VARSTVOSLOVJE,
letn. 22
št. 2
str. 197–210

Uporaba nevronskih mrež v kibernetski varnosti

Črt Uršič, Anže Mihelič, Simon Vrhovec

Namen prispevka:

Ob neprestanem izboljševanju informacijske infrastrukture so za njeno zaščito potrebni tudi novi pristopi k varnosti in razvoj novih tehnik zaznavanja kibernetskih groženj. Med te tehnologije sodijo tudi nevronske mreže, ki se že dolgo uporabljajo na različnih področjih, kot so medicina, logistika in biologija. Namen prispevka je prepoznati in predstaviti njihovo uporabnost na področju kibernetske varnosti.

Metode:

Izveden je bil sistematični pregled literature, s katerim so bile prepoznane pozitivne in negativne lastnosti nevronskih mrež kot aktualnega pristopa strojnega učenja za zaznavo kibernetskih napadov. Izvedena je bila tudi primerjava uporabe nevronskih mrež s konvencionalnimi sistemi.

Ugotovitve:

Rezultati eksperimentov so večinoma v korist nevronskim mrežam, saj je proces hitrejši, natančnejši in z manj lažnimi alarmi kot konvencionalni sistemi, ki običajno delujejo na principu statične analize. Vendar pa so nevronske mreže zaradi njihovega načina delovanja pogosto nepredvidljive in so najbolj učinkovite šele v kombinaciji s konvencionalnimi sistemi. V obstoječi literaturi primanjkujejo predvsem testiranja teh sistemov v realnih situacijah, izven kontroliranih umetnih okolij.

Omejitve/uporabnost:

V pregled literature so bili vključeni znanstveni prispevki, objavljeni v letih od 2017 do 2019 in indeksirani v bazah Web of Science in Scopus.

Izvirnost/pomembnost prispevka:

Prispevek s povzetkom osnovnih tehničnih principov delovanja nevronskih mrež predstavlja začetno točko za strokovnjake na področju kibernetske varnosti, ki z njimi še niso seznanjeni. Prispevek povzema trenutno stanje na področju uporabe nevronskih mrež v kibernetski varnosti in potencialne smeri razvoja v prihodnosti. Prispevek predstavlja enega prvih sistematičnih pregledov literature na področju uporabe nevronskih mrež v kibernetski varnosti, ki se je znanstvenoraziskovalno razcvetelo predvsem v zadnjih treh letih.

UDK: 004.056+004.032.26

Ključne besede: nevronske mreže, strojno učenje, umetna inteligenca, sistem za zaznavanje vdorov, škodljiva programska oprema

Use of neural networks in cybersecurity

Purpose:

In addition to the continuous improvement of the information infrastructure, new approaches to security and the development of new techniques for detecting cyber threats are needed to protect it. These technologies also include neural networks which have long been used in various fields such as medicine, logistics and biology. The purpose of this paper is to identify and present their applicability in the field of cybersecurity.

Design/Methods/Approach:

A systematic review of the literature was performed to identify the positive and negative properties of neural networks as a current approach to machine learning for the detection of cyberattacks. A comparison of the use of neural networks with conventional systems was also done.

Findings:

The results of the experiments are mostly in favour of neural networks, as the process is faster, more accurate and with fewer false positives than conventional systems which are typically based on static analysis. However, neural networks are often unpredictable due to their mode of operation and are most effective only in combination with conventional systems. The existing literature predominantly lacks testing these systems in real-world situations outside of controlled artificial environments.

Research Limitations/Implications:

The review of the literature included scientific papers published in the years from 2017 to 2019 and indexed in the Web of Science and Scopus databases.

Practical Implications:

The paper recapitulates the basic technical principles of neural networks and is a starting point for cybersecurity experts who are not yet familiar with them. The paper summarizes the current situation in the use of neural networks in cybersecurity and potential directions for future development.

Originality/Value:

The paper presents one of the first systematic reviews of the literature in the field of the use of neural networks in cybersecurity which has flourished scientifically in the last three years.

UDC: 004.056+004.032.26

Keywords: neural networks, machine learning, artificial intelligence, intrusion detection systems, malware

1 UVOD

Raziskava *The Security Infringement Survey* je v letih 2016 in 2017 pokazala, da je bilo 47 odstotkov vseh globalnih podjetij tarča kibernetskih napadov ali vdorov v zadnjih letih (Geluvaraj et al., 2019). Zdi se, da postaja kibernetski prostor vedno

bolj nevaren, saj je vsak dan zabeleženih skoraj 350.000 novih potencialnih groženj (Rosenberg et al., 2018). S strmim povečevanjem števila naprav, ki so povezane v kibernetski prostor, pa bodo tudi te številke še bolj narasle (Karatas et al., 2018).

