ӘОЖ 004.8

ГАЗ ҚҰБЫРЫ SCADA БАЙЛАНЫС АРНАЛАРЫ ТРАФИГІНДЕГІ КИБЕРШАБУЫЛДАРДЫ ГИБРИДТІК АНЫҚТАУ: АШЫҚ ICS ДЕРЕКТЕРІ НЕГІЗІНДЕГІ ЭКСПЕРИМЕНТТІК ЗЕРТТЕУ

Әбсаматов Әділет Әбсаматұлы
Ақпараттық технологиялар кафедрасының магистранты,
7М06139-Ақпараттық жүйелер, Қ. Құлажанов атындағы Қазақ технология және бизнес университеті, Астана қ., Қазақстан

Ғылыми жетекші: Касекеева А.Б., доцент м.а., PhD

Газ тасымалдау инфрақұрылымының тұрақты жұмысы далалық құрылғылар, бағдарламаланатын логикалық контроллерлер (PLC), SCADA серверлері және диспетчерлік жүйелер арасындағы сенімді әрі қауіпсіз деректер алмасуға тәуелді. Осы жұмыста UAH New Gas Pipeline ашық деректер жинағы негізінде газ құбыры SCADA трафигіндегі кибершабуылдарды бинарлық анықтау мәселесі зерттеледі. Зерттеуде базалық классификаторлар, сынып теңгерімін ескеретін және шешім шегі бейімделген модельдер, сондай-ақ ережеге негізделген бағалауды, бақыланатын оқытуды (Random Forest) және аномалияны анықтауды (Isolation Forest) біріктіретін гибридтік архитектура салыстырылды. Қарастырылған модельдер ішінде RF_balanced_subsample_tuned конфигурациясы неғұрлым теңгерімді нәтиже көрсетті: Accuracy = 0.7534, Precision = 0.4562, Recall = 0.6654, F1-score = 0.5413, ROC-AUC = 0.8380. Ал ұсынылған Hybrid Model Accuracy = 0.8566 және Precision = 0.9034 мәндеріне қол жеткізгенімен, Recall = 0.3853 деңгейінде қалды. Нәтижелер шабуылдарды мүмкіндігінше толық анықтау қажет болған жағдайда теңгерімсіздікке бейімделген Random Forest моделінің анағұрлым қолайлы екенін, ал гибридтік модель жалған дабылдарды шектеу маңызды болған жағдайда консервативтірек жұмыс профилін қамтамасыз ететінін көрсетеді.

Кілт сөздер: SCADA қауіпсіздігі, ICS, газ құбыры киберқауіпсіздігі, шабуылды анықтау, Random Forest, Isolation Forest, Logistic Regression, Hybrid Model, сынып дисбалансы, ROC-AUC.

Кіріспе

Газ тасымалдау инфрақұрылымы ұлттық энергетикалық жүйенің маңызды құрамдас бөлігі болып табылады. Бұл инфрақұрылымның тұрақты жұмысы сенсорлар, PLC, SCADA серверлері, операторлық жұмыс станциялары және диспетчерлік орталықтар арасындағы үздіксіз деректер алмасуға тәуелді. Байланыс арналары телеметрияны жеткізу, басқару командаларын беру, күй параметрлерін синхрондау, авариялық сигналдарды тарату және техникалық қызмет көрсету үшін қашықтан қолжетімділікті қамтамасыз етеді. Сондықтан мұндай арналардың бұзылуы тек ақпараттық қауіпсіздік инцидентіне ғана емес, технологиялық процестің тоқтауына, жабдықтың зақымдалуына және өндірістік тәуекелдердің артуына алып келуі мүмкін [1], [2].

Өнеркәсіптік басқару жүйелері (Industrial Control Systems, ICS) дәстүрлі IT-инфрақұрылымнан бірқатар ерекшеліктерімен ажыратылады. Оларға нақты уақыт режимінде жұмыс істеу талаптары, жабдықтың тоқтап қалуына жол бермеу қажеттілігі, ескірген немесе аз қорғалған хаттамаларға тәуелділік, сондай-ақ кибер және физикалық процестер арасындағы тығыз байланыс жатады [3], [4]. Осы ерекшеліктер классикалық корпоративтік қауіпсіздік шешімдерін ICS ортасына тікелей көшіруді қиындатады. Соның салдарынан жалған командаларды енгізу, телеметриялық деректерді бұрмалау, replay-шабуылдар және қызмет көрсетуден бас тарту сияқты қауіптер технологиялық процестердің бұзылуына әкелуі мүмкін [2], [4].

Stuxnet (2010) және Украина энергетикалық инфрақұрылымына жасалған шабуыл (2015) сияқты кеңінен белгілі инциденттер ICS/SCADA қауіпсіздігі мәселесінің практикалық маңызын айқын көрсетті [5]. Осыған байланысты соңғы жылдары шабуылдарды дер кезінде анықтауға арналған интеллектуалды әдістерге, әсіресе машиналық оқытуға негізделген тәсілдерге қызығушылық артты.

Осы жұмыстың мақсаты – ашық ICS деректері негізінде газ құбыры SCADA байланыс арналары трафигіндегі кибершабуылдарды бинарлық анықтауға арналған модельдерді салыстырмалы бағалау. Осы мақсатқа сәйкес келесі міндеттер қойылды:

– Базалық бинарлық классификаторлардың нәтижелерін бағалау;

– Сынып теңгерімі мен шек реттеуінің анықтау сапасына ықпалын зерттеу;

– Ережеге негізделген бағалауды, бақыланатын оқытуды және аномалия детекторын біріктіретін гибридтік модельді ұсыну және бағалау.

Әдебиетке шолу

Өнеркәсіптік басқару жүйелеріндегі киберқауіпсіздік мәселелері соңғы онжылдықта белсенді зерттеліп келеді. Morris және Gao өндірістік хаттамаларға негізделген трафик деректерінің ашық жинақтары шабуылды анықтау жүйелерін зерттеуде аса маңызды екенін көрсетті [6]. Олардың жұмыстары

Modbus негізіндегі зертханалық стендтерден алынған деректерді кейінгі IDS зерттеулеріне пайдалануға мүмкіндік берді.

Pinto және т.б. Машиналық оқытуға негізделген IDS шешімдерінің критикалық инфрақұрылымды қорғаудағы рөлін жан-жақты талдап, signature-based және anomaly-based тәсілдердің артықшылықтары мен шектеулерін сипаттады [7]. Bhamare және т.б. ICS ортасының ерекшеліктері ретінде ескірген бағдарламалық жасақтама, нақты уақыт талаптары, жабдықтың тоқтамауы және шектеулі деректер көлемі сияқты факторларды атап көрсетті [3]. Ал Knowles және т.б. Өнеркәсіптік басқару жүйелерінде қауіпсіздікті басқару тәсілдері жалпы ақпараттық қауіпсіздік тәжірибелерінен ерекшеленетінін және ICS ортасында тәуекелдерді бағалау мен қорғаныс шараларын бейімдеу қажеттігін көрсетеді [8].

