UMĚLÁ INTELIGENCE A KYBERBEZPEČNOST KRITICKÉ INFORMAČNÍ INFRASTRUKTURY

ARTIFICIAL INTELLIGENCE AND CYBERSECURITY OF THE CRITICAL INFORMATION INFRASTRUCTURE

Josef Bernátek

Abstrakt

Příspěvek pojednává o využití umělé inteligence pro zajištění kyberbezpečnosti prvků kritické informační infrastruktury a rizicích z její případné nevhodné aplikace nebo zneužití ze strany škodlivých útočníků. V první části příspěvku jsou zmíněny počátky vývoje umělé inteligence, definovány základní pojmy jako umělá inteligence, strojové učení, kyberbezpečnost a kritická informační infrastruktura. V následující části jsou zhodnoceny přínosy při aplikaci umělé inteligence pro ochranu kritické informační infrastruktury s praktickými příklady jejího využití a rovněž uvedena rizika plynoucí z její nevhodné aplikace, případně zneužití škodlivým útočníkem se stanovením doporučení pro eliminaci těchto rizik. Na závěr je zmíněna nutnost kontinuálních inovací a aplikací dostupných technologií pro ochranu kybernetického prostoru.

Klíčová slova: *umělá inteligence, kyberbezpečnost, hrozba*

Abstract

The paper deals with the use of artificial intelligence to ensure the cybersecurity of critical information infrastructure elements and the risks of its inappropriate application or misuse by malicious attackers. In the first part of the paper, the beginnings of artificial intelligence development are mentioned, basic terms such as artificial intelligence, machine learning, cybersecurity, and critical information infrastructure are defined. The following section evaluates the benefits of the application of artificial intelligence for the protection of critical information infrastructure with practical examples of its use, as well as the risks arising from its inappropriate application, or abuse by malicious attackers with the establishment of recommendations to eliminate these risks. Finally, the need for continuous innovation and application of available technologies for cyberspace protection is mentioned.

Key words: *artificial intelligence, cybersecurity, threat*

1 ÚVOD DO PROBLEMATIKY

Je to již téměř 70 let od představení Turingova testu, který měl za cíl vyhodnotit schopnosti stroje v porovnání s lidskou myslí. Rok 1950 tedy můžeme považovat za počátek vývoje umělé inteligence.[1] V současné době se v souvislosti s umělou inteligencí dostávají do popředí termíny jako strojové učení, hluboké učení a kvantové počítání. Technologické společnosti deklarují u svých produktů prvky založené na bázi umělé inteligence. Zda bude mít umělá inteligence pro lidstvo pouze přínosný charakter zůstává otázkou. Před dvěma lety se věhlasný fyzik Stephen Hawking o vzniku umělé inteligence vyjádřil jako o nejhorší události v historii civilizace, pokud společnost nenajde cestu ke kontrole jejího vývoje. S obavou před možným zneužitím autonomních zbraní, které by se mohly stát třetí revolucí ve válčení se v otevřeném dopise adresovaném Organizaci spojených národů vyjádřilo více než 100 osobností z technologického světa.[2]

Cílem příspěvku je stručné shrnutí současných poznatků o umělé inteligenci ve vztahu ke kyberbezpečnosti kritické informační infrastruktury se stanovením demonstrativního výčtu doporučení pro eliminaci rizik plynoucích z nevhodné aplikace umělé inteligence nebo jejího zneužití škodlivými útočníky. Příspěvek se nezabývá aplikací umělé inteligence v jiných než bezpečnostních procesech kritické informační infrastruktury.

1.1 Umělá inteligence

Pojem umělá inteligence zahrnuje 3 rozdílné kategorie. Umělou superinteligenci, umělou obecnou inteligenci a umělou úzkou inteligenci. Umělou superinteligenci si lze představit jako schopnost počítače překonat člověka ve všech oblastech. Výsledný systém by mohl mít schopnosti jaké je možno spatřit v Holywoodském sci-fi thrilleru I Am Mother. Umělá obecná inteligence představuje schopnost systému na inteligenční úrovni člověka se stejnými schopnostmi zvládání problémů vyžadujícími učení a uvažování. Předpokládá se, že této úrovně nebude s velkou pravděpodobností dosaženo v příštích 15–20 letech. Umělá úzká inteligence představuje schopnost systému zpracovávat široký rozsah dat a detekovat v nich vzorce a vztahy, které by byly pro člověka obtížné nebo nemožné. Ve třetím případě se tedy jedná o překonání člověka pouze ve specifických úlohách jako rozpoznání škodlivého obsahu v příloze e-mailu nebo detekce anomálií v síťovém provozu.[3] Této úrovně je v současné době u některých systémů již dosaženo a produkty na ní založené jsou v praxi využívány.