Obstoječe metode za analize in soočanje s kibernetskimi grožnjami zahtevajo relativno veliko človeškega posredovanja. Če potencialna žrtev odkrije, da ji preti kibernetska grožnja, to informacijo običajno posreduje ekipi, ki je specializirana za analizo kibernetskih groženj in ukrepanje. Če se izkaže, da je bila neka vrsta kibernetske grožnje že spoznana za zlonamerno, ker je bila npr. najdena tudi na drugih sistemih, se jo poskuša odstraniti. Če pa se izkaže, da gre za novo vrsto kibernetske grožnje, jo je treba podrobno analizirati, ugotoviti njen namen in način delovanja (Cordonsky et al., 2018). Vse pogostejše pojavljanje kibernetskih groženj lahko terja svoj davek pri hitrosti in natančnosti odziva specializiranih ekip. Določene vrste analiz, kot je npr. statična analiza, so dodatno otežene s pojavljanjem novih vrst kibernetskih groženj, kot so metamorfna in polimorfna zlonamerna programska oprema, saj je zaznava teh novih vrst kibernetskih groženj z njimi skoraj nemogoča (Cordonsky et al., 2018). Izmed perspektivnih potencialnih rešitev nekoliko izstopa uporaba nevronskih mrež. Osnovna ideja nevronskih mrež, ki so sicer družina metod strojnega učenja, je, da se najprej samodejno naučijo z vnaprej danimi primeri, nato pa na podlagi novih izkušenj svojo bazo znanja postopoma dopolnjujejo in izpopolnjujejo brez potrebe po dodatnem programiranju.

Nevronske mreže (angl. *neural networks*) spadajo v področje umetne inteligence (angl. *artificial intelligence*). Umetna inteligenca je zelo široko raziskovalno področje, ki se ukvarja s posnemanjem človeškega mišljenja, posnemanja človeškega obnašanja, racionalnega mišljenja in racionalnega obnašanja (Bratko, 1986; Russell in Norvig, 2020). V grobem se umetna inteligenca deli na podpodročja, ki se ukvarjajo z reševanjem problemov (npr. iskanje optimalnih rešitev), znanjem, sklepanjem in planiranjem (npr. logični agenti, klasično planiranje, predstavitev znanja), negotovim znanjem in sklepanjem (npr. kvantifikacija negotovosti, verjetnostno sklepanje), učenjem (npr. učenje iz primerov, znanje v učenju, verjetnostni modeli učenja, spodbujevalno učenje) in komunikacijo, zaznavo in delovanjem (npr. obdelava naravnega jezika, zaznavanje, robotika) (Russell in Norvig, 2020). Nevronske mreže spadajo v podpodročje, ki se ukvarja z učenjem oz. natančneje s strojnim učenjem (angl. *machine learning*). Poleg nevronskih mrež poznamo še več različnih metod strojnega učenja, kot so odločevalna drevesa, podporni vektorski stroji, regresijska analiza itd. Čeprav so danes nevronske mreže že uporabljene in uveljavljene na različnih področjih, kot so npr. klasifikacija in procesiranje besedila, prepoznava govora in slik, vizualizacija (Zhang et al., 2019), pa njihova vloga v kibernetski varnosti ni tako zelo jasna. Namen prispevka je zapolniti navedeno vrzel s pregledom literature na področju uporabe nevronskih mrež v kibernetski varnosti. Na podlagi prikazane raziskave poskušamo prepoznati prednosti in slabosti uporabe nevronskih mrež v kibernetski varnosti v primerjavi s klasičnimi sistemi, najpogostejša področja uporabe in zakaj se kljub izjemno dobrim rezultatom na drugih področjih nevronske mreže še niso povsem uveljavile tudi na področju kibernetske varnosti.

V prispevku so v naslednjem poglavju predstavljene nevronske mreže in njihove posebnosti, nato sledijo metodologija pregleda literature, rezultati raziskave in razprava o dobljenih rezultatih. V zadnjem poglavju prispevek zaključimo s sklepnimi mislimi.

2 TEORETIČNO OZADJE

V prispevku se osredotočamo na strojno učenje, ki se ukvarja z analizo vnaprej podanih podatkov, v katerih se iščejo ponavljajoči se vzorci. Nevronske mreže so namreč skupina algoritmov strojnega učenja, ki na poenostavljen način simulirajo delovanje nevronov v možganih (Xin et al., 2018). V splošnem se strojno učenje deli na dve fazi (Geluvaraj et al., 2019). V prvi fazi se algoritmi strojnega učenja »naučijo« nečesa iz podanih podatkov. To pomeni, da algoritem sam razvije matematični model, ki omogoča napovedovanje oz. odločanje, prilagojeno na konkretno situacijo. V drugi fazi se matematični model uporabi pri nalogi, za katero je bil razvit. Strojno učenje ima kar nekaj skupnega s statistiko, saj temelji na učenju iz podatkov iz preteklosti.

