Annual of Konstantin Preslavsky University of Shumen
Faculty of Mathematics and Informatics
vol. XXVI C, 2025, pp. 81 – 93, https://doi.org/10.46687/ZMZC7716

INTEGRATION OF ARTIFICIAL INTELLIGENCE IN INTRUSION DETECTION AND PREVENTION SYSTEMS*

STOYAN R. STOYANOV, TEODORA T. STOYANOVA, RADOSTIN E. RAFAILOV

ABSTRACT: *This article focuses on the integration of artificial intelligence (AI) into Intrusion Detection and Prevention Systems (IDPS) in response to the rapid evolution of security threats. Traditional signature-based security approaches are unable to cope with complex and advanced attacks. AI-based IDPS use machine learning, deep learning and natural language processing to analyze large volumes of data, detect anomalies and predict emerging threats in real time. The article offers practical recommendations for organizations implementing AI-based security solutions.*

KEYWORDS: *Artificial Intelligence, AI, Intrusion Detection and Prevention Systems, IDPS, Machine learning, Deep learning;*

2010 Math. Subject Classification: *68T99, 68M25*

ИНТЕГРИРАНЕ НА ИЗКУСТВЕН ИНТЕЛЕКТ В СИСТЕМИТЕ ЗА ОТКРИВАНЕ И ПРЕДОТВРАТЯВАНЕ НА ПРОНИКВАНИЯ†

СТОЯН Р. СТОЯНОВ, ТЕОДОРА Т. СТОЯНОВА, РАДОСТИН Е. РАФАИЛОВ

*This paper is (partially) supported by Scientific Research Grant RD-08-106/05.02.2025 of Konstantin Preslavsky University of Shumen

†Статията е частично финансирана по проект № РД-08-106/05.02.2025 на ШУ „Еп. Константин Преславски“

1 Въведение

С широкото използване на Интернет киберпрестъпността нараства със значителни темпове. Киберпрестъпниците, вариращи от отделни хакери до организирани групи, използват уязвимости в мрежи, системи и приложения, като създават сериозни заплахи за физически лица, организации и критични инфраструктури.

Традиционните подходи за киберсигурност (защитни стени, антивирусни програми и системи за откриване на проникване) са надеждни, но трудно успяват да се справят с непрекъснато развиващите се тактики, прилагани от атакуващите. Те работят на принципа на съвпадение на моделите със съществуващи известни модели на атаки. Ограниченията на традиционните подходи налагат възприемането на модерни технологии, които ги укрепват и дават възможност за откриване на заплахи и за реагиране в реално време. Тази революционна концепция се постига с въвеждането на изкуствения интелект (ИИ). В частност чрез алгоритмите за машинно обучение се обучават системи, които анализират големи количества данни с цел откриване на потенциални и нововъзникващи заплахи за сигурността.

Прогнозите сочат, че глобалният пазар на ИИ в киберсигурността ще достигне 46,3 млрд. долара до 2027 г. [1]. Причините за увеличаващата се употреба на ИИ са възможността му за автоматизация, прогнозиране на заплахи и краткото време за реакция. Автоматизацията позволява на системите с ИИ да задействат незабавна защита след засичане на потенциална атака. Възможността за прогнозиране на заплахи се свързва със способността на ИИ да анализира големи количества данни за кратко време, непрекъснато да се развива и обучава върху тях. Последната ключова характеристика е времето за реакция, тъй като анализаторите по киберсигурност не винаги реагират бързо на атаките за разлика от системите с ИИ.

Системите за откриване и предотвратяване на проникванията (Intrusion Detection And Prevention Systems - IDPS) се използват за

наблюдение на мрежовия трафик и системните дейности за подозрителни модели или аномалии. IDPS базирани на ИИ, използват усъвършенствани техники като машинно обучение, дълбоко обучение и обработка на естествен език за откриване, анализиране и смекчаване на сложни киберзаплахи в реално време, като по този начин значително променят облика на киберсигурността. Те са необходими в съвременната киберсигурност, заради критичната необходимост от ефективни методи за справяне с атаки от типа „нулев ден“ и напреднали постоянни заплахи (Advanced Persistent Threat - APT), в което те са по-добри от традиционните системи [2].