Oblast umělé inteligence zahrnuje široké spektrum přístupů, které je třeba aplikovat u zájmových technologiích pro získání požadovaných vlastností. Nejčastěji se lze setkat s její podmnožinou označovanou jako strojové učení, které může mít rovněž podmnožinu hlubokého učení.[3]

1.2 Strojové učení

Strojové učení je nejen podmnožinou umělé inteligence, ale i nástrojem k dosažení všech 3 výše uvedených kategorií. Systémy využívající algoritmy strojového učení nemusí být vždy v učení efektivnější než člověk, ale jsou rychlejší ve zpracování velkého objemu dat. Jedná se o techniku umožňující počítačovým systémům pomocí množiny algoritmů analýzu dat za účelem rozeznání zájmových informací bez lidské interakce. V praxi se strojové učení využívá při rozpoznávání obličejů, analýze sentimentu v novinových článcích, detekci fake news na sociálních sítích, podvodů při bankovních operacích nebo při překladu jazyků u internetového překladače.[4]

Proces analýzy dat je umožněn na základě předchozího učení počítačového systému. Primárně se ve strojovém učení užívají v podstatě 2 přístupy k učení, učení pod dohledem a učení bez dohledu. Při učení pod dohledem jsou počítačovému systému předkládána předem označená data. Člověk tedy v tomto procesu působí jako učitel pro umělou inteligenci. Výstupy jsou předem známy a data jsou označeny příslušnými odpověďmi člověkem. U učení bez dohledu se umělá inteligence učí z datových setů, které nejsou předem označeny a reaguje na případné společné znaky v těchto datech. Předchozí označení dat člověkem je mnohdy z časového hlediska téměř nemožné. Jedná se například o obsah síťového provozu. Využití učení bez dohledu je uplatnitelné například v detekci anomálií v tomto provozu. Mezi další přístupy ke strojovému učení lze zařadit kombinaci učení s učitelem a bez učitele, kde jsou počítačovému systému předkládána částečně označená data a zpětnovazebné učení, kde dochází k učení počítačového systému na základě metody pokusu a omylu.[5]

1.3 Kyberbezpečnost a kritická informační infrastruktura

Pod pojem kyberbezpečnost lze zahrnout souhrn právních, organizačních, technických a vzdělávacích opatření směřujících k zajištění ochrany kyberprostoru. Kyberprostorem je poté digitální prostředí umožňujícího vznik, zpracování a výměnu informací. Může se jednat jednak o prostor ve formě informačního systému, služby nebo sítě elektronických komunikací.[6]

Kritickou infrastrukturou jsou systémy a služby, které by v případě poruchy měly závažný dopad na bezpečnost státu, jeho ekonomiku, veřejnou správu a zajištění základních životních potřeb obyvatelstva. Kritickou informační infrastrukturou jsou prvky nebo systémy prvků kritické infrastruktury v odvětví komunikační a informační systémy. Předpokladem k určení prvků nebo systémů jako kritické informační infrastruktury je naplnění legislativou stanovených průřezových a odvětvových kritérií.[6] V praxi se může jednat o informační systém kritický pro řízení jaderné elektrárny nebo bankovní operace pro ekonomiku státu významné bankovní společnosti.

2 PŘÍNOS A HROZBA PRO KRITICKOU INFORMAČNÍ INFRASTRUKTURU

Již z hlediska obecného popisu možností umělé inteligence je zřejmý silný potenciál jejího využití, ale i zneužití. O jejím možném zneužití ze strany útočníků jako multiplikátoru jejich schopností, který umožní aktérům automatizovat velkou řadu složitých úkonů, varoval i Národní úřad pro kybernetickou a informační bezpečnost ve Zprávě o stavu kybernetické bezpečnosti České republiky za rok 2018 publikované 30. září 2019.[7] V následujících podkapitolách budou shrnuty možné způsoby využití pro kybernetickou ochranu kritické informační infrastruktury a rovněž i rizika plynoucí z nesprávného užití umělé inteligence nebo její dostupnosti škodlivě konajícím aktérům.

2.1 Pozitivní aplikace umělé inteligence

Mezi pozitivní aplikaci umělé inteligence při ochraně kritické informační infrastruktury lze uvést následující oblasti. Filtrování nevyžádané e-mailové pošty. Škodlivé přílohy v nevyžádané e-mailové poště jsou jedním z primárních vektorů útoku nejen na prvky kritické informační infrastruktury. Jejich rozlišení od legitimní komunikace je z daného důvodu klíčové. V případě standardních detekčních metod nevyžádané pošty jsou využívány seznamy škodlivých odesílatelů nebo škodlivých příloh dle kontrolních součtů nebo jejich názvů. V případě strojového učení se může detekční systém učit na základě znalosti obsahu předchozí e-mailové komunikace kategorizované jako škodlivé a legitimní. Případné budoucí e-mailové zprávy tak budou automatizovaně detekčním systémem vyhodnoceny se znalostí předchozích označených e-mailů s určitou mírou pravděpodobnosti jako škodlivé, pokud se budou podobat předchozím škodlivým e-mailům. Označení bude realizováno i přes skutečnost, že jejich odesílatel nebude na seznamu škodlivých odesílatelů a e-mailová zpráva bude obsahovat detekčnímu systému dosud neznámou škodlivou přílohu. Na obdobném principu jsou založeny i následující oblasti aplikace umělé inteligence.

