

IJTIMOY MUHANDISLIK HUJUMLARINI ANIQLASH VA ULARGA QARSHI KURASHISHNING INNOVATSION INTELLEKTUAL USULLARI VA ALGORITMLARINI ISHLAB CHIQISH

Eshmurodov Alisher Akrom o'g'li

Muhammad al-Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti

Annotatsiya

Maqolada elektron pochta, SMS, messenjerlar va ovozli qo'ng'iroqlar orqali amalga oshiriladigan ijtimoiy muhandislik hujumlarini aniqlashning besh bosqichli dialog modeli taklif etilgan. Mavjud vositalar xabarlarini alohida tahlil qilgani uchun ko'p bosqichli suhbat shaklidagi hujumlarni kech aniqlashi ko'rsatilgan. Model ko'p tilli transformer asosidagi semantik belgilar, psixologik ta'sir tamoyillari va kontekst belgilarini GRU qatlamida birlashtirib, har bir yangi navbatdan keyin 0-100 oraliqdagi xavf ko'rsatkichini yangilaydi va foydalanuvchini zararli harakatni amalga oshirishidan oldin ogohlantiradi. Xavf ko'rsatkichiga ko'ra suhbat o'tkaziladi, foydalanuvchi ogohlantiriladi yoki suhbat bloklanadi. Bank va mobil operatorlarning maxfiy ma'lumotlarini markazlashtirmasdan o'qitish uchun federativ o'qitish, tushunarli ogohlantirish uchun SHAP usuli qo'llanilgan. Tashkiliy va texnik qarshi choralar, erta aniqlash samaradorligini baholash hamda bank va operatorlarda pilot sinovlar o'tkazish bo'yicha takliflar berilgan.

Kalit so'zlar: ijtimoiy muhandislik, dialog darajasida aniqlash, erta ogohlantirish, ko'p tilli transformer modellar, psixologik ta'sir tamoyillari, federativ o'qitish, tushuntiriladigan sun'iy intellekt, o'zbek tili.

Аннотация

В статье предложена пятиэтапная диалоговая модель обнаружения атак социальной инженерии, осуществляемых через электронную почту, SMS, мессенджеры и голосовые звонки. Показано, что существующие средства анализируют сообщения по отдельности и поэтому поздно обнаруживают атаки, развивающиеся в форме многоэтапного диалога. Модель объединяет в слое GRU семантические признаки на основе многоязычного трансформера, принципы психологического влияния и контекстные признаки, после каждой новой реплики обновляет показатель риска в диапазоне 0-100 и предупреждает пользователя до совершения вредоносного действия. В зависимости от показателя риска диалог пропускается, пользователь получает предупреждение или диалог блокируется. Для обучения без централизации конфиденциальных данных банков и мобильных операторов применено федеративное обучение, для понятных предупреждений - метод SHAP. Даны предложения по организационным и техническим мерам противодействия, оценке эффективности раннего обнаружения и пилотным испытаниям в банках и у операторов.

Ключевые слова: социальная инженерия, обнаружение на уровне диалога, раннее предупреждение, многоязычные трансформеры, принципы убеждения, федеративное обучение, объяснимый искусственный интеллект, узбекский язык.

Abstract

The article proposes a five-stage dialogue-level model for detecting social engineering attacks carried out through e-mail, SMS, messengers and voice calls. It was shown that existing tools analyse messages separately and therefore detect attacks that unfold as a multi-stage conversation too late. The model combines semantic features from a multilingual transformer, psychological persuasion principles and contextual features in a GRU layer, updates a risk score on a 0-100 scale after each new turn and warns the user before a harmful action occurs. Depending on the risk score, the conversation is passed, the user is warned or the conversation is blocked. Federated learning was used to train the model without centralising the confidential data of banks and mobile operators, and the SHAP method was used to produce understandable alerts. Proposals were made on organisational and technical countermeasures, evaluation of early detection performance and pilot testing at banks and operators.