2 Системите за откриване и предотвратяване на прониквания (IDPS)

IDPS са механизми за сигурност, предназначени да наблюдават, идентифицират и смекчават неоторизиран достъп или злонамерени дейности в рамките на мрежа или система. Разработването на IDPS е претърпяло значително развитие през десетилетията, като всеки етап от еволюцията им отразява адаптацията към нарастващата сложност на киберзаплахите.

Еволюция на IDPS

Най-ранните системи, описани през 1987 г., са базирани на сигнатури и разчитат на предварително дефинирани модели на известни атаки. Ефективни са за идентифициране на предварително документирани заплахи, но трудно се справят с нови и сложни атаки. През 90-те години на миналия век се появяват системи за откриване на аномалии, които използват статистически методи за идентифициране на отклонения от нормалното поведение. Тези подходи демонстрират по-добри показатели за откриване на неизвестни заплахи, но са критикувани заради високия процент фалшиво положителни резултати. По-новите решения се фокусират върху хибридни системи, които комбинират методи, базирани на сигнатури и методи за засичане

на аномалии, за да балансират между точността на откриване и намаляването на фалшивите положителни резултати [3].

Интегрирането на изкуствения интелект допълнително трансформира IDPS, позволявайки динамични възможности за откриване и превенция в реално време. Разликите между традиционните IDPS и тези базирани на ИИ са представени в Таблица 1.

Еволюцията на ИИ в IDPS е белязана от значителен напредък както в дефанзивните, така и в офанзивните способности. Въвеждането на технологиите за машинно обучение и дълбоко обучение позволява на системите да откриват по-ефективно модели, аномалии и потенциални заплахи. С течение на времето инструментите, задвижвани от изкуствен интелект, преминават от реактивни мерки, като например идентифициране на известни сигнатури на зловреден софтуер, към проактивни стратегии, като например прогнозен анализ и автоматизирано реагиране на инциденти. Тази еволюция позиционира ИИ като крайъгълен камък на съвременните решения за киберсигурност [4].

Таблица 1: Сравнение между традиционни и ИИ-базирани IDPS

Критерий	Традиционни IDPS	IDPS базирани на ИИ
Метод	статични правила, сигнатури	поведенчески анализ, обучение от данни
Точност на откриване	висока при познати заплахи	висока както при познати, така и при непознати атаки
Фалшиви положителни резултати	често срещани (при сложни правила)	по-рядко срещани

Гъвкавост	ниска	висока
Скорост на реакция	бърза при известни шаблони	обучение се по-бавно, но реагира по-бързо при атаки
Изисквания към данните	не изисква големи обеми от данни	изисква големи набори от данни за обучение
Обяснимост	висока	често ниска
Мащабируемост	трудна при големи мрежи	подходяща за мащабиране с облачни решения
Поддръжка и актуализация	необходимост от ръчна актуализация на сигнатурите	автоматично обучение и адаптация към нови заплахи
Зависимост от експерти	висока	по-ниска
Приложение в реално време	добро за познати заплахи, ограничено за нови	ефективно при правилно обучение и оптимизация

Приложение на ИИ в IDPS

Техники за машинно обучение

Машинното обучение се е утвърдило като фундаментален компонент на съвременните IDPS, позволявайки автоматизиран анализ на мащабни данни. Алгоритмите за контролирано обучение, като например Decision Trees и Support Vector Machines (SVM), са широко използвани за откриване на проникванията, като постигат висока точност за маркирани набори от данни. Техники

като k-Means Clustering и Gaussian Mixture Models, са обещаващи при откриване на аномалии, когато липсват маркирани данни. Освен контролирано обучение IDPS използва и неконтролирани техники за обучение, например клъстериране (K-Means или DBSCAN), които идентифицират нови заплахи, с които системата не се е сблъсквала преди [3].

Приложения на дълбокото обучение

Дълбокото обучение, подмножество на машинното обучение, значително разширява възможностите на IDPS. Техники като Convolutional Neural Networks (CNNs) и Recurrent Neural Networks (RNNs) са особено ефективни за анализиране на многомерни и последователни данни, като например логове на мрежовия трафик [5]. Установено е, че Long Short-Term Memory Networks_(LSTM), които са вид RNN, се отличават с отлични постижения при откриването на времеви модели в данните за проникване, което ги прави подходящи за откриване на заплахи в реално време. За откриване на аномалии все по-често се използват autoencoders и Generative Adversarial Networks (GANs), заради тяхната способност да моделират нормалното поведение и да идентифицират отклоненията.