Škodlivý soubor určený k útoku na kritickou informační infrastrukturu nemusí být zaslán pouze e-mailem. Může se jednat i o útok cílící na prvek kritické informační infrastruktury oddělený od internetu. V praxi nejznámějším kybernetickým útokem na informační systém oddělený od internetu je Stuxnet, který za pomocí cíleně vytvořeného škodlivého kódu cílil na průmyslové systémy určené k obohacování uranu v Íránském Natanzu.[8] V takovém případě by mohlo jako nosič škodlivého obsahu sloužit přenosné paměťové médium ve formě USB disku. Mnohdy i prvky kritické informační infrastruktury vyžadují aktualizaci softwaru nebo

obousměrný manuální přenos dat mezi systémy. V takovém případě je nanejvýš vhodné mít implementovány procesy zahrnující antivirové kontroly datových nosičů, které budou připojeny do chráněných informačních systémů. Antivirové programy se primárně spoléhají na definice, které nemusí zahrnovat dosud neznámé druhy škodlivého softwaru. Pomocí strojového učení se může antivirový program naučit rozpoznat škodlivé typy souborů na základě předem označených vzorků i přes skutečnost, že se bude jednat o dosud antivirovému programu neznámý typ malwaru.

Detekce síťového provozu představuje další významnou formu pozitivní aplikace umělé inteligence při obraně prvků kritické informační infrastruktury. I přes aplikaci všech reálně aplikovatelných kontrolních opatření na vstupu, může dojít k průniku škodlivého softwaru do systému. Současné typy malwaru jsou obvykle postaveny na modulárním principu s tím, že po průniku do systému dochází k následné komunikaci na řídící server útočníka, vykonávání jeho pokynů, případně stahování dalších součástí škodlivého softwaru. Z daného důvodu je klíčová včasná detekce škodlivé komunikace malwaru na řídící server útočníka. Aplikací strojového učení v bezpečnostních prvcích síťového provozu tak může dojít k včasné detekci škodlivé komunikace i přes skutečnost, že se jedná o detekčnímu systému dosud neznámou formu komunikace na dosud neznámé řídící servery útočníka.

2.2 Rizika plynoucí z nevhodné implementace nebo zneužití umělé inteligence

Negativní aspekty umělé inteligence v souvislosti s kyberbezpečností kritické informační infrastruktury můžeme rozdělit do 2 kategorií. První kategorie bude zahrnovat rizika plynoucí z nesprávné implementace systémů založených na umělé inteligenci. Druhá kategorie bude poté zahrnovat rizika plynoucí ze zneužití systémů umělé inteligence širokým spektrem útočníků.

V případě nevhodné implementace umělé inteligence se můžeme setkat s nevhodně zvolenou, případně nevhodně označenou sadou dat pro strojové učení. Informační systém se tak i přes bezchybně konstruované algoritmy pro strojové učení může na základě nevhodných vstupních dat naučit nesprávně vyhodnocovat vstupní data. Rizikem pak mohou být upozornění na detekci škodlivého obsahu bez relevance, nedekování skutečného škodlivého obsahu a v krajním případě např. u detekčních prvků založených na umělé inteligenci s možností aktivního zásahu i blokace legitimního e-mailu, síťového provozu nebo aktualizačního souboru prvku kritické informační infrastruktury. Pro eliminaci těchto rizik jsou klíčové vhodně zvolené metody výuky autonomního systému, datový set, jeho označení a rovněž finální testování chování systému před jeho implementací do produkčního prostředí.

U rizik plynoucích ze zneužití umělé inteligence se můžeme setkat se zneužitím těchto technologií ze strany útočníků rozdělených dle kategorií jako v případě současných kybernetických útoků. Je předpoklad, že státní aktéři budou mít přístup k nejpokročilejším formám umělé inteligence. Teroristické organizace, kyberkriminálníci a hacktivisté poté k méně pokročilým formám systémů založených na umělé inteligenci s tím, že budou s největší pravděpodobností odkázáni na outsourcing těchto technologií na dark webu v souladu s dnes dostupnými službami hackerů na objednávku. Eliminaci těchto rizik lze docílit např. tím, že budou v případě kybernetické bezpečnosti chráněného prvku kritické informační infrastruktury aplikovány všechny dostupné technologie a procesy v souladu s doporučeními z předem provedené analýzy rizik. Pro ochranu každého prvku kritické informační infrastruktury je nutno v první fázi vykonat analýzu rizik a následně se rozhodnout jaká úroveň rizika je pro provozovatele daného prvku akceptovatelná a jaká již nikoliv.

