրի անվանական բաշխումն ընտրվել է՝

$$P^*(0,0) = 0.20, \quad P^*(0,1) = 0.05,$$
$$P^*(1,0) = 0.05, \quad P^*(1,1) = 0.70:$$



Նկար 6: Գաղտնիության առավելագույն հասանելի մակարդակի կախվածությունը հուսալիության ցուցչից՝ $\Delta_d = 0.15$ դեպքում:



Նկար 7: Հրապարակման նվազագույն արագության կախվածությունը հուսալիության ցուցչից՝ $\Delta_d = 0.15$ և $\Delta_e = 0.2$ դեպքում:

Ընտրվել են Հեմինգի շեղման չափը՝ $\Delta_d = 0.15$, և հրապարակման արագության հաշվարկի համար գաղտնիության $\Delta_e = 0.2$ շեմը: Հաշվարկներն իրականացվել են R միջավայրում՝ աղբյուրի և վերականգնման պայմանական բաշխումների նկատմամբ ցանցային որոնմամբ: Ստացված կախվածությունները ներկայացված են Նկար 6-ում և Նկար 7-ում:

Հուսալիության պահանջի խստացմամբ գաղտնիության առավելագույն մակարդակը նվազում է, իսկ հրապարակման նվազագույն արագությունը՝ աճում: Վերջինս 0.362-ից հասնում է 0.421 բիթ/նմուշի՝ աճելով մոտ 16.3%-ով: Դիտարկված օրինակում $E \approx 0.25$-ից հետո երկու կորերը գրանցում են հագեցում, իսկ գաղտնիության մակարդակը մնում է $\Gamma^* \approx 0.242$:

Ստացված տեսական սահմանները հիմք են տալիս գնահատելու տվյալների հրապարակման մեխանիզմների արդյունավետությունը: Դրանց թվային հաշվարկը հնարավորություն է տալիս հիմնավորելու համակարգի պարամետրերի ընտրությունը՝ հաշվի առնելով հուսալիության պահանջների հետ կապված հաղորդակցման ծախսերն ու գաղտնիության սահմանափակումները:

Աշխատանքի հիմնական արդյունքները

Աշխատանքում ստացվել են հետևյալ հիմնական արդյունքները:

1. Գաղտնի բաղադրիչով կոռելացված ելքերով դիսկրետ առանց հիշողության միակողմանի աղբյուրի համար սահմանվել և բնութագրվել են արագություն–հուսալիություն–շեղում–անորոշություն և անորոշություն–հուսալիություն–շեղում ֆունկցիաների արտահայտությունները: Դիտարկվել է հուսալիության ցուցչի՝ զրոյի ձգտման սահմանային դեպքը և հաստատվել է համապատասխան սահմանների համընկնումը Յամամոտոյի արդյունքների հետ: [1]

2. Կոդավորիչին և ապակոդավորիչին հայտնի կողմնակի ինֆորմացիայով աղբյուրի համար որոշվել է արագություն–հուսալիություն–շեղում ֆունկցիան: [2]
3. Փոխկապակցված ելքերով միակողմանի աղբյուրի համար, որտեղ կողմնակի ինֆորմացիան հասանելի է կոդավորիչին և ապակոդավորիչին, ներմուծվել և հետազոտվել են արագություն-հուսալիություն-շեղում-անորոշություն ֆունկցիան, ինչպես նաև համապատասխան անորոշություն-հուսալիություն-շեղում և արագություն-հուսալիություն-շեղում ֆունկցիաները: [3]
4. Վերջավոր այբուբենով առանց հիշողության տվյալների հենքերի հրապարակման խնդրում ներկայացվել է օգտակարության և գաղտնիության դասական ձևակերպման ընդլայնումը հուսալիության հաշվառմամբ՝ հիմնված աղբյուրի կոդավորման նախորդ արդյունքների վրա: Ստացված շրջանակը ինֆորմացիայի տեսության տեսանկյունից հիմք է ապահովում **տվյալների հենքերի մշակման խնդիրներում կապի արագության, օգտակարության, գաղտնիության և հուսալիության միջև փոխադրեցության** վերլուծության համար: [4]

