ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ

“ЧЕРНОРИЗЕЦ ХРАБЪР”

ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”

КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

АЛИ МОХАМАД САБРА

ДЪЛБОКО УЧЕНЕ ПРИ МОДЕЛИРАНЕТО НА DDOS АТАКИ В КИБЕРСИГУРНОСТТА

АВТОРЕФЕРАТ

на дисертационен труд за присъждане
на образователна и научна степен „доктор“,
професионално направление 4.6. "Информатика и компютърни науки",
докторска програма „Информационни системи и технологии,
информатика и компютърни науки“

Научен ръководител:
доц. д-р Златогор Минчев

ВАРНА
2026 г.

ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ

“ЧЕРНОРИЗЕЦ ХРАБЪР”

ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”

КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

АЛИ МОХАМАД САБРА

ДЪЛБОКО УЧЕНЕ ПРИ МОДЕЛИРАНЕТО НА DDOS АТАКИ В КИБЕРСИГУРНОСТТА

АВТОРЕФЕРАТ

на дисертационен труд за присъждане
на образователна и научна степен „доктор“,
професионално направление 4.6. "Информатика и компютърни науки",
докторска програма „Информационни системи и технологии,
информатика и компютърни науки“

Научен ръководител:
доц. д-р Златогор Минчев

Рецензенти:
проф. д-р Теодора Бакърджиева
проф. д-р Радослав Йошинов

ВАРНА
2026 г.

Дисертацията е с обем 118 страници. Състои се от въведение, изложение в четири глави, заключение, списък с използвани съкращения, 23 фигури, 6 таблици. Използваната библиография включва 77 литературни източника. По темата на дисертацията са направени шест публикации.

Авторът на дисертацията е докторант в катедра „Компютърни науки“ към факултет “Социални, стопански и компютърни науки” на Варненския свободен университет „Черноризец Храбър“.

Публичната защита ще се проведе на 04.02.2026 г. от 13,00 ч. на открито заседание на научното жури в заседателната зала на ВСУ “Черноризец Храбър“. Материалите по защитата са на разположение в катедра „Компютърни науки“, както и на сайта на университета www.vfu.bg, раздел „Докторанти“.

ГЛАВА 1:

Въведение – Преглед на DDoS атаките на приложно ниво и предизвикателствата при откриването им

1.1 Мотивация

Уеб-базираните приложения, заедно с методите за разпространение на софтуер, като „Software as a Service“ (SAS), увеличиха институционалното използване на интернет, но този преход разкри значителни уязвимости, които сега заплашват интернет услугите чрез разпределени атаки за отказ от услуга (Distributed Denial of Service).

Идентифицирането на DDoS атаки става по-сложно, тъй като обучението на модела се извършва върху нереалистични набори от данни, които водят до високи нива на фалшиво положителни резултати и намаляват точността на откриване. Предишни проучвания са изследвали множество техники за машинно обучение и дълбоко учене чрез различни набори от данни, но изследователите не са обърнали достатъчно внимание на използването на данни от реални уеб сървъри. Това изследване създава оригинален набор от данни от регистрационни файлове за достъп до уеб сървър, който съдържа нормален системен трафик, наред с DDoS трафика, генериран чрез множество инструменти за атака.

1.2 Формулиране на проблема

Редица изследвания оценяват моделите на машинно обучение (ML) и дълбоко учене (DL), използвайки набори от данни, които не съответстват на действителните модели на уеб трафик, което намалява тяхната ефективност. Това изследване решава този проблем, като създава персонализиран набор

от данни от действителни регистрационни файлове за достъп до уеб сървър, който комбинира DDoS трафик от различни инструменти за атака с безобиден трафик. Изследването прилага множество модели за класификация, за да постигне по-добра точност на откриване, като същевременно намалява фалшивите положителни резултати и оценява способността им да откриват киберзаплахи от SQL инжекции и Cross-Site Scripting (XSS). Изследването има за цел да изгради система за откриване, базирана на изкуствен интелект, която открива DDoS атаки в реално време, като същевременно укрепва защитата си срещу развиващи се заплахи.

1.3 Цел и задачи

Целта на изследването включва създаването на усъвършенствана система за откриване на DDoS атаки на приложно ниво чрез модели на машинно обучение и дълбоко учене върху данни от регистрационни файлове за достъп до уеб сървър.

Тема: Предмет на това изследване е откриването на DDoS атаки на приложно ниво.

Обект: Обект на това изследване са алгоритмите за машинно обучение и дълбоко учене за откриване на трафик на DDoS атаки и други кибератаки от регистрационни файлове за достъп до уеб сървъри.

Основна цел: Да се разработи и оцени интелигентна система за откриване на DDoS атаки на приложно ниво, използвайки модели за машинно обучение и дълбоко учене, изпробвани върху реални регистрационни файлове за достъп до уеб сървър.

Цели:

1. Да се изгради реалистичен набор от данни, комбиниращ доброкачествен и злонамерен уеб трафик, генериран от множество DDoS инструменти.
2. Да се обработят предварително и разработят подходящи функции от сървърни файлове за достъп за обучение на модел.
3. Да се оценят експериментално и сравнят алгоритмите за машинно обучение и дълбоко учене за точно откриване на DDoS атаки като същевременно се минимизират фалшивите положителни резултати.
4. Да се тестват обучените модели върху други видове атаки (напр. SQL Injection, XSS), за да се оцени обобщението.
5. Да се оцени интерпретируемостта на модела, използвайки SHAP анализ, и да се предложи рамка за интеграция в реалния свят.

1.4 Изследователски въпроси

Изследователските въпроси, които възникват от това проучване, са:

1. Кои са най-ефективните методи за машинно обучение и дълбоко учене за откриване на DDoS атаки на приложно ниво?
2. Кои модели на машинно обучение и дълбоко учене се представят най-добре при откриване на DDoS атаки от регистрационни файлове на уеб сървъри?
3. Могат ли моделите за откриване на DDoS атаки да се използват за идентифициране на атаки на други уеб приложения с висока точност?

Всички горепосочени въпроси определят целта на настоящия дисертационен труд.

1.5 Изследователски хипотези

Целите и въпросите на изследването формират основата за създаване на множество проверими хипотези, които ще ръководят емпиричното изследване. Предложените изследователски хипотези показват, че използването на усъвършенствани техники за обучение с автентични източници на данни ще подобри способностите за откриване на DDoS атаки и ще позволи на моделите да откриват различни уеб-базирани заплахи.