Random Forest ICS трафигіндегі шабуылдарды анықтауда жиі қолданылатын және тәжірибеде жақсы нәтиже беретін алгоритмдердің бірі болып саналады. Mubarak және т.б. Әртүрлі алгоритмдерді салыстыра отырып, Random Forest-тің бірқатар ICS деректер жинақтарында сенімді нәтиже көрсететінін атап өтті [9]. Duque Anton және т.б. Өндірістік Modbus/TCP деректер жиынында машиналық оқыту алгоритмдерін бағалай отырып, F1-score және ROC-AUC сияқты метрикалардың сынып дисбалансы жағдайында анағұрлым мазмұнды екенін көрсетті [10].

Терең оқытуға негізделген гибридтік тәсілдер де зерттелуде. Adibhatla және т.б. Стектелген терең архитектураларды пайдаланып, өнеркәсіптік жүйелерге шабуылдарды анықтауда жоғарырақ Recall алуға болатынын көрсетті [11]. Сонымен қатар, оқытусыз аномалияны анықтау тәсілдері, соның ішінде Isolation Forest, белгісіз немесе сирек шабуылдарды табу үшін қолданылып келеді, алайда олар supervised модельдермен салыстырғанда жиі төменірек толықтық береді [12].

SCADA/ICS деректер жинақтарында қалыпты бақылаулар саны шабуылдарға қарағанда басым болуы жиі кездеседі. Elsheikh және т.б. Мұндай дисбаланс IDS жұмысына кері әсер ететінін және теңгерімдеу тәсілдері Recall пен F1-score-ды арттыра алатынын көрсетті [13]. Ensemble және oversampling негізіндегі тәсілдер де дисбалансты деректерде пайдалы болуы мүмкін [14].

Гибридтік архитектуралар дәстүрлі supervised және anomaly-based тәсілдердің шектеулерін азайту мақсатында ұсынылуда. Kabore және т.б. SCADA желілері үшін гибридтік аномалия анықтау жүйелерінің әлеуетін көрсетті [15]. Сонымен қатар, ансамбльдік және soft-voting негізіндегі модельдер жалған оң нәтижелерді азайтуға және тұрақтылықты арттыруға мүмкіндік беретіні көрсетілген [16].

ICS қауіпсіздігіне арналған ашық деректер жинақтары саны жағынан шектеулі, ал олардың едәуір бөлігі нақты өндірістік ортаның барлық

ерекшеліктерін толық қамтымайды [18]. UAH New Gas Pipeline деректер жинағы газ құбыры жүйесінің Modbus RTU трафигін қамтитын және шабуылды анықтау зерттеулерінде кеңінен қолданылатын жинақтардың бірі болып табылады [17]. Бұл жинақта желілік хаттама өрістерімен қатар технологиялық процесс параметрлері де қамтылған, сондықтан ол SCADA қауіпсіздігіне арналған модельдерді салыстыру үшін ыңғайлы эксперименттік орта береді.

Зерттеу әдістемесі

Осы зерттеудің эксперименттік бөлігі ашық қолжетімді UAH New Gas Pipeline деректер жинағы негізінде жүргізілді [17]. Аталған деректер жиыны өнеркәсіптік басқару жүйелеріндегі, соның ішінде газ құбыры инфрақұрылымындағы SCADA байланыс трафигін талдауға арналған және шабуылды анықтау модельдерін салыстырмалы бағалау үшін жиі қолданылатын ашық ICS деректер көздерінің бірі болып табылады. ARFF форматындағы файл жүктелгеннен кейін деректер жиыны 274 628 жазбадан және 20 атрибуттан тұратыны анықталды. Зерттеу барысында шабуылдарды анықтау есебі бинарлық жіктеу түрінде қарастырылды, яғни мақсатты белгі ретінде `binary result` пайдаланылып, барлық бақылаулар екі класқа бөлінді: 0 – қалыпты күй (Normal) және 1 – шабуыл немесе аномалиялық күй (Attack)**.**

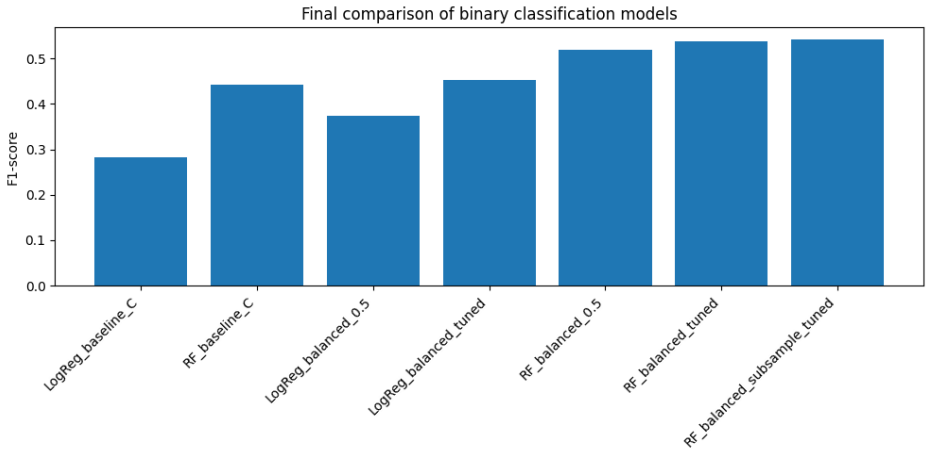
Сыныптардың үлестірімі 1-кестеде көрсетілген. Кестеден көрінгендей, деректер жиыны теңгерімсіз сипатқа ие: қалыпты бақылаулар үлесі шабуыл жазбаларынан едәуір жоғары. Мұндай дисбаланс жағдайында модель тек басым класты көбірек болжау арқылы да жоғары жалпы дәлдікке қол жеткізуі мүмкін, бірақ шабуылдарды нақты анықтау толықтығы төмен болып қалады. Сондықтан осы зерттеуде модельдердің сапасын бағалау кезінде Accuracy көрсеткішімен шектелмей, ең алдымен Recall**,** F1-score және ROC-AUC метрикаларына басымдық берілді. Мұндай тәсілдің негізділігі 1-суретте келтірілген F1-score бойынша салыстырмалы көріністен де байқалады: модельдердің жалпы дәлдігі ұқсас болғанымен, олардың шабуылдарды анықтау сапасы айтарлықтай өзгеше болуы мүмкін.

1-кесте. Бинарлық анықтау міндетіндегі сынып үлестірімі

Класс	Жазбалар саны	Үлес, %
Норма (Normal)	214 580	78.13
Шабуыл (Attack)	60 048	21.87

Модельдердің белгілерге сезімталдығын және ықтимал қызметтік немесе жанама утечкаға жақын атрибуттардың әсерін бағалау үшін белгілердің үш конфигурациясы қарастырылды. А нұсқасы мақсатты бағандарды қоспағандағы

барлық 17 белгіні қамтыды және толық ақпарат жағдайындағы модель жұмысын сипаттады. В нұсқасында time белгісі алынып тасталып, 16 белгі қалдырылды, өйткені уақыт айнымалысы кейбір жағдайларда эксперименттік сценарийдің ұйымдастырылу ерекшеліктерімен жанама байланыста болуы мүмкін. С нұсқасындаtime, command response және crc rate белгілері алынып тасталып, 14 белгі пайдаланылды. Бұл конфигурация ең консервативті нұсқа ретінде қарастырылды, себебі онда ықтимал қызметтік сипаттағы атрибуттардың әсері барынша азайтылды. Мұндай салыстыру модель сапасының тек артефактілік ақпаратқа емес, шын мәніндегі желілік және процестік заңдылықтарға сүйенетінін бағалау үшін қажет болды.