Ստացված արդյունքների կիրառության փորձարկման և բնութագրիչների կախվածության վարքը պատկերելու համար դիտարկվել է խելացի էլեկտրացանցի **(smart grid)** երկուական մոդելի թվային օրինակը: Գնահատվել են գաղտնիության առավելագույն հասանելի մակարդակը և տվյալների հրապարակման փոքրագույն արագությունը՝ որպես հուսալիության ցուցչի ֆունկցիաներ: Ստացված թվային արդյունքները ցույց են տալիս **հուսալիության պահանջների ազդեցությունը** ինչպես հասանելի գաղտնիության մակարդակի, այնպես էլ տվյալների բացահայտման նվազագույն արագության վրա:

Հրապարակված աշխատանքների ցանկ

1. Haroutunian M. E., Santrosyan J. S., Hakobyan P. M., “Reliability criteria in source coding problem with secret component”, Mathematical Problems of Computer Science, vol. 63, pp. 14–24, 2025, doi: 10.51408/1963-0128.

2. Haroutunian M., Hakobyan P., Santrosyan J., “Rate-Reliability-Distortion function for source with side information known to encoder and decoder”, Proceedings of International Conference on Computer Science and Information Technologies (CSIT), Yerevan, Armenia, 2025, pp. 125–128. https://doi.org/10.51408/csit2025_38

3. Santrosyan J. S., “Rate-Reliability-Distortion-Equivocation Tradeoffs for Source with Secret Component and Side Information”, Mathematical Problems of Computer Science, vol. 65, pp. 16–29, 2026, doi: 10.51408/1963-0143.

4. Hakobyan P., Santrosyan J., Haroutunian M., “Utility–Privacy Tradeoff with Reliability Criteria in Databases: An Information-Theoretic Formulation with IoT Applications”, CODASSCA 2026, Yerevan, Armenia, 2026.
DOI 10.1109/CODASSCA69992.2026.00052

ИНФОРМАЦИОННО-ТЕОРЕТИЧЕСКИЙ ПОДХОД К ПРОБЛЕМЕ КОМПРОМИССА МЕЖДУ КОНФИДЕНЦИАЛЬНОСТЬЮ И ПОЛЕЗНОСТЬЮ В БАЗАХ ДАННЫХ

Абстракт

Современные информационные системы всё в большей степени основываются на сборе, обработке и публикации крупномасштабных баз данных. Эти базы данных представляют значительную ценность, поскольку обеспечивают возможности для анализа, прогнозирования, управления, персонализации и принятия решений. Вместе с тем публикация полезных данных может привести к утечке конфиденциальной информации вследствие статистической зависимости между общедоступными и конфиденциальными признаками.

Анализ собираемых данных требует их публикации в незашифрованном виде, что может привести к утечке конфиденциальной информации, то есть к нарушению неприкосновенности частной жизни. В отличие от задачи обеспечения секретности (secrecy problem), в задаче защиты конфиденциальности персональных данных (privacy problem) раскрытие данных позволяет получить информационную полезность, одновременно создавая потенциальный риск нарушения конфиденциальности. Вследствие статистических зависимостей между общедоступными и конфиденциальными данными передача полезных данных может раскрывать также сведения, относящиеся к конфиденциальной компоненте.

В литературе предложен ряд подходов к решению данной проблемы, однако они применимы не во всех случаях и не охватывают всех возникающих задач. Совместное обеспечение полезности и конфиденциальности информации в базах данных связано с требованиями к точности восстановления публикуемых данных, скорости, необходимой для их передачи, и сохранению конфиденциальной информации.

Цель диссертации состоит в исследовании проблемы компромисса между конфиденциальностью и полезностью информации в базах данных посредством введения компоненты надёжности и в решении этой проблемы с использованием инструментов теории информации.

