FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Tomáš Čejka, Ph.D.
Student:	Bc. Jan Peřina
Název práce:	Detekce síťových anomálií na základě dat z traceroute
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	4. února 2024

Hodnotící kritéria

1. Splnění zadání

- ▸ [1] **zadání splněno**
  [2] zadání splněno s menšími výhradami
  [3] zadání splněno s většími výhradami
  [4] zadání nesplněno

Práce se věnovala analýze dat z měření dostupnosti a kvality spojení pomocí traceroute mezi uzly komunikační infrastruktury. Konkrétně bylo cílem navrhnout modely detekce anomálií, které by mohly odhalit potenciální problémy s propustností a spolehlivostí komunikačních cest. Práce obsahuje důkladnou analýzu zkoumané problematiky, návrh detekčních mechanismů založených na známých statistických metodách a nakonec i implementaci softwarového prototypu v jazyce Python a experimenty včetně vizualizace dat pomocí python notebooků.

2. Písemná část práce 75 /100 (C)

Práce je vypracovaná v anglickém jazyce a celkově je na poměrně vysoké úrovni. Textová část práce však představuje poměrně velký prostor ke zlepšení. Dokument obsahuje řadu typografických nedostatků a text je místy méně srozumitelný. Pořadí Sekcí 2.4 a 2.3 je poměrně zvláštní a možná zbytečně vede ke zmatení čtenáře. Některé obrázky-grafy jsou poměrně obtížně interpretovatelné a bylo by užitečné v textu vysvětlit, čeho konkrétně by si měl čtenář všimnout. Celá práce je založena na zpracování velkého množství naměřených dat a v tomto ohledu je zpracována precizně. Podle mého názoru však v závěru chybí shrnutí dosaženého stavu a zhodnocení, zda se podařilo dosáhnout definovaných cílů: 1) vyvinutí vhodných nástrojů pro analýzu dat měření pomocí traceroute a detekci anomálií; 2) str. 4 poslední věta: vylepšit stávající heuristický algoritmus nebo navrhnout alternativní přístup k detekování problémů na síti; potenciálně i 3) strana 7, Sekce 2.1.1: ověřit, zda je IPv6 vhodnější pro datové přenosy.

3. Nepísemná část, přílohy 100 /100 (A)

Výsledek práce je zveřejněný v Github repozitáři. Jedná se především o zdrojové kódy v jazyce Python. Autor pečlivě připravil a zdokumentoval prostředí a nastavil automatické procesy pro kontrolu kvality zdrojových kódů. Dokumentace a testy jsou součástí odevzdané práce. Všechny tyto body svědčí o vysoké kvalitě vytvořených výstupů.

4. Hodnocení výsledků, jejich využitelnost 100 /100 (A)

Práce je zaměřena na důležitou problematiku monitorování stavu rozsáhlé infrastruktury využívající veřejnou komunikační síť - internet - pro přenos velkých objemů dat z CERN LHC k dalším vědeckým pracovištím po světě. Z dat získávaných měřením pomocí traceroute byla vytvořena datová sada, nad kterou byla následně provedena analýza a experimenty kolem detekce anomálií. Autor se zabýval různými metodami a vytvořil softwarové prototypy detekčních modelů, které lze použít pro dávkové zpracování dat. Textová část práce sice nepopisuje výsledek implementačních prací příliš detailně, ale v rámci práce vznikla integrace měření a detekce do systému ELK, který zároveň umožňuje operátorům vizualizace zjištěních informací.

Celkové hodnocení 90 /100 (A)

Textová část práce sice představuje prostor pro zlepšení, ale celkově je diplomová práce kvalitně zpracovaná a výstupy jsou užitečné v praxi.

Otázky k obhajobě

1. Jakým způsobem je možné využít výsledky navržených a implementovaných modelů detekce anomálií pro řešení detekovaných problémů?
2. Bylo by možné porovnat (kvalitativně, kvantitativně) vytvořený nástroj s předchozím řešením? (tzn. Jsou výsledky přesnější? Je "alertů" méně?)
3. Podařilo se vyhodnotit, jestli je nějaký významný rozdíl mezi chování traceroute u IPv4 vs. IPv6, jak je zmíněno v Sekci 2.1.1?

Instrukce

Splnění zadání

Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.