Откриване на аномалии и идентифициране на заплахи

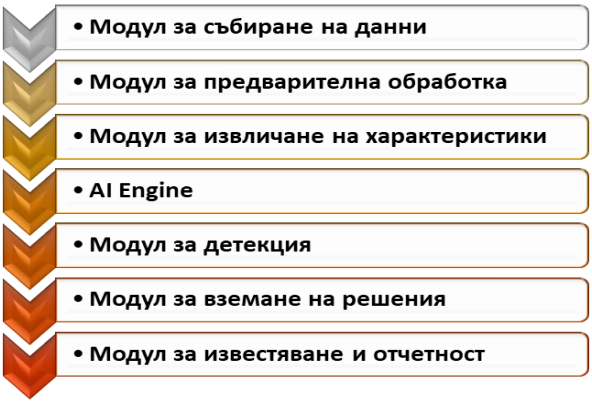
ИИ също така улеснява поведенческия анализ, като позволява на системите да разбират нормалната работа на потребителите и приложенията. Тази способност позволява откриването на вътрешни заплахи, напреднали постоянни заплахи (APT) и експлойти от типа „нулев ден“ [2]. Освен това интегрирането на обработката на естествен език (NLP) позволява на системите с ИИ да анализират текстови данни, като например информационни канали за заплахи и системни логове, което допълнително засилва способността им да идентифицират потенциални уязвимости [3].

Архитектура на IDPS

Архитектурата на една система за откриване и предотвратяване на проникванията, базирана на ИИ, е критичен

компонент, който определя нейната ефективност и ефикасност при идентифицирането и смекчаването на заплахите за сигурността. Тя интегрира различни модули, всеки от които е предназначен за изпълнение на специфични задачи, вариращи от събиране на данни до вземане на решения. IDPS, базирана на ИИ, обикновено се състои от няколко взаимосвързани компонента, които работят заедно за откриване и предотвратяване на проникванията. Основните компоненти включват [3]:

1. **Модул за събиране на данни:** Събира необработени данни от различни източници, като например логове на мрежовия трафик, системни логове и външни информационни канали за заплахи.



Фигура 1: **Архитектура на ИИ-базирана IDPS**

2. **Модул за предварителна обработка:** Почиства, нормализира и форматира данните за анализ, като гарантира, че те са подходящи за алгоритмите на изкуствения интелект.

3. **Модул за извличане на характеристики:** Извлича значими атрибути от необработените данни, като например IP хедъри, размери на пакети и модели на потребителска активност, с цел по-ефективен анализ.

4. **AI Engine:** Съдържа модели за машинно обучение и дълбоко обучение, които анализират извлечените характеристики.

5. **Модул за детекция:** Обработва резултатите от AI engine, за да открие аномалии, сигнатури или злонамерено поведение.

6. **Модул за вземане на решения:** Предприема проактивни мерки, като например блокиране на подозрителен трафик и изолиране на компрометирани системи.

7. **Модул за известяване и отчетност:** Генерира предупреждения за администраторите за откритите заплахи и предприетите действия, като предоставя подробни регистри за по-нататъшен анализ.

Този модулен дизайн, показан на Фигура 1, гарантира, че системата е мащабируема, адаптивна и ефективна, способна да обработва големи обеми от данни и да се развива заедно със средата на заплахите.

3 Предизвикателства и ограничения на ИИ в областта на киберсигурността

Въпреки че интегрирането на изкуствения интелект в системите за откриване и предотвратяване на проникванията носи значителни ползи, от съществено значение е да се признаят и преодолеят предизвикателствата и ограниченията, свързани с тези технологии. Разбирането на тези пречки е решаващо за разработването на стабилни стратегии за киберсигурност.

Употреба на ИИ от киберпрестъпници

Използването на ИИ освен в позитивен, може да се разглежда и негативен аспект. Причината е, че киберпрестъпниците също използват ИИ, но с цел манипулиране на системите за сигурност [1]. Потенциалните противникови атаки включват:

• *Отравяне на данни*: В данните за обучение атакуващите инжектират манипулирани данни, което води до неправилна оценка на заплахите. Това може да стане с безобидно изглеждащи файлове, в които има подвеждаща информация. В резултат, на това системата губи своята ефективност и пропуска повече зловредни програми.