Keywords: Social Engineering, Dialogue-level Detection, Early Warning, Multilingual Transformers, Persuasion Principles, Federated Learning, Explainable AI, Uzbek Language.

KIRISH

Ijtimoiy muhandislik - inson ishonchi va qarorlariga ta'sir ko'rsatish orqali maxfiy ma'lumotlarni olish, zararli havolalarni ochish yoki pul mablag'larini o'tkazishga undaydigan hujumlar majmuasidir. Bunday hujumlarni faqat texnik vositalar yordamida to'liq bartaraf etish qiyin [4; 9].

O'zbekistonda kiberjinoyatlar soni besh yilda 11 barobar oshib, 62 mingdan ortgan [13]. 2025-yilning 11 oyida fuqarolar qariyb 1,9 trln so'm zarar ko'rgan [14]. 2025-yilda kiberjinoyatlar barcha jinoyatlarning 49,6 foizini tashkil etgan, kiberog'irlik va kiberfiribgarlik holatlarining 96 foizi elektron to'lov tizimlari orqali sodir etilgan. Firibgarlik usullari orasida sotuvchi yoki xaridor nomidan aldash (18%) va firibgar qo'ng'iroqlar (7%) qayd etilgan [15]. Ushbu yo'nalishdagi tadqiqotlar "Kiberxavfsizlik to'g'risida"gi Qonun [11] hamda 2030-yilgacha sun'iy intellekt texnologiyalarini rivojlantirish strategiyasiga [12] asoslanadi.

Mavjud aniqlash vositalari xabarlarini alohida tahlil qiladi, biroq ijtimoiy muhandislik hujumlari ko'pincha bir necha bosqichli suhbat shaklida rivojlanadi. Dastlabki xabarlar zararsiz ko'rinib, xavfli so'rov suhbat oxirida paydo bo'lishi mumkin. Shu sababli, xabar darajasidagi tizimlar hujumni kech aniqlashi ehtimoli mavjud.

Ushbu ishda suhbat davomida dalillarni to'plovchi, maqsadli harakatdan oldin ogohlantiruvchi, turli aloqa kanallari hamda o'zbek va rus tillarida ishlaydigan, maxfiy ma'lumotlarni markazlashtirmasdan o'qitiladigan intellektual algoritmnini ishlab chiqish maqsad qilingan.

ASOSIY QISM

Ijtimoiy muhandislik hujumlarining hayotiy sikli modellari [7] va R. Chaldinining ta'sir tamoyillari [2] asosida hujum dialogi besh bosqich bilan tavsiflandi. Bosqichlar

odatda ketma-ket keladi, biroq hujumchi ularni o'tkazib yuborishi yoki orqaga qaytishi mumkin, shu sababli algoritm bosqichni qat'iy belgilamaydi, balki har bir navbatda bosqichlar bo'yicha ehtimollik taqsimotini baholaydi. "Maqsadli harakat" bosqichida ilova (APK) o'rnatish alohida ajratilgan, chunki zararli dasturlar O'zbekistondagi kiberjinoyat usullarining eng katta ulushini (38 foiz) [15] tashkil etadi.

Taklif etilayotgan algoritm suhbatni $D = (m_1, \dots, m_T)$ navbatlar ketma-ketligi sifatida qabul qiladi va har bir yangi navbatdan keyin xavf ko'rsatkichini qayta hisoblaydi (1-rasm):

1) oldindan ishlov berish - matnni normallashtirish (lotin/kirill transliteratsiyasi), tilni aniqlash, havola, telefon raqami va akkaunt yoshi kabi metama'lumotlarni ajratish; ovozli qo'ng'iroqlar nutqni matnga aylantirish (ASR) modeli orqali matnga o'tkaziladi [8];