3 ZÁVĚR

V první části příspěvku byly zmíněny počátky vývoje umělé inteligence, definovány základní pojmy jako umělá inteligence, strojové učení, kyberbezpečnost a kritická informační infrastruktura. V následující části byly zhodnoceny přínosy při aplikaci systémů založených na umělé inteligenci pro ochranu prvků kritické informační infrastruktury s praktickými příklady jejího využití. Rovněž byla uvedena rizika plynoucí z nevhodné implementace umělé inteligence, případně jejího zneužití škodlivým útočníkem se stanovením doporučení pro eliminaci demonstrativně uvedeného výčtu rizik.

Je zcela zřejmé, že 100% kybernetické bezpečnosti není reálné dosáhnout stejně jako není možno všechna rizika pro prvky kritické informační infrastrukturu plynoucí z umělé inteligence eliminovat na nulovou úroveň. Vhodnou aplikací doporučení z kvalifikovaně provedené analýzy rizik však můžeme rizika plynoucí z nevhodně implementovaných autonomních systémů nebo zneužití umělé inteligence zmírnit na akceptovatelnou úroveň. Umělá inteligence tak pro nás nebude představovat jen riziko, ale i nástroj k obraně technologií klíčových pro bezpečnost státu, životy a zdraví obyvatelstva a jeho ekonomiku.

Tyto nově dostupné technologie nelze z dlouhodobého hlediska opomíjet. Pokud škodliví útočníci začnou pro usnadnění neoprávněných přístupů do počítačových systémů a dalších kriminálních činů využívat přidané hodnoty, kterou jim umělá inteligence bude nabízet, nezbývá nic jiného než recipročně posílit i kybernetickou bezpečnost chráněných prvků kybernetického prostoru. Je to stejný princip jako v případě konkurenčního boje v ekonomické sféře. Pokud má obchodní společnost zájem na zachování udržitelného růstu a eliminaci rizik z případného bankrotu, musí sledovat kroky konkurence a inovovat nabízené výrobky, aby byly na trhu stále žádané. Pokud při ochraně kritické infrastruktury před kybernetickými útoky zaznamenáme, že útočníci užívají umělou inteligenci k průnikům do námi chráněných systémů, je na místě vybavit příslušné detekční systémy rovněž funkcemi založenými na umělé inteligenci. Je to nekončící boj mezi obráncem a útočníkem.

Použitá literatura

1. COWLEY, Mike. *Machine Learning: A Complete Guide for Beginners to Mastering the Fundamentals of ML. Learn about Machine Learning, Artificial Intelligence, Deep Learning and their Application in Finance & Business*. 2019. ISBN 978-1700195128.
2. Stephen Hawking says A.I. could be 'worst event in the history of our civilization'. *CNBC* [online]. 6. 11. 2017 [cit. 2019-11-23]. Dostupné z: www.cnbc.com.
3. *Introduction to Artificial Intelligence for Security Professionals* [online]. The Cylance Press, 2017 [cit. 2019-11-23]. ISBN 978-0-9980169-2-4.
4. HALDER, Soma a Sinan ODEZIMIR. Hands-On Machine Learning for Cybersecurity: Safeguard your system by making your machines intelligent using the Python ecosystem. Packt Publishing, 2018. ISBN 978-1788992282.
5. Understanding the Strategic and Technical Significance of Technology for Security: Implications of AI and Machine Learning for Cybersecurity [online]. The Hague Security Delta, 2019 [cit. 2019-11-24].
6. JIRÁSEK, Petr, Luděk NOVÁK a Josef POŽÁR. *Výkladový slovník kybernetické bezpečnosti: Cyber security glossary*. Třetí aktualizované vydání. Praha: Policejní akademie ČR v Praze, 2015. ISBN 978-80-7251-436-6.
7. Zpráva o stavu kykybernetické bezpečnosti České republiky za rok 2018 [online]. Národní úřad pro kybernetickou a informační bezpečnost, 2019 [cit. 2019-11-24].

8. LANGNER, Ralph. Stuxnet: Dissecting a Cyberwarfare Weapon: Dissecting a Cyberwarfare Weapon. *IEEE Security & Privacy*, 2011. ISSN 1558-4046.

Kontaktní údaje
Ing. Josef Bernátek
České vysoké učení technické v Praze, Fakulta biomedicínského inženýrství
nám. Sítná 3105, 272 01 Kladno
email: bernajo1@fbmi.cvut.cz