Обща хипотеза: Моделите за дълбоко учене, изпробвани върху реални данни от регистрационни файлове на уеб сървър, постигат по-добра обобщаваща способност и точност на прогнозиране за откриване на DDoS на приложно ниво, отколкото моделите, обучени единствено върху синтетични или симулирани данни.

Подхипотеза 1: Реалните данни подобряват устойчивостта на модела и намаляват фалшивите положителни резултати в сравнение със синтетичните набори от данни.

Подхипотеза 2: Проверката с помощта на допълнителни атаки (SQL Injection, XSS) потвърждава способността на модела да обобщава, но не представлява валидиране на бъдещи невидими заплахи.

Експерименталният дизайн се основава на тези хипотези, които установяват проверими предположения за оценка на моделите на машинно обучение и дълбоко учене.

1.6 Структура на дисертацията

Тази дисертация е структурирана в логическа последователност от четири глави, всяка от които постепенно се развива към разработването и валидирането на интелигентна система за откриване на DDoS атаки на приложно ниво, използваща методи за машинно обучение и дълбоко учене. Подредбата на главите е проектирана така, че да насочва читателя от основните аспекти на изследователския проблем до окончателните заключения и предложените бъдещи насоки.

Глава 1: **Въведение**, полага основите на изследването, като контекстуализира нарастващото разпространение на киберзаплахите, по-специално DDoS атаките, в по-широката област на киберсигурността. Тази глава представя мотивацията за изследването, формулира специфичния проблем, който се разглежда, и очертава основната цел заедно с оперативните задачи. Тя също така определя централните изследователски въпроси, които са ръководили целия процес на разследване. Тази глава служи не само за обосноваване на необходимостта от изследването, но и за очертаване на неговия обхват и фокус.

Глава 2: **Преглед на литературата**, предоставя задълбочен преглед на съществуващите изследвания и теоретични основи, свързани с DDoS атаките и приложението на машинно обучение и дълбоко учене в системите за откриване на проникванията. Тази глава изследва еволюцията на атаките на приложно ниво, ограниченията на традиционните защитни механизми и нарастващата роля на подходите, базирани на изкуствен интелект, в киберсигурността. Тя включва критични оценки на публично достъпни набори от данни, алгоритмични техники и реални приложения. Прегледът

идентифицира съществени пропуски в настоящите изследвания, особено по отношение на липсата на реални набори от данни и ограничената обобщаемост на съществуващите модели за откриване, проблеми, които тази дисертация се стреми да реши.

Глава 3: **Методология**, описва подробно дизайна на емпиричното изследване и процедурите стъпка по стъпка, следвани по време на проучването. Тази глава представя конструирането на нов набор от данни, базиран на реални регистрационни файлове за достъп до уеб сървър Apache, които включват както легитимен, така и злонамерен трафик, генериран чрез осем различни DDoS инструмента. Тя очертава експерименталната настройка, техниките за предварителна обработка на данни и процесите на инженерство на характеристики, използвани за подготовка на набора от данни за обучение на модела. Освен това, тя представя структурата и конфигурацията на различни модели за машинно обучение и дълбоко учене, включително Random Forest, Support Vector Machine, изкуствени невронни мрежи и мрежи с дълга краткосрочна памет. Главата завършва със сравнителна оценка на производителността на тези модели и фаза на оценка в реални условия, където обучените модели се тестват срещу други форми на уеб-базирани атаки, като SQL injection и Cross-Site Scripting.

Глава 4: **Заключение**, синтезира ключовите открития и приноси на изследването. Главата разглежда техническите изисквания за интеграция на модели, оценява рисковете и ограниченията от внедряването на системата

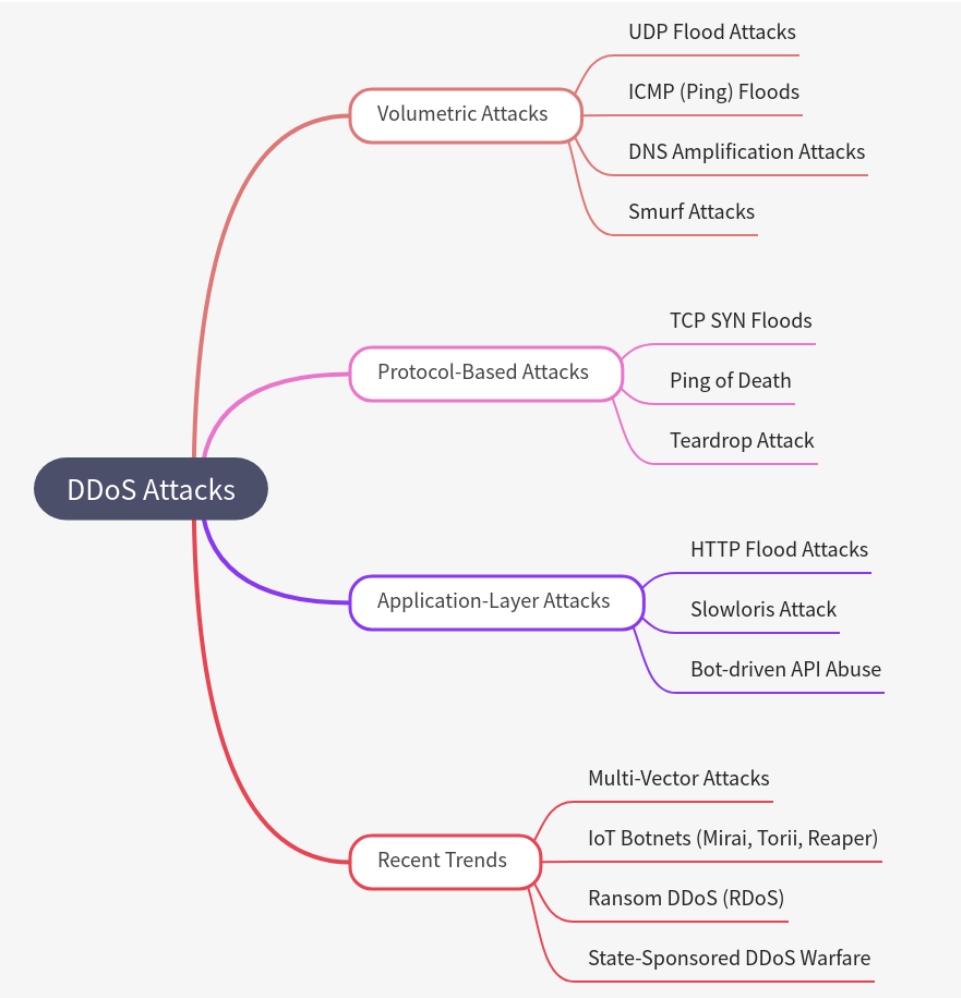
в реални производствени среди и анализира нейната икономическа ефективност, като сравнява потенциалните спестявания с търговските услуги за защита от DDoS атаки, базирани на изкуствен интелект. В нея се разглеждат практическите и теоретичните последици от резултатите и се формулира как изследването се насочва към съществуващите пропуски в литературата. В допълнение към обобщаването на резултатите от оценяваните модели, тази глава предлага няколко насоки за бъдеща работа. Те включват разработването на хибридни архитектури за откриване, интегрирането на механизми за онлайн обучение, използването на федеративно обучение за откриване, запазващо поверителността, и прилагането на анализ на големи данни в реално време в производствени среди.