2) belgilarni shakllantirish - har bir navbat uchun uch guruh belgi hisoblanadi: ko'p tilli transformer (XLM-R) orqali semantik vektor e_t [3]; Chaldini tamoyillari bo'yicha ko'p belgili (multi-label) klassifikator chiqaradigan olti o'lchamli psixologik vektor $p_t \in [0, 1]^6$ (obro', shoshiltirish, qo'rqitish, manfaat, yaqinlik, maxfiylik) [2]; kanal, yuboruvchi yangiligi, so'rov turi va navbatlar orasidagi vaqtni o'z ichiga olgan kontekst vektori c_t ;

3) suhbat holatini kuzatish - takroriy GRU qatlami [1] navbatlar bo'yicha yig'ilgan dalillarni yashirin holat h_t da saqlaydi (1), uning asosida esa besh bosqich ehtimollari q_t baholanadi (2);

4) xavfni baholash - navbat t dagi xavf ko'rsatkichi 0-100 oraliqda hisoblanadi (3);

5) qaror qabul qilish - $R_t < 40$ bo'lsa suhbat o'tkaziladi, $40 \leq R_t < 70$ bo'lsa foydalanuvchiga tushuntirish bilan real vaqtda ogohlantirish beriladi, $R_t \geq 70$ bo'lsa suhbat bloklanadi va bank yoki SOC xizmatiga uzatiladi; chegaralar validatsiya to'plamida soxta signallar ulushiga qarab sozlanadi.

Suhbat holati, bosqich ehtimollari va xavf ko'rsatkichi quyidagi formulalar bilan hisoblanadi:

$$h_t = \text{GRU}([e_t \parallel p_t \parallel c_t], h_{t-1}) \quad (1)$$

bu yerda h_t - t -navbatdagi suhbatning yashirin holati; e_t , p_t , c_t - semantik, psixologik va kontekst vektorlari; $\parallel$ - vektorlarni ketma-ket birlashtirish (konkatenatsiya) amali; h_{t-1} - oldingi navbatdagi yashirin holat. Formula har bir yangi xabar suhbatning oldingi dalillari bilan birga baholanishini ifodalaydi.

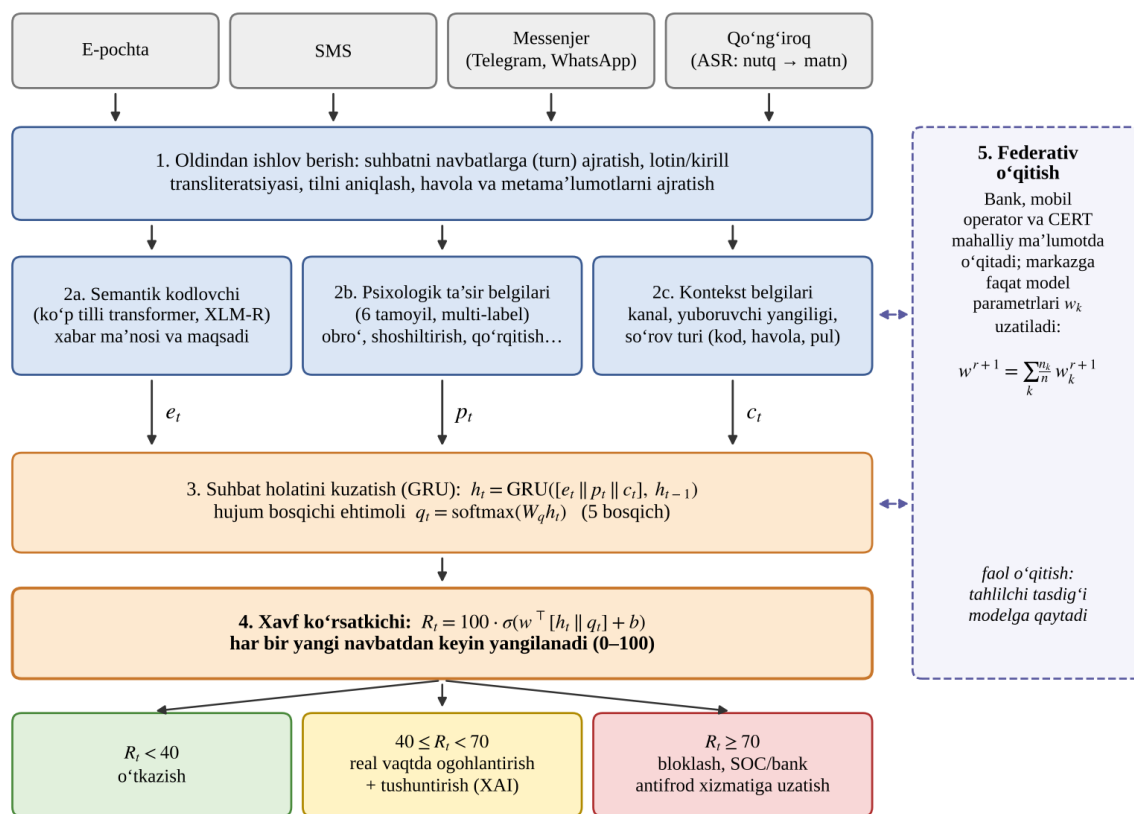
$$q_t = \text{softmax}(W_q h_t) \quad (2)$$