1-сурет. Бинарлық классификация модельдерінің F1-score көрсеткіші - бойынша салыстырмасы

Эксперименттер алдында деректер алдын ала өңделді. Жетіспейтін мәндер медианалық толтыру тәсілі арқылы өңделді, өйткені бұл әдіс шеткі мәндерге төзімді және өнеркәсіптік өлшеулердегі табиғи ауытқулар жағдайында орнықтырақ нәтиже береді.

Сандық белгілер standardscaler көмегімен стандартталды. Бұл әсіресе логистикалық регрессия сияқты ауқымға сезімтал модельдер үшін маңызды, себебі белгілерді біртекті шкалаға келтіру параметрлерді неғұрлым тұрақты бағалауға мүмкіндік береді. Бұдан кейін деректер стратификацияны сақтай отырып, 80/20 қатынасында оқыту және тестілеу жиындарына бөлінді. Стратификация қолдану арқылы екі жиында да қалыпты және шабуыл кластарының арақатынасы жалпы деректер жиынының құрылымына жақын күйде сақталды.

Бастапқы салыстырмалы бағалау үшін екі supervised классификатор таңдалды: Logistic Regression және Random Forest. Логистикалық регрессия интерпретациясы салыстырмалы түрде қарапайым және сызықтық

тәуелділіктерді модельдеуге арналған базалық әдіс ретінде қолданылды. Ал Random Forest белгілер арасындағы сызықтық емес байланыстарды және күрделі өзара әрекеттесулерді анықтауға қабілетті ансамбльдік тәсіл ретінде қарастырылды. Осы екі модельдің нәтижелерін салыстыру өнеркәсіптік трафикті талдауда қарапайым сызықтық тәсіл мен күрделірек ансамбльдік тәсілдің айырмашылығын бағалауға мүмкіндік берді. Бинарлық модельдердің F1-score көрсеткіші бойынша салыстырмасы 1-суретте берілген, ал шабуылдарды анықтау толықтығы, яғни Recall көрсеткіші бойынша айырмашылықтар 2-суретте көрсетілген.

Деректер жиынының теңгерімсіз сипатын ескере отырып, зерттеуде модельдерді жетілдірудің екі негізгі бағыты қарастырылды. Біріншісі – кластар салмақтарын түзету, яғни class_weight="balanced" және class_weight= "balanced_subsample" параметрлерін қолдану болды. Бұл тәсіл модельді кіші класты, яғни шабуылдарды, көбірек ескеруге ынталандырады. Екіншісі – шешім қабылдау шегін бейімдеу, яғни әдепкі 0.5шегінің орнына Precision мен Recall арасындағы анағұрлым қолайлы тепе-теңдікті қамтамасыз ететін мәнді таңдау болды. Мұндай threshold tuning әсіресе өнеркәсіптік қауіпсіздік міндеттерінде маңызды, өйткені кейбір жағдайларда шабуылдарды барынша толық анықтау басым болса, ал басқа жағдайларда жалған дабылдарды азайту маңызды болуы мүмкін. Бұл тәсілдердің тиімділігі 1-сурет пен 2-суретте анық көрінеді: class balancing және threshold tuning қолданылған модельдерде F1-score мен Recall көрсеткіштері baseline нұсқаларымен салыстырғанда едәуір жақсарған.

Зерттеудің келесі кезеңінде бірнеше түрлі ақпарат көздерін біріктіруге бағытталған Hybrid Model қарастырылды. Ұсынылған гибридтік архитектура үш негізгі компоненттен тұрды. Бірінші компонент – rule-based модуль, ол күдікті хаттамалық және процестік үлгілер негізінде нормаланған ереже баллын есептейді. Екінші компонент – supervised модуль, ол Random Forest моделі арқылы есептелген шабуыл ықтималдығын береді. Үшінші компонент – anomaly модуль, ол тек қалыпты жазбаларда оқытылған Isolation Forest моделі арқылы алынған аномалия баллын береді. Осындай құрылым сараптамалық ережелерді, бақыланатын оқытудың дискриминациялық қабілетін және аномалияны анықтау тәсілінің жаңа немесе сирек кездесетін ауытқуларға сезімталдығын бір жүйеге біріктіруге бағытталды.

Жалпы гибридтік балл келесі түрде есептелді:

$$S_{\text{hybrid}} = w_1 S_{\text{rules}} + w_2 P_{\text{RF}} + w_3 S_{\text{anomaly}} \quad (1)$$

мұнда: S_{rules}— ережелерге негізделген модульдің нормаланған балы,
P_{RF} — Random Forest моделінің шабуыл ықтималдығы,

S_{anomaly} — Isolation Forest аномалия балы,
w_1, w_2, w_3 — салмақ коэффициенттері.
Салмақ коэффициенттері:

$$w_1 + w_2 + w_3 = 1. \quad (2)$$

шартын қанағаттандырды.

Осы жұмыста эмпирикалық түрде w_1=0.20, w_2=0.55, w_3=0.25 мәндері қолданылды. Бұл баптауда негізгі үлес supervised компонентке берілді, себебі алдынғы тәжірибелер Random Forest моделінің ең тұрақты нәтижелер көрсеткенін көрсетті. Гибридтік балл есептелгеннен кейін финалдық шешім осы мәнді шекті көрсеткішпен салыстыру арқылы қабылданды: егер $S_{\text{hybrid}} \geq$ болса, объект шабуыл ретінде, ал $S_{\text{hybrid}} <$ болса, қалыпты күй ретінде белгіленді.

Модельдердің сапасын бағалау үшін Accuracy, Precision, Recall, F1-score және ROC-AUC метрикалары пайдаланылды. Accuracy барлық бақылаулардың ішінде дұрыс жіктелгендерінің үлесін сипаттайды, алайда теңгерімсіз деректер жағдайында бұл көрсеткіш жалғыз өзі жеткіліксіз. Precision шабуыл ретінде белгіленген объектілердің ішінде шынымен шабуыл болғандарының үлесін сипаттайды және жалған оң нәтижелердің деңгейін жанама түрде көрсетеді. Recall нақты шабуылдардың қаншасы дұрыс табылғанын сипаттайды және өнеркәсіптік қауіпсіздік міндеттері үшін ерекше маңызды болып табылады, өйткені нақты шабуылды өткізіп жіберу жалған дабылдан әлдеқайда ауыр салдарға әкелуі мүмкін. Precision мен Recall арасындағы теңгерімді сипаттау үшін F1-score қолданылды:

$$\mathrm{F}_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Ал ROC-AUC метрикасы модельдің әртүрлі шекті мәндер жағдайында шабуыл мен қалыпты бақылауларды бір-бірінен ажырату қабілетін сипаттайды. Модельдердің осы метрикалар бойынша салыстырмалы нәтижелері кейінгі бөлімдерде кестелер мен суреттер арқылы талданады. Атап айтқанда, 1-суретте бинарлық классификация модельдерінің F1-score көрсеткіші бойынша салыстырмасы берілсе, 2-суретте олардың шабуылдарды анықтау толықтығы, яғни Recall бойынша айырмашылықтары көрсетілген. Ал гибридтік модель мен ең тиімді жеке модельдің негізгі сапа метрикалары бойынша салыстырмасы 3-суретте келтірілген.

