Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Bc. David Surma
Název práce	Analýza blockchainu používaného pro Bitcoin
Rok odevzdání	2023
Studijní program	Informatika **Studijní obor** Umělá inteligence

Autor posudku	Mgr. Pavel Hubáček, Ph.D. **Role** oponent
Pracoviště	Informatický ústav Univerzity Karlovy

Text posudku:

Práce kolegy Surmy se zabývá analýzou struktury veřejného transakčního grafu kryptoměny Bitcoin. Specificky se studuje problém přiřazení psedonymních identifikátorů objevujících se v transakcích k reálným entitám odpovídajícím těmto identifikátorům – snahou autora je na základě pochopení praktického fungování studovaného systému definovat heuristiky, které by umožnily robustně sdružovat identifikátory jednotlivých entit. Tento problém byl již ve vědecké literatuře studován, ale většina prací využívá kromě samotných transakcí také externí zdroje informací spojující identifikátory a reálné entity, jejichž spolehlivost však nelze s jistotou určit.

Po důkladném představení systému Bitcoin a známých výsledků relevantních k tématu práce autor představuje známé a nové heuristiky a popisuje výsledky experimentů aplikujících tyto heurustiky na transakční graf kryptoměny Bitcoin.

Silné stránky práce:

- Oceňuji, že si autor práce vytyčil ambiciózní cíl a zaměřil se na heuristiky, které nevyužívají externí informace mimo samotnou množinu transakcí.
- Text práce ukazuje autorovo pochopení netriviálního reálného systému a schopnost jasně popsat důležité aspekty takovýchto systémů.

Vlastní příspěvek:

- Úvod práce může jistě posloužit jako dobrý přehledný materiál pro další studenty a studentky, kteří by měli zájem věnovat se podobné problematice.
- Autor vysvětluje, proč některé známé heuristiky nemusí být smysluplné.
- Autor navrhuje nové heuristiky vycházející z analýzy celého aktuálního transakčního grafu.

Nedostatky:

- Práce by dle mého názoru mohla být ve čtvrté kapitole strukturovanější a obsáhlejší. Ocenil bych explicitní shrnutí nových poznatků již ve čtvrté kapitole před závěrem.
- Práce do značné míry opomíjí výpočetní aspekty studované problematiky. Pro čtenáře by bylo jistě přínosné, pokud by autor nastínil, jaký hardware byl použit pro praktické experimenty. Bylo by také možné diskutovat relativní složitost jednotlivých heuristik.
- Úvod čtvrté kapitoly zmiňuje interakce s reálnými směnárnami a zpětnou analýzu výsledných transakcí ukazující strukturu „peeling chainu“. Pokud správně rozumím odpovídajícím diagramům, tyto experimenty ale vlastně nepotvrzují autorovu hypotézu o původu transakcí utracených ve stejném bloku, ve kterém byly vytvořeny. Ocenil bych, pokud by autor býval více zdůvodnil, proč formuloval tuto hypotézu a uvedl případně návrhy přístupů, jak ji testovat.

Drobné nepřesnosti a překlepy:

- Poznámka o jednosměrnosti SHA256 nedává smysl. Bezpečnost Proof-of-Work založeného na SHA256 je pouze heuristická a je motivována analýzou idealizovaných hašovacích funkcí. Nejsem si vědom alternativních konstrukcí Proof-of-Work schémat, pro které by jako předpoklad postačovala existence jednosměrných funkcí.
- „Lightning network“ je uvedena bez reference.
- V sekci o útocích je zmíněna základní analýza double-spend útoku od Nakamota, z které zdánlivě vyplývá bezpečnost konsenzu v Bitcoin protokolu za předpokladu, že žádný uživatel neovládá nadpoloviční zlomek výpočetního výkonu v systému. Jsou však známé útoky, které nevyžadují ani polovinu výpočetního výkonu v systému, jako například „selfish mining“ v článku Eyala a Sirera z roku 2014.
- V definici klastrování by bylo vhodné uvést explicitně, že $\mathcal{A}$ značí množinu všech adres.
- Motivace podklastrování: „Obvyklým postupem je zklastrování jen takové podmnožiny adres, u kterých je jejich příslušnost ke společnému majiteli jistým způsobem prokazatelná. Adresy jedné entity se tak mohou vyskytovat ve více klastrech.“ je pro mě trochu zavádějící, protože následná definice vyžaduje pokrytí celé množiny adres.
- O multi-input heuristice: „Je nejpoužívanější a nejefektivnější metodou.“ Bylo by vhodné dodat v jakém smyslu.
- ... použil*i* Meiklejon a kol. ...

- Nemělo by C v definici relativního úbytku adres být formálně *podklastrování*?
- Pojem „kolaps klastrů“ je uveden bez reference.
- Kapitola 4.2 zmiňuje Louvainovu metodu pro komunitní detekci bez vysvětlení o jakou metodu se jedná a proč by měla být relevantní.
- Pozor na správné použití zájmen jenž a jež v celém textu.

Studentovi se rozhodně podařilo zadání práce splnit. Celkově mi práce přijde jako velmi koherentní a zajímavá. Nedostatky a drobné nepřesnosti uvedené výše nejsou zásadního charakteru.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 25. 5. 2023

Podpis: