https://doi.org/10.53360/2788-7995-2026-3(23)-27

BY 4.0
Check for updates

ҒТАХР: 81.93.29

Н.А. Бажаев*, А.К. Токкулиева
Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
010000, Қазақстан Республикасы, Астана қ., Қ. Сәтпаев көшесі, 2
*e-mail: shaikhanova_ak@enu.kz

ЖЕЛІЛІК ТРАФИКТІҢ УАҚЫТТЫҚ ҚАТАРЛАРЫН СЕГМЕНТАЦИЯЛАУ НЕГІЗІНДЕ IOT-ҚҰРЫЛҒЫЛАРДЫҢ АҚПАРАТТЫҚ ҚАУІПСІЗДІГІН МОНИТОРИНГТЕУ ҮШІН БЕЛГІЛЕР КЕҢІСТІГІН ҚАЛЫПТАСТЫРУ ЖӘНЕ КЕҢЕЙТУ ӘДІСІ

Аңдатпа: *Зерттеу пәні. Заттар интернеті (IoT) құрылғыларының ақылды үй жүйелерінде, өнеркәсіптік автоматтандыруда, көлікте, денсаулық сақтауда және қалалық инфрақұрылымда кеңінен таралуы рұқсат етілмеген қолжетімділіктің ықтимал нүктелерінің санының өсуімен қатар жүреді. Дәстүрлі есептеу жүйелерінен айырмашылығы, IoT-құрылғылар шектеулі аппараттық ресурстармен – жад көлемімен, процессор өнімділігімен, байланыс арналарының өткізу қабілетімен және энергия тұтынуымен сипатталады, бұл ақпараттық қауіпсіздікті (АҚ) мониторингтеудің күрделі есептеу және криптографиялық әдістерін ресурстары шектеулі жағдайларға алдын ала бейімдеусіз тікелей қолдануға мүмкіндік бермейді. Әдіс. IoT-құрылғылардың АҚ күйінің ақпараттық белгілер сипаттамасын қалыптастыру үшін желілік трафик параметрлерінің уақыттық қатарларын статистикалық, жиіліктік және шектік критерийлер негізінде сегментациялауды, сондай-ақ бақыланатын белгілердің мәндеріне әсер ететін сыртқы және ішкі факторларды ескере отырып, белгілер кеңістігін кеңейтуді қолдану ұсынылады. Күйлерді жіктеу әртүрлі күрделіліктегі алты машиналық оқыту алгоритмін – логистикалық регрессияны, наивті Байес классификаторын, тірек векторлар әдісін, кездейсоқ орманды, көпқабатты персептронды және конволюциялық нейрондық желіні қолдану арқылы орындалады. Негізгі нәтижелер. Эксперименттік бағалау IoT-түйіндерден (IEEE 802.15.4, TelosB, ESP32) тұратын зертханалық стендте және UNSW-NB15 деректер жиынтығында жүргізілді. Трафиктің уақыттық қатарларын сегментациялау және контекстік факторларды ескеру неғұрлым ықшам және ажыратылатын белгілер кеңістігін қалыптастыруға мүмкіндік беретіні, сонымен қатар классификаторларды оқытуға жұмсалатын есептеу шығындарын төмендететіні көрсетілді. Практикалық маңыздылығы. Әзірленген тәсіл ресурстары шектеулі IoT-инфрақұрылымдарға арналған АҚ оқиғаларын мониторингтеудің жеңілдетілген жүйелерін құру кезінде қолданылуы мүмкін: қарапайым модельдер шеткі түйіндерде, ал дәлірек модельдер шлюздерде және бұлттық платформаларда орындалуы мүмкін, бұл жалған іске қосылулар санын азайтып, шабуылдарды анықтаудың жеделдігін арттырады.*

Түйін сөздер: *ақпараттық қауіпсіздік, заттар интернеті, IoT, уақыттық қатарларды сегментациялау, белгілер кеңістігі, машиналық оқыту, шабуылдарды анықтау, желілік трафик, мониторинг.*

1 Кіріспе

«Индустрия 4.0» технологияларының дамуы және заттар интернеті (IoT) құрылғыларының кеңінен енгізілуі аса маңызды инфрақұрылымдарға ықтимал ақпараттық шабуылдардың бетін айтарлықтай кеңейтті [1, 2]. IoT-құрылғылар ақылды үй жүйелерінде, өнеркәсіптік автоматтандыруда, көлікте, денсаулық сақтауда және қалалық инфрақұрылымда барған сайын кеңінен қолданылады [3], бұл ретте олардың өндірістік-технологиялық және ақпараттық жүйелермен тығыз интеграциясы ақпараттық қауіпсіздікті (АҚ) мониторингтеу міндетін классикалық есептеу жүйелеріне қарағанда күрделірек етеді.

Осы саладағы зерттеулердің елеулі бөлігі машиналық оқыту алгоритмдерін қолдана отырып, IoT-құрылғылардың желілік трафигін талдау негізінде шабуылдар мен аномалияларды анықтау әдістеріне арналған [4-6]. Желілік трафик белгілерінің мәндерін бөлу уақыт бойынша тұрақты болып қалмайды: ол классификатор, әдетте, тікелей бақылай алмайтын факторлардың әсерінен өзгереді (желінің құрамының өзгеруі, жүктеменің тәуліктік және маусымдық циклділігі және т. б.), бұл штаттық оқиғаларды аномалиялық оқиғаларға қате жатқызуға әкеледі.

©Н.А. Бажаев, А.К. Шайханова, А.К. Токкулиева, 2026

Осыған байланысты IoT-желілеріндегі желілік трафиктің стационарлы еместігіне тұрақты және бақыланатын күйлердің контекстін ескеретін ықшам әрі ақпараттық белгілер кеңістігін қалыптастыру міндеті өзекті болып табылады. Осы жұмыстың мақсаты ресурстары шектеулі IoT-инфрақұрылымдарға тән есептеу шығындарын төмендету кезінде шабуылдарды анықтау дәлдігін арттыруды қамтамасыз ететін желілік трафиктің уақыттық қатарларын сегментациялау негізінде IoT-құрылғылардың АҚ мониторингі үшін белгілер кеңістігін қалыптастыру және кеңейту әдісін әзірлеу болып табылады.

