Posudek práce

předložené na Přírodovědecké fakultě JU

❑ posudek vedoucího	☑ posudek oponenta
☑ bakalářské práce	❑ diplomové práce

Autor/ka:	**Daniel Cába**
Název práce:	**Analýza síťového provozu pomocí metod data miningu s cílem detekce nestandardního chování uživatelů**
Studijní program a obor:	Aplikovaná informatika
Rok odevzdání:	2012
Jméno a tituly vedoucího/oponenta:	Ing. Jaroslav Kothánek, Ph.D. Ing. Rudolf Vohnout; Ing. Miroslav Skrbek, Ph.D.
Pracoviště:	Ústav aplikované informatiky
Kontaktní e-mail:	jkothanek@prf.jcu.cz

Odborná úroveň práce:
❑ vynikající ❑ velmi dobrá ☑ průměrná ❑ podprůměrná ❑ nevyhovující

Věcné chyby:
❑ téměř žádné ❑ vzhledem k rozsahu přiměřený počet ☑ méně podstatné ❑ četné ❑ závažné

Výsledky:
☑ originální ❑ původní i převzaté ❑ netriviální kompilace ❑ citované z literatury ❑ opsané

Rozsah práce:
❑ velký ☑ standardní ❑ dostatečný ❑ nedostatečný

Grafická, jazyková a formální úroveň:
❑ vynikající ❑ velmi dobrá ❑ průměrná ☑ podprůměrná ❑ nevyhovující

Tiskové chyby:
☑ téměř žádné ❑vzhledem k rozsahu a tématu přiměřený počet ❑ četné

Celková úroveň práce:
❑ vynikající ❑ velmi dobrá ❑ průměrná ☑ podprůměrná ❑ nevyhovující

Slovní vyjádření, komentáře a připomínky vedoucího/oponenta:
(oponent Ing. Jaroslav Kothánek, Ph.D.)
Autor se ve sveé práci zabývá analýzou síťového provozu s detekcí nestandardního chování uživatele. Autor se zde zaobírá myšlenkou detekce napadení systému a následné adekvátní reakce správce systému. Myšlenka a náplň práce je velice aktuální a byla by přínosná pro bezpečnost informačních systémů. Vzhledem k tomu, že jako oponent jsem v některých pasážích měl pochybnosti, požádal jsem doplňující oponenty, viz níže, aby se též vyjádřili k odborným otázkám z jejich odbornosti, které mé domněnky potvrdil.

V rámci bakalářské práce je nutné autorovi vytknout značné množství chyb a překlepů. Taktéž autor používá chybně napsaná anglická slova např. „soket", pakety apod. Dále je nutné poukázat též na popis jednotlivých obrázků.

První část práce (oponent Ing. Rudolf Vohnout)
Po odborné stránce má předložená práce velký potenciál. Hodnocena však bude pouze její první část věnována popisu a realizaci síťových útoků. V teoretické části jsou dostatečným způsobem popsány později realizované útoky. Kde ovšem práce naprosto nenaplňuje očekávání, je část praktická. Ta se především vyznačuje velmi vágním popisem provedených experimentů. Není exaktně vysvětlen postup u jednotlivých pokusů, stejně jako přesný počet opakovaných pokusů pro validaci. Závěr u všech pokusů je prosté konstatování, že se vše zdařilo. Není precizně definován použitý hardware, a když ano tak chybně (například na straně 7 zmíněný produkt „Cisco Linksys E9000" jsem nebyl schopen na internetu nalézt) a minimálně v příloze bych očekával specifikaci toho aktivního prvku. Dále mi v práci vadí vysoký počet domněnek (například strana 8) a pak především závěr, který je zjednodušeně řečeno výčtem problému, se kterými se autor během vypracovávání práce setkal.

Druhá část práce (oponent Ing. Miroslav Skrbek, Ph.D.)
Student založil mezodu detekce útoků na algoritmu shlukové analýzy, využívá tří algoritmů (K-menoids, K-means, X-means). Vycházel z předpokladu, že tyto algoritmy odhalí oddělené shluky v datech pro útok a pro běžný provoz a bude tak možné odhalit anomálie v síťovém provozu. Pro analýzu využívá dataminingového nástroje Rapid Miner. S tímto přístupem lze principiálně souhlasit, nicméně se student dopustil řady chyb, které znevažují platnost výsledku. Shlukovací algoritmy využívají metriky typu Euklidovská vzdálenost, které nelze doporučit pro srovnávání hodnot typu MAC adresa. Dále používá shlukovací algoritmus tam kde by stačilo pouhé porovnání (data viz. Fig3, Fig4). Experimenty byly provedeny na velmi malých vzorcích dat, takže výsledek nelze považovat za průkazný. Výsledek práce lze pouze považovat za prvotní studii, která musí být revidována z hlediska vstupů (dat) pro shlukovací algoritmy, které musí mít charakter spojitých veličin, například četností, intenzity provozu apod.

Všichni oponenti se shodují na *doporučujícím* verdiktu a práci hodnotí stupněm *dobře*.

Případné otázky při obhajobě a náměty do diskuze:
Kolik přesně bylo potřeba falešných MAC adres, aby došlo k zahlcení CAM tabulky použitého switche?
Prozkoumal student jakým způsobem Rapid Miner měří podobnost MAC adres, protože je lze chápat jako řetězec znaků nebo číslo typu integer?

Práci
☑ doporučuji
❑ nedoporučuji
uznat jako diplomovou/bakalářskou.

Navrhuji hodnocení stupněm:
❑ výborně ❑ velmi dobře ☑ dobře ❑ neprospěl/a

Místo, datum a podpis vedoucího/oponenta:
V Českých Budějovicích dne 10 1. 2013