410—425

Mezinárodněprávní přičitatelnost aktů způsobených nasazením autonomních kapacit kybernetické obrany

Jakub Vostoupal*

Abstrakt: Rozmach AI technologií se dotkl mnoha odvětví, kybernetickou bezpečnost nevyjímaje. Nekonečný boj mezi útočníky a obránci klade stále vyšší nároky na rychlost a nutná interakce mezi člověkem a softwarem je proto limitující. A přestože se zatím není nutné bát systémů po vzoru Skynet z filmu Terminátor, autonomní kapacity schopné kybernetických útoků již existují. Je tak na místě se zaměřit mimo jiné i na to, jak může jejich nasazení ovlivnit stávající právní řád. Autor se v tomto článku konkrétně zaměřuje na otázku, jakou výzvu představuje nasazení autonomních kapacit kybernetické obrany pro možnost přičítání kybernetických útoků státům podle mezinárodního práva. V prvé řadě se zaměřuje na rozvoj stávajících kapacit a vylučuje, alespoň prozatím, nutnost zaobírat se plně autonomními systémy, které by si samy volily cíle. Místo toho se více zaměřuje na povahu kontroly nad autonomní kapacitou a dochází k názoru, že navzdory tomu, že autonomní systémy mohou činit dílčí kroky samostatně, stále se jedná jenom o pokročilý nástroj, který slouží potřebám lidí, případně států. Na základě této myšlenky se pak zabývá odpovědností za akty způsobené nasazením autonomních kapacit a jejím potenciálem sloužit jako základ přičtení, což nakonec kombinuje s přičítacími pravidly dle mezinárodněprávní odpovědnosti států a shledává, že tato kombinace je kompatibilní. Zapojení nestátních aktérů ovšem zůstává problematické i v tomto kontextu, stejně jako v případě „tradičních" kybernetických útoků.

Klíčová slova: kybernetická bezpečnost, autonomní kapacity, kybernetické útoky, přičitatelnost, mezinárodní odpovědnost

Úvod

Historické městečko Natanz ve středním Íránu, konkrétně tamější zařízení na obohacení uranu, se v roce 2010 zarylo hluboko do dějin kybernetické bezpečnosti jako místo, které se stalo cílem jednoho z nejsofistikovanějších kybernetických útoků digitálního věku – červa jménem Stuxnet.[1] Jednalo se o mimořádně komplexní malware, který stál za značným nárůstem poruchovosti centrifug na obohacování uranu (uvádí se, že tento červ stál až za tisíci zničenými centrifugami, což mělo zásadní finanční dopad na tamější jaderný program[2]). Kromě komplexity a sofistikovanosti spočívající mj. v počtu zneužitých zranitelností nultého dne a ukradených digitálních klíčů se Stuxnet vyznačoval i určitou (na tu dobu pokročilou) mírou autonomie.[3] Interní síť většiny jaderných zařízení totiž

* Mgr. Jakub Vostoupal, Ph.D., výzkumník, Ústav práva a technologií, Právnická fakulta Masarykovy univerzity. E-mail: jakub.vostoupal@law.muni.cz. ORCID: https://orcid.org/0000-0002-1669-9931. Autor by tímto rád poděkoval anonymním recenzentům za podnětné připomínky k článku.

1 ZETTER, K. *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. New York, NY, USA: Crown Publishers, 2014, kap. 1.

2 Ibidem, kap. 17.

3 HAATAJA, S. Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility. In: LIIVOJA, R. – VÄLJATAGA, A. (eds.). *Autonomous Cyber Capabilities under International Law*. Tallinn, Estonia: NATO CCD COE Publications, 2021, s. 261–262.

bývá chráněna tzv. *air gap*, tedy plným odpojením od internetu, což v tomto konkrétním případě sice nezabránilo infekci, ale znemožňovalo to interakci s operátorem v reálném čase.[4]

Jenže přestože se jednalo o přelomový útok, který změnil vnímání kyberprostoru a jeho možností pro mnoho státních i nestátních aktérů a fakticky odstartoval nové závody ve zbrojení, zdůraznil též zásadní slabinu systému mezinárodní odpovědnosti – problém přičitatelnosti v kyberprostoru. Navzdory dopadům, které tento útok měl (ostatně se místy mluvilo i o aktivaci čl. 51 Charty OSN, tedy práva na sebeobranu), totiž Irán nikdy nepřistoupil k oficiálnímu přičtení na mezinárodní scéně.[5] A to ani v momentě, kdy došlo ve Spojených státech k úniku utajovaných dokumentů, které potvrzovaly, že Stuxnet vzešel ze společné operace USA a Izraele jménem Olympijské hry (což bylo posléze potvrzené i Edwardem Snowdenem v rozhovoru pro Der Spiegel[6]).[7]

Je ovšem nutné si přiznat, že přičítání kybernetických útoků v souladu s pravidly mezinárodněprávní odpovědnosti je problematické samo o sobě (v odborné literatuře se používá zmíněný „problém přičitatelnosti“[8]).[9] V tomto článku se tak zaměřím na otázku, zda současný rozvoj umělé inteligence a autonomních kapacit a jejich rozšířené nasazování i v oblasti kybernetické bezpečnosti může představovat další či dokonce přelomovou výzvu pro možnost přičítat kybernetické útoky státům podle mezinárodního práva. Při hledání odpovědi se nejdříve zaměřuji na relevantní vlastnosti autonomních kapacit kybernetické obrany, aspekty jejich kontroly (a vzhledem k častým rozdílům v terminologii i na samotný pojem „autonomie“), dále na odpovědnost za škodu způsobenou autonomními systémy jakožto potenciální základ pro přičtení, načež analyzuji jednotlivé módy přičitatelnosti podle práva mezinárodní odpovědnosti a jak může jejich funkci ovlivnit nasazení autonomních kapacit.

---

4 Ibidem.

5 Gholamreza Jalali, šéf íránské civilní obrany, prohlásil v rozhovoru pro Kayhan Daily, že dle vyšetřování stojí za útokem Stuxnet USA a „Sionistický režim“ (Izrael). Veřejně přitom obvinil také společnost Siemens ze zapojení do tohoto útoku. Na mezinárodní scéně se však Irán nedovolával mezinárodní odpovědnosti ani jednoho aktéra a útok tak zůstal nepřičten. Více viz HACK, J. – MOSTAFAVI, R. Iran accuses Siemens over Stuxnet virus attack. *Reuters*. 2011. Dostupné z: https://www.reuters.com/article/idUSTRE73G0NB/.

6 LOPEZ, P. „Als Zielobjekt markiert“. *Der Spiegel*. 2013, Nr. 28.

7 NAKASHIMA, E. – WARRICK, J. Stuxnet was work of U. S. and Israeli experts, officials say. *Washington Post*. 2012. Dostupné z: https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html. SANGER, D. E. Obama Order Sped Up Wave of Cyberattacks Against Iran. *The New York Times*. 2012. Dostupné z: https://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html. GROLL, E. Obama's General Pleads Guilty to Leaking Stuxnet Operation online. *Foreign Policy*. 17. 10. 2016 [cit. 26. 12. 2023]. Dostupné z: https://foreignpolicy.com/2016/10/17/obamas-general-pleads-guilty-to-leaking-stuxnet-operation/.

8 V anglickém jazyce se stále častěji používá pojem *„the problem of cyber-attribution“*. Viz např. PAYNE, C. – FINLAY, L. Addressing Obstacles to Cyber-Attribution: A Model based on State Response to Cyber-Attack. *The George Washington International Law Review*. 2017, Vol. 49. BERGHEL, H. On the Problem of (Cyber) Attribution. *Computer – IEEE Computer Society*. 2017, Vol. 50, No. 3.

9 BERGHEL, H. *On the Problem of (Cyber) Attribution*, s. 84–85. BOEBERT, W. E. A Survey of Challenges in Attribution. *Proceedings of a Workshop on Deterring CyberAttacks: Informing Strategies and Developing Options for U.S. Policy*. Washington, D.C.: The National Academies Press, 2010, s. 43. Dostupné z: http://www.nap.edu/catalog/12997.html. BANKS, W. Cyber Attribution and State Responsibility. *International Law Studies*. 2021, Vol. 1039, No. 97, s. 1040–1042.

1. Význam autonomních kapacit pro kybernetickou bezpečnost a obranu

Kybernetická bezpečnost[10] je aktivitou často reaktivní povahy.[11] Schopnost rychle reagovat a adaptovat se na změny v prostředí má tím pádem obrovskou hodnotu, a to jak pro útočníky, tak pro obránce.[12] Nutná interakce mezi softwarem a člověkem je ovšem v mnoha ohledech limitující, což nevyhnutelně vede ke snaze ji omezit či přesunout do fáze vývoje technologie, a nikoliv do jejího nasazení a realizace.[13] To vede k útokům komplexnějším (např. Stuxnet), samostatnějším a proměnlivějším (mezi současné hrozby v souvislosti s umělou inteligencí patří tzv. „malware generovaný AI" či adaptivní kód[14]).[15] Navíc vzestup generativních umělých inteligencí zásadně snížil hranici dovedností a schopností, které musí útočník ovládat, aby provedl úspěšný kybernetický útok (i v případě ChatGPT je stále možné donutit tento model k přípravě škodlivého kódu), což samozřejmě pozici obránců výrazně stěžuje.