Глава 2: Преглед на литературата – Подходи за машинно обучение (ML) и дълбоко учене (DL) за откриване на DDoS атаки

2.1 Общ преглед на киберсигурността и DDoS атаките

Сложността на киберзаплахите, заедно с нарастващия им мащаб, се е увеличила през последните двадесет години поради технологичния напредък и разширяващото се използване на свързани с интернет устройства, както и развиващите се киберпрестъпни методи. DDoS атаките представляват сериозна разрушителна кибератака, която причинява значителни финансови и оперативни щети на организации по целия свят.

DDoS атаките могат да бъдат класифицирани в три основни категории въз основа на техните методологии и очакваното въздействие, както е показано на фигурата по-долу (Класификация на DDoS атаките):



Фигура 1: Класификация на DDoS атаките

Съвременните защитни механизми трябва да внедряват многопластови адаптивни подходи, тъй като DDoS атаките продължават да се развиват по сложност.

2.2 Машинно обучение и дълбоко учене в киберсигурността

Технологията на машинното обучение е основен компонент за защита на мрежите чрез откриване на прониквания. Моделите на машинно обучение превъзхождат традиционните системи, базирани на сигнатури, защото те

изучават модели на атаки от исторически данни, за да откриват аномалии в реално време.

2.3 Предишни проучвания на откриването на DDoS атаки на приложно ниво

Таблицата по-долу (Сравнение на литературни обзори) предоставя общ преглед на най-съвременните методи за откриване на DDoS атаки в IoT, софтуерно дефинирани мрежи и интелигентни мрежи. Таблицата предоставя общ преглед на изследователския принос, включително използваните методологии, използваните набори от данни, получените резултати, ограниченията и предложенията за бъдеща работа. Докладите сравняват различни подходи за машинно обучение (ML) и дълбоко учене (DL) за идентифициране на DDoS атаки и техните слабости, включително зависимост от данни, изчислителна сложност и способност за работа в реални среди.

Таблица 1 : Сравнение на литературни обзори

Автор(и)	**Изследовател-ски фокус**	**Методология**	**Използван набор от данни**	**Ключови констатации**	**Ограничения**
Мачака и др. (2021)	Откриване на DDoS атаки в IoT	Модел на оркестрация, базиран на машинно обучение	Данни за трафика на IoT	Висока точност с ANN, SVM	Необходима е оптимизация на класификатора
Хамарше и др. (2023)	Откриване на DDoS атаки в SDN	Сравнение на модели за машинно обучение (RF, SVM, DT, XGBoost)	Симулиран набор от DDoS данни	Изборът на характеристики е от решаващо значение за точността	Зависимост от набор от данни
Накви и др. (2024)	Откриване на DDoS атаки в интелигентни мрежи	DCAE, MLP, XGBoost	CICIDS2017, CICDDoS2019	Разнообразни ефективни модели на машинно обучение	Ограничени тестове в реални условия
Праджапати и др. (2019)	HTTP-базирани DDoS атаки	Различни рамки за машинно обучение и дълбоко учене	Не е посочено	Ефективни рамки като HADEC и D-FACE	Предизвикателства при имитирането на истински HTTP
Канбер и др. (2022)	Многостепенно машинно обучение за DDoS атаки	XGBoost , LightGBM	CICDDoS2019	99,7% точност	Изчислителни разходи
Розей и др. (2019 г.)	Откриване на мрежови прониквания	MLP модел	CICIDS2017	99% точност, нисък брой фалшиво положителни	Ограничен обхват на набора от данни
Алашхаб и др. (2022)	DDoS атаки в облака	Класификация на типовете атаки	Облачна среда	Фокус върху икономически отказ от устойчивост (EDoS)	Уязвимости, специфични за облака

Ксие и др. (2023)	DDoS на приложно ниво	Рекурентни невронни мрежи	CICIDS2017	Ефективен при HTTP flood атаки	Висока изчислителна мощност
Гунаван и др. (2023)	Небалансирано обработване на набори от данни	SMOTE, ADASYN	CICDDoS2019	Подобрена точност при дисбаланс на класовете	Рискове от преобучение
Мустафа и др. (2023)	Обучение за DDoS атаки с враждебна цел	Състезателни невронни мрежи	CICDDoS2019	Подобрено откриване на напреднали атаки	Проблеми с обобщението
Ахмед и др. (2023)	Откриване на DDoS атаки в мрежи	MLP срещу RF, SVM	CICIDS2017	MLP намалява фалшиво положителните	Високо изчислително търсене
Рамола и др. (2023)	Модели на машинно обучение за DDoS	RF, DT, линейна регресия	CICIDS2017	Random Forest с най-висока точност	DT е по-добър в реално време
Заргар и др. (2013 г.)	DDoS защитни механизми	Проучване	баба	Проактивна/реактивна категоризация	Ограничена емпирична валидация
Али и др. (2023)	ML/DL за SDN-базирани DDoS атаки	XGBoost, LightGBM, CNN	SDN набор от данни	CNN превъзхожда другите	Ограничени видове атаки, проучени
Алджухани и др. (2021)	Техники за машинно обучение в мрежите	Под наблюдение и без наблюдение	CICIDS2017, CICDDoS2019	Хибридните модели са по-ефективни	Тежки изчисления
Гао и др. (2023)	Класификация на DDoS атаките	KNN, DT, LR, NB	CICDDoS2019	LR, NB най-добра прецизност	Резултати, специфични за набора от данни
Лий и др. (2011 г.)	DDoS профилиране на приложно ниво	Статистическо мрежово профилиране	Данни за уеб трафика	Ефективен за HTTP DDoS	Ограничено покритие на атаките
Бхайо и др. (2023)	Откриване на DDoS в SD- IoT	Decision Tree, SVM	SDN- IoT набор от данни	Висока степен на откриване в IoT	Ограничена мащабируемост
Джоши и Харибабу (2023)	Модели на машинно обучение в SDN	k-NN, DT, SVM, ANN	Избор на характеристики: mRMR , NCA	Ансамбловото обучение подобрява точността	Изчислителни разходи