• *Техники за избягване*: Атакуващите променят минимално всички аспекти на зловредния код, за да се избегне откриването му от ИИ и да се класифицира като добронамерен.

• *Атаки за обръщане на модела*: Атакуващите извличат информация от моделите на ИИ, за да разберат какви са механизмите за сигурност [6].

Фалшиви резултати

Пристрастните или небалансирани набори от данни могат да доведат до неточни прогнози. Моделите с прекомерно приспособяване могат да дадат неточни прогнози, което води до фалшиви положителни резултати и ненужни предупреждения.

Фалшивите положителни резултати (доброкачествени дейности, маркирани като заплахи) и фалшивите отрицателни резултати (злонамерени дейности, класифицирани като доброкачествени) са критични фактори, влияещи върху използваемостта и надеждността на IDPS [3].

Фалшивите положителни резултати представляват сериозно предизвикателство при прогнозирането на заплахи в реално време. Решенията за сигурност, базирани на ИИ, генерират големи обеми от сигнали, които често изискват ръчна проверка от човешки анализатори. Умората от множеството сигнали води до пропускане на реални заплахи. Докладите показват, че екипите по сигурността прекарват до 40% от времето си в разследване на фалшиви сигнали, генерирани от ИИ [7, 8].

4 Бъдещи тенденции и предизвикателства

Тъй като заплахите стават все по-сложни, решенията, базирани на ИИ, се развиват, за да отговорят на изискванията за усъвършенствано откриване и предотвратяване на заплахи. Наред с тези възможности обаче трябва да се преодолеят значителни предизвикателства, свързани с етиката, неприкосновеността на личния живот, мащабируемостта и прилагането в реално време [3, 2]. Сред водещите технологични тенденции изпъкват:

• *Обединено обучение (Federated Learning)* - позволява на моделите да се учат съвместно на децентрализирани устройства, без да се компрометират чувствителни данни, подобрявайки способността си да предвиждат и противодействат на атаки.

• *Graph Neural Networks (GNNs)* набират популярност като ефективен метод за откриване на скрити връзки и аномалии между субектите в сложния мрежов трафик, което ги прави особено подходящи за приложения в системите за откриване и предотвратяване на кибератаки.

5 Препоръки за успешна интеграция на изкуствен интелект в IDPS

Този раздел предлага препоръки за организации, които планират или вече внедряват системи за откриване и предотвратяване на проникванията, базирани на ИИ.

ИИ е сравнително нова технология и съществува значителна липса на осведоменост за потенциала й. Препоръчва се преди мащабното разгръщане да се стартира с поетапно внедряване в конкретни области с висок риск, за да се оцени ефикасността на решението. Добра стратегия е и комбинирането на традиционни IDPS, с такива базирани на ИИ, като постепенно се увеличава зависимостта от ИИ компоненти.

Инвестициите в ИИ от страна на организациите са значителни. Приоритет трябва да са обезпечаването на достатъчна изчислителна мощност за обучение и работа на модели в реално време, инвестиране в събиране, обработка и структуриране на качествени данни за обучение на моделите и наемане и обучение на специалисти в ИИ и киберсигурност.

Препоръките за избор на подходящи техники и модели са в три основни насоки. За откриване на вече известни заплахи да се използва Supervised Learning алгоритми като Random Forest, SVM или невронни мрежи, обучени на маркирани данни. За откриване на непознати или нови заплахи се препоръчва прилагането на Unsupervised Learning техники като K-Means, DBSCAN или

автоенкодери за идентификация на аномалии. За поведенчески анализ ефективни са RNN или LSTM [8].

Значителен процент от времето на екипите по сигурността преминава в разследване на фалшиви сигнали, генерирани от ИИ. За справяне с този проблем се препоръчва непрекъснато актуализиране на праговете за откриване на атаки, вземайки предвид обратната връзка от екипите по сигурността. Оценките на заплахите да зависят и от контекстуална информация, като час от денонощието, потребителска роля и др. Необходимо е приоритизиране на предупрежденията въз основа на тяхната критичност [8].

С цел осигуряване на безпристрастност в работата на IDPS базирани на ИИ, се препоръчва при обучение да се осигурят разнообразни и проверени данни, както и редовно тестване на моделите за наличие на пристрастия.