Neustálý boj o rychlejší a efektivnější útočné i obranné kapacity v kyberprostoru vede ke snaze o přinejmenším semiautomatizaci, příp. plnou autonomii daných technologií.[16] Není ale nutné si ihned představit umělou inteligenci po vzoru Skynetu z kultovního filmu Terminátor. Jak Tammet upozorňoval v roce 2021, kyberobranné autonomní kapacity se nenacházely na takovém stupni sofistikovanosti, aby bylo nutné řešit potenciální právní dopady jejich kompletní nezávislosti na lidské kontrole a související odpovědnost za škody.[17] I dnes pak platí, že značná část kyberobranných aktivit (ať už čistě obranných, nebo útočných) je stále příliš komplexní na to, aby mohla být efektivně plně nahrazena autonomním softwarem (značná část technologií je v současné době polo-automatizovaná).[18]

---

10 *„Kybernetická bezpečnost představuje souhrn organizačních, politických, právních, technických a vzdělávacích opatření a nástrojů směřujících k zajištění zabezpečeného, chráněného a odolného kyberprostoru v České republice, a to jak pro subjekty veřejného a soukromého sektoru, tak pro širokou českou veřejnost. Kybernetická bezpečnost pomáhá identifikovat, hodnotit a řešit hrozby v kyberprostoru, snižovat kybernetická rizika a eliminovat dopady kybernetických útoků, informační kriminality, kyberterorismu a kybernetické špionáže ve smyslu posilování důvěrnosti, integrity a dostupnosti dat, systémů a dalších prvků informační a komunikační infrastruktury. Hlavním smyslem kybernetické bezpečnosti je pak ochrana prostředí k realizaci informačních práv člověka."* Viz Národní bezpečnostní úřad. *Národní strategie kybernetické bezpečnosti České republiky na období let 2015 až 2020*, s. 5. Dostupné z: https://nukib.cz/download/publikace/strategie_akcni_plany/narodni_strategie_kb_2015-2020.pdf.

11 GRAVES, D. J. Reactive vs. Proactive Cybersecurity: Which Strategy is Best? online. *Fortinet Blog*. 22. 2. 2019 [cit. 26. 12. 2023]. Dostupné z: https://www.fortinet.com/blog/industry-trends/reactive-vs--proactive-cybersecurity--5-reasons-why-traditional-.

12 What is Security Automation? [online]. *Palo Alto Networks*. 2022 [cit. 26. 12. 2023]. Dostupné z: https://www.paloaltonetworks.com/cyberpedia/what-is-security-automation.

13 MCFARLAND, T. The Concept of Autonomy. In: Liivoja, R., Väljataga, A. (eds). *Autonomous Cyber Capabilities under International Law*. Tallinn, Estonia: NATO CCD COE Publications, 2021, s. 18–19.

14 V obou případech se jedná o produkt umělé inteligence, který se může v průběhu času měnit a vyhnout se tak detekci, která byla dosud založena na rozeznávání částí kódu. Místo toho bude tak nutné využít umělou inteligenci i na straně obránců a spoléhat se na behaviorální malware analýzu. Více viz ANGELO, D. D. The Dark Side of AI in Cybersecurity – AI-Generated Malware [online]. *Palo Alto Networks Blog*. 15. 5. 2024 [cit. 6. 2. 2025]. Dostupné z: https://www.paloaltonetworks.com/blog/2024/05/ai-generated-malware/.

15 TAMMET, T. Autonomous Cyber Defence Capabilities. In: LIIVOJA, R. – VÄLJATAGA, A. (eds.). *Autonomous Cyber Capabilities under International Law*. Tallinn, Estonia: NATO CCD COE Publications, 2021, s. 36–37.

16 Security Automation: 5 Best Practices & More online. *CrowdStrike*. 2023 [cit. 26. 12. 2023]. Dostupné z: https://www.crowdstrike.com/cybersecurity-101/security-automation/. What is Security Automation? [online]. *Palo Alto Networks*. 2022 [cit. 26. 12. 2023]. Dostupné z: https://www.paloaltonetworks.com/cyberpedia/what-is-security-automation.

17 TAMMET, T. *Autonomous Cyber Defence Capabilities*, s. 42–46.

18 LOHN, A. et al. *Autonomous Cyber Defense*. Center for Security and Emerging Technology, 2023. Dostupné z: https://

Pro úplné odstranění potřeby lidské interakce je buď nutné, aby bylo tvůrci autonomního softwaru dobře známé prostředí, do kterého se software může dostat, a toto prostředí nebylo nadmíru proměnlivé (interakci lze poté nahradit relativně jednoduchou sérií příkazů „pokud–tak", přičemž příkladem může být i jednoduchý firewall), nebo je potřeba již pokročilejších algoritmických funkcí, přičemž dostupné kapacity zatím na tento úkol nestačí.[19] Určitou vizí budoucnosti jsou v tomto ohledu kombinace dnešních AI kapacit s již staršími technologiemi založenými na tzv. *kybernetických rozhodovacích systémech*, které *„kombinují řadu nástrojů, technik a expertních znalostí za účelem vytvoření autonomního systému, který je schopen automatické detekce zranitelností, vytváření exploitů a bezpečnostních záplat bez potřeby lidského zásahu."*[20] Zjednodušeně řečeno se tedy jedná o kapacity, které kombinují obranné i útočné aspekty, nejedná se však zatím o efektivní náhražku lidských odborníků.[21] To si ale Tammet vysvětluje zejména tím, že je daleko jednodušší vytvořit autonomní nástroj sloužící k útoku než k obraně (či hybridní).[22]

Vzhledem k tomuto aspektu a s přihlédnutím k rychlému rozvoji AI technologií, u nějž Stroppa upozorňuje, že by bylo chybné se již nyní nevěnovat právním konotacím jejich širšího nasazení,[23] se v tomto článku zaměřím na následky nasazení autonomních „kyberkapacit" primárně ofenzivního charakteru. Takto obecný pojem používám z důvodu, že není nutné ani účelné analyzovat všemožné konkrétní druhy kyberbezpečnostních technologií souvisejících s umělou inteligencí či autonomií, a místo toho využiji skutečnosti, že v zahraniční odborné literatuře se pro kategorii obdobných technologií začalo používat označení, které by bylo možné do češtiny přeložit jako *„autonomní kapacity kybernetické obrany."*[24] Pro lepší představu, co se skrývá pod tímto označením, je možné uvést příklad (přestože již relativně dávný), který jsem již zmínil v úvodu tohoto článku – Stuxnet.[25]

---

cset.georgetown.edu/publication/autonomous-cyber-defense/. MILES, I. et al. *Reinforcement Learning for Autonomous Resilient Cyber Defence*. Las Vegas, USA: Informa PLC, 2024. Dostupné z: https://i.blackhat.com/BH-US-24/Presentations/US-24-MilesFarmer-ReinforcementLearningForAutonomousResilientCyberDefence-wp.pdf. MARCOS, W. Self-Defending Networks: The Promise of Autonomous Cyber Defense Systems [online]. *CyberMatter*. 21. 9. 2024 [cit. 6. 2. 2025]. Dostupné z: https://cybermatters.info/ai-cybersecurity/autonomous-cyber-defense-systems/.

19 VYAS, S. et al. *Automated Cyber Defence: A Review*. arXiv, 2023, s. 111:6-111:8. Dostupné z: http://arxiv.org/abs/2303.04926. MILES, I. et al. *Reinforcement Learning for Autonomous Resilient Cyber Defence*. MARCOS, W. *Self-Defending Networks: The Promise of Autonomous Cyber Defense Systems*. LOHN, A. et al. *Autonomous Cyber Defense*.

20 BROOKS, T. N. *Survey of Automated Vulnerability Detection and Exploit Generation Techniques in Cyber Reasoning Systems*. Cham: Springer International Publishing, 2019, s. 1083.

21 JIANG, H. – CHOI, T. – KO, R. K. L. Pandora: A Cyber Range Environment for the Safe Testing and Deployment of Autonomous Cyber Attack Tools. In: THAMPI, S. M. et al. (eds.). *Security in Computing and Communications*. Singapore: Springer Singapore, 2021, s. 3–5. Dostupné z: https://link.springer.com/10.1007/978-981-16-0422-5_1. MCFARLAND, T. *The Concept of Autonomy*, s. 17. LOHN, A. et al. *Autonomous Cyber Defense*.

22 TAMMET, T. *Autonomous Cyber Defence Capabilities*, s. 37–38.

23 STROPPA, M. Legal and Ethical Implications of Autonomous Cyber Capabilities: A Call for Retaining Human Control in Cyberspace. *Ethics and Information Technology*. 2023, Vol. 25, No. 1, s. 2–3.