Алкахтани и Абдула (2024)	Машинно обучение/дълбоко учене за DDoS класификация	SVM, CNN, анализ на набор от данни	NSL-KDD, CICIDS2017, CICDDoS2019	Анализ на тенденциите в машинното обучение спрямо дълбокото учене	Предизвикателства при еволюцията на наборите от данни

Таблицата по-долу (Сравнение на набори от данни за откриване на DDoS атаки в изследванията на машинното обучение и дълбокото учене) предоставя обобщено сравнение на пет набора от данни, често използвани от изследователите за откриване на DDoS атаки чрез методи на машинно обучение и дълбоко учене. Таблицата представя информация за набора от данни по година на издаване, приблизителен брой записи, издателска институция или организация, видове включени атаки, ключови бележки или контекст на употреба. Оценката на пригодността за изследване зависи от този сравнителен формат, който помага да се определи съвместимостта на набора от данни за различни изследователски нужди, включително диапазон на типа атака, количество данни, съвременност и приложимост в реалния свят.

Таблица 2 : Сравнение на набори от данни за откриване на DDoS атаки в изследванията на машинното обучение и дълбокото учене

Име на набор от данни	**Година**	**Записи**	**Организация**	**Видове атаки**	**Бележки**
CICIDS2017	2017 г.	~2 830 743 потока	Канадски институт за киберсигурност (CIC)	Brute Force FTP/SSH, DoS, DDoS, Heartbleed, Web Attack, Infiltration, Botnet, Port Scan	79 характеристики на потока; обхваща ~5 дни трафик
CICDDoS2019	2019 г.	12 794 627 записа	Канадски институт за киберсигурност (CIC)	Рефлективен и директен DDoS: NTP, DNS, LDAP, NetBIOS, SNMP, SSDP, SYN, UDP, WebDDoS, TFTP, PortScan	Балансиран с ~50% атака/доброкачествен; ~6.3 GB
NSL-KDD	2009 г.	Приблизително 125 973	Университет на Ню Брънзуик	DOS, Probe, U2R, R2L	Остарял, често критикуван за липса на съвременни заплахи
UNSW-NB15	2015 г.	Приблизително 2,5 милиона	Австралийски център за киберсигурност	Fuzzers, Analysis, Backdoors, DoS, Exploits, etc.	Може да липсва динамика на последните заплахи

2.4 Сравнителен преглед на индустриалните системи за откриване на DDoS атаки

Въпреки че академичните изследвания са създали множество модели за машинно обучение и дълбоко учене за откриване на DDoS атаки, също толкова важно е да се оцени как тези подходи се сравняват със съществуващите системи от индустриален клас. Търговски и решения с отворен код като Cloudflare DdoS Protection, AWS Shield, Akamai Kona Site Defender, Google Cloud Armor и Imperva Incapsula използват хибридни архитектури за откриване, комбиниращи филтриране, базирано на правила, ограничаване на скоростта и откриване на поведенчески аномалии чрез бази данни със сигнатури и анализ на потока. Инструменти с отворен код като Snort, Suricata и ModSecurity по подобен начин разчитат на предварително дефинирани сигнатури и евристики, за да идентифицират обемни и протоколни атаки.

2.5 Пропуски в съществуващите изследвания

Въпреки значителното развитие все още съществуват множество важни пропуски в изследванията в областта на използването на техники за машинно обучение за откриване на киберзаплахи. В настоящите изследвания все още липсват надеждни набори от данни от реалния свят, които правилно да представят променящия се характер на кибератаките. Повечето настоящи набори от данни се състоят от изкуствени или старомодни примери, включително NSL-KDD и CICIDS2017, които ограничават универсалното приложение и оперативната стойност на моделите за откриване. Сложността и разнообразието на трафика в реалния

свят надвишават възможностите на настоящите набори от данни, включително UNSW-NB15, според Moustafa & Slay.

Глава 3: Методология – Проектиране на ML/DL рамка и предварителна обработка на уеб логове

3.1 Дизайн на изследването

Това изследване използва емпирична методология на проектиране, която комбинира усъвършенствани техники в машинното обучение и дълбокото учене за откриване на DDoS атаки на приложно ниво. Стратегията зависи както от теоретични концепции, така и от текущи оперативни проблеми в киберсигурността. Изследователският дизайн използва рамката на Междуиндустриалния стандартен процес за извличане на данни (CRISP-DM), която започва с разбиране на бизнеса, последвано от разбиране на данните, след това подготовка на данните, преди да се премине към моделиране, оценка и накрая внедряване.

3.2 Изграждане на набор от данни и анализ на характеристики

Контролирана лабораторна конфигурация хостваше WordPress на Apache. Два интерфейса позволяваха доброкачествени интернет заявки, докато атаките се генерираха вътрешно. В продължение на 48 часа бяха регистрирани 344 054 записа, с ≈62,8% атака и ≈37,2% доброкачествен трафик – в съответствие с целевите кампании. Осем инструмента с отворен код създадоха разнообразни L7 поведения: GoldenEye , HULK, Raven-Storm, Torshammer , CC-Attack, MHDDoS , DDoSX и Eagle Eye.

Цялата дейност е в съответствие с институционалната политика. Записване на използван комбиниран формат на лог файлове със синхронизирани времеви отметки. Не са засегнати системи на трети страни.

3.3 Анализ на регистрационните файлове за достъп до уеб сървъра

Запазени са седем полета: IP адрес, времева маркировка, ред на заявката, статус, размер на отговора, референт, потребителски агент. Временната плътност, концентрацията на крайни точки, аномалиите в отговорите, липсващите референти и повтарящите се потребителски агенти сигнализират за автоматизирано поведение. Прекратяването на HTTPS оставя семантиката на лога непокътната, което позволява еднаква обработка в HTTP версиите.