Metode strojnega učenja se glede na povratne informacije delijo na nadzorovano, nenadzorovano, delno nadzorovano in spodbujevalno učenje glede na to, iz kakšnih podatkov se učijo (Geluvaraj et al., 2019; Russell in Norvig, 2020). Pri *nadzorovanem učenju* dobi algoritem strojnega učenja označene podatke, iz katerih se uči. Označeni podatki pomenijo, da je vnaprej točno jasno, kaj določeni podatki pomenijo. Tipičen primer bi lahko bila skupina slik psov – vsaka slika v tej skupini je označena kot pes. Algoritem strojnega učenja nima nobenega vnaprejšnjega znanja o lastnostih psa, a se iz slik sam nauči prepoznati psa tako, da prepozna njegove ključne lastnosti. Pri *nenadzorovanem učenju* dobi algoritem strojnega učenja neoznačene podatke, iz katerih se uči. Neoznačeni podatki poenostavljeno pomenijo, da vnaprej ni jasno, kaj določeni podatki pomenijo. Tipičen primer bi bila skupina slik mačkov in psov, za katere algoritem ne bi vedel, na kateri sliki je maček in na kateri je pes. Samo iz skupine neoznačenih slik bi se nato algoritem naučil ločiti med slikami mačkov in psov. *Delno nadzorovano* učenje je kombinacija nadzorovanega in nenadzorovanega, kjer se mreža delno nadzorovano uči z nekaj označenimi podatki, ki niso niti nujno najbolj natančno označeni, delno pa se uči tudi z neoznačenimi podatki. *Spodbujevalno učenje* je nekoliko drugačno, saj algoritem strojnega učenja ne dobi podatkov vnaprej. Algoritem bi takoj začel z naključnim odločanjem, nato pa bi se učil na podlagi svojih izkušenj, tj. uspehov in napak, za katere je nagrajen oz. kaznovan. Da bi tak algoritem lahko deloval, je treba zagotoviti, da algoritem dobi (točne) povratne informacije, ali je bila odločitev pravilna ali ne. Posebnost nevronskih mrež je, da ne glede na vrsto učenja potrebujejo ogromno število podatkov za uspešno učenje (Homayoun et al., 2018).

3 METODE

V okviru predstavljene raziskave je bil opravljen sistematičen pregled literature v bazah znanstvenih publikacij *Scopus* in *Web of Science* – WoS. V tabeli 1 so povzete poizvedbe po posameznih bazah, omejitve in število zadetkov.

Baza	Poizvedba	Omejitve (leto izida)	Zadetkov
Scopus	cybersecurity AND (neural AND networks) AND (intrusion AND detection AND system) AND (machine AND learning)	2017–2019	621
WoS	cybersecurity AND (neural AND networks)	2017–2019	49

Tabela 1: Poizvedbe po bazah znanstvenih publikacij

Pregled literature je bil izveden med 25. marcem 2019 in 30. aprilom 2019. Skupaj je bilo najdenih 670 zadetkov. Po izločitvi dvojnikov in pregledu naslovov ter povzetkov glede na vključitvene in izključitvene kriterije je ostalo 46 znanstvenih del. Po podrobnem branju prispevkov smo v pregledu literature ohranili 40 znanstvenih del. Kriteriji za vključitev oz. izključitev iz pregleda literature so predstavljeni v tabeli 2.

Vključitveni kriteriji	**Izključitveni kriteriji**
Prispevek v angleškem jeziku	Prispevek ni v angleškem jeziku
Tematska ustreznost	Tematska neustreznost
Izvirni ali pregledni znanstveni članki	Prispevki, ki niso znanstveni članki
Objavljen med 2017 in 2019	Objavljen v 2016 ali prej

Tabela 2: Vključitveni in izključitveni kriteriji

4 REZULTATI

V tabeli 3 so predstavljeni rezultati sistematičnega pregleda literature. Za vsak članek, ki je bil vključen v pregled, so podane raziskovalne metode, vzorec in ključne ugotovitve.

Vir	**Metodologija**	**Vzorci**	**Ugotovitve**
(Lu in Wong, 2019)	Eksperiment	CERT Insider Threats Dataset	Nevronske mreže se lahko učijo, prepoznavajo dejanja uporabnikov ter preverjajo njihova odstopanja. Tako lahko ugotovijo neavtorizirano uporabo računov.
(Geluvaraj et al., 2019)	Pregledni članek	-	Čez leta se bo povečala uporaba nevronskih mrež v kibernetski varnosti, saj se lahko uporabijo za samodejno odkrivanje ranljivosti ali zaznavanje napadov.
(Karatas et al., 2018)	Eksperiment	NSL-KDD, KDD-99, CIC IDS 2017	Predstavitev in primerjava učinkovitosti različnih pristopov učenja globokih nevronskih mrež in učinkovitost različnih podatkovnih baz.
(Qamar et al., 2019)	Pregledni članek	-	Mobilni telefoni postajajo vse pametnejši in vse pomembnejši vektor napadov na sisteme, zato narašča uporaba nevronskih mrež v mobilnih protivirusnih sistemih.