2 Міндет қою

IoT-құрылғылардың есептеу ресурстары шектеулі жағдайларда ең аз есептеу шығындары кезінде желілік трафикті талдаудың жоғары дәлдігін қамтамасыз ететін ақпараттық қауіпсіздік күйін бағалау әдістерін әзірлеу қажеттілігі туындайды. IoT-инфрақұрылымдардың ерекшелігі желілік өзара әрекеттесуді көпөлшемді уақыттық қатарлар түрінде ұсыну болып табылады, олардың сипаттамалары құрылғылардың жұмыс режиміне және сыртқы әсерлерге байланысты уақыт өте өзгереді [7].

Желілік трафик параметрлерінің уақыттық қатарларының жиынтығы бойынша IoT-құрылғының ақпараттық қауіпсіздігінің ағымдағы күйін анықтауға мүмкіндік беретін белгілер кеңістігін қалыптастыру және кеңейту әдісін әзірлеу талап етіледі [8]. Ақпараттық қауіпсіздік күйі бақыланатын желілік трафиктің келесі сыныптардың біріне тиесілілігі ретінде түсіндіріледі: қалыпты жұмыс істеу, істен шығу, аномальды мінез-құлық немесе желілік шабуылдардың бір түрі.

3. Ұсынылатын әдіс

Ұсынылып отырған зерттеуде IoT-құрылғылардың күйін бағалау үшін бастапқы деректер ретінде құрылғылардың әртүрлі жұмыс режимдеріндегі мінез-құлқын сипаттайтын уақыттық қатарлар түрінде ұсынылған желілік параметрлердің реттіліктері қарастырылады.

Ғылыми жаңалық уақыт қатарларын сегментациялауды және атрибут кеңістігін контексттік кеңейтуді бірлесіп қолдану болып табылады: әр сегмент желілік трафикке әсер ететін белсенді фактор туралы ақпаратты салыстырады, бұл атрибуттардың ішкі кеңістігін қалыптастыруға және IOT құрылғыларының АҚ күйін жіктеудің тұрақтылығын арттыруға мүмкіндік береді.

Барлық бақыланатын күйлер екі ішкі жиынға бөлінеді: қауіпсіз және қауіпті. Күйлер жиынын формалды түрде келесідей анықтауға болады:

C={C1, C2}, C1⊂C – қауіпті күйлер (шабуылдар, ақаулар, ауытқулар), C2⊂C – қауіпсіз күйлер (қалыпты жұмыс).

Құрылғылардың бақыланатын күйлерін сипаттайтын көптеген уақыт тізбектері келесідей берілген:

$$X = \{X_{1,\dots}X_{c,\dots}X_C\} \in X$$

мұндағы әрбір Xc тізбегі желілік трафик параметрлерінің немесе құрылғының c күйіндегі өзара әрекеттесулерінің уақыт қатары болып табылады.

Өңдеудің есептеу ресурстарын азайту, алгоритмдердің оқу жылдамдығын арттыру мақсатында осы уақыт қатарларынан ақпараттық сипаттамаларды алу үшін сегменттік бөлу функцияларын пайдалану ұсынылады:

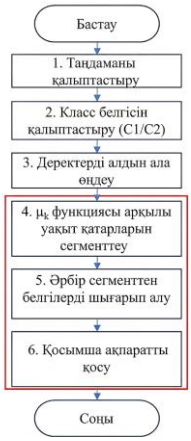
$$\mu = \{\mu_{1,\dots}\mu_{k,\dots}\mu_K\},\ \mu_k : X_c \to \{X^{ck}{}_1, \dots X^{ck}_m\}$$

мұндағы әрбір μk функциясы сыртқы факторлардың әсерінен болатын трафиктің жергілікті мінез-құлық ерекшеліктерін сипаттайтын Xckm тізбегін бөледі. Оларға оқыту үлгісінде алдын ала белгілі шабуыл, жұмыс режимдері, трафиктің уақыт қатарына әсер ететін сыртқы жағдайлардың өзгеруі кіруі мүмкін.

Жоғарыда аталған факторларды дәл анықтау мүмкін болмаған немесе олар болмаған жағдайда, сегментация әдістері статистикалық сипаттамаларға (орташа мән, дисперсия, эксцесс, асимметрия), бақылаулар арасындағы метрикалық қашықтықтарды талдауға (мысалы, Евклид қашықтығы, DTW), жиіліктік декомпозицияға, сондай-ақ сигналдың шектік сипаттамалары бойынша сүзуге негізделуі мүмкін [9].

μk функциясы арқылы қалыптастырылатын сегменттер құрылымдалған белгілер кеңістігін одан әрі құру үшін негізді анықтайды. Алынған сегменттерден белгілерді бөліп алу трафиктің өзгеруіне тұрақты деректерді қалыптастыруға бағытталған.

Оқыту таңдауын қалыптастыру қадамдарының жалпы реттілігі ұсынылған әдістің сызбасы 1-суретте көрсетілген, онда сегментация, белгілерді бөліп алу, таңдауларды қалыптастыру және классификаторды оқыту кезеңдерінің өзара байланысы көрсетілген.



Сурет 1 – Оқыту үлгісін қалыптастыру қадамдарының реттілігі

Белгілер кеңістігін кеңейту және IoT құрылғыларының ақпараттық қауіпсіздік күйлерін жіктеу міндеті үшін келесі деректерді қарастырайық. X – объектілер сипаттамаларының жиыны, $x \in X$ – алдын ала берілген векторлық кеңістіктегі белгілердің d-өлшемді кортежі $X = R^d$. x сипаттамаларына сәйкес қойылған, $\{c_1, c_2, \dots c_l\} \in C$ белгілерімен белгіленген кластар жиыны күйлердің бинарлық жиынына бөлінеді, яғни біріктірілген кластар $\{C_1, C_2\} \in C$. Белгілердің мәндеріне әсер ететін факторлардың V жиыны бар. X жиыны $f: X \to D_f$ белгілер мәндерінің жиынымен сипатталады.

Қазіргі уақыттағы әсер етуші $v \in V$ факторын ескере отырып, берілген $f_1, \dots, f_n$ кезінде бақылау объектісінің $X_i \in X$ белгілер бойынша сипаттамасы $X_i = \left(f_1(v,x), \dots, f_n(v,x)\right)$ түрінде анықталады.

Оқыту жиыны таңбаланған таңдау түріне ие болады $\left\{v_j, \{(x_i, c_i)\}_{i=1}^{N}\right\}_{j=1}^{M}$, мұндағы i – қарастырылып отырған j-фактор v_j әсер еткен сәттегі кортеждер саны.

v факторының әсерін ескеретін жіктеу алгоритмін құру қажет, $\alpha: X \to C$. Алгоритм кіріс белгілер сипаттамасы бойынша сәйкес класты анықтайды. Мұндай модельді оқыту белгілер кеңістігінің үлкен өлшемділігімен ғана емес, сонымен қатар айнымалылардың бір бөлігінің объект күйінің өзгеруіне байланысты емес, классификатор тікелей бақылай алмайтын сыртқы және ішкі факторлардың әсерінен өзгеруімен де күрделенеді.