Текстовите полета носят дискриминационна сила за L7 атаки. По този начин, каналът токенизира заявката, реферера и потребителския агент и нормализира числовите характеристики за съвместно моделиране.

3.4 Канал за предварителна обработка на данни

Липсващите стойности за „размер“ (3,8%) бяха обяснени чрез средна стойност; липсващият „референт“ (35,25%) - чрез токена „Неизвестен референт“. Стратифицираните разделяния запазиха съотношенията на етикетите: 72% влак, 15% тест, 13% валидация. StandardScaler нормализира числовите полета. Експлораторните корелации предполагат силна връзка между размера на заявката и отговора, но кодовете за състояние остават до голяма степен ортогонални, което подпомага допълващите се сигнали.

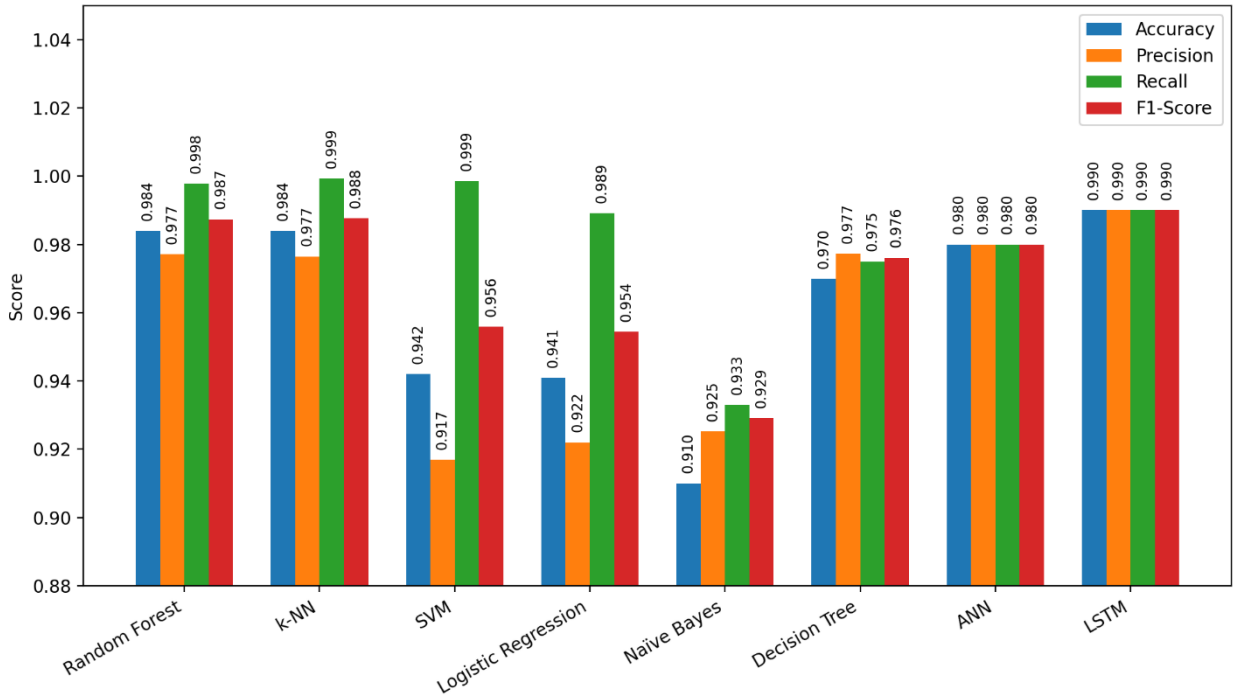
3.5 Структури на модели за машинно обучение

Класически модели: Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, SVM, k- NN. Дълбоки модели: ANN с три скрити слоя и LSTM с вграждания за текст, конволюционни блокове за локални модели и темпорални клетки за моделиране на последователности, последвани от плътна класификация.

Оптимизацията използва търсене в мрежа за машинно обучение и Adam с ранно спиране за дълбоко учене. Претеглянето на класовете и отпадането от обучение са насочени към дисбаланс и пренареждане. Показатели: точност, прецизност, извиканост, F1, подкрепени от матрици на объркване и криви на обучение.

3.6 Обучение и оценка на модела

Фигурата по-долу показва стълбовидна диаграма, сравняваща производителността на осем модела за машинно обучение, използвайки четири показатели за оценка: Точност, Прецизност, Припомняне и F1-оценка. Както е показано, LSTM моделът постига най-високите резултати по всички показатели, докато Naïve Bayes регистрира най-ниската обща производителност.



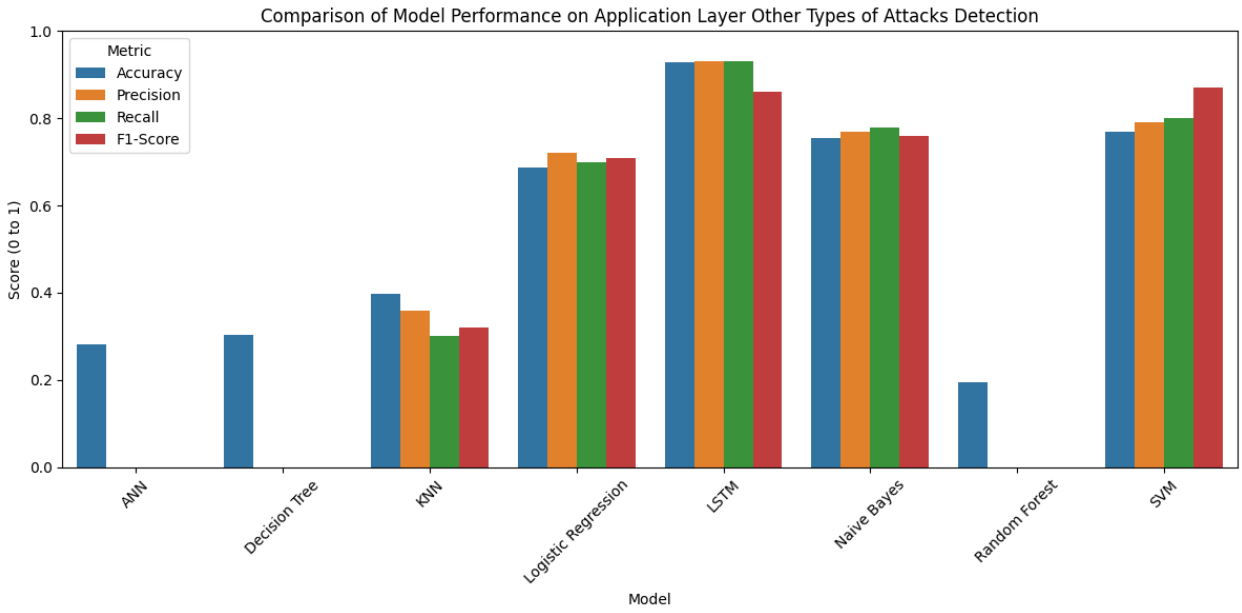
Фигура 2 : Сравнение на производителността на осем модела за машинно обучение

3.7 Проверка и обобщение на модела

Модели, обучени върху DDoS лог файлове, бяха използвани повторно за класифициране на други заплахи на приложно ниво, като например SQL injection и XSS, записани в лог файловете за достъп. LSTM запази висока точност, докато класическите модели се влошиха, разкривайки свръхадаптиране към специфични за DDoS сигнатури.