bu yerda q_t - besh bosqich bo'yicha ehtimollar vektori; W_q - o'qitiladigan vaznlar matritsasi; softmax - qiymatlarni yig'indisi birga teng ehtimolliklarga aylantiruvchi funksiya. Shu sababli algoritm bosqichni qat'iy belgilamaydi, balki uni ehtimollik taqsimoti orqali baholaydi.

$$R_t = 100 \cdot \sigma(w^T [h_t \parallel q_t] + b) \quad (3)$$

bu yerda R_t - t -navbatdagi xavf ko'rsatkichi (0-100); σ - sigmoid funksiya; w - vaznlar vektori, T - transponirlash belgisi; b - siljish koeffitsiyenti. Ko'rsatkich 40 va

70 chegaralari bilan solishtirilib, suhbatni o'tkazish, foydalanuvchini ogohlantirish yoki suhbatni bloklash to'g'risida qaror qabul qilinadi.



1-rasm. Ijtimoiy muhandislik hujumlarini dialog darajasida aniqlash algoritmining sxemasi¹

1-rasmda turli kanallardan kelgan suhbat navbatlariga oldindan ishlov berilishi, semantik, psixologik va kontekst belgilari GRU qatlamida birlashtirilishi, xavf ko'rsatkichi R_t qiymatiga ko'ra uch xil qaror qabul qilinishi hamda model federativ va faol o'qitish orqali yangilanib borishi ko'rsatilgan.

Haqiqiy firibgarlik suhbatlari bank va mobil operatorlarda saqlanadi, ammo ular shaxsga doir ma'lumot hisoblangani uchun bir joyga yig'ilishi mumkin emas [10]. Shu sababli model federativ tarzda o'qitiladi: har bir ishtirokchi k modelni o'z ma'lumotlarida o'qitadi, markazga esa faqat model parametrlari uzatiladi va ular og'irlashtirilgan o'rtacha usulida (FedAvg) [6] birlashtiriladi (4), bunda n_k - ishtirokchi k dagi namunalar soni, $n = \sum n_k$.

$$w^{r+1} = \sum_k \frac{n_k}{n} \cdot w_k^{r+1} \quad (4)$$

bu yerda w^{r+1} - $(r + 1)$ -raunddagi umumiy model parametrlari; w_k^{r+1} - k -ishtirokchi o'z ma'lumotlarida yangilagan parametrlar; n_k/n - k -ishtirokchi namunalarining umumiy namunalar sonidagi ulushi. Ko'proq ma'lumotga ega ishtirokchi umumiy modelga kattaroq ta'sir ko'rsatadi, xom ma'lumotlar esa markazga uzatilmaydi.

¹ Manba: muallif ishlanmasi.