Бұл IoT желілеріне қатысты машиналық оқыту әдістерінің принциптік шектеуінен тұрады: жүйе динамикада жұмыс істейді, қоршаған орта факторларының әсерінен бір режимнен екіншісіне үнемі ауысып отырады, ал олардың әсерін есепке алмай оқытылған алгоритм мұндай ауысуларды нақты шабуылдардан немесе іркілістерден, сондай-ақ желілік трафиктің өзгерістерін ақпараттық қауіпсіздіктің бұзылу белгілері ретінде тұрақты түрде ажырата алмайды.

Мысалы, желі сегментіне жаңа құрылғының қосылуы қызметтік хабарламалар санының, конфигурациялық командалардың және кідірістің ұлғаюына заңды түрде әкеледі – бұл әдетте аномальды белсенділікті анықтау үшін қолданылатын белгілердің дәл өзі. Өзгерістің себебін ескермей, модель штаттық оқиғаны ықтимал қауіп ретінде қабылдайды.

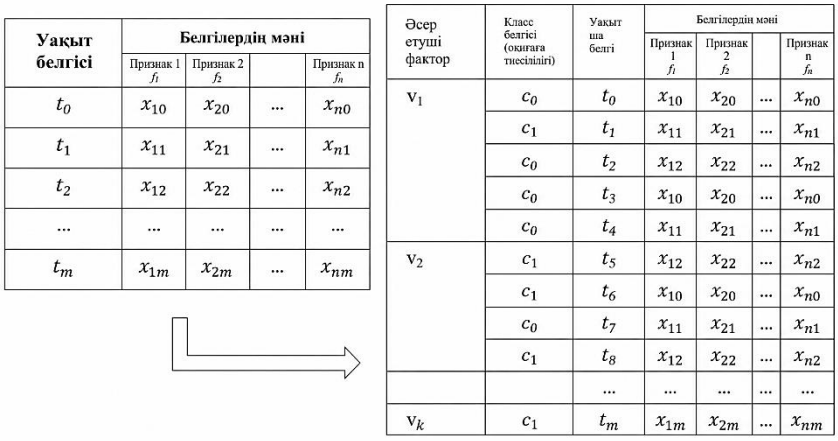
Жүйенің жұмыс істеуі туралы жинақталған статистика матрица түрінде ұсынылады, онда әрбір уақыт белгісіне тіркеуші құрылғылар арқылы тіркелген белгілердің мәндері сәйкес келеді. Осы мәндерге әсер ететін факторлардың бір бөлігі алдын ала белгілі – мысалы, жүктеменің тәуліктік немесе маусымдық циклділігі. Бұл олардың әсерімен шу ретінде күреспей, олар туралы ақпаратты оқыту ішкі жиынының таңбалануына тікелей енгізуге, әрбір жазбаға сәйкес өрісті қосуға мүмкіндік береді.

Оқыту үлгілерін қалыптастыру 1-кестеде келтірілген белгілерге негізделеді, онда штаттық және қауіпті күй үшін трафиктің негізгі параметрлері салыстырылған [10].

Кесте 1 – Объектілердің штаттық және қауіпті күйдегі белгілер бойынша сипаттамасы

Қауіпсіз (қалыпты жұмыс)		Қауіпті жағдай (шабуыл, сәтсіздік, аномалия)	
Белгі	Сипаттама	Белгі	Сипаттама
proto	TCP немесе UDP, тұрақты қосылым түрі қолданылады	proto	Ерекше хаттамалар (ICMP flood, стандартты емес UDP)
service	Қызметтердің тұрақты жиынтығы (HTTP, MQTT, CoAP)	service	Типтік емес қызметтерге қол жеткізу (telnet, ftp, rlogin)
sbytes, dbytes	Түйіндер арасындағы берілген байттар саны	state, swin, dwin, tcprtt, synack, ackdat	TCP аномальды комбинациялары бар жалаулар
state	Қосылу күйі	sbytes, dbytes, Spkts, Dpkts	Деректер теңгерімсіздігі
Spkts, Dpkts	Түйіндер арасындағы пакеттер саны	is_ftp_login, ct_ftp_cmd	Сәтсіз кірістер сериясы
Sload, Dload	Деректерді беру жылдамдығы	label	Желі күйінің мақсатты екілік белгісі
ct_srv_src, ct_srv_dst, ct_src_ltm, ct_dst_ltm, ct_src_dport_ltm, ct_dst_sport_ltm, ct_dst_src_ltm	Уақыт терезесі ішінде сол IP-мекенжаймен немесе портпен қосылымдар саны	attack_cat	Шабуыл санаты: Fuzzers, Analysis, Backdoor, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms. Қауіпті жағдайдың түрін анықтауға мүмкіндік береді.
Spkts, dsport	Бір порттан пакеттердің қайталануы	Sload	Дереккөзден тағайындалған жерге дейінгі деректерді беру жылдамдығы. Жоғары немесе күрт өзгеретін мәндер желінің шамадан тыс жүктелуін, flood-шабуылдарды және басқа да аномалияларды анықтау үшін пайдаланылуы мүмкін.
Sintpkt, Dintpkt	Телеметрия хабарламалары арасындағы орташа аралық	ct_dst_ltm, ct_srv_dst, ct_dst_src_ltm	Сол хостпен қосылымдар саны күрт өсті
tcprtt, synack, ackdat	TCP өзара әрекеттесуі және қосылымды орнату кідірісі	dst_host_srv_diff_host_rate	Тағайындалған хосттардың әртүрлілігі

2-суретте деректер жиынтығының бастапқы жазбалары әсер етуші фактор туралы ақпаратпен толықтырылған жазбаларға қалай түрлендірілетіні көрсетілген: осындай таңбалаудан кейін әрбір жазба нақты факторға байланыстырылған болады, бұл өз кезегінде белгілер кеңістігін осы фактордың мәніне сәйкес ішкі кеңістіктерге одан әрі бөлуге мүмкіндік береді.