В работе получены следующие основные результаты

1. Для дискретного одностороннего источника без памяти с коррелированными выходами и секретной компонентой даны определения и получены выражения для функций «скорость–надёжность–искажение–неопределённость» и «неопределённость–надёжность–искажение». Рассмотрен предельный случай, когда показатель надёжности стремится к нулю, и установлено совпадение соответствующих пределов с результатами Ямамото. [1]
2. Для источника с побочной информацией, известной кодеру и декодеру, определена функция «скорость–надёжность–искажение». [2]
3. Для одностороннего источника с взаимосвязанными выходами, в котором побочная информация доступна кодеру и декодеру, введены и исследованы функция «скорость–надёжность–искажение–неопределённость», а также соответствующие функции «неопределённость–надёжность–искажение» и «скорость–надёжность–искажение». [3]

4. В задаче публикации баз данных, описываемых моделью источника без памяти с конечным алфавитом, представлено расширение классической постановки задачи обеспечения полезности и конфиденциальности с учётом надёжности, основанное на полученных ранее результатах по кодированию источников. Предложенный подход обеспечивает теоретико-информационную основу для анализа взаимодействия между скоростью передачи данных, полезностью, конфиденциальностью и надёжностью в задачах обработки баз данных. [4]

Для апробации полученных результатов и иллюстрации поведения исследуемых характеристик рассмотрен численный пример для бинарной модели интеллектуальной электросети **(smart grid)**. Оценены максимально достижимый уровень конфиденциальности и минимальная скорость публикации данных как функции показателя надёжности. Полученные численные результаты показывают **влияние требований к надёжности** как на достижимый уровень конфиденциальности, так и на минимальную скорость раскрытия данных.

AN INFORMATION-THEORETIC APPROACH TO THE PRIVACY–UTILITY TRADEOFF PROBLEM IN DATABASES

Abstract

Modern information systems increasingly rely on the collection, processing, and release of large databases. These databases provide substantial value by supporting analysis, prediction, control, personalization, and decision-making. At the same time, releasing useful data may lead to the leakage of sensitive information due to statistical dependencies between public and private attributes.

Analyzing the collected data requires their release in unencrypted form, which may lead to the leakage of sensitive information and thus to a violation of privacy. Unlike the secrecy problem, the personal data privacy problem involves disclosing data to obtain utility while simultaneously creating a potential risk of privacy violation. Due to statistical dependencies between public and private data, transmitting useful data may also reveal information about the private component.

Various approaches to this problem have been proposed in the literature; however, they are not applicable in all settings and do not address all the issues that arise. Jointly ensuring utility and privacy in databases involves requirements concerning the reconstruction accuracy of the released data, the rate required for their transmission, and the protection of private information.

The aim of this dissertation is to investigate the utility–privacy tradeoff in databases by incorporating a reliability component and to solve this problem using information-theoretic tools.

The following main results have been obtained in this work:

1. For a single-terminal discrete memoryless source with correlated outputs and a private component, the rate–reliability–distortion–equivocation and equivocation–reliability–distortion functions have been defined, and expressions for them have been derived. The limiting case in which the reliability exponent tends to zero has been considered, and the corresponding limits have been shown to coincide with Yamamoto's results. [1]
2. For a source with side information known to both the encoder and the decoder, the rate–reliability–distortion function has been determined. [2]

3. For a single-terminal source with correlated outputs and side information available to both the encoder and the decoder, the rate–reliability–distortion–equivocation function, as well as the corresponding equivocation–reliability–distortion and rate–reliability–distortion functions, have been introduced and studied. [3]
4. For the problem of releasing databases modeled as a memoryless source over a finite alphabet, the classical utility–privacy formulation has been extended to incorporate reliability, building on the preceding source coding results. The proposed framework provides an information-theoretic basis for analyzing **the interaction among communication rate, utility, privacy, and reliability in database processing problems**. [4]

To assess the applicability of the results and illustrate the behavior of the quantities under study, a numerical example involving a binary **smart grid** model has been considered. The maximum achievable privacy level and the minimum data release rate have been evaluated as functions of the reliability exponent. The numerical results show **the impact of reliability requirements** on both the achievable privacy level and the minimum data disclosure rate.