Осылайша, зерттеу әдістемесі бірнеше модельді бірдей деректер құрылымында, бірдей алдын ала өңдеу және бағалау қағидаттары негізінде салыстыруға мүмкіндік берді. Мұндай тәсіл газ құбыры SCADA байланыс

арналары трафигіндегі кибершабуылдарды анықтау үшін неғұрлым қолайлы модельді таңдауға, сынып теңгерімсіздігінің әсерін бағалауға және гибридтік тәсілдің практикалық орны туралы неғұрлым негізді қорытынды жасауға жағдай жасады.

Тәжірибе нәтижелері және талқылау

Baseline нәтижелері

Белгілердің C нұсқасы бойынша алынған baseline нәтижелері 2-кестеде келтірілген. Бұл конфигурацияда `time`, `command response` және `crc rate` белгілері алынып тасталғандықтан, ол ең консервативті және методологиялық тұрғыдан анағұрлым қатаң нұсқа ретінде қарастырылды. 2-кестеден көрінгендей, Random Forest моделі Logistic Regression моделіне қарағанда жоғарырақ нәтиже көрсетті: оның Accuracy, F1-score және ROC-AUC көрсеткіштері айтарлықтай жоғары болды. Атап айтқанда, Random Forest үшін F1-score = 0.4427**,** ал Logistic Regression үшін F1-score = 0.2816 болды; сонымен қатар ROC-AUC көрсеткіші тиісінше 0.8312 және 0.6765 мәндерін көрсетті. Бұл нәтиже Random Forest моделінің шабуыл мен қалыпты күй арасындағы сызықтық емес заңдылықтарды жақсырақ ұстайтынын көрсетеді.

2-кесте. C нұсқасындағы baseline бинарлық классификация нәтижелері

Модель	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.8123	0.8636	0.1682	0.2816	0.6765
Random Forest	0.8435	1.0000	0.2843	0.4427	0.8312

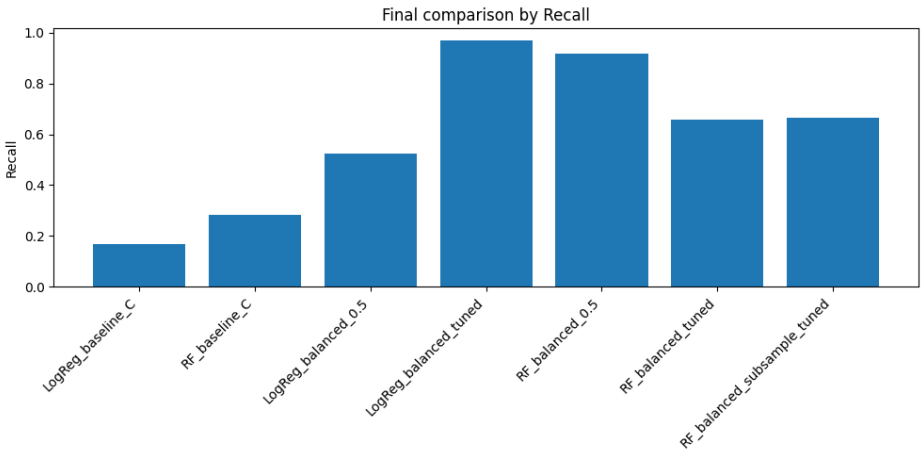
Сонымен бірге 2-кестедегі нәтижелер екі модельдің де шабуылдарды анықтау толықтығы тұрғысынан шектеулі екенін көрсетеді. Logistic Regression моделінің Recall = 0.1682**,** ал Random Forest моделінің Recall = 0.2843 болуы baseline конфигурациядағы екі классификатордың да шабуылдарды анықтауда тым сақ жұмыс істейтінін білдіреді. Басқаша айтқанда, модельдер шабуыл деп белгілеген бақылаулардың басым бөлігі расында да шабуыл болғанымен, нақты шабуылдардың едәуір бөлігі анықталмай қалады. Бұл ерекшелік әсіресе өнеркәсіптік қауіпсіздік міндеттері үшін маңызды, өйткені нақты шабуылды өткізіп алу жалған дабылдан әлдеқайда қауіпті. Қарастырылған модельдердің шабуылдарды анықтау толықтығы, яғни Recall көрсеткіші бойынша салыстырмасы 2-суретте көрсетілген. 2-суреттен baseline модельдердің шабуылдарды табу қабілеті шектеулі екені анық байқалады.

Белгілер сезімталдығын талдау

Белгілердің әртүрлі конфигурациялары бойынша модельдердің жұмыс нәтижелері 3-кестеде көрсетілген. Бұл талдау модель сапасының нақты белгілерге, әсіресе қызметтік немесе жанама утечкаға ұқсас атрибуттарға тәуелділігін бағалау мақсатында жүргізілді. Кестеден көрінгендей, `time`

белгісін алып тастау модель сапасына айтарлықтай әсер етпеді: A және B нұсқаларындағы нәтижелер өте жақын болды. Мысалы, Random Forest үшін ROC-AUC мәні 0.8758-ден 0.8486-ға ғана төмендеді, ал Recall көрсеткіші 0.2983 пен 0.2963 аралығында қалды. Бұл уақыт белгісінің модель үшін шешуші рөл атқармайтынын білдіреді.

Ал `command response` және `crc rate` белгілерін қосымша алып тастау кейбір метрикалардың төмендеуіне алып келді. Дегенмен бұл нашарлау апатты сипатта болған жоқ. Мысалы, Random Forest моделінің ROC-AUC көрсеткіші 0.8312 деңгейінде сақталды, бұл модельдің қызметтік сипаттағы атрибуттарсыз да шабуыл мен қалыпты трафикті жеткілікті деңгейде ажырата алатынын көрсетеді. Сол сияқты Logistic Regression үшін де кейбір метрикалар өзгергенімен, жалпы қорытынды өзгерген жоқ: сызықтық модель шабуылдарды анықтауда әлсіздеу болып қалды.



2-сурет. Бинарлық модельдердің шабуылдарды анықтау толықтығы (Recall) бойынша салыстырмасы

Осы нәтижелер модель жұмысының тек жасанды немесе қызметтік артефактілерге толық тәуелді емес екенін көрсетеді. Бұл өз кезегінде зерттеу нәтижелерінің белгілі бір деңгейде методологиялық орнықтылығын қолдайды. Яғни алынған жіктеу сапасы тек эксперименттік сценарийдің қосымша белгілеріне ғана емес, желілік және процестік деректердегі шынайырақ заңдылықтарға да негізделген деп пайымдауға болады.