Сурет 2 – IoT құрылғыларының нысан белгілері кортежінің ресми көрінісі

Мұндай таңбалау белгілер кеңістігін әсер етуші фактордың мәніне сәйкес ішкі кеңістіктерге бөлуге және классификаторды бүкіл таңдау жиынтығында емес, олардың әрқайсысында жеке-жеке оқытуға мүмкіндік береді. Объектіні белгілі бір күйге жатқызу міндеті осылайша контекстен бөлек емес, сол сәтте тіркелген нақты ақпараттық әсерді ескере отырып шешіледі.

Мұндай тәсіл сыртқы факторлардан туындаған деректердің өзгергіштігін объект күйінің өз өзгеруін көрсететін өзгергіштіктен ажыратуға және бөлу нәтижесінде алынған таңдау жиынтықтарының қасиеттерін әрбір сегмент үшін жеке бағалауға мүмкіндік береді. Бұл, өз кезегінде, деректердің нақты сегменті үшін өңдеудің ең тиімді алгоритмін таңдауға негіз қалыптастырады.

4. Эксперимент жүргізу

Эксперименттік бағалау әзірленген әдісті кешенді тәсілмен жүргізілді: физикалық стендтегі зертханалық сынақтар симуляциялық орталарда масштабталатын модельдеумен толықтырылды. Бұл ресурстық шектеулерге байланысты нақты жабдықта іске асыру мүмкін емес үлестірілген желілер сценарийлерін қайта жаңғыртуға мүмкіндік берді.

Зертханалық қондырғы IEEE 802.15.4 қолдау көрсететін микроконтроллерлер негізіндегі IoT түйіндерінен (TelosB, ESP32), трафикті талдау модулі бар шлюзден және жүктемені басқару контроллерінен тұрды. Желі аралас құрылымға ие болды: статистиканы жинауға арналған сымды сегменттер және коллизияларға, шуылдарға әрі кеңтаралымды трафиктің шамадан тыс жүктелуіне бейім сымсыз арналар қолданылды. Бұл өнеркәсіптік IoT ортасына тән жағдайларды, атап айтқанда өткізу қабілетінің шектеулілігі мен қосылымдардың тұрақсыздығын қайта жаңғыртуға мүмкіндік берді.

Эксперименттік тексеру үшін 2,5 миллионнан астам желілік оқиғаны қамтитын UNSW-NB15 деректер жиынтығы пайдаланылды. Оның құрамында қалыпты трафикпен қатар желілік шабуылдардың заманауи түрлері: Fuzzers, Analysis, Backdoor, DoS, Exploits, Generic, Reconnaissance, Shellcode және Worms қамтылған. Осы деректер жиынтығын пайдалану ұсынылған әдістің желілік трафиктің өзгеруінің әртүрлі сценарийлеріндегі тиімділігін бағалауға мүмкіндік берді.

Бастапқы деректер алдын ала тазартылып, нормализацияланды, содан кейін уақыттық қатарларды сегментациялау және белгілер кеңістігін қалыптастыру орындалды. Ұсынылып отырған тәсілдің ерекшелігі – әрбір объектіге IoT құрылғысының жұмыс істеу жағдайларын сипаттайтын әсер етуші фактор v туралы ақпараттың қосылуы. Нәтижесінде оқыту үлгісінде желілік трафик белгілерінің мәндерімен қатар олардың қалыптасу контексті де қамтылады. Бұл сыртқы әсерлерден туындайтын өзгерістерді ақпараттық қауіпсіздіктің бұзылуынан туындайтын өзгерістерден ажыратуға мүмкіндік береді.

Эксперименттік бағалау үшін машиналық оқытудың алты алгоритмі пайдаланылды. Логистикалық регрессия (Logistic Regression) және Наивті Байес классификаторы (Naive Bayes) – перифериялық, ресурстары шектеулі IoT түйіндерінде қолданылатын жылдам, ресурсты аз қажет ететін базалық модельдер ретінде қарастырылды. Тірек векторлар әдісі (SVM) және кездейсоқ орман әдісі (Random Forest) – күрделілігі орташа классикалық алгоритмдер ретінде қолданылды. Көпқабатты персептрон (MLP) және конволюциялық нейрондық желі (CNN) – терең оқыту модельдері ретінде пайдаланылды. Алгоритмдердің мұндай жиынтығы әртүрлі күрделілік деңгейі мен есептеу шығындарына ие модельдерді бірізді түрде салыстыруға мүмкіндік береді.

Базалық модельдердің тиімділігін сандық бағалау үшін Accuracy, Precision, Recall, F1-score және ROC-AUC көрсеткіштері бойынша салыстыру жүргізілді. Алынған нәтижелер 2-кестеде ұсынылған.

Кесте 2 – Трафикті талдау кезіндегі машиналық оқыту алгоритмдерін салыстырмалы бағалау

Алгоритм	Accuracy	Precision	Recall	F1-score	ROC-AUC
Тірек векторлар әдісі (SVM)	92%	94%	91%	90%	93%
Кездейсоқ орман әдісі (Random Forest)	93%	94%	92%	92%	95%
Логистикалық регрессия (Logistic Regression)	88%	89%	87%	88%	89%
Наивті Байес классификаторы (Naive Bayes)	82%	83%	80%	81%	83%
Көп қабатты перцептрон (MLP)	95%	95%	94%	95%	96%
Конволюциялық нейрондық желі (CNN)	96%	96%	95%	96%	97%

Алынған нәтижелер қарастырылған базалық модельдердің ішінде ең жақсы көрсеткіштерді конволюциялық нейрондық желі (CNN) көрсететінін дәлелдейді: Accuracy – 95,8 %, Precision – 96 %, Recall – 0,95, F1-score – 0,96 және ROC-AUC – 0,97. Random Forest моделі де жоғары тиімділікті көрсетеді (Accuracy – 93,1 %, F1-score – 0,92, ROC-AUC – 0,95), ал Logistic Regression және Naive Bayes модельдерінің көрсеткіштері төменірек. Бұл олардың желілік трафиктің күрделі сызықтық емес тәуелділіктерін модельдеу мүмкіндігінің шектеулілігімен байланысты. Одан әрі CNN ұсынылған әдіспен салыстыру үшін негізгі базалық модель ретінде пайдаланылады.

Гиперпараметрлерді баптау алдын ала жүргізілген есептеу эксперименттері және IoT желілік трафигін талдау бойынша заманауи әдебиеттердегі ұсынымдар негізінде эксперименттік түрде орындалды. Модельдердің негізгі параметрлері 3-кестеде келтірілген.