Фигурата по-долу (Сравнение на производителността на модела при откриване на други видове атаки на приложно ниво) представя сравнителна визуализация на показателите за производителност, като Точност, Прецизност, Припомняне и F1-оценка за осем модела за машинно обучение и дълбоко учене, оценени върху набор от данни, съдържащ други видове

атаки на приложно ниво. Графиката подчертава превъзходната способност за обобщаване на моделите LSTM и SVM, докато традиционните модели, като Decision Tree, Random Forest и ANN, не успяха да идентифицират правилно доброкачествени извадки, което доведе до нулеви резултати за Прецизност, Припомняне и F1-оценка.



Фигура 3 : Сравнение на производителността на модела на приложно ниво при откриване на други видове атаки

3.8 Интерпретируемост на LSTM модела с помощта на SHAP

Откриването на DDoS атаки на приложно ниво в съвременните системи за киберсигурност все повече зависи от моделите за дълбоко учене. LSTM моделът постигна отлични резултати в това изследване, като достигна 99,0% точност при откриване на DDoS атаки и 92,89% точност при откриване на други видове атаки на приложно ниво, като SQL injection и атаки чрез Cross

Site Scripting. Резултатите показват колко добре могат да се представят последователните модели за дълбоко учене при анализ на регистрационни файлове за достъп до уеб сървъра.

Полето на обяснимия изкуствен интелект (XAI) въведе няколко метода за обяснение на модели, за да разреши проблемите с интерпретируемостта, с които се сблъскват моделите за дълбоко учене, включително LSTM.

Приложихме SHAP към нашата обучена LSTM мрежа чрез KernelExplainer, който е моделно-агностичен SHAP вариант, който може да се приложи към всеки модел за машинно обучение, включително модели за дълбоко учене, внедрени в TensorFlow и Keras.

Чрез добавяне на интерпретируемост, базирана на SHAP, към нашата рамка за откриване на LSTM, ние установяваме връзка между прецизността на модела и човешкото разбиране. Общността по сигурност развива повишено доверие чрез този метод, което също така позволява:

- Анализ на първопричината: Подходът позволява на анализаторите да определят причината за откриване на атака, която стои зад маркирана заявка.

- Усъвършенстване на характеристиките: SHAP показва кои характеристики допринасят най-много за фалшиво положителни или истински положителни резултати, което помага за усъвършенстване на характеристиките.

- Одитируеми канали за решения: Моделът предоставя одитируеми канали за решения, които позволяват валидиране и оспорване чрез визуални доказателства.

3.9 Методологично заключение

Резултатите от това проучване съвпадат с предишни изследвания, тъй като точността на откриване достига между 98% и 99% при използване на алгоритми LSTM, ANN и Random Forest. Резултатите съответстват на най-високите нива на производителност на моделите от изследванията на Xie et al. (2023) и Ahmed et al. (2023). Основният принос на това изследване се състои в използването на съвременен набор от данни, базиран на реални регистрационни файлове на уеб сървъри, който представя действителни модели на трафик вместо синтетични или остарели бенчмаркове.

Резултатите в таблицата по-долу показват, че третирането на записите в лога като отделни събития в предишни методи е създало разлика в производителността, като същевременно демонстрира, че методите, които са наясно с последователността, са жизненоважни за откриване на атаки на приложно ниво.

Таблица 3 : Сравнение на алгоритми за машинно обучение и дълбоко учене в предишна работа спрямо настоящо проучване

Статия (Автор, Година)	**Използван(и) алгоритъм(и)**	**Точност, докладвана в тяхното проучване**
Гунаван и др. (2023)	DL (ADASYN, SMOTE)	F1-резултат ≈ 0.9997
Лий С. и др. (2011 г.)	Откриване на аномалии, базирано на PCA	86,7% процент на откриване

Ксие Б. и др. (2023)	EDRN (Мрежа с повтарящи се данни с експлицитна продължителност)	Точност ≈ 99,6%
Бхайо Дж. и др. (2023)	NB, DT, SVM (в SDN- IoT)	NB: 97,4%, SVM: 96,1%, DT: 98,1%
Гао Й. и др. (2023)	KNN, DT, LR, NB	LR: 99,9%, DT: 99,8%
Ахмед С. и др. (2023)[5]	MLP (дълбоко учене)	98,99%

Глава 4: Заключение – Обобщение на резултатите и бъдещи насоки на изследването

4.1 Приноси

Дисертацията представя както научно-приложни, така и практически ориентирани приноси към интелигентната киберсигурност, съчетавайки теоретични иновации с оперативно внедряване.

Научни приноси:

- Формулиран е нов методологичен подход за откриване на DDoS атаки на приложно ниво, който обединява дълбоко учене със структурирана предварителна обработка и инженерство на функции, съобразено с уеб трафика.
- Изследването създава нови аналитични методи и предоставя потвърждаващи доказателства, че моделите, обучени върху реални регистрационни файлове на уеб сървъри, превъзхождат тези, използващи синтетични данни.
- Сравнителната оценка на моделите за машинно обучение и дълбоко учене изяснява техните силни страни, ограничения и обобщение за различните видове атаки.
- Интерпретируемостта, базирана на SHAP, въвежда обясним AI слой, който разкрива влиянието на характеристиките върху резултатите от модела, увеличавайки прозрачността и доверието.