Tabela 3: Raziskave, vključene v sistematični pregled literature

Vir	Metodologija	Vzorci	Ugotovitve
(Handa et al., 2019)	Eksperiment	K-OCSVM	Nevronske mreže so lahko tarče napadov, ki z manipulacijo zadnje plasti mreže spremenijo njeno končno odločitev. Zaradi sestave klasifikatorjev se taki napadi težko odkrijejo.
(Xin et al., 2018)	Eksperiment	DARPA, KDD-99	Primerjava različnih pristopov k učenju in uporaba različnih podatkovnih nizov. Najučinkovitejša metoda odločanja še ne obstaja.
(Al Hawawreh et al., 2018)	Eksperiment	UNSW-NB15	Razvoj nevronske mreže, ki preprečuje notranji napad porazdeljene ohromitve storitve (angl. *distributed denial-of-service* – DDoS) na storitev v oblaku z analizo prometa med virtualnimi napravami.
(Clark et al., 2018)	Eksperiment	Spodbujevalno učenje brez učne množice podatkov	Nevronske mreže so lahko tarče napadov, kar je predstavljeno s sistemom za avtonomno vožnjo.
(Singh in Hofmann, 2017)	Eksperiment	249 mobilnih aplikacij	Nevronska mreža je uspešno določila namen aplikacije le s štetjem klicev jedra operacijskega sistema.
(Viet et al., 2018)	Eksperiment	NSL-KDD, UNSW-NB15	Globoka nevronska mreža je lahko zaznala zbiranje informacij z enako natančnostjo kot programi s statično analizo prometa, a z manj računalniškimi viri.
(Cakir in Dogdu, 2018)	Eksperiment	Microsoft Malware Classification Challenge Dataset	Nevronska mreža s 96-odstotno natančnostjo razloči viruse s funkcijo Word2Vec, ki niz spremeni v vektor, ta pa je kasneje uporabljen kot vnosni podatek.
(Zhu et al., 2018)	Eksperiment	1200 mobilnih aplikacij	Delovanje protivirusnih programov na platformi Android deluje na dveh stopnjah: izbira zaupljivih lastnosti aplikacije in preverjanje teh lastnosti ter odločitev.
(Bulavas, 2018)	Eksperiment	NSL-KDD	Kombinacija odločevalnih dreves in nevronskih mrež lahko z visoko natančnostjo zaznava viruse, ob enem pa ima tudi možnost preproste širitve in nadgradnje.
(Baykara in Gürel, 2018)	Eksperiment	Anti Phishing Simulator	Phishing je še vedno najpogostejši vektor napada, zato lahko sistemi za odkrivanje phishing strani znatno vplivajo na ranljivost organizacij.
(Chattopadhyay et al., 2018)	Pregledni članek	-	Primerjava uporabe različnih metod učenja/učnih podatkov/modelov in lastnosti. Najučinkovitejša je kombinacija KDD, mehke logike (angl. *fuzzy logic*) in genetskih algoritmov.
(Cordonsky et al., 2018)	Eksperiment	Kombinacija obstoječih družin zlonamerne programske opreme	Nevronska mreža razloči zlonamerne programe od nenevarnih. Hkrati lahko uspešno določi tudi njihovo družino in točno vrsto.

Vir	Metodologija	Vzorci	Ugotovitve
(Ait Tchakoucht in Ezziyyani, 2018)	Eksperiment	KDD-99	Metoda za zaznavo DDoS napada na sistemih s kratkim odzivnim časom.
(Rosenberg et al., 2018)	Eksperiment	Cuckoo APT (Rusija in Kitajska)	Metoda za preprečevanje napadov na sisteme v določeni državi. Lastnosti analize so metapodatki v glavi, uporabljene knjižnice in izhodni podatki programa.
(Hai in Hwang, 2018)	Eksperiment	DMOZ, MLD, Mac0De, CleanMX	Nevronska mreža, ki lahko določi namen spletne strani le z analizo spletne povezave (npr. domena, zaporedje črk, nizov in številk).
(Ponkarthika in Saraswathy, 2018)	Eksperiment	NSL-KDD	Primerjanje konvolucijskih nevronskih mrež z drugimi algoritmi za zaznavanje vdorov. Nevronske mreže imajo dobro natančnost, kratek čas učenja in porabo virov, zaradi česar so najugodnejša metoda za zaznavanje vdorov.
(Vinayakumar et al., 2017)	Eksperiment	KDD-99	Primerjava različnih modelov nevronskih mrež za zaznavanje vdorov. Najbolje se odrežejo mreže z dolgim kratkoročnim spominom (angl. long short-term memory), najslabše pa ponavljajoče se nevronske mreže.
(Potluri et al., 2017)	Eksperiment	NSL-KDD	Ocena različnih vrst učenja na hibridnih sistemih za zaznavanje vdorov. Kljub dobri zaznavi nekaterih razredov, so bile zaznave drugih nenatančne zaradi pomanjkanja baz podatkov.
(Spaulding in Mohaisen, 2018)	Eksperiment	Alexa Top 1M Websites	Razvoj nevronskih mrež, ki se učijo na lastnostih phishing strani. Zaznavanje phishing strani glede na metapodatke spletnih strani.
(Moustafa et al., 2018)	Eksperiment	UNSW-NB 15, NIMS	Analiza protokolov MQTT, DNS in http ter razvoj učinkovitega sistema za zaznavanje vdorov za internet stvari (angl. *internet of things – IoT*).
(Sergio Ordoñez in Cesar Guerra, 2018)	Eksperiment	KDD-99	Hibridni sistem nevronskih mrež in genetskih algoritmov za zaznavo napadov na omrežja. Analizira dnevnike dogodkov, glave paketkov TCP, mrežni promet in programska vrata.
(Parameshwarappa et al., 2018)	Eksperiment	-	Odkrivanje napak s pomočjo nevronskih mrež v primerjavi s klasičnim sistemom za zaznavanje vdorov.
(Demidov et al., 2018)	Pregledni članek	-	Predstavitev uporabe nevronskih mrež v okoljih VANET, FANET in MARINET.
(Mouhoub et al., 2018)	Eksperiment	UNSW-NB15	Zaradi njihove narave so sistemi za zaznavanje vdorov za naprave IoT neučinkoviti, zato je predstavljen sistem nevronskih mrež v oblaku.
(Vinayakumar et al., 2018)	Eksperiment	Alexa, DMOZ, MalwareURL	Poskus napada na nevronske mreže za zaznavanje phishing spletnih naslovov.