3-кесте. A, B және C нұсқалары бойынша модель жұмысын салыстыру

Нұсқа	Модель	Белгілер саны	Accuracy	Precision	Recall	F1	ROC-AUC
A	Logistic Regression	17	0.8152	0.9844	0.1572	0.2711	0.7104

A	Random Forest	17	0.8465	0.9994	0.2983	0.4595	0.8758
B	Logistic Regression	16	0.8151	0.9843	0.1570	0.2709	0.7097
B	Random Forest	16	0.8461	1.0000	0.2963	0.4571	0.8486
C	Logistic Regression	14	0.8123	0.8636	0.1682	0.2816	0.6765
C	Random Forest	14	0.8435	1.0000	0.2843	0.4427	0.8312

Теңгерілген және шегі бейімделген модельдер

Сынып дисбалансының әсерін төмендету және шабуылдарды анықтау сапасын жақсарту мақсатында class balancing және threshold tuning тәсілдері қолданылған модельдердің нәтижелері 4-кестеде келтірілген. 4-кестеден көрінгендей, қарастырылған нұсқалар ішінде RF_balanced_subsample_tuned

Моделі неғұрлым теңгерімді нәтиже көрсетті. Бұл модель үшін Accuracy = 0.7534**,** Precision = 0.4562**,** Recall = 0.6654**,** F1-score = 0.5413**,** ROC-AUC = 0.8380 болды.

Егер осы нәтижені baseline Random Forest моделімен салыстырсақ, шабуылдарды анықтау толықтығының айтарлықтай өскені байқалады. Атап айтқанда, Recall көрсеткіші 0.2843-тен 0.6654-ке дейін артты, ал F1-score 0.4427-ден 0.5413-ке дейін жақсарды. Бұл class balancing және шешім шегін бейімдеу тәсілдерінің шабуылды анықтау сапасын едәуір жақсарта алатынын дәлелдейді. Мұндай өзгеріс әсіресе өнеркәсіптік киберқауіпсіздік міндеттері үшін маңызды, өйткені модель тек нақты шабуылдарды көбірек таба бастайды, сонымен бірге жалған дабылдар да толық бақылаудан шықпайды.

4-кестеде көрсетілгендей, `RF_balanced (threshold=0.5)` конфигурациясы

Recall = 0.9158 мәніне дейін жетті, яғни шабуылдарды өте жоғары деңгейде анықтай алды. Алайда бұл жағдайда Precision төмендеп, Accuracy де айтарлықтай кеміді. Мұндай нәтиже шабуылдарды барынша толық табуға бағдарланған, бірақ жалған оң нәтижелері көп модельге тән. Ал `RF_balanced_tuned` және `RF_balanced_subsample_tuned` нұсқалары Precision, Recall және F1-score арасындағы неғұрлым ұтымды тепе-теңдікті қамтамасыз етті. 1-суретте F1-score көрсеткіші бойынша барлық бинарлық модельдердің салыстырмасы берілген, ал 2-суретте олардың Recall бойынша салыстырмасы көрсетілген. Осы суреттерден теңгерімсіздікті ескеру мен шекті мәнді бейімдеу baseline модельдермен салыстырғанда анағұрлым жақсы нәтижелер беретінін анық көруге болады.

4-кесте. Baseline және жетілдірілген бинарлық классификаторлардың салыстырмасы

Модель	Accuracy	Precision	Recall	F1-score	ROC-AUC
RF_balanced_subsample_tuned	0.7534	0.4562	0.6654	0.5413	0.8380
RF_balanced_tuned	0.7518	0.4536	0.6592	0.5374	0.8306
RF_balanced (threshold=0.5)	0.6297	0.3627	0.9158	0.5196	0.8306
Logreg_balanced_tuned	0.4891	0.2958	0.9680	0.4531	0.6735
RF_baseline_C	0.8435	1.0000	0.2843	0.4427	0.8312
Logreg_baseline_C	0.8123	0.8636	0.1682	0.2816	0.6765

Hybrid Model нәтижелері

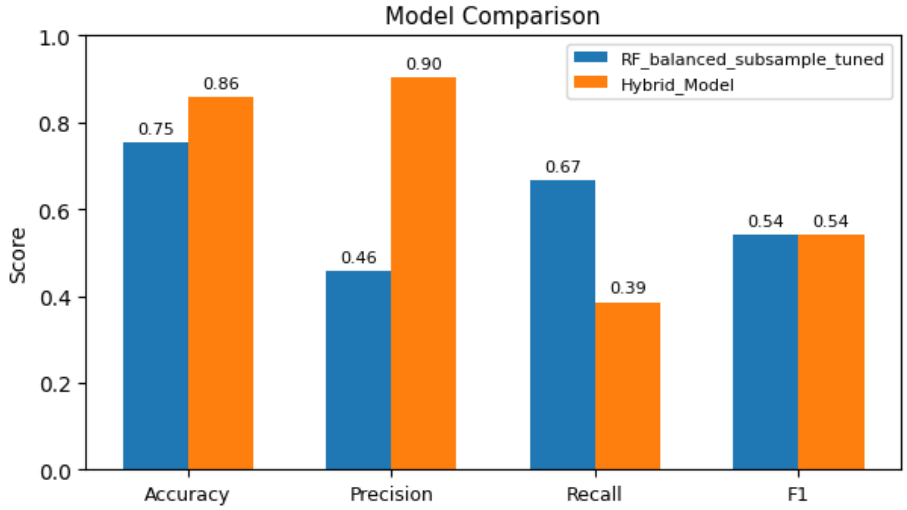
Ұсынылған гибридтік модельдің нәтижелері 5-кестеде және 3-суретте келтірілген. Салыстыру үшін 5-кестеде ең тиімді жеке модель ретінде анықталған `RF_balanced_subsample_tuned` және ұсынылған Hybrid Model қатар берілген. Кестеден көрінгендей, Hybrid Model үшін Accuracy = 0.8566 және Precision = 0.9034 мәндері алынды, бұл сәйкес Random Forest моделінен едәуір жоғары. Сонымен қатар оның F1-score = 0.5402 мәні `RF_balanced_subsample_tuned` моделінің F1-score = 0.5413 нәтижесіне өте жақын болды.

Алайда бұл ұқсастыққа қарамастан, екі модельдің қателік профилі айтарлықтай өзгеше. Hybrid Model жағдайында Recall = 0.3853 болды, яғни ол нақты шабуылдардың анағұрлым аз бөлігін анықтады. Сонымен қатар ROC-AUC көрсеткіші 0.7663 деңгейіне дейін төмендеді, бұл гибридтік модельдің жалпы ранжирлеу қабілеті қарастырылған Random Forest конфигурациясынан әлсіздеу екенін білдіреді. 3-суретте осы айырмашылықтар көрнекі түрде көрсетілген: Hybrid Model жоғары Precision мен Accuracy береді, бірақ Recall және ROC-AUC бойынша ұтылады.