Ogohlantirish matni foydalanuvchi uchun tushunarli bo'lishi shart: aks holda u signalga e'tibor bermaydi. Shuning uchun xavfni oshirgan navbatlar va ta'sir usullari SHAP kabi usullar yordamida ajratib ko'rsatiladi [5] (masalan: "Sizga bank xodimi nomidan murojaat qilinmoqda, SMS-kod so'ralmoqda va shoshiltirilmoqda"). Tahlilchilar tasdiqlagan yoki rad etgan holatlar faol o'qitish orqali modelni muntazam yangilab turadi.

Algoritmning ilmiy yangiligi quyidagilardan iborat: aniqlash xabar darajasidan dialog darajasiga ko'tarilib, hujum bosqichlari ehtimoli bilan ifodalanadi; hujum jabrlanuvchi harakatidan oldin aniqlanadi; psixologik belgilar kanalga bog'liq emas, shuning uchun havolasiz va texnik jihatdan "toza" hujumlarni ham qamrab oladi; model bir nechta tashkilotning maxfiy ma'lumotlarini ochmasdan o'qitiladi.

Ijtimoiy muhandislik hujumlariga qarshi kurashda sun'iy intellekt asosidagi aniqlash algoritmi tashkiliy va texnik choralar bilan birgalikda qo'llanishi lozim. Bunda FIDO2/WebAuthn va passkey kabi xavfsiz autentifikatsiya usullarini joriy etish, xavfli vaziyatlarda foydalanuvchini ogohlantirish, shubhali moliyaviy operatsiyalarni qo'shimcha tasdiqlash, banklar va operatorlar o'rtasida tahdidlar haqidagi ma'lumotlarni almashish hamda raqamli savodxonlikni oshirish muhimdir.

Sun'iy intellekt asosidagi tizimlar ham yangi hujumlarga duch kelishi mumkin. Shu sababli ularni raqib dialoglari yordamida sinash, muntazam yangilash va xavfsizligini baholab borish zarur.

XULOSA VA TAKLIFLAR

Tadqiqot natijalariga ko'ra, ijtimoiy muhandislik hujumlarini faqat alohida xabarlar orqali aniqlash yetarli emas. Suhbatning semantik, psixologik va kontekstual belgilarini tahlil qiluvchi algoritmi hujumni maqsadli harakatdan oldin aniqlashga yordam beradi. Federativ o'qitish esa bank va mobil operatorlarning maxfiy ma'lumotlarini markazlashtirmasdan modelni takomillashtirish imkonini beradi.

Tizim samaradorligini oshirish uchun quyidagilar taklif etiladi:

- 1) o'zbek va rus tillaridagi anonimlashtirilgan ma'lumotlar to'plamini yaratish;
- 2) erta aniqlash va soxta ogohlantirish ko'rsatkichlarini baholash;
- 3) bank va operatorlarda pilot sinovlarni o'tkazish;
- 4) real vaqt ogohlantirish mexanizmlarini joriy etish.

Ijtimoiy muhandislik hujumlariga qarshi samarali himoya dialogni tahlil qiluvchi intellektual algoritmlar va foydalanuvchilarning raqamli savodxonligini uyg'unlashtirish orqali ta'minlanadi.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1724-1734). <https://doi.org/10.3115/v1/D14-1179>