Vir	Metodologija	Vzorci	Ugotovitve
(Kim in Aminanto, 2017)	Pregledni članek	-	Uporaba računalniške gruče v kombinaciji z nevronskimi mrežami.
(Homayoun et al., 2018)	Eksperiment	ISCX Botnet Dataset	Zaznavanje botnetov z rezultati projekta BoTShark.
(Qabajeh et al., 2018)	Eksperiment	Phishtank	Primerjanje nevronske mreže s klasičnimi sistemi za prepoznavanje phishinga.
(Nguyen et al., 2018)	Eksperiment	KDD-99	Predstavitev postopka normalizacije podatkovnih množic in različnih rezultatov za sistem za zaznavanje vdorov, ki temelji na konvolucijskih nevronskih mrežah.
(Liu et al., 2018)	Pregledni članek	-	Predstavitev projektov z nevronskimi mrežami in kibernetsko varnostjo skozi čas ter njihovi rezultati.
(Diro in Chilamkurti, 2018)	Eksperiment	NSL-KDD	Sistem za zaznavanje vdorov za pametna mesta, ki temelji na povezovanju naprav IoT v gručo.
(Tran et al., 2018)	Eksperiment	NGIDS-DS, ADFA--LD	Sistem za zaznavanje vdorov, ki deluje na vseh računalnikih določenega omrežja in tako preprečuje, da bi onesposobitev glavnega sistema omogočila napadalcu prost dostop.
(Jones in Straub, 2017)	Pregledni članek	-	Sistem nevronskih mrež, ki preprečuje prevzetje nadzora nad roboti.
(Moradpoor et al., 2017)	Eksperiment	SpamAssassin, 8000 e-mailov	Razločevanje phishing sporočil elektronske pošte glede na njihove metapodatke.
(Alom in Taha, 2017)	Eksperiment	KDD-99	Gruča naprav, namenjenih sistemu za zaznavanje vdorov z nevronskimi mrežami v računalniškem oblaku.
(Wang et al., 2017)	Eksperiment	MNIST, CIFAR-10	Razvoj nevronskih mrež, odpornih na vse znane napade.

Iz rezultatov, predstavljenih v zgornji tabeli, je razvidno, da je velika večina raziskav empiričnih (33 eksperimentalnih in 7 preglednih člankov). Pri vzorcih eksperimentov so prevladovali KDD-99, NSL-KDD in UNSW-NB15. KDD-99 je najstarejša baza napadov, iz leta 1999. Sestavljena je iz štirih napadov, in sicer IPSweep, Nmap, PortSweep in Satan, ter vsebuje 41 lastnosti, ki jih nevronske mreže lahko analizirajo. NSL-KDD je posodobljena različica KDD-99, z nekaj novejšimi podatki in brez podvojenih primerov. UNSW-NB15 je najnovejša baza, pripravljena leta 2015. Vsebuje posodobljene metode napadov in boljšo normalizacijo podatkov. Sestavljena je iz 49 značilnosti. V pregledanih raziskavah so bili ti vzorci običajno razbiti na učne množice za učenje algoritma in testne množice za preverjanje uspešnosti algoritma. V nekaj raziskavah se je za ocenjevanje uspešnosti algoritma uporabilo točkovanje F1. To je sistem točkovanja, ki ocenjuje natančnost (tj. število pravilnih odločitev, deljenih z vsemi pravilnimi rezultati) in možnost odpoklica (tj. število pravilnih rezultatov, deljenih z vsemi relevantnimi rezultati).

5 RAZPRAVA

Čeprav raziskovalni področji nevronskih mrež in kibernetske varnosti ločeno obstajata že vrsto let, je njuno združevanje postalo aktualno šele v zadnjih treh letih. Lastnosti nevronskih mrež, kot so fleksibilnost, delovanje z veliko podatki in prilagodljivost, jim nudijo občutne prednosti pred klasičnimi sistemi, kot so npr. sistemi za statično analizo kode. *Fleksibilnost* se nanaša na to, da se lahko nevronske mreže uporabi tako za regresijske (npr. ocenjevanje stopnje tveganj) kot klasifikacijske (npr. sistemi za prepoznavo zlonamerne programske opreme) probleme, saj se enaki modeli uporabljajo za reševanje drugačnih problemov. V nasprotju s klasičnimi sistemi nevronske mreže zelo dobro delujejo z *veliko količino podatkov*. Klasični sistemi namreč ob naraščajočem številu naprav, prometa in vrst napadov niso več sposobni zagotavljati zadovoljivo natančnih odločitev v realnem času na podlagi pogosto nelinearnih podatkov. Nevronske mreže naslavljajo ta problem z razdeljevanjem problemov v plasti, s čimer se bistveno izboljša delovanje z veliko količino podatkov. *Prilagodljivost* se kaže v prilagajanju modela na nove situacije. *Če* so rezultati premalo natančni ali se potrebuje nov vnosni podatek (npr. zaradi nove vrste kibernetskega napada), se model med uporabo spremeni in ponovno nauči, s čimer je tudi takoj ponovno pripravljen za uporabo.