5-кесте. RF_balanced_subsample_tuned және Hybrid Model салыстырмасы

Модель	Accuracy	Precision	Recall	F1-score	ROC-AUC
RF_balanced_subsample_tuned	0.7534	0.4562	0.6654	0.5413	0.8380
Hybrid Model	0.8566	0.9034	0.3853	0.5402	0.7663

Осы нәтижелер Hybrid Model-ді «жақсырақ» модель ретінде емес, басқа операциялық профиль беретін модель ретінде сипаттауға мүмкіндік береді. Егер қолданбалы сценарийде жалған оң нәтижелерді барынша азайту маңызды болса, онда Hybrid Model пайдалы болуы мүмкін. Ал егер басты мақсат – шабуылдарды мүмкіндігінше толық анықтау болса, онда `RF_balanced_subsample_tuned` моделі анағұрлым орынды таңдау болып табылады. Яғни екі модель әртүрлі практикалық талаптарға жауап береді, және олардың қайсысы тиімді екені нақты операциялық контекстке тәуелді.



3-сурет. RF_balanced_subsample_tuned және Hybrid Model модельдерінің негізгі метрикалар бойынша салыстырмасы

Талқылау

Алынған нәтижелер бірнеше маңызды ғылыми тұжырым жасауға мүмкіндік береді. Біріншіден, ICS қауіпсіздігі міндеттерінде Accuracy жалғыз өзі жеткілікті көрсеткіш емес**.** Бұл 5-кесте мен 3-суреттен анық көрінеді: Hybrid Model жоғары Accuracy көрсеткенімен, оның Recall мәні төмен болды. Демек, модельді бағалауда көпметрикалық тәсіл қолдану қажет, әсіресе шабуылды өткізіп алмау маңызды болған жағдайда [7], [10].

Екіншіден, сынып дисбалансы өнеркәсіптік трафикті талдауда шешуші рөл атқарады**.** 4-кесте мен 2-суреттен class balancing және threshold tuning қолдану Recall пен F1-score-ды айтарлықтай жақсартатыны байқалады. Бұл тәжірибелік тұрғыдан тек алгоритмді таңдаумен шектелмей, шешім қабылдау логикасын да бейімдеу қажет екенін көрсетеді [13], [14].

Үшіншіден, белгілер сезімталдығын талдау модельдердің нәтижесі тек қызметтік артефактілерге тірелмейтінін көрсетті**.** 3-кесте деректері бойынша ықтимал утечкаға жақын белгілерді алып тастағаннан кейін де Random Forest моделінің сапасы толық күйремеген. Бұл алынған нәтижелердің белгілі бір деңгейде әдіснамалық жарамдылығын қолдайды.

Төртіншіден, гибридтік тәсіл жалған дабылдарды азайтуға бағытталған шешім ретінде қызығушылық тудырады**,** алайда осы жұмыста қарастырылған конфигурацияда ол шабуылдарды толық анықтау тұрғысынан Random Forest моделінен басым түспеді. Сондықтан гибридтік архитектураны болашақта салмақ коэффициенттерін автоматты түрде баптау, валидациялық деректерді пайдалану немесе мета-модель қосу арқылы жетілдіру орынды.

Зерттеудің шектеулері

Осы жұмыстың нәтижелерін интерпретациялау кезінде бірнеше шектеуді ескеру қажет. Біріншіден, зерттеу бір ғана ашық деректер жинағына негізделді, сондықтан қорытындыларды басқа өнеркәсіптік орталарға толықтай тікелей көшіру шектеулі болуы мүмкін. Екіншіден, модельдер бір train/test бөлу аясында бағаланды, сол себепті алынған нәтижелер белгілі бір дәрежеде қолданылған бөлу схемасына тәуелді болуы ықтимал. Үшіншіден, шешім шегін бейімдеу бөлек сыртқы валидациялық жиынтықсыз жүргізілді, сондықтан бұл баптаулардың тұрақтылығы қосымша тексеруді қажет етеді. Төртіншіден, алынған нәтижелер нақты өндірістік газ тасымалдау инфрақұрылымында апробациядан өткен жоқ**,** яғни бұл зерттеу тікелей енгізуге дайын шешімді емес, ашық ICS деректерінде модельдерді салыстырмалы бағалаудың эксперименттік нәтижелерін көрсетеді.

Осыған қарамастан, жүргізілген тәжірибелер машиналық оқытуға негізделген тәсілдердің газ құбыры SCADA байланыс арналары трафигіндегі кибершабуылдарды анықтауда қолданбалы әлеуеті бар екенін көрсетеді. Әсіресе сынып дисбалансын ескеретін және шешім шегі бейімделген Random Forest моделі осы зерттеу шеңберінде неғұрлым орнықты әрі практикалық тұрғыдан пайдалы нәтиже көрсетті, ал гибридтік модель жалған дабылдарды шектеуге басымдық берілетін сценарийлер үшін қызығушылық тудырады.

Қорытынды

Бұл жұмыста UAH New Gas Pipeline ашық деректер жинағы негізінде газ құбыры SCADA байланыс арналары трафигіндегі кибершабуылдарды бинарлық анықтау міндеті зерттелді. Зерттеу шеңберінде базалық supervised модельдер, сынып теңгерімін ескеретін және шешім шегі бейімделген конфигурациялар, сондай-ақ ережеге негізделген бағалауды, бақыланатын оқытуды және аномалияны анықтауды біріктіретін гибридтік тәсіл салыстырмалы түрде бағаланды.

Алынған нәтижелер бірнеше маңызды қорытынды жасауға мүмкіндік берді. Біріншіден, baseline режиміндегі Logistic Regression және Random Forest модельдері жалпы дәлдік бойынша қанағаттанарлық нәтиже көрсеткенімен, шабуылдарды анықтау толықтығы шектеулі болып қалды. Бұл ICS қауіпсіздігі міндеттерінде Accuracy көрсеткішін жеке қарастыру жеткіліксіз екенін және модельді бағалауда Recall**,** F1-score және ROC-AUC сияқты метрикалардың анағұрлым маңызды екенін көрсетті. Екіншіден, сынып дисбалансын ескеру және шешім шегін бейімдеу Random Forest моделінің нәтижелерін едәуір жақсартты. Қарастырылған нұсқалар ішінде `RF_balanced_subsample_tuned` конфигурациясы шабуылдарды анықтау толықтығы мен жалған дабылдар деңгейі арасындағы неғұрлым қолайлы тепе-теңдікті қамтамасыз етті.

Үшіншіден, белгілер сезімталдығын талдау модель сапасының тек қызметтік сипаттағы атрибуттарға тәуелді еместігін көрсетті. `Time`, `command response` және `crc rate` белгілерін алып тастағаннан кейін де Random Forest моделінің сапасы толық күйремеген, бұл алынған нәтижелердің белгілі бір деңгейде әдіснамалық орнықтылығын қолдайды. Төртіншіден, ұсынылған Hybrid Model жоғары Precision және Accuracy мәндерін қамтамасыз еткенімен, шабуылдарды толық анықтау тұрғысынан `RF_balanced_subsample_tuned` моделінен басым түспеді. Осыған байланысты гибридтік модельді жалған дабылдарды шектеу маңызды болған сценарийлер үшін консервативтірек балама ретінде қарастыруға болады.