Кесте 3 – Модельдердің гиперпараметрлері

Модель	Негізгі параметрлер
Логистикалық регрессия	C=1.0; class_weight=balanced
Наивті Байес классификаторы	GaussianNB (без настройки)
Кездейсоқ орман әдісі	n_estimators=300; max_depth=25; min_samples_leaf=2
Тірек векторлар әдісі	kernel=RBF; C=1.5; γ=0.01
Көп қабатты перцептрон	hidden=(256,128,64); dropout=0.3; Adam lr=0.001
Конволюциялық нейрондық желі	Conv1D(64-128-256); kernel=3; Dropout=0.5

Тиімділікті бағалау жалпыға қабылданған Accuracy, Precision, Recall, F1-score және ROC-AUC [11] метрикалары бойынша жүргізілді. Нәтижелердің тұрақтылығын растау мақсатында деректер оқыту және тестілеу үлгілеріне 80/20 арақатынасында бөлінді, сондай-ақ машиналық оқыту модельдерін оқыту кезінде 5 еселік кросс-валидация қолданылды. Мұндай тәсіл қайта оқытылу ықтималдығын азайтып, алынған нәтижелердің қайталанымдылығын қамтамасыз етуге мүмкіндік берді.

Эксперимент екі кезеңде жүргізілді. Бірінші кезеңде әсер етуші факторлар туралы қосымша ақпарат ескерілмей, жіктеу жүзеге асырылды. Екінші кезеңде ұсынылған кеңейтілген белгілер кеңістігі қолданылып, оған әсер етуші фактор v туралы ақпарат қосылды. Мұндай салыстыру ұсынылған әдістің жіктеу сапасына қосатын үлесін сандық тұрғыдан бағалауға мүмкіндік берді.

Ұсынылған әдістің әсерін егжей-тегжейлі бағалау үшін Random Forest алгоритмы мысалында қосымша салыстыру жүргізілді. Базалық нұсқа ретінде UNSW-NB15 бастапқы белгілер кеңістігі бойынша жіктеу қолданылды, ал ұсынылған нұсқада уақыттық қатарларды сегменттеу және әрекет етуші фактор туралы ақпаратпен белгілер кеңістігін кеңейту жүзеге асырылды. Салыстырмалы эксперименттің нәтижелері 4-кестеде ұсынылған.

Кесте 4 – Random Forest негізіндегі базалық және ұсынылған тәсілдерді салыстыру

Тәсіл	Accuracy	Precision	Recall	F1-score	ROC-AUC
Бастапқы белгілер кеңістігі бар Random Forest	93%	94%	92%	92%	95%
Уақыттық қатарларды сегменттеуі бар Random Forest	94%	95%	93%	94%	96%
Әрекет етуші факторды ескере отырып сегменттелген уақыттық қатарлары бар Random Forest	96%	96%	95%	95%	97%

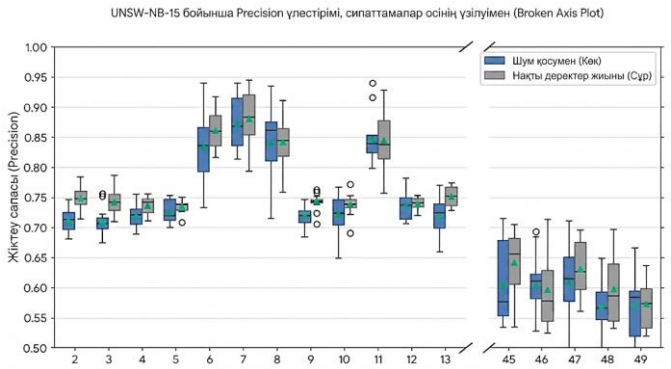
Нәтижелер уақыттық қатарларды сегменттеуді және әрекет етуші факторды есепке алуды дәйекті түрде қолдану бастапқы белгілер кеңістігін пайдаланумен салыстырғанда жіктеу сапасын арттыратынын көрсетеді. Random Forest үшін Accuracy мәні 0,93-тен 0,96-ға дейін, Precision – 0,94-тен 0,96-ға дейін, Recall – 0,92-ден 0,95-ке дейін, F1-score – 0,92-ден 0,95-ке дейін, ал ROC-AUC – 0,95-тен 0,97-ге дейін артты. Осылайша, IoT-құрылғысының жұмыс істеу контекстін ескере отырып белгілер кеңістігін кеңейту желілік трафик сипаттамалары өзгерген кезде жіктеудің тұрақтылығын арттыруды қамтамасыз етеді.

Сонымен қатар, сегменттеу параметрлерінің жіктеу сапасына әсерін бағалау үшін сегменттердің әртүрлі саны бойынша зерттеу жүргізілді. Эксперимент барысында сегменттер саны 2-ден 20-ға дейінгі аралықта өзгертіліп, содан кейін қолданылған классификаторлардың сапа көрсеткіштері салыстырылды. Алынған нәтижелер 5-8 сегмент аралығында жіктеу сапасы мен есептеу жүктемесі арасындағы ең қолайлы арақатынасқа қол жеткізілетінін көрсетті: Precision 93-94 %, Recall – 0,91–0,93, ал F1-score – шамамен 0,92 құрайды. Сегменттер санын одан әрі арттырған кезде өңдеу уақытының бір мезгілде өсуімен қатар сапа көрсеткіштерінің тұрақтануы байқалады.

Белгілер кеңістігін қалыптастыру үшін уақыттық қатарлар сегменттерінің статистикалық, уақыттық және жиіліктік сипаттамалары талданды. Сегменттердің статистикалық сипаттамалары ретінде орташа мән, дисперсия, стандартты ауытқу және медиана қарастырылды. Бұл сипаттамаларды талдау желілік трафиктің статистикалық қасиеттерінің жекелеген уақыттық сегменттерде өзгеруін ескеруге мүмкіндік береді. 3-суретте dload белгісін шуылдандырудың Precision метрикасының мәніне әсері көрсетілген. Экспериментте предикторлардың біріне дәйекті түрде гаусстық шуыл қосылып, содан кейін қауіпті күйді анықтау кезінде Precision көрсеткішінің өзгеруі бағаланды. UNSW-NB15 деректер жиыны үшін 49 предиктор қарастырылды. Осылайша, графиктегі әрбір нүкте зерттелетін белгілердің бірінің күйі өзгерген кездегі жіктеу нәтижесіне сәйкес келеді. Мұндай тәсіл жекелеген белгілердің жергілікті ауытқуларға сезімталдығын бағалауға және сегменттелген ұсынудың желілік трафиктегі кездейсоқ өзгерістердің әсерін қаншалықты төмендететінін анықтауға мүмкіндік берді.