Čeprav so se nevronske mreže dobro izkazale na določenih področjih, kot so npr. prepoznava glasu (Alom et al., 2019), medicina (Tran et al., 2018) in klasifikacija besedila (Vinayakumar et al., 2018), pa se njihova uporaba širi relativno počasi. Prepoznali smo dva pomembnejša razloga. Prvi razlog je *lastnost črne škatle*, ki pomeni, da je praktično nemogoče razumeti, kako je nevronska mreža prišla do svoje rešitve. To močno oteži odkrivanje napak in njihovo nadaljnje raziskovanje. Drugi razlog je *pomanjkanje podatkov za učenje* nevronskih mrež. Kot pri prej omenjenih področjih ima področje kibernetske varnosti ogromno število podatkov, ki bi lahko bili uporabljeni pri učenju mrež. Vendar pa je večina teh neoznačenih in nekategoriziranih, bi lahko pripisali pomanjkanju strokovnjakov na tem področju. V medicini je vsak primer bolezni dokumentiran in evidentiran, kar posledično pomeni ogromno število vzorcev, ki so pripravljeni za učenje. V kibernetski varnosti se poleg tega vsak dan pojavi na tisoče novih vzorcev in primerov že obstoječih groženj, kar oteži nadaljnjo pripravo ustreznih vzorcev. Uporabljeni vzorci pri pregledanih znanstvenih prispevkih se pogosto ponavljajo, saj primanjkuje kakovostnih javnih baz, ki bi jih bilo mogoče uporabiti za učenje nevronskih mrež. Ker učenje nevronskih mrež zahteva posebno prilagojene in normalizirane baze podatkov, ki morajo biti hkrati tudi dovolj velike in relevantne, imajo nevronske mreže v boju z novejšimi vrstami napadov pogosto slabše rezultate kot klasični pristopi. Deloma bi lahko pomanjkanje kakovostnih podatkovnih baz za učenje pripisali relativno nedavni popularizaciji nevronskih mrež v kibernetski varnosti, deloma pa težavam pri zajemu podatkov iz realnih kibernetskih napadov. Zaradi pomanjkanja novejših in bolj kakovostnih baz je iz predstavljenega pregleda literature težko oceniti dejansko uporabnost nevronskih mrež v realnem svetu, saj je večina raziskav narejenih na starih podatkih in v kontroliranem okolju. Za boljšo oceno uporabnosti nevronskih mrež v kibernetski varnosti bi bile zelo koristne raziskave in testi v realnih kontekstih (Vinayakumar et al., 2017).

Rezultati pričujočega pregleda literature nudijo tako teoretične kot tudi praktične implikacije. Med teoretične lahko štejemo enega novejših sistematičnih pregledov literature na področju, ki je znanstvenoraziskovalno v razcvetu predvsem v zadnjih treh letih. Pomembnejša praktična implikacija predstavlja vpogled v trenutno stanje uporabe tehnologije umetne inteligence v področju kibernetske varnosti. Predstavlja tudi smer razvoja, trenutno problematiko in bralcu predstavi osnovne tehnične principe delovanja te tehnologije.

Prikazana raziskava ima nekatere omejitve. Prvič, pregled literature je bil omejen na dela, ki so bila objavljena v zadnjih treh letih. Drugič, omejili smo se na pregled znanstvenih člankov, ki so indeksirani zgolj v angleškem jeziku in dveh najobsežnejših bazah znanstvenih del (WoS in Scopus). Pregled v več bazah, v daljšem časovnem razponu in dodatnih jezikih bi lahko razširil vpogled v trenutno stanje raziskav na tem področju.

6 ZAKLJUČEK

Nevronske mreže imajo potencial na različnih področjih kibernetske varnosti, kot so zaznava zlonamerne programske opreme, zaznava in preprečevanje vdorov, samodejno iskanje ranljivosti v sistemih, če naštejemo le nekatere. V okviru raziskave smo ugotovili, da je ena izmed največjih ovir za adopcijo nevronskih mrež na področju kibernetske varnosti pomanjkanje ustreznih baz za učenje, saj potrebujejo nevronske mreže ogromno število podatkov za uspešno delovanje.

Neposredna primerjava uporabe nevronskih mrež s klasičnimi pristopi na področju kibernetske varnosti je težavna, saj različne raziskave kljub uporabi enakih podatkovnih baz uporabljajo različne metode za pridobitev lastnosti teh vzorcev. Prav tako ne obstaja standardni sistem evaluacije teh sistemov. Čeprav je najpogosteje primerjana natančnost sistema, bi s primerjavo te in klasičnimi pristopi, ali celo drugimi raziskavami z uporabo nevronskih mrež, dobili enostranski rezultat, ki ne bi upošteval drugih lastnosti, kot so kompleksnost in čas učenja, poraba procesne moči, možnost integracije (Xin et al., 2018) ipd.

Nevronske mreže so kljub temu pomembne za prihodnost kibernetske varnosti. Ne omogočajo le učinkovitejših obstoječih pristopov (zaznava in analiza zlonamerne kode, zaznava vdorov in preprečevanje ipd.), temveč omogočajo tudi nove metode pristopov k varnosti, ki prej niso bili mogoči (zaznava lažnih novic, zaznave phishing spletnih strani, prepoznava notranjih groženj ipd.).