2. Cialdini, R. B. (2021). Influence, new and expanded: The psychology of persuasion. New York: Harper Business.

3. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 8440-8451). <https://doi.org/10.18653/v1/2020.acl-main.747>
4. Federal Bureau of Investigation, Internet Crime Complaint Center (IC3). (2026). Internet Crime Report 2025. Washington, D.C.: FBI. https://www.ic3.gov/AnnualReport/Reports/2025_IC3Report.pdf
5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
6. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of Machine Learning Research, 54, 1273-1282. <https://proceedings.mlr.press/v54/mcmahan17a.html>
7. Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. Computers & Security, 59, 186-209. <https://doi.org/10.1016/j.cose.2016.03.004>
8. Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. Proceedings of Machine Learning Research, 202, 28492-28518. <https://proceedings.mlr.press/v202/radford23a.html>
9. Verizon. (2025). 2025 Data Breach Investigations Report (DBIR). Verizon Business. <https://www.verizon.com/business/resources/reports/dbir/>
10. O'zbekiston Respublikasining 2019-yil 2-iyul O'RQ-547-son "Shaxsga doir ma'lumotlar to'g'risida"gi Qonuni. <https://lex.uz/docs/4396419>
11. O'zbekiston Respublikasining 2022-yil 15-aprel O'RQ-764-son "Kiberxavfsizlik to'g'risida"gi Qonuni. <https://lex.uz/docs/5960604>
12. O'zbekiston Respublikasi Prezidentining 2024-yil 14-oktyabr PQ-358-son "Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasini tasdiqlash to'g'risida"gi Qarori. <https://lex.uz/docs/-7158604>
13. Ichki ishlar vazirligi ma'lumotlari: kiberjinoyatlar soni va zarar hajmi // Spot.uz, 27.01.2026. <https://www.spot.uz/oz/2026/01/27/cybercrime/>
14. Joriy yilda kiberjinoyatchilar o'zbekistonliklarning 1,9 trln so'm mablag'larini o'g'irladi // Spot.uz, 23.12.2025. <https://www.spot.uz/oz/2025/12/23/cyber-crime/>
15. O'zbekistonda 10 mingdan ortiq shubhali kartalar bloklandi, 1,72 trln so'm saqlab qolindi // Spot.uz, 16.06.2026. <https://www.spot.uz/oz/2026/06/16/cybersecurity/>



Marketing

ilmiy, amaliy va ommabop jurnali

Muharrir:
Ingliz tili muharriri:
Rus tili muharriri:
Musahhih:
Sahifalovchi va dizaynerlar:

Turobova Hulkar Rustamovna
Oripov Mahmud Ashurovich
Xidirova Gavxar Rustamovna
Qudratov Aziz Dehqonboyevich
Sadikov Shoxrux Shuxratovich
Abidjonov Nodirbek Odijon o'g'li

2026-YIL, 24-25-SENTABR, MAXSUS-SON, III-IV-SEKSIYALAR

© Materiallar ko'chirib bosilganda "Marketing" ilmiy, amaliy va ommabop jurnali manba sifatida ko'rsatilishi shart. Jurnalda bosilgan material va reklamalardagi dalillarning aniqligiga mualliflar mas'ul. Tahririyat fikri har vaqt ham mualliflar fikriga mos kelavermasligi mumkin. Tahririyatga yuborilgan materiallar qaytarilmaydi.

Mazkur jurnalda maqolalar chop etish uchun quyidagi havolalarga murojaat qilish mumkin. Ilmiy maqola, ommabop maqola, reklama, hikoya va boshqa ilmiy-ijodiy materiallar yuborishingiz mumkin.

Materiallar va reklamalar pullik asosda chop etiladi.

Elektron pochta: info@marketingjournal.uz
Bot: [@marketinjournalbot](https://t.me/@marketinjournalbot)
Tel.: +99890 5125464, +998939266610

Jurnalning rasmiy sayti: <https://marketingjournal.uz>

Marketing jurnali O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi **Oliy attestatsiya komissiyasi** rayosatining **2024-yil 04-oktabrdagi 362/5 sonli qarori** bilan milliy ilmiy nashrlar ro'yxatiga kiritilgan



"Marketing" ilmiy, amaliy va ommabop jurnali 2024-yil 15-martdan O'zbekiston Respublikasi Prezidenti Administratsiyasi huzuridagi Axborot va ommaviy kommunikatsiyalar agentligi tomonidan **C-5669517** reyestr raqami tartibi bo'yicha ro'yxatdan o'tkazilgan. **Litsenziya raqami: №240874**



"Marketing" ilmiy, amaliy va ommabop jurnalining xalqaro darajasi: **9710**. GOCT 7.56-2002 "Seriya nashrlarning xalqaro standart raqamlanishi" davlatlaro standartlari talablari. **Berilgan ISSN tartib raqami: 3060-4621**