Негізгі салыстырмалы эксперимент үшін ауытқудың орташа деңгейі ретінде $\sigma = 0.04$, мәні қолданылды. Бұл мән желілік трафиктің нормаланған сипаттамаларындағы шағын ауытқуларды олардың бастапқы үлестірімін елеулі түрде бұзбай модельдеуге мүмкіндік береді. Аталған деңгейді таңдау алдыңғы зерттеулердің нәтижелерімен қосымша негізделді, оларда UNSW-NB15 деректер жиыны үшін гаусстық шуылдың 0.02–0.08 деңгейлері қолданылып, шабуылдарды анықтау модельдерінің шуыл деңгейінің артуына төзімділігі бағаланған [12].

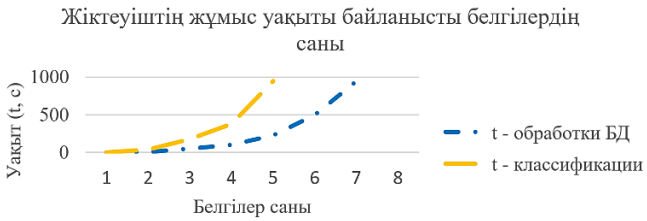
Белгілер кеңістігін қалыптастыру үшін уақыттық қатарлар сегменттерінің статистикалық, уақыттық және жиіліктік сипаттамалары талданды. 3-суретте dload белгісіне шу қосылуының Precision метрикасының мәніне әсері көрсетілген. Алынған нәтижелер сегменттелген ұсынуды қолдану жергілікті шу әсерлерінің қауіпті күйлерді анықтау сапасына ықпалын төмендетуге мүмкіндік беретінін растайды.



Сурет 3 – Шусыз және шу қолданылған кездегі precision дәлдігі бойынша жіктеу сапасы

Белгілерді іріктеу әдістерін қолдану ақпараттық маңызы төмен белгілерді алып тастау және деректердің өлшемділігін азайту есебінен есептеу шығындарын төмендетуге мүмкіндік береді. Классификатордың жұмыс уақытының пайдаланылатын белгілер санына тәуелділігі 4-суретте көрсетілген.

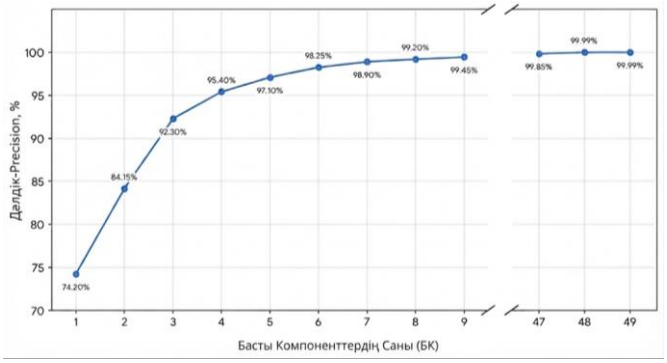
Есептеу шығындарын сандық бағалау үшін эксперимент барысында пайдаланылатын белгілер саны әртүрлі болған кезде деректерді өңдеу уақыты және классификатордың жұмыс істеу уақыты өлшенді. Осылайша, есептеу тиімділігі тікелей өлшенетін сипаттама – деректерді өңдеу және жіктеуді орындау уақыты бойынша бағаланды. Бұл белгілер кеңістігінің өлшемділігі азайған кезде есептеу жүктемесінің өзгеруін бағалауға мүмкіндік береді.



Сурет 4 – Белгілердің санына байланысты жіктеуіштің жұмыс уақыты

Белгілер кеңістігінің өлшемділігінің әсерін бағалау үшін негізгі компоненттер әдісі (PCA, Principal Component Analysis) қолданылды. Бастапқы белгілер негізгі компоненттер кеңістігіне түрлендірілді, содан кейін сақталатын компоненттер саны әртүрлі болған жағдайда жіктеу жүргізілді. Компоненттер санының әрбір мәні үшін жіктеу сапасына жеке эксперименттік бағалау орындалды, ал алынған Precision мәндері 5-суретте көрсетілген тәуелділікті құру үшін пайдаланылды. Осылайша, графиктегі әрбір нүкте сақталатын PCA-компоненттердің берілген саны кезіндегі эксперименттік жолмен алынған Precision мәніне сәйкес келеді.

5-суретте пайдаланылатын негізгі компоненттердің санына байланысты дәлдік сапасының қол жеткізілген көрсеткішінің мәндері келтірілген. Өсу белгілі бір нүктеге дейін жалғасады, содан кейін мәндер үстіртке шығады және компонентті одан әрі қосу айтарлықтай әсер етпейді.



Сурет 5 – Дәлдік мәнінің PCA әдісімен алынған негізгі компоненттер санына тәуелділігі

Алынған сегменттер әртүрлі статистикалық қасиеттермен және белгілердің әртүрлі таралуымен сипатталады. Бұл қолданыстағы факторлар туралы қосымша ақпаратты есепке алу біртекті оқыту үлгілерін қалыптастыруға және желілік трафиктің тұрақсыздығының әсерін азайтуға мүмкіндік беретіндігін көрсетеді. Сонымен қатар, әртүрлі классификаторларды қолдану қолтаңба кеңістігін қалыптастырудың ұсынылған әдісінің әмбебаптығын растайды.

Есептеу тиімділігін талдау нәтижелері сегменттеуді зерттеу нәтижелерімен сәйкес келеді: жіктеу сапасы тұрақтану аймағына жеткеннен кейін сегменттер санын одан әрі арттыру негізінен өңделетін деректер көлемі мен есептеу уақытының ұлғаюына әкеледі, алайда сапаның соған сәйкес елеулі өсуін қамтамасыз етпейді. Сондықтан сегменттеу параметрлерін таңдау кезінде жіктеу сапасы мен есептеу шығындарын бір мезгілде ескеру қажет.

Алынған нәтижелер белгі кеңістігін қалыптастыру кезінде қолданыстағы факторлар туралы қосымша ақпаратты пайдалану IoT-құрылғылардың ақпараттық қауіпсіздігінің жай-күйін айқындау сапасын арттыратынын және желілік трафиктің өзгермелі сипаттамалары жағдайында неғұрлым тұрақты жіктеуді қамтамасыз ететінін растайды.

