Oponentský posudek disertační práce

Název disertační práce: *Reputace zdrojů škodlivého provozu*
Autor: Ing. Václav Bartoš

Disertační práce se zabývá metodami a systémy pro sběr, zpracování, vyhodnocení a sdílení hlášení o kyberbezpečnostních incidentech a dalších souvisejících informací. Tento aspekt ochrany proti kybernetickým útokům je sice méně efektní než třeba detekce malwaru nebo analýza jeho kódu, při praktické ochraně počítačových sítí, služeb a zdrojů má ale naprosto zásadní význam.

V kapitole 2 (Úvod do problematiky) jsou přehledně popsány tři oblasti, které s tématem disertační práce bezprostředně souvisejí: monitorování síťového provozu, detekce bezpečnostních hrozeb a postupy sdílení a zpracování kybernetických hrozeb. V kapitole 3 (Přehled související literatury) jsou pak rozebrány a kriticky zhodnoceny předchozí publikované výsledky, na něž práce bezprostředně navazuje, především v oblastech klasifikace zdrojů škodlivého provozu, stanovení reputace síťových entit a predikce útoků.

Ve 4. kapitole je popsán a definován návrh nové metody vyhodnocování reputace síťových entit a výpočtu prediktivního indexu FMP (Future Misbehaviour Probability). Na něm oceňuji zejména jeho komplexnost, neboť umožňuje zahrnout a zohlednit všechny dostupné zdroje informací, které mohou přispět k hodnocení bezpečnostního rizika spojeného s konkrétními síťovými entitami (jednotlivými zařízeními, DNS doménami, IP prefixy nebo autonomními systémy). Návrh bere v úvahu i časový aspekt – starší informace postupně ztrácejí svoji relevanci.

Kapitola 5 (Detekce bezpečnostních událostí) popisuje framework NEMEA, na jehož vývoji ing. Bartoš spolupracuje v rámci sdružení CESNET. Jde o modulární systém, který umožňuje na základě monitorování a analýzy síťového provozu generovat formalizovaná hlášení o bezpečnostních incidentech. Výstupy tohoto systému získané z provozu národní výzkumné sítě CESNET2 slouží jako vstupní data pro další zpracování, které je vlastním jádrem disertační práce.

V 6. kapitole je popsán návrh reputační databáze síťových entit, který autor implementoval jako systém NERD (Network Entity Reputation Database). Z hlediska softwarového inženýrství považuji jeho architekturu i implementaci za dobře promyšlené a zvládnuté. Fungující části softwaru NERD prozatím pokrývají potřeby sítě CESNET2, takže jako jediný primární zdroj dat je k dispozici systém Warden. Návrh je ale dostatečně obecný a škálovatelný, takže bude možné jej využít v bodoucnu i v jiných sítích a s jinými primárními zdroji dat.

Kapitola 7 (Použitá data a jejich charakteristiky) obsahuje analýzu dat ze systému Warden, která jsou dále použita ve výzkumné části disertační práce. Tato analýza je nezbytná pro poučenou volbu parametrů predikčních algoritmů, je ale zajímavá (a částečně deprimující) i sama o sobě, neboť ukazuje, jak rozsáhlým a globálním problémem je kybernetická bezpečnost.

V kapitole 8 (Predikce škodlivého chování IP adres) jsou popsány zvolené atributy feature vektoru, který slouží jako vstup predikčních algoritmů a zdůvodněna

volba klíčových parametrů použitých pro predikci. predikční algoritmy. Jsou také popsány postupy předzpracování vstupních dat, trénování predikčních algoritmů a jejich vlastní použití.

Kapitola 9 (Vyhodnocení) je věnována kvantitativním výsledkům výpočtu FMP a jejich dalšímu využití. Jsou zde popsány oba použité predikční algoritmy (neuronové sítě a rozhodovací stromy GBDT) a analyzována kvalita dosažených výsledků, zejména z hlediska výskytu falešně pozitivních i negativních výstupů a v porovnání s běžně používanými postupy (jednoduchými blacklisty). Konečně jsou zde popsány možné způsoby dalšího využití reputačního skóre FMP – prediktivní blacklisty a zařízení na ochranu proti DDoS útokům.

Disertační práci celkově hodnotím jako velmi obsáhlou a kvalitní. Celý systém (včetně částí, které nejsou předmětem této práce) je úspěšně využíván ve vysokorychlostní síti.

Neuronové sítě (NN) a podobné algoritmy jsou dnes prezentovány jako metody umělé inteligence nebo strojového učení, a jsou nasazovány na vhodné i méně vhodné problémy. Disertační práce ukazuje, byť na základě relativně omezeného vzorku dat, že použití pro predikci nepezpečného chování síťových entit vede k prokazatelně lepším výsledkům než běžně používané metody při vcelku mírných nárocích na výpočetní výkon. Nevýhodou algoritmů typu NN je, že fungují jako černá skříňka a vlastně nic nevypovídají o podstatě řešení. Je proto otázka, zda by se obdobných výsledků nedalo dosáhnout mnohem jednodušším způsobem, například klouzavým váženým průměrem stejného feature vektoru.

Po formální stránce je disertační práce napsána pečlivě a přehledně. Upozorňuji pouze na to, že podstatná jména typu *software*, *hardware* apod. se v češtině normálně skloňují podle vzoru hrad.

Závěry

1. Téma disertační práce odpovídá zvolenému oboru studia a je aktuální z hlediska současného stavu vědy a technologií.
2. Původní přínos autora spatřuji především ve
   — významném podílu na vývoji komplexního systému pro sběr, analýzu a publikaci dat o bezpečnostních incidentech;
   — návrhu a implementaci reputační databáze síťových entit;
   — vývoji a testování predikčních metod pro vyhodnocování kyberbezpečnostních rizik.
3. Všechny původní přínosy autora byly publikovány formou prestižních mezinárodních publikací, které jsou v práci citovány.
4. Disertační práce dostatečně dokládá vědeckou erudici Ing. Václava Bartoše a **odpovídá obecně uznávaným požadavkům pro udělení akademického titulu PhD**.

V Litvínovicích, 17. dubna 2019 Ing. Ladislav Lhotka, CSc.