24 Z angl.: *autonomous cyber defence capabilities* (v literatuře se též vyskytuje *autonomous cyber defense* či *autonomous cyber defence systems*, u poslední zmíněné se však jedná o poněkud zužující výběr). Daná kategorie byla upřednostněna před dalšími kategoriemi zaužívanými v oblasti kybernetické obrany, jako jsou *autonomous systems*, *autonomous weapons systems* či *autonomous agent*, neboť na základě analýzy provedené autorským týmem knihy *Autonomous Cyber Capabilities under International Law* pod NATO CCD COE tyto kategorie nedostatečně reflektují kombinaci současné vyspělosti daných technologií, míru dosažené autonomie a specifika kybernetické obrany. Ibidem, s. 3–4. TAMMET, T. *Autonomous Cyber Defence Capabilities*, s. 36–38. VYAS, S. et al. *Automated Cyber Defence: A Review*, s. 111:1–111:3. LOHN, A. et al. *Autonomous Cyber Defense*. MILES, I. et al. *Reinforcement Learning for Autonomous Resilient Cyber Defence*. MARCOS, W. *Self-Defending Networks: The Promise of Autonomous Cyber Defense Systems*.

25 TAMMET, T. *Autonomous Cyber Defence Capabilities*, s. 37.

1.1 „Automatické, automatizované, autonomní“ aneb co je autonomie?

Pro lepší pochopení fungování autonomních kapacit je v prvé řadě vhodné se zaměřit na to, co vlastně znamená pojem „autonomie“. Ovšem navzdory tomu, že tento pojem je jedním z ústředních prvků materie umělé inteligence a autonomních systémů, nemálo autorů se jeho definici buďto úplně vyhýbá, nebo se autonomní chování snaží klasifikovat na základě určitých objektivních kritérií, u nichž může neopatrný přístup vést k dezinterpretaci celého pojmu. Příkladem první skupiny může být Matthew Scherer, který se ve svém článku otázky autonomie sice dotýká, ale přistupuje k ní jako k notorietě a více se zaměřuje na pochopení pojmu „inteligence“.[26] Paul Scharre, který zastupuje druhou skupinu, pak ve svém pojednání o autonomních zbraních přistupuje k rozlišení relevantních technologických kapacit na *„automatické, automatizované a autonomní“*,[27] ke kterým přistupuje na základě binárního „zapadá/nezapadá do dané kategorie“.[28] Takový závěr pak činí na základě posuzování specifických kritérií, což však v kombinaci s jeho „binárním“ přístupem může vést k nechtěným závěrům. Pokud bych např. posuzoval autonomii technologií na základě kritéria schopnosti samostatné reakce na změnu v prostředí, ve kterém kapacita operuje, naplní jej všechny technologie regulované skrze zpětnou vazbu (*closed-loop* systém, např. termostat).[29] Jako vhodnější se mi tedy jeví McFarlandův přístup, dle kterého je nutné autonomii vnímat spíše jako otázku míry.[30]

Co se pak týká samotné definice autonomie, je nutné uznat, že v kontextu autonomních kapacit kybernetické obrany je řada definic poněkud nevyhovujících.[31] Pro tyto účely je za jednu z nejfunkčnějších považována[32] následující definice amerického ministerstva obrany: *„Autonomie je schopnost či soubor schopností, které umožňují, aby byly určité systémy či aktiva automatizovány či, v rámci naprogramovaných hranic, samosprávné.“*[33, 34] Dobře totiž ilustruje, že přestože klíčovým atributem autonomní kapacity je schopnost seberegulace či samosprávy v určitém prostředí a na základě určitých vstupů operátora, zásadní roli hrají i nadále člověkem určené hranice. Nejedná se tedy o technologie, které by byly mimo lidskou kontrolu či na člověku plně nezávislé.[35]

---

26 SCHERER, M. U. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*. 2016, Vol. 29, No. 2, s. 361–363.

27 SCHARRE, P. Autonomous Weapons and Operational Risk. *Ethical Autonomy Project*. 2016, s. 12.

28 Ibidem, s. 13–14.

29 MCFARLAND, T. *The Concept of Autonomy*, s. 16.

30 Ibidem, s. 16 a 31.

31 To se např. týká i definice, kterou používá unijní návrh Aktu o umělé inteligenci (v době psaní článku však nemám přístup k verzi schválené po trialogu a závěry se tak mohou lišit).

32 Např. HUGUES, J. et al. *Assuring Increasingly Autonomous Cyber-Physical Systems*. [online]. [cit. 26. 12. 2023]. Dostupné z: https://www.sei.cmu.edu/publications/annual-reviews/2022-year-in-review/year_in_review_article.cfm?customel_datapageid_315013=494108. MCFARLAND, T. *The Concept of Autonomy*, s. 17–19. STROPPA, M. *Legal and Ethical Implications of Autonomous Cyber Capabilities: A Call for Retaining Human Control in Cyberspace*, s. 3.

33 Z anglického originálu: *„Autonomy is a capability (or a set of capabilities) that enables a particular action of a system to be automatic or, within programmed boundaries, 'self-governing'.“* Viz US Department of Defense, Defense Science Board. *Task Force Report: The Role of Autonomy in DoD Systems*. 2012. Dostupné z: https://irp.fas.org/agency/dod/dsb/autonomy.pdf.

34 Je také možné vycházet z rámce vytvořeného pod záštitou Severoatlantické aliance, který je ale v zásadě analogický. Více viz KOTT, A. et al. *Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture, Release 2.0*. DEVCOM – Army Research Laboratory, 2019. Dostupné z: https://apps.dtic.mil/sti/pdfs/AD1080471.pdf.

35 MCFARLAND, T. *The Concept of Autonomy*, s. 14.

1.2 Autonomie, předvídatelnost, kauzalita a kontrola

Pokud je tedy autonomní kapacita stále závislá na člověku, další klíčovou otázkou je právě povaha a rozsah kontroly, kterou je možné nad takovou kapacitou vykonávat. Jestliže autonomii interpretuji jako škálu dle McFarlandova modelu, různá míra autonomie a metody dosažení této autonomie s sebou přinášejí různé stupně předvídatelnosti chování dané kapacity, možnosti sledování příčinné souvislosti a míry kontroly nad danou kapacitou, což ovlivňuje jak otázky odpovědnosti, tak potenciální přičitatelnost.[36]

Při pohledu na sofistikovanost samostatného chování např. u firewallu či šachových botů, jež jsou v zásadě konfigurovány na základě příkazů „pokud–tak", je patrné, že se kontrola nad chováním takovéto kapacity pouze „přesouvá" předdefinováním z reálného času do přípravné fáze, nikoliv úplně odstraňuje či přechází na kapacitu samotnou.[37] Míra nejistoty ohledně chování takto nadefinovaného software je pak minimální, a tím i náročnost sledování příčinné souvislosti pro odpovědnostní účely.[38]

I u komplexnějších algoritmických řešení sice dochází v zásadě jenom k „přesunu" kontroly, míra nejistoty ale v závislosti na komplexitě roste. Jak McFarland popisuje, primárním důvodem nejistoty ohledně výsledku chování takové technologie jsou chyby, kolize, komplexnost použitého algoritmu či otrávené datasety.[39] Scherer pak doplňuje, že i samotná lidská schopnost předvídat chování algoritmů je limitována nižším „výpočetním výkonem" než jakým vládne autonomní kapacita.[40] Ve výsledku však oba docházejí k závěru, že míra kontroly člověka nad technologií tímto nemusí být (poněkud paradoxně) citelně zasažena.[41]

V daném aspektu jsou totiž autonomní kapacity podobné performativním pravidlům.[42] Přestože způsob realizace je flexibilní, rámcově nejistý (či pro vnějšího pozorovatele přímo náhodný) a pod vlivem autonomní kapacity, cíl je stále stanoven člověkem, v daném kontextu kombinací vstupů tvůrce a operátora, a daná kapacita jej bude sledovat a přizpůsobovat tomu jednotlivé dílčí aktivity.[43] Předvídatelnost jednotlivých kroků autonomní kapacity tak může být sice nízká, ale pokud se nejedná o následek chyby,[44] kauzálně bude stále možné provázat tyto aktivity s cílem, ke kterému směřují, a tím, kdo jej stanovil a rozhodl o nasazení dané kapacity.

---

36 SCHERER, M. *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*. 2015, s. 364–368. Dostupné z: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777.

37 MCFARLAND, T. *The Concept of Autonomy*, s. 21–22.

38 STROPPA, M. *Legal and Ethical Implications of Autonomous Cyber Capabilities: A Call for Retaining Human Control in Cyberspace*, s. 3–4.

39 MCFARLAND, T. *The Concept of Autonomy*, s. 37.

40 SCHERER, M. *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, s. 364.

41 MCFARLAND, T. *The Concept of Autonomy*, s. 37. SCHERER, M. *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, s. 365.

42 K performativním pravidlům více viz COGLIANESE, C. *Performance-Based Regulation: Concepts and Challenges*. National Academy of Sciences, Transportation Research Board, Washington, DC, 2016. Dostupné z: http://onlinepubs.trb.org/onlinepubs/pbr/coglianese041416.pdf.

43 VYAS, S. et al. *Automated Cyber Defence: A Review*, s. 111:1–111:3.

44 I chyby je možné přičíst prostřednictvím příčinné souvislosti konkrétnímu subjektu, jenom by se mohlo jednat o jinou entitu, než kdo o nasazení kapacity rozhodnul. V mezinárodním kontextu by měl nasazující subjekt nést odpovědnost i za chyby autonomní kapacity (přestože na národní úrovni by tato odpovědnost mohla spíše jít za výrobcem, což je však vhodnější řešit případným regresem), neboť to přispěje k odpovědnějšímu a stabilnějšímu prostředí. Více k tomu v další kapitole.

2. Odpovědnost autonomní kapacity jako základ přičtení

Při nasazení autonomních kapacit v rámci kybernetických útoků mezi státy dává otázka přičitatelnosti smysl v zásadě ze dvou pohledů. V první řadě je otázkou, jak by bylo možné přičíst akt způsobený nasazením autonomní kapacity státu. Na tom je pak závislá otázka druhá, a to zda by autonomní kapacita byla schopna takové přičtení provést pro účely potenciální odplaty. To je otázka sice komplexnější, ale odpověď je paradoxně jednodušší, neboť jak Stroppa upozorňuje ve svém článku, pokud by autonomní kapacita přičítala na základě technických stop,[45] bylo by možné tento proces relativně snadno zneužít pro tzv. útok pod falešnou vlajkou (*false-flag attack*) a zatáhnout tak do konfliktu nevinný stát.[46] Navíc je nutné počítat s tím, že přičtení samotné je nesmírně komplexní proces, jehož podmínky by musely být *a priori* očekávány a zaneseny do autonomní kapacity, která by zároveň musela vládnout situačním povědomím na nesmírné úrovni.[47] Navíc, jak upozorňuji a více rozebírám ve své disertační práci (a navazuji tím na závěry profesora Tsagouriase),[48] právní přičtení a související dovolání se mezinárodněprávní odpovědnosti za kybernetický útok dosud nebylo nikdy ze strany jakéhokoliv státu učiněno.[49] Z těchto důvodů je nyní nerealistické, že by autonomní kapacita přičítala sama, či že by to vůbec mělo být bez určitého stupně lidské kontroly možné. Považuji však za důležité zdůraznit, že problém s přičtením by stíhal obě strany fiktivního konfliktu, čehož by bylo možné využít. Na útok, který je problematicky přičitatelný, tak může být odpovězeno útokem, který je problematicky přičitatelný. Tento přístup sice destabilizuje mezinárodní společenství a pohybuje se mimo limity mezinárodního práva, bylo by však naivní jej zde nezmínit.

S tím se vrátím zpět k první otázce, tedy zda je možné přičíst akt autonomní kapacity státu. Jedním z primárních důvodů existence problému přičitatelnosti jsou problémy spojené s dokazováním činů provedených v kyberprostoru a nalezením spojitosti mezi konkrétním aktérem a státem. Zásadní roli tak bude hrát situace, v jaké dojde k nasazení autonomní kapacity, a zejména subjekt, který o daném rozhodne (více v podkapitole 3.1). Zapojení těchto kapacit však nemusí nutně představovat výraznější komplikaci, spíše jenom další krok v přičítacím řetězci. Pokud je totiž možné stanovit odpovědnost za škody způsobené nasazením autonomních kapacit, je možné identifikovat subjekt, jehož specifické postavení vůči státu je posléze zkoumáno a na základě kterého je možné uvažovat o přičtení. Stanovení odpovědnosti je tím pádem možné vnímat jako jeden z úvodních kroků přičítacího procesu.

Bohužel otázky spojené se stanovením takovéto odpovědnosti stále nejsou úplně a uspokojivě vyřešeny. Je např. možné se bavit o odpovědnosti tvůrce, operátora/vlastníka, další zúčastněné strany či dokonce autonomního systému (což přichází velmi teoreticky v úvahu v případě, že by byla takovému systému připsána subjektivita). Nejistým je i samotný režim odpovědnosti, tedy zda vyžadovat zavinění nebo trvat naopak na režimu

45 Zohlednění dalších aspektů na operační či přímo strategické úrovni není při současné sofistikovanosti autonomních kapacit kybernetické obrany realistické.

46 STROPPA, M. *Legal and Ethical Implications of Autonomous Cyber Capabilities: A Call for Retaining Human Control in Cyberspace*, s. 3–4.

47 Ibidem, s. 4.

48 TSAGOURIAS, N. – FARRELL, M. Cyber Attribution: Technical and Legal Approaches and Challenges. *European Journal of International Law*. 2020, Vol. 31, No. 3, s. 946.

49 Dostupné z: https://is.muni.cz/auth/th/lfryu/.

objektivní odpovědnosti. Report CERRE o odpovědnosti umělé inteligence[50] tento názorový střet detailně analyzuje a uvádí různá pozitiva a negativa. Obdobné postřehy a závěry pak zahrnují i Report Evropské komise o bezpečnosti umělé inteligence, internetu věcí a robotiky[51] a Bílá kniha o umělé inteligenci.[52]

Vzhledem k tomu, že unijní Akt o umělé inteligenci se nevztahuje na vojenské technologie,[53] pod které by autonomní kapacity kybernetické obrany mohly být, alespoň teoreticky, zařazeny,[54] rozhodl jsem se pro kombinaci závěrů výše zmíněných reportů, principů, na kterých je založen Akt o umělé inteligenci, a doporučení Evropského parlamentu k režimům civilní odpovědnosti umělé inteligence.[55, 56] Přikláním se tak k názoru, že odpovědnost za akty autonomních kapacit by měla fungovat v režimu objektivní odpovědnosti s možností liberace[57] zacílené na osobu „deployera", tedy subjekt rozhodující o nasazení autonomní kapacity, kontrolující tuto kapacitu a mající z jejího provozu prospěch.[58] A zvláště v kontextu problémů s přičitatelností a mj. též rizika eskalace a „přetečení" útoku mimo kýžený cíl (to se ostatně stalo i v případě Stuxnetu, jehož agresivnější verze nakonec unikla z jaderného zařízení v Natanzu a začala se šířit po světě, což vedlo k odhalení celé operace), je nezbytné, aby si dané kapacity jak z právních, tak z etických důvodů[59] i *pro futuro* zachovaly aspekt lidské kontroly.[60]

Aplikací tohoto přístupu je možné se vyhnout odpovědnosti výrobce, která v kontextu autonomních kapacit kybernetické obrany není vhodná, a zároveň se pojistit, že kdo podobné nástroje implementuje, bude za jejich akty odpovědný, ať už se bude jednat o následky úmyslné, nedbalostní, či přímo nechtěné (pokud nedojde k liberaci). Vzhledem k sofistikovanosti a samostatnosti daných kapacit zatím není nutné se vůbec zabývat debatami o odpovědnosti kapacity jako takové.

---

50 BUITEN, M. – DE STREEL, A. – PEITZ, M. *EU Liability Rules for the Age of Artificial Intelligence: Report*. CERRE, 2021. Dostupné z: https://cerre.eu/wp-content/uploads/2021/03/CERRE_EU-liability-rules-for-the-age-of-Artificial-Intelligence_March2021.pdf.

51 *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*. European Commission, 2020. Dostupné z: https://ec.europa.eu/info/sites/default/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf.

52 *White Paper on Artificial Intelligence – A European approach to excellence and trust*. European Commission, 2020. Dostupné z: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

53 Čl. 2, odst. 3 Aktu o umělé inteligenci.

54 V případě těchto technologií bude záležet na povaze dané technologie a jejího využívání – může se jednat o čistě vojenskou technologii, technologii dvojího užití či plně civilní. Osobně však nečekám zařazení mezi civilní technologie, zejména kvůli regulatorním limitacím a transparentnosti, které nejsou v souladu s povahou nasazování autonomních kapacit kybernetické obrany. Dle mého názoru je tak nepravděpodobné, že by členské státy nechaly takovéto technologie pod rozsah Aktu o umělé inteligenci vůbec spadnout, neboť by tím došlo k určitému omezení schopností států zajišťovat národní bezpečnost (další sekce citovaného odstavce). Ani vojenské technologie však nejsou z tohoto pohledu bezproblémové, a to ani na úrovni mezinárodního práva, více viz KUDLÁČKOVÁ, I. – WALLACE, D. – HARAŠTA, J. *Cyber Weapons Review in Situations Below the Threshold of Armed Conflict*. Estonia, Tallinn: NATO CCDCOE Publications, 2020. Dostupné z: https://ieeexplore.ieee.org/document/9131728/.

55 Dostupné z: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html.

56 Ani toto doporučení by se na autonomní kapacity kybernetické obrany pravděpodobně nevztahovalo, je však založeno na přístupu, který hodnotím v daném kontextu jako vhodnou kombinaci kontroly a efektivity autonomních kapacit.

57 Pro ilustraci: subjekt rozhodující o nasazení autonomní kapacity kybernetické obrany se zprostí odpovědnosti za negativní následky plynoucí z jejího nasazení, pokud prokáže, že vynaložil veškeré úsilí, které bylo možné požadovat, aby daným negativním následkům předešel.

58 Čl. 4.1 a 8.1 doporučení Evropského parlamentu k režimům civilní odpovědnosti umělé inteligence.