Научно-практически приноси:

- Разработена и тествана е система с висока точност за откриване на нисък процент фалшиво положителни резултати за интегриране в реални среди за мониторинг и реагиране.
- Универсален експериментален протокол поддържа възпроизводимо обучение и тестване за различни класове атаки (напр. SQL Injection, XSS).
- Реалистичен набор от данни, базиран на инфраструктурата на реални сървъри, предоставя еталон за изследвания и обучение по киберсигурност.
- Методологията установява основа за бъдещи подобрения, включително хибридни архитектури за откриване, онлайн и федеративно обучение и анализи в реално време за адаптивни системи за киберсигурност.

4.2 Оценка на хипотезите

Резултатите от изследването позволяват директна проверка на хипотезите, представени във въведението.

Подхипотеза 1 – Доказана: Моделите, обучени върху реални данни от лог файлове на уеб сървър, превъзхождат тези, обучени върху синтетични набори от данни, потвърждавайки, че автентичният трафик осигурява превъзходно представяне на поведението в реалния свят и води до по-надеждно откриване с по-ниски нива на фалшиво положителни резултати.

Сравнителните доказателства са показани в Таблица 6 (Сравнение на алгоритми за машинно обучение и дълбоко учене в предишна работа

спрямо настоящо проучване), където моделите, обучени върху реален набор от данни от уеб сървърни логове, разработен в това изследване, са постигнали по-висока точност (LSTM = 99,0%) от сравними проучвания, използващи синтетични набори от данни като CICIDS2017 или CICDDoS2019 (обикновено 94–97%). Това количествено сравнение потвърждава, че обучението върху автентичен трафик води до превъзходна надеждност на откриване и способност за обобщаване в сравнение със синтетичните бенчмаркове за данни.

Подхипотеза 2 – Частично доказана: Последователният LSTM модел демонстрира силна способност за проверка при тестване върху допълнителни типове атаки, като SQL Injection и Cross-Site Scripting, което показва добра генерализация. Традиционните модели за машинно обучение обаче не поддържат сравнима производителност. Тези резултати показват, че генерализацията зависи от специфични за архитектурата свойства, а не от универсално споделени характеристики на модела.

4.3 Обобщение и заключение

Резултатите в Таблица 6 (Сравнение на алгоритми за машинно обучение и дълбоко учене в предишна работа спрямо настоящо проучване) показват, че моделите, обучени с реални данни от лог файлове на уеб сървъра, са дали по-добри резултати при откриване и по-малко неправилни предупреждения, отколкото моделите, обучени със синтетични данни от CICIDS2017 и CICDDoS2019. Експерименталните резултати показват, че автентичните данни за трафика съдържат по-точни модели, които съответстват на действителните DDoS атаки на приложно ниво в реални сценарии. Моделите, обучени с реални данни, показват по-добра точност на

прогнозиране и обобщение на типа атака, отколкото моделите, обучени със синтетични данни, които служат главно за целите на проверка и сравнителен анализ. Изследването потвърждава основния изследователски въпрос, като същевременно установява доказан метод за предстоящи изследвания в областта на прогнозната сигурност, който комбинира реални и синтетични източници на данни.

4.4 Поведение и ограничения на модела

Експерименталните данни показват, че както LSTM, така и SVM класификаторите постигат висока точност, но изследователите са открили специфични оперативни ограничения. Невронната мрежова система е генерирала случайни фалшиво положителни резултати, които са изглеждали реалистични, но са липсвали доказателства от модели на лог данни. Моделът генерира фалшиво положителни резултати чрез вътрешно усилване на модели на редки точки от данни и шумови елементи, което води до уверени, но неправилни класификации. Предсказващият характер на извода от дълбокото учене води до този тип халюцинации, защото дава резултати, които не са напълно детерминистични.

Методът на опорните вектори поддържаше стабилна и висока точност, но неговите граници на решения въведоха изкривявания в представянето на данните от реалния свят. Фиксираните трансформации на ядрото, използвани от SVM, създават опростени представяния на поведението на уеб трафика, които съдържат сложни динамични модели. Невронната мрежа постига по-добро контекстуално моделиране от SVM, но SVM поддържа по-добра структурна стабилност в резултатите си.

Двата модела демонстрират различни силни и слаби страни, защото SVM предоставя надеждни структурни резултати, докато LSTM проследява модели, базирани на време, като същевременно генерира потенциални интерпретационни грешки.

4.5 Практически съображения и рискове при внедряването

Въпреки че разработеният модел постигна висока точност на откриване в контролирана експериментална среда, внедряването му в реални производствени системи изисква внимателно обмисляне на оперативните ограничения и рискове. Моделът работи с регистрационни файлове за достъп до уеб сървъра, което означава, че анализира само HTTP заявки, които вече са достигнали приложното ниво и са били записани от уеб сървъра след преминаване през периметърни защити, като защитни стени, системи за откриване на проникванията, защитни стени за уеб приложения (WAF), филтри за маршрутизиране и облачни услуги за смекчаване на риска. Следователно, той допълва, а не замества съществуващите механизми за защита на мрежовото ниво. Тъй като зависи от трафик, който вече е проникнал през първоначалните защити, моделът е особено ценен за откриване на DDoS атаки на приложно ниво (L7), които имитират легитимно потребителско поведение и често заобикалят конвенционалните системи.

Наборът от данни, използван за обучение на модела, е изграден от реален доброкачествен трафик, генериран от стотици легитимни потребители, и злонамерен трафик, генериран от осем добре познати инструмента за атака, което гарантира реалистични поведенчески модели и нива на шум. Въпреки това пълномащабното внедряване в корпоративни среди въвежда допълнителни рискове, като например променливост във форматите на лог

файловете, разлики в инфраструктурата и изчислителни разходи при обработка в реално време. Бъдещо разширение на тази работа може да включва тестване на модела в симулирана или хибридна корпоративна инфраструктура, интегрирането му с потоци от лог файлове на живо и оценка на мащабируемостта, латентността и съвместимостта с корпоративната SIEM система. или SOC рамки. Такова тестване би осигурило по-ясно разбиране на предизвикателствата пред интеграцията и би насочило разработването на готови за производство реализации.

4.6 Технически изисквания за интегриране на модела

Разработеният модел за откриване е проектиран да може да се внедрява на стандартен хардуер от сървърен клас и не изисква специализирана инфраструктура. В експерименталната конфигурация, описана в Глава 3, всички стъпки на обучение и оценка бяха изпълнени на работна станция, оборудвана с процесор Intel Core i7 и 16 GB RAM, работеща под Windows 10 и Apache HTTP Server 2.4.59. Тази конфигурация може да се счита за практическа долна граница за пълно обучение на модела и пакетна оценка върху набори от данни с размера, използван в тази дисертация (≈344 000 записа в лога).