UPORABLJENI VIRI

Ait Tchakoucht, T. in Ezziyyani, M. (2018). Building a fast intrusion detection system for high-speed-networks: Probe and DoS attacks detection. *Procedia Computer Science, 127*, 521–530. doi:10.1016/j.procs.2018.01.151

Al Hawawreh, M., Rawashdeh, A. in Alkasassbeh, M. (2018). An anomaly-based approach for DDoS attack detection in cloud environment. *International Journal of Computer Applications in Technology, 57*(4), 312–324. doi:10.1504/IJCAT.2018.10014729

Alom, M. Z. in Taha, T. M. (2017). Network intrusion detection for cyber security using unsupervised deep learning approaches. V *2017 IEEE National Aerospace and Electronics Conference* (str. 63–69). IEEE.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S. in Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics, 8*(3), 292–358. doi:10.3390/electronics8030292

Baykara, M. in Gürel, Z. Z. (2018). Detection of phishing attacks. V *2018 6th International Symposium on Digital Forensic and Security* (str 1–5). NYC: IEEE.

Bratko, I. (1986). *Prolog programming for artificial intelligence* (First ed.). Addison-Wesley.

Bulavas, V. (2018). Investigation of network intrusion detection using data visualization methods. V *59th International Scientific Conference on Information Technology and Management Science of Riga Technical University* (str. 1–6). IEEE.

Cakir, B. in Dogdu, E. (2018). Malware classification using deep learning methods. V *Proceedings of the ACMSE 2018 Conference on – ACMSE* (str. 1–5). ACM.

Chattopadhyay, M., Sen, R. in Gupta, S. (2018). A comprehensive review and meta-analysis on applications of machine learning techniques in intrusion detection. *Australasian Journal of Information Systems, 22*(1995), 1–27. doi:10.3127/ajis.v22i0.1667

Clark, G., Doran, M. in Glisson, W. (2018). A malicious attack on the machine learning policy of a robotic system. V 1*7th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/ 12th IEEE International Conference on Big Data Science and Engineering* (str. 516–521). IEEE.

Cordonsky, I., Rosenberg, I., Sicard, G. in David, E. O. (2018). DeepOrigin: End-to-end deep learning for detection of new malware families. 2018 International Joint Conference on Neural Networks (str. 1–7). IEEE.

Demidov, R. A., Pechenkin, A. I., Zegzhda, P. D. in Kalinin, M. O. (2018). Application model of modern artificial neural network methods for the analysis of information systems security. *Automatic Control and Computer Sciences, 52*(8), 965–970. doi:10.3103/S0146411618080072

Diro, A. A. in Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems, 82*, 761–768. doi:10.1016/j.future.2017.08.043

Geluvaraj, B., Satwik, P. M. in Ashok Kumar, T. A. (2019). The future of cybersecurity: Major role of artificial intelligence, machine learning, and deep learning in cyberspace. V S. Smys, R. Bestak, J. I.-Z. Chen in I. Kotuliak (ur.), *International Conference on Computer Networks and Communication Technologies* (str. 739–747). Springer.

Hai, Q. T. in Hwang, S. O. (2018). Detection of malicious URLs based on word vector representation and ngram. *Journal of Intelligent & Fuzzy Systems, 35*(6), 5889–5900. doi:10.3233/JIFS-169831

Handa, A., Sharma, A. in Shukla, S. K. (2019). Machine learning in cybersecurity: A review. *WIREs Data Mining and Knowledge Discovery, 9*(4), 1–7. doi:10.1002/widm.1306

Homayoun, S., Ahmadzadeh, M., Hashemi, S., Dehghantanha, A. in Khayami, R. (2018). BoTShark: A deep learning approach for botnet traffic detection. V A. Dehghantanha, M. Conti in T. Dargahi (ur.), *Cyber Threat Intelligence* (str. 137–153). Springer. doi.org/10.1007/978-3-319-73951-9_7

Jones, A. in Straub, J. (2017). Using deep learning to detect network intrusions and malware in autonomous robots. V I. V. Ternovskiy in P. Chin (ur.), *Cyber Sensing 2017*. doi:10.1117/12.2264072

Karatas, G., Demir, O. in Koray Sahingoz, O. (2018). Deep learning in intrusion detection systems. V *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism* (str. 113–116). IEEE.

Kim, K. in Aminanto, M. E. (2017). Deep learning in intrusion detection perspective: Overview and further challenges. V *2017 International Workshop on Big Data and Information Security* (str. 5–10). IEEE.

Liu, Q., Li, P., Zhao, W., Cai, W., Yu, S. in Leung, V. C. M. (2018). A survey on security threats and defensive techniques of machine learning: A data driven view. *IEEE Access, 6,* 12103–12117. doi:10.1109/ACCESS.2018.2805680

Lu, J. in Wong, R. K. (2019). Insider threat detection with long short-term memory. V *Proceedings of the Australasian Computer Science Week Multiconference on – ACSW 2019* (str. 1–10). doi:10.1145/3290688.3290692

Moradpoor, N., Clavie, B. in Buchanan, B. (2017). Employing machine learning techniques for detection and classification of phishing emails. V *2017 Computing Conference* (str. 149–156). IEEE.

Mouhoub, M., Sadaoui, S., Ait Mohamed, O. in Ali, M. (2018). Erratum to: Recent trends and future technology in applied intelligence. V M. Mouhoub, S. Sadaoui, O. Ait Mohamed in M. Ali (ur.), *Recent trends and future technology in applied intelligence* (str. E1–E1). Springer International Publishing. doi:10.1007/978-3-319-92058-0_87

Moustafa, N., Turnbull, B. in Choo, K.-K. R. (2018). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. *IEEE Internet of Things Journal, 6*(3), 4815–4830. doi:10.1109/JIOT.2018.2871719

Nguyen, S.-N., Nguyen, V.-Q., Choi, J. in Kim, K. (2018). Design and implementation of intrusion detection system using convolutional neural network for DoS detection. V *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing - ICMLSC '18* (str. 34–38). doi:10.1145/3184066.3184089

Parameshwarappa, P., Chen, Z. in Gangopadhyay, A. (2018). Analyzing attack strategies against rule-based intrusion detection systems. V *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking – Workshops ICDCN '18* (str. 1–4). ACM.