Жалпы алғанда, жүргізілген эксперименттер ашық ICS деректері негізінде машиналық оқыту тәсілдерін қолдану газ құбыры SCADA байланыс арналары трафигіндегі кибершабуылдарды анықтауда перспективалы бағыт екенін көрсетті. Осы зерттеу шеңберінде шабуылдарды неғұрлым толық анықтау міндеті үшін теңгерімсіздікке бейімделген және шешім шегі реттелген Random Forest моделі анағұрлым қолайлы нәтиже берді.

Зерттеудің шектеулері ретінде оның бір ғана ашық деректер жинағына негізделгенін, нәтижелердің бір train/test бөлу схемасында алынғанын және нақты өндірістік газ тасымалдау инфрақұрылымында апробациядан өтпегенін атап өткен жөн. Болашақ зерттеу бағыттарына шабуылдарды көпсыныпты жіктеу, гибридтік модельдің салмақ коэффициенттерін автоматты түрде баптау, SMOTE/ADASYN сияқты теңгерімдеу әдістерін пайдалану, cost-sensitive тәсілдерді енгізу және нақты өндірістік ортада валидация жүргізу жатады.

Пайдаланылған әдебиеттер тізімі

1. Morris T. H., Srivastava A., Reaves B., Gao W., Pavurapu K., Reddi R. A control system testbed to validate critical infrastructure protection concepts // *International Journal of Critical Infrastructure Protection*. – 2011. – Vol. 4, No. 2. – P. 88–103. – DOI: 10.1016/j.ijcip.2011.06.005.

2. Gao W., Morris T., Reaves B., Richley D. On SCADA control system command and response injection and intrusion detection // *Proceedings of the ecrime Researchers Summit (ecrime)*. – Dallas, 2010.

3. Bhamare D., Zolanvari M., Erbad A., Jain R., Khan K., Meskin N. Cybersecurity for industrial control systems: A survey // *Computers & Security*. – 2020. – Vol. 89. – Article 101677. – DOI: 10.1016/j.cose.2019.101677.

4. Nankya M., Chataut R., Akl R. Securing industrial control systems: Components, cyber threats, and machine learning-driven defense strategies // *Sensors*. – 2023. – Vol. 23, No. 21. – Article 8840. – DOI: 10.3390/s23218840.

5. US DHS-CISA. Cyber-attack against Ukrainian critical infrastructure [Electronic resource]. – 2021. – ICS Alert IR-ALERT-H-16-056-01. – Access mode: https://www.cisa.gov/uscert/ics/alerts/IR-ALERT-H-16-056-01 (accessed: 16.03.2026).

6. Morris T., Gao W. Industrial control system traffic data sets for intrusion detection research // *Critical Infrastructure Protection VIII* / eds. J. Butts, S. Shenoi. – Berlin: Springer, 2014. – Vol. 441. – P. 65–78. – DOI: 10.1007/978-3-662-45355-1_5.

7. Pinto A., Herrera L.-C., Donoso Y., Gutierrez J. A. Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure // *Sensors*. – 2023. – Vol. 23, No. 5. – Article 2415. – DOI: 10.3390/s23052415.

8. Knowles W., Prince D., Hutchison D., Disso J. F. P., Jones K. A survey of cyber security management in industrial control systems // *International Journal of Critical Infrastructure Protection*. – 2015. – Vol. 9. – P. 52–80.

9. Mubarak S., Habaebi M. H., Islam M. R., Rahman F. D. R., Tahir M. Anomaly detection in ICS datasets with machine learning algorithms // *Computer Systems Science and Engineering*. – 2021. – Vol. 37, No. 1. – P. 33–46.

10. Duque Anton S., Kanoor S., Fraunholz D., Schotten H. D. Evaluation of machine learning-based anomaly detection algorithms on an industrial Modbus/TCP data set // *Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES)*. – Hamburg, 2018.

11. Adibhatla V. Et al. A stacked deep learning approach to cyber-attacks detection in industrial systems: Application to power system and gas pipeline systems // *Cluster Computing*. – 2022. – Vol. 25. – P. 561–578. – DOI: 10.1007/s10586-021-03426-w.

12. Nguyen T. D., Tran K. P., Thomassey S., Hamad M. Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques // *International Journal of Information Management*. – 2021. – Vol. 57. – Article 102282. – DOI: 10.1016/j.ijinfomgt.2020.102282.

13. Elsheikh A., Islam M. R., Suliman F. M. The effect of dataset imbalance on the performance of SCADA intrusion detection systems // *Sensors*. – 2023. – Vol. 23, No. 2. – Article 758. – DOI: 10.3390/s23020758.

14. Guo A., Li M., Pu Z. A robust intrusion detection approach for imbalanced datasets using ADASYN-based ensemble [Electronic resource] // *HAL Open Science*. – 2024. – Access mode: https://hal.science/hal-05041761(accessed: 16.03.2026).

15. Kabore R., Tao H., Sere S., Ouedraogo K. Hybrid deep neural network anomaly detection system for SCADA networks // *Far East Journal of Mathematical Sciences*. – 2021. – Vol. 128, No. 2. – P. 141–191.

16. Khalid N. A., Habaebi M. H., Zolkapli A. A., Yusoff A., Islam M. R. A deep learning/machine learning approach for anomaly-based network intrusion detection // *Frontiers in Computer Science*. – 2025. – DOI: 10.3389/fcomp.2025.1587073.

17. Morris T. H., Thornton Z., Turnipseed I. P. Industrial control system simulation and data logging for intrusion detection system research // *Proceedings of the 7th Annual Southeastern Cyber Security Summit*. – Huntsville, AL, 2015.

18. Conti M., Donadel D., Turrin F. A survey on industrial control system testbeds and datasets for security research // *IEEE Communications Surveys & Tutorials*. – 2021. – Vol. 23, No. 4. – P. 2248–2294. – DOI: 10.1109/COMST.2021.3101843.

ГИБРИДНОЕ ОБНАРУЖЕНИЕ КИБЕРАТАК В ТРАФИКЕ SCADA-КАНАЛОВ СВЯЗИ ГАЗОПРОВОДА: ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НА ОСНОВЕ ОТКРЫТЫХ ДАННЫХ ICS

Абсаматов Адилет Абсаматович
магистрант кафедры информационных технологий,
7М06139 – Информационные системы, Казахский университет технологии и бизнеса имени К. Кулажанова, г.Астана, Казахстан

Научный руководитель: Касекеева А.Б., и.о. Доцента, PhD

Стабильная работа газотранспортной инфраструктуры зависит от надежного и безопасного обмена данными между полевыми устройствами, программируемыми логическими контроллерами (PLC), SCADA-серверами и диспетчерскими системами. В данной работе исследуется задача бинарного обнаружения кибератак в SCADA-трафике газопровода на основе открытого набора данных UAH New Gas Pipeline. В исследовании были сопоставлены базовые классификаторы, модели с учетом дисбаланса классов и адаптацией порога принятия решения, а также гибридная архитектура, объединяющая оценку на основе правил, контролируемое обучение (Random Forest) и обнаружение аномалий (Isolation Forest). Среди рассмотренных моделей конфигурация RF_balanced_subsample_tuned показала наиболее сбалансированный результат: Accuracy = 0.7534, Precision = 0.4562, Recall = 0.6654, F1-score = 0.5413, ROC-AUC = 0.8380. В то же время предложенная Hybrid Model, достигнув Accuracy = 0.8566 и Precision = 0.9034, продемонстрировала Recall = 0.3853. Полученные результаты показывают, что в условиях, когда приоритетом является максимально полное обнаружение атак, более подходящей оказывается модель Random Forest, адаптированная к дисбалансу классов, тогда как гибридная модель обеспечивает более консервативный режим работы в сценариях, где важно ограничить количество ложных срабатываний.