За интеграция в производствени среди се препоръчват следните минимални хардуерни и софтуерни параметри:

Хардуер (обучение и периодична преквалификация):

- Многоядрен процесор (≥4 физически ядра, например Intel Core i7 или еквивалент на AMD).

- RAM памет: поне 16 GB, за да се поддържа зареждането на предварително обработения набор от данни от лога, структурите за токенизация и параметрите на модела в паметта без размяна.
- Съхранение: ≥100 GB свободно дисково пространство за съхранение на сурови регистрационни файлове за достъп, предварително обработени CSV файлове, контролни точки на модели и резултати от SHAP анализ.
- Специализиран графичен процесор (напр. NVIDIA карта с ≥6 GB VRAM) е по избор, но е полезен за по-бързо преобучение на по-големи набори от данни или разширени търсения на хиперпараметри .

Хардуер (заключение / внедряване):

- Многоядрен процесор (≥4 ядра) и 8–16 GB RAM са достатъчни за изпълнение на обучения модел в пакетен режим или режим почти в реално време върху входящи регистрационни файлове за достъп.
- Не е строго необходим графичен процесор за извод при типични обеми на уеб трафик; ускорението на графичния процесор става релевантно само за много високи нива на трафик или множество внедрявания.

Софтуер:

- Операционна система: Windows 10/11 или скорошна Linux дистрибуция (напр. Ubuntu Server LTS).

- Уеб сървър: Apache HTTP Server 2.4.59 (или съвместим), конфигуриран да генерира комбинирани регистрационни файлове за достъп в стабилен формат.
- Програмна среда: Python 3.9+ със следните основни библиотеки:
    - TensorFlow / Keras (за изпълнение на LSTM и Conv1D модели)
    - scikit -learn (за предварителна обработка, разделяне и оценка на показатели),

4.7 Икономическа ефективност и оценка на разходите

Търговските услуги за защита от DDoS атаки като AWS Shield Advanced, Cloudflare, Google Cloud Armor и Akamai Kona обикновено начисляват високи повтарящи се такси – от около 200 до 5000 щатски долара на месец (само AWS Shield Advanced струва приблизително 3000 долара плюс таксите за превод).

За разлика от това, предложеният в дисертацията модел за интелигентна защита изисква само еднократни разходи за хардуер от 2000 долара (≈ 33 долара на месец, амортизирани за 5 години) и използва изцяло софтуер с отворен код.

Тази локално внедрена система, базирана на изкуствен интелект, би могла да спести от 2000 до 35 000 долара годишно в сравнение с алтернативите на пазара, като същевременно запази поверителността на данните, намали зависимостта от външни доставчици и подобри прозрачността чрез обясним изкуствен интелект (SHAP).

Той е особено подходящ за средни организации и академични или изследователски институции с ограничен бюджет.

4.8 Бъдеща работа

Изследването представя основна рамка за откриване на DDoS атаки на приложно ниво, използвайки модели за машинно обучение и дълбоко учене, обучени върху реалистични уеб лог файлове, и идентифицира няколко бъдещи насоки на изследване:

1. Хибридни обучителни архитектури: комбинирайте подходи, базирани на правила, и подходи, базирани на аномалии, или групирайте методи, за да откривате както известни, така и zero-day атаки, подобрявайки точността и интерпретируемостта.

2. Онлайн и поетапно обучение: Използвайте адаптивно обучение в реално време, за да се справите с променящите се модели на атака и отклонението на концепциите.

3. Интеграция на големи данни в реално време: Интегрирайте машинно обучение/дълбоко учене с инструменти за големи данни като Apache Spark или Flink за мащабируемо откриване в корпоративни среди.

4. Федерирано и запазващо поверителността обучение: Използвайте федерално обучение, за да се даде възможност за съвместно обучение на модели без споделяне на сурови данни.

5. Устойчивост на състезания: Разработване на защитни механизми като обучение на състезания, дефанзивна дестилация и генеративни модели за противодействие на тактиките за избягване.

6. Мултимодален и междуслоен анализ: Обединяване на данни от различни мрежови слоеве (DNS, API, лог файлове), използвайки модели, базирани на внимание, за по-богата информация за заплахите.

7. Сравнителен анализ с данни от реалния свят: Създаване на развиващи се набори от данни с отворен код от реални корпоративни среди, за да се подобри обобщаемостта на модела.

8. Разширяване към IoT среди: Адаптиране на рамките за откриване към IoT трафика и леките протоколи чрез персонализирано извличане на характеристики и преобучение.

Списък с публикации

1. A. Sabra, N. Rmeiti, and M. Atieh, “Using Machine Learning Techniques to Detect Cyberattacks in Smart Homes: A Survey,” in 2023 International Scientific Conference on Computer Science (COMSCI), IEEE, Sep. 2023, pp. 1–6. doi: 10.1109/COMSCI59259.2023.10315831.

2. M. ATIEH, O. MOHAMMAD, A. SABRA, and N. RMAYTI, “IdO, apprentissage profond et cybersécurité dans la maison connectée : une étude,” in Cybersécurité des maisons intelligentes, ISTE Group, 2024, pp. 215–256. doi: 10.51926/ISTE.9086.ch6.

3. M. Atieh, O. Mohammad, A. Sabra, and N. Rmayti, “IOT, Deep Learning and Cybersecurity in Smart Homes: A Survey,” in Cybersecurity in Smart Homes, Wiley, 2022, pp. 203–244. doi: 10.1002/9781119987451.ch6.

4. G. Momcheva, A. Sabra, and N. Rmeiti, MACHINE LEARNING in CYBERSECURITY. Varna Free University “Chernorizetz Hrabar” University Publishing House, 2022.

5. A. Sabra, N. Rmeiti, and Z. Minchev, “Machine Learning Approaches to Detect Application Layer DDoS Attacks”, Applications of Mathematics in Engineering and Economics (AMEE’24); AIP Publishing (accepted) .

6. Rmeiti, N., Sabra, A., Shibbi, A., & Marinova, R. (2025). Detecting Human Emotions Using Machine Learning Techniques: A Comprehensive Approach. 2025 International Conference on Computer Systems and Technologies (CompSysTech), 01–09.

https://doi.org/10.1109/CompSysTech65493.2025.11137007