Ponkarthika, M. in Saraswathy, V. R. (2018). Network intrusion detection using deep neural networks. *Asian Journal of Applied Science and Technology, 2*(2), 665–673.

Potluri, S., Henry, N. F. in Diedrich, C. (2017). Evaluation of hybrid deep learning techniques for ensuring security in networked control systems. V *22nd IEEE International Conference on Emerging Technologies and Factory Automation* (ETFA) (str. 1–8). IEEE.

Qabajeh, I., Thabtah, F. in Chiclana, F. (2018). A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review, 29*, 44–55. doi:10.1016/j.cosrev.2018.05.003

Qamar, A., Karim, A. in Chang, V. (2019). Mobile malware attacks: Review, taxonomy & future directions. *Future Generation Computer Systems, 97*, 887–909. doi:10.1016/j.future.2019.03.007

Rosenberg, I., Sicard, G. in David, E. (2018). End-to-End deep neural networks and transfer learning for automatic analysis of nation-state malware. *Entropy, 20*(5), 390. doi:10.3390/e20050390

Russell, S. J. in Norvig, P. (2020). *Artificial intelligence: A modern approach* (Third ed.). Pearson.

Sergio Ordoñez, G. in Cesar Guerra, T. (2018). Prototype of a security system with artificial intelligence using neural networks and evolutionary algorithms. V F. Torres Guerrero, J. Lozoya-Santos, E. Gonzalez Mendivil, L. Neira-Tovar, P. G. Ramírez Flores in J. Martin-Gutierrez (ur.), *Smart technology* (str. 31–39). Springer. doi.org/10.1007/978-3-319-73323-4_4

Singh, L. in Hofmann, M. (2017). Dynamic behavior analysis of android applications for malware detection. V *2017 International Conference on Intelligent Communication and Computational Techniques (ICCT)* (str. 1–7). IEEE.

Spaulding, J. in Mohaisen, A. (2018). Defending internet of things against malicious domain names using D-FENS. V *2018 IEEE/ACM Symposium on Edge Computing (SEC)* (str. 387–392). doi:10.1109/SEC.2018.00051

Tran, N. N., Sarker, R. in Hu, J. (2018). An approach for host-based intrusion detection system design using convolutional neural network. V N. N. Tran, R. Sarker in J. Hu (ur.), *Mobile networks and management* (str. 116–126). Springer. doi:10.1007/978-3-319-90775-8_10

Viet, H. N., Van, Q. N., Trang, L. L. T. in Nathan, S. (2018). Using deep learning model for network scanning detection. V *Proceedings of the 4th International Conference on Frontiers of Educational Technologies* (str. 117–121). ACM.

Vinayakumar, R., Soman, K. P. in Poornachandran, P. (2017). Evaluation of recurrent neural network and its variants for intrusion detection system (IDS). *International Journal of Information System Modeling and Design, 8*(3), 43–63. doi:10.4018/IJISMD.2017070103

Vinayakumar, R., Soman, K. P. in Poornachandran, P. (2018). Evaluating deep learning approaches to characterize and classify malicious URL's. *Journal of Intelligent & Fuzzy Systems, 34*(3), 1333–1343. doi:10.3233/JIFS-169429

Wang, Q., Guo, W., Zhang, K., Ororbia, A. G., Xing, X., Liu, X. in Giles, C. L. (2017). Adversary resistant deep neural networks with an application to malware detection. V *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (str. 1145–1153). ACM.

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H. in Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. IEEE *Access, 18*(6), 35365–35381. doi:10.1109/ACCESS.2018.2836950

Zhang, Q., Zhang, M., Chen, T., Sun, Z., Ma, Y. in Yu, B. (2019). Recent advances in convolutional neural network acceleration. *Neurocomputing, 323*, 37–51. doi:10.1016/j.neucom.2018.09.038

Zhu, H.-J., You, Z.-H., Zhu, Z.-X., Shi, W.-L., Chen, X. in Cheng, L. (2018). Droid-Det: Effective and robust detection of android malware using static analysis along with rotation forest model. *Neurocomputing, 272*, 638–646. doi:10.1016/j.neucom.2017.07.030

O avtorjih:

Črt Uršič, študent programa Informacijska varnost na Fakulteti za varnostne vede Univerze v Mariboru. Raziskovalno se ukvarja s kibernetsko varnostjo in strojnim učenjem.

Anže Mihelič, doktorski kandidat na Fakulteti za računalništvo in informatiko ter Pravni fakulteti Univerze v Ljubljani; asistent, Fakulteta za varnostne vede Univerze v Mariboru; raziskovalec, Fakulteta za matematiko in računalništvo, Fern Universität in Hagen. Njegovi raziskovalni interesi obsegajo tehnične, zasebnostne ter psihološke vidike informacijske in kibernetske varnosti.

Dr. Simon Vrhovec, docent, Fakulteta za varnostne vede Univerze v Mariboru. Raziskovalno se ukvarja s kibernetsko varnostjo.