5 Қорытынды

Ұсынылған IoT құрылғыларының ақпараттық қауіпсіздігін мониторингілеу үшін желілік трафиктің уақыттық қатарларын сегментациялауға негізделген белгілер кеңістігін қалыптастыру және кеңейту әдісі деректерді өңдеудің қолайлы есептеу күрделілігін сақтай отырып, желі күйлерінің белгілер арқылы сипатталуының ақпараттылығын арттыруға мүмкіндік береді. Сегменттік талдауды қолдану желілік трафиктің мінез-құлқындағы жергілікті ерекшеліктерді анықтауға, бақыланатын процестердің стационарлық емес әсерін азайтуға және қауіпсіз күйлерді, ақауларды және желілік шабуылдардың әртүрлі түрлерін дәлірек ажыратуға мүмкіндік береді.

Ұсынылған тәсілдің айрықша ерекшелігі-оқыту үлгілерін қалыптастыру кезінде желілік трафиктің сипаттамаларына әсер ететін факторлар туралы қосымша ақпаратты пайдалану. Мұндай ақпаратты есепке алу атрибуттық кеңістікті кеңейтуге, бақылау объектілерінің атрибуттық сипаттамасының ақпараттылығын арттыруға және IoT құрылғыларының жұмыс істеу режимдеріне байланысты өзгерістердің әсерін азайтуға мүмкіндік береді, бұл машиналық оқыту модельдерінің желілік трафик сипаттамаларының өзгеруіне тұрақтылығын арттыруға ықпал етеді.

Жүргізілген эксперименттік бағалау әртүрлі есептеу күрделілігіндегі машиналық оқыту алгоритмдерін пайдалану кезінде ұсынылған әдістің тиімділігін растайды. Нәтижелер кеңейтілген қолтаңба кеңістігін қолдану пайдаланылған деректерді өңдеу үлгісіне қарамастан ақпараттық қауіпсіздік күйлерін жіктеу сапасын арттыруға ықпал ететінін көрсетеді. Әзірленген тәсілді ресурспен шектелген IoT-инфрақұрылымдардың ақпараттық қауіпсіздігін мониторингтеудің бейімделген зияткерлік жүйелерін құру үшін әдістемелік негіз ретінде пайдалануға болады.

Әдебиеттер тізімі

1. Lee J. A Cyber-Physical Systems Architecture for Industry 4.0 / J. Lee, B. Bagheri, H.A. Kao // Manufacturing Letters. – 2015. – Vol. 3. – P. 18-23.
2. Atzori L. The Internet of Things: A Survey / L. Atzori, A. Iera, G. Morabito // Computer Networks. – 2010. – Vol. 54, № 15. – P. 2787-2805.
3. Threat analysis of IoT networks using artificial neural network intrusion detection system / E. Hodo et al // 2016 International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, 2016. – P. 1-6, https://doi.org/10.1109/ISNCC.2016.7746067.
4. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets and Comparative Study / M.A. Ferrag et al // Journal of Information Security and Applications. – 2020. – Vol. 50.
5. Survey of intrusion detection systems: techniques, datasets and challenges / A. Khraisat et al // Cybersecur 2, 20 (2019). https://doi.org/10.1186/s42400-019-0038-7.
6. Alrawashdeh K. Toward an Online Anomaly Intrusion Detection System Based on Deep Learning / K. Alrawashdeh, C. Purdy // 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 2016. – P. 195-200. https://doi.org/10.1109/ICMLA.2016.0040.
7. Sukhoparov M.E. Improving the Quality of the Identification of the Information Security State Based on Sample Segmentation / M.E. Sukhoparov, I.S. Lebedev // Aut. Control Comp. Sci. – 2023. – № 57. – P. 1071-1075. https://doi.org/10.3103/S0146411623080321.
8. Semenov V.V. Identification of Abnormal Functioning of Devices of Cyber-Physical Systems / M.E. Sukhoparov, I.S. Lebedev // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN ruSMART 2020. Lecture Notes in Computer Science. – 2020. – vol 12525. Springer, Cham. https://doi.org/10.1007/978-3-030-65726-0_1.
9. Segmenting Time Series: A Survey and Novel Approach / E. Keogh et al // Data Mining in Time Series Databases. – World Scientific, 2004.
10. Moustafa N. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) / N. Moustafa, J. Slay // 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia. – 2015. – P. 1-6. https://doi.org/10.1109/MilCIS.2015.7348942.
11. Fawcett T. An Introduction to ROC Analysis / T. Fawcett // Pattern Recognition Letters. – 2006. – Vol. 27, № 8. – P. 861-874.

Алғыс білдіру

Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитетінің қаржылық қолдауымен орындалды (грант № AP25794699).

Н.А. Бажаев*, А.К. Токкулиева

Евразийский национальный университет им. Л.Н. Гумилева,
010000, Республика Казахстан, г. Астана, улица К. Сатпаева, 2,
*e-mail: shaikhanova_ak@enu.kz

МЕТОД ФОРМИРОВАНИЯ И РАСШИРЕНИЯ ПРИЗНАКОВОГО ПРОСТРАНСТВА ДЛЯ МОНИТОРИНГА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ IOT-УСТРОЙСТВ НА ОСНОВЕ СЕГМЕНТАЦИИ ВРЕМЕННЫХ РЯДОВ СЕТЕВОГО ТРАФИКА

Предмет исследования. Широкое распространение устройств Интернета вещей (IoT) в системах умного дома, промышленной автоматизации, на транспорте, в здравоохранении и городской инфраструктуре сопровождается ростом числа возможных точек несанкционированного доступа. В отличие от традиционных вычислительных систем IoT-устройства характеризуются ограниченными аппаратными ресурсами – объёмом памяти, производительностью процессора, пропускной способностью каналов связи и энергопотреблением, что делает невозможным прямое применение сложных вычислительных и криптографических методов мониторинга информационной безопасности (ИБ) без их предварительной адаптации к ресурсоограниченным условиям. Метод. Для формирования информативного признакового описания состояния ИБ IoT-устройств предложено использовать сегментацию временных рядов параметров сетевого трафика на основе статистических, частотных и пороговых критериев, а также расширение признакового пространства с учётом внешних и внутренних факторов, влияющих на наблюдаемые значения признаков. Классификация состояний выполняется с применением шести алгоритмов машинного обучения различной сложности – логистической регрессии, наивного байесовского классификатора, метода опорных векторов, случайного леса, многослойного персептрона и свёрточной нейронной сети. Основные результаты. Экспериментальная оценка выполнена на лабораторном стенде из IoT-узлов (IEEE 802.15.4, TelosB, ESP32) и на наборе данных UNSW-NB15. Показано, что сегментация временных рядов трафика и учёт контекстных факторов позволяют формировать более компактное и различимое признаковое пространство, снижая при этом вычислительные затраты на обучение классификаторов. Практическая значимость. Разработанный подход применим при построении облегчённых систем мониторинга событий ИБ для ресурсоограниченных IoT-инфраструктур: простые модели могут выполняться на периферийных узлах, а более точные – на шлюзах и облачных платформах, что снижает число ложных срабатываний и повышает оперативность обнаружения атак.