За да функционира ефективно една IDPS, тя трябва безпроблемно да се интегрира със съществуващата инфраструктура за сигурност на организацията. Изисква се съвместимост със защитните стени, SIEM системи, интеграция с външни бази данни за възникващи заплахи в реално време и системи за управление на политики [9].

Внедряването на ИИ в IDPS е непрекъснат процес на усъвършенстване и адаптация. Посочените препоръки са отправна точка за организациите в процеса на интегриране на тези системи за сигурност.

6 Заключение

Това проучване подчертава решаващата роля на ИИ при откриването, анализа и предотвратяването на проникванията, като показва как усъвършенствани техники като машинно обучение, дълбоко обучение и обработка на естествен език трансформират традиционните системи за сигурност.

Задвижваните от ИИ IDPS се отличават от традиционните по способността си да анализират огромни масиви от данни, да

откриват аномалии и да предвиждат бъдещи заплахи, което ги прави незаменими. Въпреки това, проучването разкрива и значителни предизвикателства, които трябва да бъдат преодолени. Киберпрестъпниците все по-активно използват техники за противниково машинно обучение, за да заобиколят защитните механизми. Независимо от тези предизвикателства, бъдещето на IDPS базирани на ИИ, е с голям потенциал.

Бързата еволюция на киберзаплахите налага иновативни подходи за подобряване на работата на IDPS системите. Основните инструменти за постигане на това са ИИ и машинното обучение, които повишават ефективността, точността и способността за противодействие на нововъзникващи заплахи в реално време. Инвестициите в тези технологии са не просто желателни, а необходими за всяка организация, която се стреми да защити своите цифрови активи и да запази доверието на своите клиенти и партньори.

ЛИТЕРАТУРА:

[1] Dash, B., Ansari, M. F., Sharma, P and Ali, A. Threats and Opportunities with AI-Based Cyber Security Intrusion Detection: A Review. – International Journal of Software Engineering & Applications (IJSEA), Vol.13, No.5, 2022.

[2] Amin, M. A. Artificial Intelligence and Machine Learning Algorithms Are Used to Detect and Prevent Cyber Threats as Well as Their Potential Impact on the Future of Cybersecurity Practices. – International Journal of Computer Science and Information Technology, vol. 17, 2025.

[3] Raghavan, S. S. A comprehensive study of artificial intelligence applications in intrusion detection and prevention. – International Journal of Artificial Intelligence Research and Development (IJAIRD), vol. 3(1), pp. 16–37, 2025.

[4] Damola, P., Miracle, A., Hoover, R. Assessing the Weaknesses of AI-Powered Cybersecurity Solutions AUTHOR. – Cybersecurity and Law, 2024.

[5] Shone, N., Ngoc, T. N., Phai, V. D., Shi, Q. A deep learning approach to network intrusion detection. – IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2(1), pp. 41-50, 2018.

[6] Arefin, S., Zannat, N. T. AI vs Cyber Threats: Real-World Case Studies on Securing Healthcare Data. – International Journal of Advanced Research in Education and Technology, vol. 12 (2), pp. 396-404, 2025.

[7] Ajala, O, Okoye, Ch., Arinze, Ch., Daraojimba, O. Review of AI and machine learning applications to predict and Thwart cyber-attacks in real-time. – Magna Scientia Advanced Research and Reviews, vol. 10, pp. 312-320, 2024.

[8] Kannan, Y., AI and machine learning for network security: applications and case studies. – International Journal of Artificial Intelligence & Machine Learning (IJAIML), vol. 3(2), pp. 1-13, 2024.

[9] Venukumar, Ch., Enhancing cloud computing security with multi-layered protection and advanced threat detection. – International Journal of Computer Engineering and Technology (IJCET), vol. 15(3), pp. 193-202, 2024.

Stoyan Stoyanov
Konstantin Preslavski University of Shumen
9712 Shumen, Bulgaria
s.stoyanov@shu.bg

Teodora Stoyanova
Konstantin Preslavski University of Shumen
9712 Shumen, Bulgaria
t.stoyanova@shu.bg

Radostin Rafailov
Konstantin Preslavski University of Shumen
9712 Shumen, Bulgaria
r.rafailov@shu.bg