59 STROPPA, M. *Legal and Ethical Implications of Autonomous Cyber Capabilities: A Call for Retaining Human Control in Cyberspace*, s. 4–6.

60 Po vzoru čl. 14 návrhu Aktu o umělé inteligenci (verze ze dne 14. 6. 2023).

3. Pravidla přičitatelnosti

Jak jsem zmínil i výše, z pohledu mezinárodního práva je přičitatelnost kybernetických útoků relativně komplikovanou otázkou, a to nejen při použití autonomních kapacit. První komplikací je stále ještě nedořešená otázka aplikace mezinárodního práva v kyberprostoru (přestože zvláště západní státy se snaží jít v tomto ohledu příkladem[61]). To, ve spojení s relativně nízkou chutí států vyjadřovat se k závaznosti konkrétních pravidel v kyberprostoru a ještě nižší chutí i některých členů Rady bezpečnosti limitovat své kybernetické operace,[62] nutně vede k závěru, že vytvoření specifických pravidel pro účely kyberprostoru je nereálné a místo toho je nutné přizpůsobit výklad stávajícího práva mezinárodní odpovědnosti nové realitě.[63] To je ovšem v některých oblastech, přičitatelnost kybernetických útoků nevyjímaje, přinejmenším neideální a zároveň pomalé (to pak způsobuje stav, kdy na některé situace není možné).

Zmínil jsem, že možnosti přičtení i v případě nasazení autonomních kapacit záleží na situaci a postavení subjektu, který o nasazení rozhodne. To je způsobené tím, že stát je *de facto* toliko konstruktem, kterému nezbývá než jednat prostřednictvím fyzických osob, jejichž jednání je mu pak dle určitých pravidel přičteno.[64] A přestože stejně jako zbytek mezinárodního práva i tato přičítací pravidla jsou primárně obyčejové povahy, jejich systém a fungování jsou detailně popsány v kodifikaci Komise OSN pro mezinárodní právo – v Návrhu článků o odpovědnosti státu za mezinárodně protiprávní činy (dále jen „Návrh článků" nebo „ARSIWA").[65] Je sice nutné uznat, že závaznost ARSIWA je poněkud problematická (a přesahující limity tohoto článku), detailní práce s judikaturou mezinárodních soudů, státní praxí i projevy *opinio juris* ze strany Komise pro mezinárodní právo ale umožňují, přinejmenším v rozsahu, který se týká pravidel přičitatelnosti, považovat ARSIWA za spolehlivý a primární pramen.[66]

---

61 OSULA, A.-M. – KASPER, A. – KAJANDER, A. EU Common Position on International Law and Cyberspace. *Masaryk University Journal of Law and Technology*. 2022, Vol. 16, No. 1.

62 Patrné, poněkud paradoxně, např. v rámci posledního pokusu Ruské federace o předložení návrhu na mezinárodněprávní zakotvení pravidel bezpečného kyberprostoru. Více viz KAJANDER, A. *Unnecessary Repetition: Russia's Latest Attempt at a New UN Convention on Cyberspace*. NATO CCD COE, 2023. Dostupné z: https://ccdcoe.org/library/publications/unnecessary-repetition-russias-latest-attempt-at-a-new-un-convention-on-cyberspace/.

63 SCHMITT, M. – WATTS, S. The Decline of International Humanitarian Law Opinio Juris and the Law of Cyber Warfare. *Texas International Law Journal*. 2015, Vol. 50, No. 2–3, s. 209–211, 220–222.

64 Přičitatelnost je občas nevhodně nazývána subjektivním prvkem státní odpovědnosti, porušení primárního pravidla mezinárodního práva pak prvkem objektivním. Tento aspekt je v kontextu kyberútoků problematický sám o sobě, neboť nepanuje shoda, jak intenzivní porušení může kyberútok představovat. Převažující názor, který je zachycen i v Tallinnském manuálu, je srovnatelnost následků s kinetickým útokem z pohledu použití síly. Polemika nad naplněním objektivního prvku je ovšem mimo limity tohoto článku a pro zbytek mého pojednání budu považovat tento aspekt za naplněný. Viz článek 2 Návrhu článků o odpovědnosti států za mezinárodně protiprávní činy a CRAWFORD, J. *State Responsibility: The General Part*. Cambridge University Press, 2013, s. 113. Dostupné z: https://www.cambridge.org/core/product/identifier/9781139033060/type/book.

65 Navzdory tomu, že se jedná toliko o „Návrh", články jsou široce uznávané jako závazné zachycení mezinárodněprávních zvyků, viz *Responsibility of States for Internationally Wrongful Acts – Comments and information received from Governments and Report of the Secretary-General (A/71/79)*. General Assembly of the United Nations, 2016. Dostupné z: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N16/112/74/PDF/N1611274.pdf?OpenElement.

66 Více viz např. PELLET, A. Ch. 9 The ILC's Articles on State Responsibility for Internationally Wrongful Acts and Related Texts. In: CRAWFORD, J. – OLLESON, S. – PARLETT, K. (eds.). *The Law of International Responsibility*. Oxford: Oxford University Press, 2010.

3.1 Vazba jednajících na stát

Návrh článků rozeznává v kontextu přičitatelnosti v zásadě tři různé konstelace aktérů. Prvním je jednání státních orgánů (včetně územní samosprávy, viz čl. 4) a osob pověřených výkonem státní moci (viz čl. 5), které, pokud učiněno v rámci oficiální kapacity, je připisováno danému státu.[67, 68] Rozsah oficiálních kapacit je pak rozšířen článkem 7 ARSIWA, který zakotvuje přičitatelnost pro jednání státních orgánů i v případech *ultra vires* (dané se uplatní i v případě vojenských aktivit oficiálních armád, a tím pádem i nasazení vojenských technologií). To samé se vztahuje i na jakoukoliv osobu či entitu, která je státní mocí nadána (viz čl. 5 a 7 ARSIWA).[69] Pouze v případech, kdy by tato osoba sledovala svým jednáním čistě soukromé zájmy, by se dané akty relevantnímu státu nepřipsaly. Mezinárodní soudní dvůr pak v rozsudku *Bosenské genocidy* postavil na roveň státnímu orgánu i entity, které jsou *de iure* nestátními aktéry, ale operují v úplné závislosti a pod kompletní kontrolou daného státu.[70]

V pomyslné kontrapozici ke státní moci pak stojí jednání nestátních aktérů. Doktrína mezinárodněprávní odpovědnosti došla napříč historií k pozici obecné nepřičitatelnosti jednání nestátních aktérů (oproti dobám např. starého Říma), neboť státy nemohou (a nechtějí) být bez dalšího odpovědné za jednání jakéhokoliv jednotlivce, jehož chování nekontrolují (tento názor byl akceptován již i za doby absolutismu, kdy se sice předpokládala absolutní moc suveréna nad poddanými, ale zároveň byly akceptovány faktické limitace tohoto ideálu).[71] V současnosti se navíc dá relativně spolehlivě předpokládat, že zakotvení obecné odpovědnosti za jednání nestátních aktérů by vedlo k erozi celého institutu státní odpovědnosti.[72]

Mezi zmíněnými krajními pozicemi se nachází stupeň nejvíce komplikovaný – nestátní aktéři, kteří buď operují v závislosti na státu, s *ex post* uznáním a adopcí jejich aktivit ze strany státu,[73] či se jedná o aktivity polostátních útvarů v případech padlých vlád či utváření nových států. Takové jednání za splnění daných podmínek přičitatelné daným státům je, přičemž v kontextu kybernetických útoků je nejpodstatnějším článek 8 ARSIWA, který stanovuje, že akt nestátního aktéra, který jedná na základě instrukcí, řízení či kontroly[74] určitého státu, je takovému státu přičitatelný.[75]

---

67 To bylo podpořeno např. i v arbitrážím nálezu *Salvador Commercial Company* (s. 477) či v poradním posudku Zvláštního zpravodaje v případu Mezinárodního soudního dvora *Difference Relating to Immunity from Legal Process of a Special Rapporteur of the Commission on Human Rights* (s. 62). Více viz CRAWFORD, J. *State Responsibility: The General Part*, s. 117.

68 Pro ilustraci se může jednat o akt soudu, armády či zpravodajské služby.

69 Podle Tallinnského manuálu se může jednat např. o soukromou společnost pověřenou prováděním kyber-špionáže. Viz SCHMITT, M., NATO Cooperative Cyber Defence Centre of Excellence (eds.). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. Cambridge: Cambridge University Press, 2017, s. 89–90. Dostupné z: https://doi.org/10.1017/9781316822524.

70 *Rozsudek Mezinárodního soudního dvora ve věci Bosenské genocidy (Bosna a Hercegovina v. Srbsko a Černá Hora)*. 2007, odst. 391–393.

71 HESSBRUEGGE, J. The Historical Development of the Doctrines of Attribution and Due Diligence in International Law. *New York University Journal of International Law and Politics*. 2004, Vol. 36, No. 4, s. 277–281.