Ключевые слова: *информационная безопасность, интернет вещей, IoT, сегментация временных рядов, признаковое пространство, машинное обучение, обнаружение атак, сетевой трафик, мониторинг.*

N. Bazhayev*, A. Tokkuliyeva
L.N. Gumilyov Eurasian National University,
2 Satbayev Street, Astana, 010000, Republic of Kazakhstan
*e-mail: shaikhanova_ak@enu.kz

A METHOD FOR FEATURE SPACE CONSTRUCTION AND EXPANSION FOR INFORMATION SECURITY MONITORING OF IoT DEVICES BASED ON NETWORK TRAFFIC TIME SERIES SEGMENTATION

Subject of the study. The widespread adoption of the Internet of Things (IoT) in smart homes, industrial automation, transportation, healthcare, and smart city infrastructures has led to a significant increase in the number of potential attack surfaces and unauthorized access points. Unlike conventional computing systems, IoT devices are characterized by limited hardware resources, including restricted memory capacity, low computational performance, limited communication bandwidth, and low power consumption. These constraints hinder the direct application of computationally intensive and cryptographic information security monitoring techniques without adaptation to resource-constrained environments. Method. To construct an informative feature representation of IoT device security states, a method based on network traffic time series segmentation using statistical, frequency-domain, and threshold-based criteria is proposed. The feature space is further expanded by incorporating external and internal factors affecting the observed feature values. Security state classification is performed using six machine learning algorithms of varying computational complexity: Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Random Forest, Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN). Main results. The proposed method was experimentally evaluated using both a laboratory testbed based on IoT nodes (IEEE 802.15.4, TelosB, ESP32) and the UNSW-NB15 dataset. The results demonstrate that network traffic time series segmentation combined with contextual information enables the construction of a more compact and discriminative feature space while reducing the computational cost of classifier training. Practical significance. The proposed approach can be applied to the development of lightweight information security monitoring systems for resource-constrained IoT infrastructures. Lightweight machine learning models can be deployed on edge devices, whereas more computationally intensive models can operate on gateways or cloud platforms, thereby reducing false alarm rates and improving the timeliness of cyberattack detection.

Key words: *information security, Internet of Things, IoT, time series segmentation, feature space, machine learning, attack detection, network traffic, monitoring.*

Сведения об авторах

Нуржан Аманкулулы Бажаев* – постдокторант, Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Республика Казахстан; e-mail: nurzhan_nfs@hotmail.com. ORCID: https://orcid.org/0009-0009-4820-6132.

Айжан Конурбаевна Токкулиева – докторант, Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Республика Казахстан; e-mail: tokkuliyeva_ak@enu.kz. ORCID: https://orcid.org/0000-0002-5019-2413.

Авторлар туралы мәліметтер

Нуржан Аманкулулы Бажаев* – постдокторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, «Мемлекеттік техникалық қызмет» АҚ, Астана қ., Қазақстан Республикасы; e-mail: nurzhan_nfs@hotmail.com. ORCID: https://orcid.org/0009-0009-4820-6132.

Айжан Конурбаевна Токкулиева – докторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан Республикасы; e-mail: tokkuliyeva_ak@enu.kz. ORCID: https://orcid.org/0000-0002-5019-2413.

Information about the authors

Nurzhan Bazhayev* – postdoctoral fellow; L.N. Gumilyov Eurasian National University, State Technical Service, Astana, Republic of Kazakhstan; e-mail: nurzhan_nfs@hotmail.com. ORCID: https://orcid.org/0009-0009-4820-6132.

Aizhan Tokkuliyeva – doctoral student; L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan; e-mail: tokkuliyeva_ak@enu.kz. ORCID: https://orcid.org/0000-0002-5019-2413.

Редакцияға енуі 20.06.2026
Өңдеуден кейін түсуі 04.08.2026
Жариялауға қабылданды 04.09.2026

https://doi.org/10.53360/2788-7995-2026-3(23)-28

МРНТИ: 68.85.85

CC BY 4.0
Check for updates

Ш.Ж. Сеилов[1], Б.Ш. Журсинбек[1], М.У. Байдельдинов[1], А.К. Майданов[*], А.А. Сейлов[1,2]
[1]Евразийский национальный университет им. Л.Н. Гумилева,
010008, Республика Казахстан, г. Астана, ул. Сатпаева, 2
[2]ТОО «Ауыл телеком» (дочерняя компания АО «Казахтелеком»)
*e-mail: makeadil@mail.ru

ИНТЕЛЛЕКТУАЛЬНАЯ СЕЛЬСКАЯ ТЕРРИТОРИЯ КАК МОДЕЛЬ ПРОСТРАНСТВЕННОГО РАЗВИТИЯ КАЗАХСТАНА В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ АГРОПРОМЫШЛЕННОГО КОМПЛЕКСА

Аннотация: *Устойчивое развитие сельских территорий остаётся одним из ключевых условий продовольственной безопасности и роста экономики Казахстана. Несмотря на значительные государственные инвестиции в цифровую инфраструктуру и сельское хозяйство, сохраняются неравномерность территориального развития, низкая производительность труда, ограниченная диверсификация сельской экономики и миграционный отток населения.*

В работе предложена концепция «Интеллектуальная сельская территория» (Intelligent Rural Territory, IRT), рассматривающая сельскую территорию как интегрированную социально-экономическую систему, в которой цифровые технологии служат инструментом повышения эффективности агропромышленного комплекса, роста экспортного потенциала и качества жизни населения. В отличие от концепций Smart Village и Digital Village, ориентированных на телекоммуникационную инфраструктуру и отдельные сервисы, предложенная модель рассматривает цифровую трансформацию как основу новой экономики сельских территорий. Предложена шестиуровневая архитектура, связывающая цифровую инфраструктуру, данные, интеллектуальную аналитику, платформы, новые виды экономической деятельности и результаты устойчивого развития восходящим потоком данных и нисходящим контуром управления.

©Ш.Ж. Сеилов, Б.Ш. Журсинбек, М.У. Байдельдинов, А.К. Майданов, А.А. Сейлов, 2026