Ключевые слова: безопасность SCADA, ICS, кибербезопасность газопровода, обнаружение атак, Random Forest, Isolation Forest, Logistic Regression, Hybrid Model, дисбаланс классов, ROC-AUC.

HYBRID CYBERATTACK DETECTION IN GAS PIPELINE SCADA COMMUNICATION TRAFFIC: AN EXPERIMENTAL STUDY BASED ON OPEN ICS DATA

Adilet A. Absamatov
Master’s student, Department of Information Technologies,
7M06139 – Information Systems, K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan

Scientific supervisor: A. B. Kasekeeva, Acting Associate Professor, phd

The stable operation of gas transportation infrastructure depends on reliable and secure data exchange between field devices, programmable logic controllers (plcs), SCADA servers, and dispatching systems. This paper investigates the problem of binary cyberattack detection in gas pipeline SCADA traffic using the open UAH New Gas Pipeline dataset. The study compares baseline classifiers, class imbalance-aware and threshold-tuned models, as well as a hybrid architecture that combines rule-based scoring, supervised learning (Random Forest), and anomaly detection (Isolation Forest). Among the evaluated models, the RF_balanced_subsample_tuned configuration demonstrated the most balanced performance, with Accuracy = 0.7534, Precision = 0.4562, Recall = 0.6654, F1-score = 0.5413, and ROC-AUC = 0.8380. At the same time, the proposed Hybrid Model achieved Accuracy = 0.8566 and Precision = 0.9034, but its Recall remained at 0.3853. The results indicate that when the primary objective is to detect attacks as completely as possible, the class imbalance-adapted Random Forest model is more suitable, whereas the hybrid model provides a more conservative operating profile in scenarios where reducing false positives is more important.

Keywords: SCADA security, ICS, gas pipeline cybersecurity, attack detection, Random Forest, Isolation Forest, Logistic Regression, Hybrid Model, class imbalance, ROC-AUC.

REFERENCES

1. Morris T. H., Srivastava A., Reaves B., Gao W., Pavurapu K., Reddi R. A control system testbed to validate critical infrastructure protection concepts // *International Journal of Critical Infrastructure Protection*. – 2011. – Vol. 4, No. 2. – P. 88–103. – DOI: 10.1016/j.ijcip.2011.06.005.
2. Gao W., Morris T., Reaves B., Richley D. On SCADA control system command and response injection and intrusion detection // *Proceedings of the ecrime Researchers Summit (ecrime)*. – Dallas, 2010.
3. Bhamare D., Zolanvari M., Erbad A., Jain R., Khan K., Meskin N. Cybersecurity for industrial control systems: A survey // *Computers & Security*. – 2020. – Vol. 89. – Article 101677. – DOI: 10.1016/j.cose.2019.101677.
4. Nankya M., Chataut R., Akl R. Securing industrial control systems: Components, cyber threats, and machine learning-driven defense strategies // *Sensors*. – 2023. – Vol. 23, No. 21. – Article 8840. – DOI: 10.3390/s23218840.
5. US DHS-CISA. Cyber-attack against Ukrainian critical infrastructure [Electronic resource]. – 2021. – ICS Alert IR-ALERT-H-16-056-01. – Access mode: https://www.cisa.gov/uscert/ics/alerts/IR-ALERT-H-16-056-01 (accessed: 16.03.2026).
6. Morris T., Gao W. Industrial control system traffic data sets for intrusion detection research // *Critical Infrastructure Protection VIII* / eds. J. Butts, S. Shenoi. – Berlin: Springer, 2014. – Vol. 441. – P. 65–78. – DOI: 10.1007/978-3-662-45355-1_5.
7. Pinto A., Herrera L.-C., Donoso Y., Gutierrez J. A. Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure // *Sensors*. – 2023. – Vol. 23, No. 5. – Article 2415. – DOI: 10.3390/s23052415.
8. Knowles W., Prince D., Hutchison D., Disso J. F. P., Jones K. A survey of cyber security management in industrial control systems // *International Journal of Critical Infrastructure Protection*. – 2015. – Vol. 9. – P. 52–80.

9. Mubarak S., Habaebi M. H., Islam M. R., Rahman F. D. R., Tahir M. Anomaly detection in ICS datasets with machine learning algorithms // *Computer Systems Science and Engineering*. – 2021. – Vol. 37, No. 1. – P. 33–46.

10. Duque Anton S., Kanoor S., Fraunholz D., Schotten H. D. Evaluation of machine learning-based anomaly detection algorithms on an industrial Modbus/TCP data set // *Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES)*. – Hamburg, 2018.

11. Adibhatla V. Et al. A stacked deep learning approach to cyber-attacks detection in industrial systems: Application to power system and gas pipeline systems // *Cluster Computing*. – 2022. – Vol. 25. – P. 561–578. – DOI: 10.1007/s10586-021-03426-w.

12. Nguyen T. D., Tran K. P., Thomassey S., Hamad M. Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques // *International Journal of Information Management*. – 2021. – Vol. 57. – Article 102282. – DOI: 10.1016/j.ijinfomgt.2020.102282.

13. Elsheikh A., Islam M. R., Suliman F. M. The effect of dataset imbalance on the performance of SCADA intrusion detection systems // *Sensors*. – 2023. – Vol. 23, No. 2. – Article 758. – DOI: 10.3390/s23020758.

14. Guo A., Li M., Pu Z. A robust intrusion detection approach for imbalanced datasets using ADASYN-based ensemble [Electronic resource] // *HAL Open Science*. – 2024. – Access mode: https://hal.science/hal-05041761(accessed: 16.03.2026).

15. Kabore R., Tao H., Sere S., Ouedraogo K. Hybrid deep neural network anomaly detection system for SCADA networks // *Far East Journal of Mathematical Sciences*. – 2021. – Vol. 128, No. 2. – P. 141–191.

16. Khalid N. A., Habaebi M. H., Zolkapli A. A., Yusoff A., Islam M. R. A deep learning/machine learning approach for anomaly-based network intrusion detection // *Frontiers in Computer Science*. – 2025. – DOI: 10.3389/fcomp.2025.1587073.

17. Morris T. H., Thornton Z., Turnipseed I. P. Industrial control system simulation and data logging for intrusion detection system research // *Proceedings of the 7th Annual Southeastern Cyber Security Summit*. – Huntsville, AL, 2015.

18. Conti M., Donadel D., Turrin F. A survey on industrial control system testbeds and datasets for security research // *IEEE Communications Surveys & Tutorials*. – 2021. – Vol. 23, No. 4. – P. 2248–2294. – DOI: 10.1109/COMST.2021.3101843.