72 CRAWFORD, J. *State Responsibility: The General Part*, s. 115–116.

73 Při tvorbě článku 11 ARSIWA Komise pro mezinárodní právo chybně interpretovala podkladové rozhodnutí, neboť v původním rozhodnutí je akceptace a uznání vykládáno jako příklad důkazu o zapojení státu do dané operace. Jedná se tedy o jeden z mnoha nutných podkladů pro přičtení, nikoliv o samostatný základ přičtení. Více se tomuto tématu věnuji v rámci své disertační práce.

74 Z anglického „*The conduct of a person or group of persons shall be considered an act of a State under international*

Jestliže se článek 8 stal klíčovým pro vztah problému přičitatelnosti a kybernetických útoků, pak pojmy *instrukce, řízení* a *kontrola* jsou středobodem tohoto problému.[76] Zatímco Návrh článků explicitně uvádí, že se jedná o tři rozdílné kategorie,[77] praxe zatím pojmy *řízení* a *kontrola* příliš nerozlišuje.[78] Podle Crawforda jsou instrukce specifická situace, ve které stát autorizuje, instruuje a pověří nestátního aktéra provedením určité operace v postavení *de facto auxiliaries* (může se jednat např. o soukromou společnost pověřenou podpořením probíhajících vojenských operací[79]).[80] V rámci řízení a kontroly se jedná o obecnější vztah, přičemž je to právě míra kontroly nad jednáním nestátního aktéra, která je určující pro to, zda je možné akt přičíst, či nikoliv. Mezinárodní soudní dvůr se vymezením dostatečné úrovně kontroly pro přičtení mezinárodně protiprávního aktu zabýval v případu *Některých vojenských a polovojenských aktivit v Nikarague a proti ní*. Zde formuloval relativně přísný test *efektivní kontroly*, dle kterého nestačí aktivity toliko podporovat, ale je nutné, aby byl stát aktivně zapojen do plánování (začátku), průběhu i konce operace a byl mimo jiné schopen operaci sám ukončit.[81] Nestátní aktér pak musí být v takovém postavení závislosti, že bez státního zapojení by pro něj nebylo možné, aby danou operaci buď úplně provedl, nebo aby ji provedl na srovnatelné úrovni.[82] Je na místě uvést, že dle Mačáka je tento test téměř nemožné splnit pro účely přičítání kybernetických útoků.[83] Osobně tak považuji tento test za nevyhovující pro dobu digitálního věku, neboť přispívá k možnostem států „schovávat" svou vůli za akty nestátních aktérů. Na to ostatně upozorňoval i soudce Al-Khasewneh ve svém disentu[84] k rozsudku ve věci Bosenské genocidy, ve kterém Mezinárodní soudní dvůr test efektivní kontroly potvrdil.[85]

Test, který je občas označován jako příhodnější, tedy test *celkové kontroly*, byl formulován Mezinárodním soudním tribunálem pro bývalou Jugoslávii v případu *Tadić*. Přístup, který tento test volí, je alespoň zdánlivě vhodnější v kontextu kybernetických útoků, bohužel však byl Mezinárodním soudním dvorem v případu Bosenské genocidy odmítnut.[86] Kvůli tomuto postoji Mezinárodního soudního dvora se v následujícím textu nebudu rozsudkem ve věci *Tadić* zabývat (přestože v některých národních pozicích států k aplikaci

law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct."

75 CRAWFORD, J. *State Responsibility: The General Part*, s. 144–147.

76 CRAWFORD, J. *The International Law Commission's Articles on State Responsibility: Introduction, Text, and Commentaries*. Cambridge, U.K. New York: Cambridge University Press, 2002, s. 110–113.

77 Odst. 7, čl. 8 ARSIWA.

78 CRAWFORD, J. *State Responsibility: The General Part*, s. 145.

79 SCHMITT, M., NATO Cooperative Cyber Defence Centre of Excellence (eds). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2017, s. 95–96.

80 CRAWFORD, J. *The International Law Commission's Articles on State Responsibility: Introduction, Text, and Commentaries*, s. 110.

81 MAČÁK, K. Decoding Article 8 of the International Law Commission's Articles on State Responsibility: Attribution of Cyber Operations by Non-State Actors. *Journal of Conflict and Security Law*. 2016, Vol. 21, No. 3, s. 413.

82 *Rozsudek Mezinárodního soudního dvora ve věci Některých vojenských a polovojenských činností v Nikaragui a proti ní (Nikaragua v. Spojené státy americké)*. 1986, odst. 109.

83 MAČÁK, K. *Decoding Article 8 of the International Law Commission's Articles on State Responsibility: Attribution of Cyber Operations by Non-State Actors*, s. 423–426.

84 *Rozsudek Mezinárodního soudního dvora ve věci Bosenské genocidy (Bosna a Hercegovina v. Srbsko a Černá Hora)*, 2007, s. 217.

85 Ibidem, odst. 391–393.

86 Rád bych podotknul, že MSD tak učinil s velmi chabým odůvodněním, což je zvláště s přihlédnutím k až zarážejícímu kontrastu v práci s citačním aparátem a státní praxí mezi oběma soudy velmi smutné. Ibidem, odst. 402–404.

mezinárodního práva v kyberprostoru a k otázkám přičitatelnosti už došlo k odmítnutí aplikace testu efektivní kontroly).[87]

Ve výsledku tak spočívá přičtení kybernetických útoků provedených nestátními aktéry (pod což je možné podřadit většinu hackerských skupin, které nejsou přímo součástí zpravodajských služeb) na testu, který je právně neaplikovatelný, a testu, který je tak nerealisticky nastavený, že kybernetický útok dle něj pravděpodobně vůbec nelze přičíst. To je ovšem problém přesahující tento článek a v zásadě neovlivňující otázku přičitatelnosti aktů autonomních kapacit kybernetické obrany (protože ta narazí na stejný problém v nezměněné podobě).

Na závěr jenom dodám, že možnost přičitatelnosti aktů státní moci *ultra vires*, kterou zakládá článek 7 ARSIWA, se na nestátní aktéry, kteří by překročili instrukce nebo kontrolu státu, nevztahuje. Crawford k tomuto uvádí, že státy využívající nestátní aktéry většinou nepřebírají riziko za překročení vlastních instrukcí a takové akty pak „utečou" z dosahu přičitatelnosti a mezinárodní odpovědnosti.[88] Pokud by se ovšem nejednalo o čistě náhodné překročení, případně by stát i nadále vystupoval v postavení efektivní kontroly (a jenom by činu z nějakého důvodu nezabránil), pak i v takovém případě je možné daný čin přičíst.[89]

4. Přičitatelnost aktů autonomních kapacit

V prvé řadě je nutné vyjasnit, že využití současných autonomních kapacit kybernetické obrany (ať už by se jednalo o jednoduché technologie, či pokročilé kybernetické rozhodovací systémy) představuje sice přidání dalšího důkazního kroku do přičítacího procesu, nikoliv však dalšího subjektu. Tyto kapacity totiž v zásadě nejsou ničím jiným než pouhými nástroji, které efektivněji, rychleji a snadněji plní přání a cíle lidí.[90]

Bylo by sice možné uvažovat o specifické aplikaci přičítacích kritérií pro nestátního aktéra pod kontrolou státu na samotnou autonomní kapacitu, takový postup by však těmto nástrojům přiznával podstatně vyšší míru subjektivity, než jakou jsou jim státy ochotny přiznat (a než jakou fakticky vládnou).[91] S ohledem na jejich sofistikovanost a samostatnost je v tuto chvíli možné uzavřít, že technologický stav si tento přístup jednoduše nežádá a celou situaci by jenom zbytečně komplikoval. Je také na místě podotknout, že pravidla přičítání mezinárodního humanitárního práva se v zásadě příliš neliší od jejich obdoby v normálním režimu.[92]

V tomto ohledu tedy navazuji na model představený Jonathanem Kwikem na mezinárodní konferenci CyCon 2024, ve kterém se zabývá fungováním útoků ve smyslu mezinárodního humanitárního práva, které jsou způsobeny autonomními zbraňovými systémy.

---

87 Více viz CASSESE, A. The Nicaragua and Tadić Tests Revisited in Light of the ICJ Judgment on Genocide in Bosnia. *European Journal of International Law*. 2007, Vol. 18, No. 4, s. 655–663.

88 CRAWFORD, J. *The International Law Commission's Articles on State Responsibility: Introduction, Text, and Commentaries*, s. 110.

89 Ibidem, s. 113.

90 HAATAJA, S. *Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility*, s. 260–262.

91 *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, 2020.

92 I tomuto aspektu jsem se detailněji věnoval ve své dizertační práci, viz kapitola 3.5. Pro více detailů, viz také CONDORELLI, L. – KRESS, C. Ch. 18 The Rules of Attribution: General Considerations. In: CRAWFORD, J. et al. (eds.). *The Law of International Responsibility*. Oxford: Oxford University Press, 2010, s. 227–228.

Na základě individuální konzultace jsme dospěli k názoru, že jeho model může koncepčně fungovat i v prostředí kyberprostoru, přestože jsou zde samozřejmě odlišnosti, které však přesahují limit tohoto článku (např. rozsah individuálního kybernetického útoku v rámci širší kampaně, která může být spuštěna jedním rozhodnutím/aktem útočníka). Přičítací proces by ale v případě aktů způsobených nasazením autonomních kapacit kybernetické obrany mohl fungovat v zásadě tradičně, jen byl obohacen o kombinaci odpovědnostních kritérií (pro účely stanovení odpovědnosti za akty autonomní kapacity)[93] s přičítacími pravidly.[94] Posuzovala by se konkrétně možnost rozhodovat o nasazení autonomní kapacity,[95] kontrola nad autonomní kapacitou a prospěch z nasazení autonomní kapacity (který také označuji jako cíl).

V případě, že by o nasazení autonomního systému kybernetické obrany rozhodoval státní orgán (např. zpravodajská služba pověřená výkonem kybernetické obrany) či subjekt pověřený výkonem státní moci (soukromá společnost legálně zmocněná k výkonu kybernetické špionáže a obrany), mohou být akty takto nasazených autonomních kapacit státu připsány.[96] V případě státních orgánů ani není nutné se v tomto kontextu příliš zabývat cílem, který je nasazením autonomních aktivit sledován. Z povahy věci by totiž měl být v souladu s politickými a strategickými cíli daného státu, a i kdyby tomu tak nebylo, na možnosti přičíst dané jednání se to nijak neprojeví.[97] Je v zásadě na státu samotném, aby si zajistil kontrolu nad vlastními orgány, což ostatně potvrdil Mezinárodní soudní dvůr v případu Ozbrojených aktivit na území Konga.[98] V případě zmocněných subjektů (čl. 5 ARSIWA) má ale otázka, komu nasazení kapacity prospěje, již zásadnější pozici, neboť pokud z jednání autonomních kapacit netěží stát, ale toliko soukromý subjekt, jednání je nepřipsatelné a odpovědným zůstane jenom samotný subjekt.[99]

V případě nestátních aktérů se posuzování otázky, kdo má pravomoc rozhodnout o nasazení autonomních kapacit projeví mj. do hodnocení závislosti a pouta mezi nestátním aktérem a státem. Výsledek se bude lišit dle toho, zda půjde o jednání na základě instrukcí či kontroly, přičemž prokázání vazby mezi nestátním aktérem a státem bude pravděpodobně ještě obtížnější než v případě kybernetického útoku učiněného člověkem. Pro ilustraci – v případě, že hackerská skupina nasazující autonomní kapacitu jedná pod

93 Jak jsem psal v kapitole 2, je vhodné, aby byla tato odpovědnost koncipována jako objektivní s možností liberace. Minimalizuje se tím hrozba právních mezer a zjednodušuje to i přičítací proces jako takový, přičemž na státy to neklade neúměrné požadavky, vzhledem k tomu, že v procesu rozhodování o nasazení autonomních kapacit jsou buď zapojeny nebo za něj nejsou odpovědní (viz dále). Naopak to může přispívat ke zvýšení snahy nasazovat kvalitní a řádně otestované kapacity.

94 Jsem si vědom, že se jedná o kombinaci národního a mezinárodního právního systému, ale v tomto ohledu existuje již pro analogickou kombinaci precedens, mj. přičtení ransomware WannaCry americkým ministerstvem spravedlnosti. Více viz BENDIEK, A. – SCHULZE, M. Attribution: A Major Challenge for EU Cyber Sanctions: An Analysis of WannaCry, NotPetya, Cloud Hopper, Bundestag Hack and the Attack on the OPCW. *SWP Research Paper*. 2021. Dostupné z: https://www.swp-berlin.org/10.18449/2021RP11/. BOSSERT, T. Press Briefing on the Attribution of the WannaCry Malware Attack to North Korea – The White House [online]. *The White House – Press Briefings*. 19. 12. 2017 [cit. 22. 11. 2023]. Dostupné z: https://trumpwhitehouse.archives.gov/briefings-statements/press-briefing-on-the-attribution-of-the-wannacry-malware-attack-to-north-korea-121917/.

95 Jedná se o odraz osoby *deployera*.

96 HAATAJA, S. *Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility*, s. 269.

97 CRAWFORD, J. *State Responsibility: The General Part*, s. 136–138.

98 *Rozsudek Mezinárodního soudního dvora ve věci Ozbrojených aktivit na území Konga (Kongo v. Uganda)*. 2005, s. 213–214.

99 HAATAJA, S. *Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility*, s. 272.

odpovídající kontrolou státu, rozhodovat o nasazení kapacit bude stát, kontrolu nad hackerskou skupinou (která kontroluje kapacitu samotnou) má stát a z nasazení kapacity těží též stát, bude možné akty takové kapacity přičíst. V případě, že v jednom z těchto kritérií nebude odpovědí „stát", pravděpodobně nebude jeho míra kontroly nad nestátním aktérem využívajícím autonomní kapacity dostatečná pro přičtení dle testu efektivní kontroly, a to ani v případě, že by dané autonomní kapacity hackerské skupině sám poskytnul.[100] Důkazní situaci pak ovlivní i stále častější praxe některých států fakticky integrovat původně nestátní hackerské skupiny pod zpravodajské služby (příkladem může být APT 28[101] či skupina stojící za útokem WannaCry[102]), přestože *de iure* zůstávají tyto skupiny stále samostatné (přinejmenším do určitého rozsahu).[103]

Jestliže by o nasazení kapacit rozhodovali nestátní aktéři, kteří nemají dostatečnou vazbu na stát, a nasazení by sledovalo pouze jejich cíle, není možné akty autonomních kapacit připsat. Výjimkou by mohl být princip *due diligence*, který zavazuje státy nenechat vědomě zneužívat své území k aktivitám poškozujícím jiné aktéry.[104] Ten by bylo ostatně možné využít i při nepříznivé důkazní situaci, která by neumožňovala přičtení aktů autonomních kapacit. V případě, že by byl stát informován o tom, že z jeho území či infrastruktury dochází k útoku (tento útok může být iniciován autonomní kapacitou), má povinnost učinit vše, aby tento útok zastavil, pokud k tomu má odpovídající kapacity, nebo za tímto účelem strpět zásah poškozeného státu. V první řadě je nutné přiznat, že princip *due diligence* v zásadě obchází přičtení a nezakládá odpovědnost za útok samotný, pouze za možnost zneužít území a infrastrukturu k jeho provedení (dle toho se budou lišit možnosti reakce, reparací ad.).[105] V druhé řadě pak považuji za vhodné poukázat na přinejmenším částečné vyhoření tohoto mezinárodního obyčeje, jehož aplikaci v kyberprostoru odmítlo několik velmi vlivných států (mj. USA) a které může praktickou aplikaci tohoto „záchranného" řešení zhatit.[106]

Oproti postavení samotného nasazujícího aktéra, u kterého neshledávám zásadnější odchylky oproti tradičnímu přičítacímu režimu, si specifickou pozornost zaslouží, a to zejména v kontextu nestátních aktérů, kritérium kontroly. Nikaragujský test Mezinárodního soudního dvora by totiž i zde představoval zásadní překážku, a to zejména kvůli

---

100 Srov. *Rozsudek Mezinárodního soudního dvora ve věci Některých vojenských a polovojenských činností v Nikaragui a proti ní (Nikaragua vs. Spojené státy americké)*, 1986, odst. 109–11.

101 MCWHORTER, D. APT28 Malware: Russia's Cyber Espionage Operations Report [online]. *Mandiant*. 10. 8. 2023 [cit. 15. 2. 2024]. Dostupné z: https://www.mandiant.com/resources/blog/apt28-a-window-into-russias-cyber-espionage-operations. Fancy Bear Hackers (APT28): Targets & Methods [online]. *CrowdStrike*. 12. 2. 2019 [cit. 30. 10. 2023]. Dostupné z: https://www.crowdstrike.com/blog/who-is-fancy-bear/.

102 *Criminal Complaint: United States of America v. Park Jin Hyok, No. MJ 18-1479*. United States District Court for the Central District of California, 2018. Dostupné z: https://www.justice.gov/media/967836/dl.

103 Nevzácně se pak stává, že vedení takovýchto skupin je přinejmenším částečně složeno z příslušníků zpravodajských služeb, kdy už se dle mého názoru přestává jednat o nestátního aktéra, ale spíše o státní orgán de facto.

104 CHIRCOP, L. A Due Diligence Standard of Attribution in Cyberspace. *International and Comparative Law Quarterly*. 2018, Vol. 67, No. 3, s. 643–644.

105 Aplikace tohoto principu v kyberprostoru je zatím spíše sporadická a např. Izrael otevřeně zpochybňuje jeho aplikovatelnost jakožto součást obyčejového práva dopadajícího na kyberprostor. Více viz SCHÖNDORF, R. Israel's Perspective on Key Legal and Practical Issues Concerning the Application of International Law to Cyber Operations. *International Law Studies*. 2021, Vol. 97, No. 1. Dostupné z: https://digital-commons.usnwc.edu/ils/vol97/iss1/21.

106 K tomu více viz VOSTOUPAL, J. Stuxnet vs WannaCry and Albania: Cyber-attribution on trial. *Computer Law & Security Review*. 2024, roč. 54.

nemožnosti zohlednění rozdílného fungování kybernetických útoků a konkrétně autonomních kapacit kybernetické obrany od konvenčních či *guerilla* operací. Jak jsem uvedl v podkapitole 1.2, kontrola se autonomizací nemusí nutně snižovat, spíše se přesouvá do jiné fáze životního cyklu dané technologie. Pokud tedy operátor či tvůrce stanoví ultimátní cíl, kterému kapacita podřizuje dílčí aktivity, a vybaví autonomní kapacitu prostředky, jak ho dosáhnout (ať už odpovídajícím kódem, nebo např. datasety pro její vycvičení), kontrolu si zachovává. Konkrétně si bezpochyby udržuje kontrolu nad začátkem a průběhem aktivit autonomních kapacit, ovšem schopnost další dílčí útoky zastavit a tím ukončit celou operaci, kterou Mezinárodní soudní dvůr vyžadoval též, může být komplikovaná. I tuto schopnost je sice možné částečně přesunout, např. Stuxnet obsahoval příkaz, kterým se měl ze všech zařízení vymazat s rokem 2012 (čímž by byla celá operace zastavena),[107] ale schopnost zastavit průběh v reálném čase jeho tvůrci neměli.[108, 109] Oproti tomu např. ransomware WannaCry[110] obsahoval *killswitch*, který umožňoval zablokovat infekci u kteréhokoliv zařízení připojeného k internetu prostřednictvím registrování specifické domény.[111] I v tomto případě sice nebylo možné vzít zpět již proběhlou infekci, ale vyžadovat takovýto krok by bylo i v porovnání s přístupem ke konvenčním operacím přehnané a neproporcionální.[112] Je ovšem také důležité zdůraznit, že tento *killswitch* se stal příčinou pádu tohoto ransomware a není tak dle mého realisticky očekávatelné, že by státy u autonomních kapacit kybernetické obrany ve větším využívaly podobnou metodu jenom kvůli uspokojení testu efektivní kontroly.

Na základě výše uvedeného jsem tak nucen k závěru, že v případě, že by nestátní aktér (a to i v případě, pokud by vystupoval v závislém postavení na státu a byl pod jeho kontrolou) nasadil autonomní kapacitu, kterou není možné zastavit v reálném čase, je velmi pravděpodobné, že by dané „uteklo" z dosahu mezinárodněprávní odpovědnosti států a nebylo by možné přičíst akty z tohoto nasazení vzešlé, neboť absentující kontrola nad ukončením aktivit autonomní kapacity (tedy nástroje) by negativně ovlivnila míru kontroly státu nad operací a nestátním aktérem.[113] Zvláště v porovnání se závěrem o nasazení autonomní kapacity státním orgánem (u kterého není kontrola nad ukončením o nic větší) pak tento závěr jen potvrzuje slova někdejšího místopředsedy Mezinárodního soudního dvora Al-Khasewneha, že přístup zvolený Mezinárodním soudním dvorem umožňuje státům schovávat své aktivity za prostředníky a zneužívat tak této právní kličky.[114]

---

107 FALLIERE, N. – O MURCHU, L. – CHIEN, E. *W32.Stuxnet Dossier*. Symantec, 2011. Dostupné z: https://archive.org/details/w32_stuxnet_dossier.

108 Počítačový červ Stuxnet byl s největší pravděpodobností nasazen státními orgány a zajišťovat kontrolu nad koncem operace pro účely přičtení by tak nebylo nutné; tento útok zde uvádím pouze pro ilustraci technických schopností kontrolovat konec operace. Connect the Dots on State-Sponsored Cyber Incidents – Stuxnet online. *Council on Foreign Relations* [cit. 26. 12. 2023]. Dostupné z: https://www.cfr.org/cyber-operations/stuxnet.

109 ZETTER, K. *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*, kap. 16.

110 Tento ransomware nemusí být nutně podřazen pod autonomní kapacity kybernetické obrany, jeho uvedení slouží jenom k ilustraci kontroly nad koncem operace.

111 BERRY, A. – HOMAN, J – EITZMAN, R. WannaCry Malware Profile [online]. *Mandiant*. 23. 5. 2017 [cit. 28. 12. 2023]. Dostupné z: https://www.mandiant.com/resources/blog/wannacry-malware-profile.

112 Srov. *Rozsudek Mezinárodního soudního dvora ve věci Některých vojenských a polovojenských činností v Nikaragui a proti ní (Nikaragua vs. Spojené státy americké)*, 1986, odst. 109 an.

113 Obdobně viz HAATAJA, S. *Autonomous Cyber Capabilities and Attribution in the Law of State Responsibility*, s. 278–280.

114 *Rozsudek Mezinárodního soudního dvora ve věci Bosenské genocidy (Bosna a Hercegovina vs. Srbsko a Černá Hora)*, 2007, s. 217.

Závěr

Na základě provedené analýzy docházím k názoru, že stávající autonomní kapacity kybernetické obrany nepředstavují pro přičítací proces následků jejich nasazení výzvu zásadnější než je tomu v případě přičítání kybernetických útoků obecně. To bohužel v praxi znamená, že i v jejich případě bude případné přičtení toliko politického charakteru a s velkou pravděpodobností nedojde v dohledné době k dovolání se mezinárodněprávní odpovědnosti ze strany některého z poškozených států. V případě autonomních kapacit je to částečně způsobeno technickými obtížemi s efektivním nahrazením řady lidských aktivit v rámci kybernetické obrany v plné míře autonomními kapacitami, ale i skutečností, že autonomizace nezbavuje člověka kontroly nad danou kapacitou, toliko danou kontrolu mění a přesouvá. Autonomní kapacity je tak nutné, alespoň ještě v tuto chvíli, vnímat toliko jako nástroje sloužící zájmům lidí a případně států. Na základě této myšlenky a modelu J. Kwika k autonomním zbraňovým systémům je pak možné považovat odpovědnost za akty autonomních kapacit (kterou je dle mého názoru v daném kontextu vhodné koncipovat alespoň jako odpovědnost objektivní s možností liberace[115]) za funkční základ přičítacího řetězce.

I v tomto kontextu ovšem představuje problém otázka efektivní kontroly v případě zapojení nestátních aktérů a zvláště kontrola nad ukončením operace, kterou Mezinárodní soudní dvůr ve věci *Některých vojenských a polovojenských aktivit v Nikaragui a proti ní* vyžadoval. Dané totiž i v kontextu autonomních kapacit představuje možnost států schovat své aktivity za prostředníka a vyhnout se tím mezinárodněprávní odpovědnosti.

Bylo by sice možné uvažovat o využití principu *due diligence* jako o potenciálním řešení této situace, ovšem chabá rozšířenost využívání tohoto postupu a otevřené zpochybňování vhodnosti či aplikovatelnosti ze strany klíčových států (mj. Izrael či USA[116]) příliš naděje nenabízí. Jedná se tak v zásadě jenom o další důkaz nevhodnosti testu efektivní kontroly na otázky spojené s kybernetickými útoky a kyberprostorem jako takovým.

115 Přestože by bylo teoreticky možné uvažovat o aplikaci *liability* režimu pro autonomní kapacity (a tedy odklonu od tradičního přístupu ke kybernetickým útokům jakožto doméně *responsibility* režimu), zůstala by takováto debata dle mého názoru spíše na úrovni akademické rozpravy, a to zejména kvůli obtížím s identifikací odpovídajícího mezinárodního obyčeje a neexistenci relevantní mezinárodní smlouvy. Je však možné, že v této oblasti dojde k dalšímu posunu, jakmile budou útoky autonomních kapacit kybernetické obrany častější.

116 *Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies*. General Assembly of the United Nations, 2021, s. 136. Dostupné z: https://front.un-arm.org/wp-content/uploads/2021/08/A-76-136-EN.pdf.

Attributability of Acts of Autonomous Cyber-Defence Capabilities

Jakub Vostoupal (https://orcid.org/0000-0002-1669-9931)

Abstract: The rise of AI technologies has impacted many industries, not least cybersecurity. The never-ending battle between attackers and defenders places ever greater demands on speed, and the necessary interaction between humans and software is limiting. And while there is no need to fear systems such as Skynet from the Terminator movie as of yet, autonomous capabilities capable of cyber-attacks already exist. It is thus appropriate to consider how they may affect the existing legal order. Specifically, the present article examines the challenges that the deployment of autonomous cyber-defence capabilities poses to the attributability of cyber-attacks under international law. The author focuses first on the existing capabilities and rules out, at least for now, the need to deal with fully autonomous systems. Instead, he turns his attention to the nature of control over autonomous capabilities, concluding that despite the fact that autonomous systems can take individual sub-actions on their own, they are still only advanced tools serving the needs of people or states. Based on this idea, he then discusses accountability for acts of autonomous capabilities and its potential to serve as a basis for attribution, which he eventually combines with attribution rules under the law of international responsibility and finds them to be compatible. However, the involvement of non-state actors remains problematic even in this context, as in the case of 'traditional' cyber-attacks.

Keywords: cybersecurity, autonomous capabilities, cyber-attacks, attribution, international responsibility