

Social Networks



Jiří Novotňák

Brno University of Technology, Faculty of Information Technology
Božetěchova 2, 612 00 Brno, CZ
www.fit.vutbr.cz/~inovotnak



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tři články

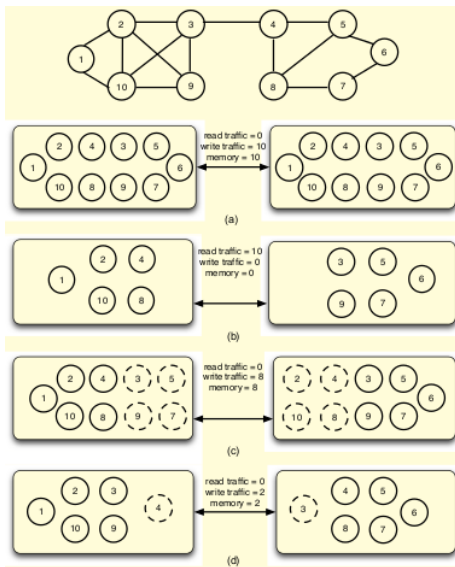
- The Little Engine(s) That Could: Scaling Online Social Networks
- Crowdsourcing Service-Level Network Event Monitoring
- An Analysis of Social Network-Based Sybil Defenses

Zaměření

- Optimalizace přístupu k datům v sociálních sítích

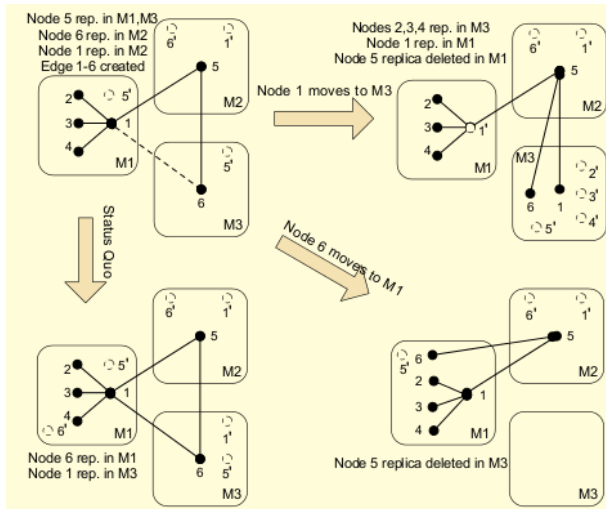
Řešené problémy

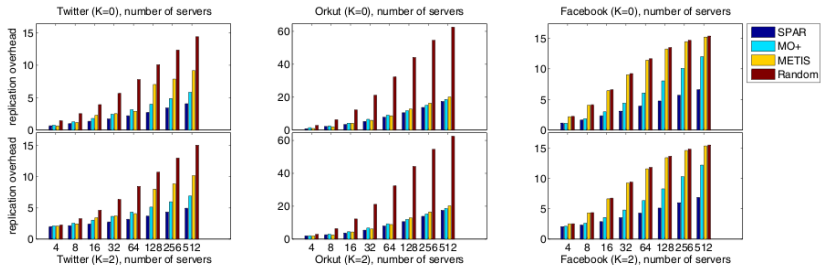
- Lokálnost dotazů
- Replikace
- Loadbalancing



Min_replica

- Algoritmus nalezení vhodného rozdělení s minimální nutností replikace
- NP-těžký problém
- Algoritmy na dělení grafů nejsou příliš vhodné
 - Většinou offline algoritmy (nutnost provést celý výpočet znovu při každé změně)
 - Algoritmy na detekci komunit jsou citlivé na vstupní podmínky - malá změna v grafu může způsobit přesouvání velkého množství uzlů.
- Nový uzel je přidán na server s nejmenším počtem master uzlů.
- Rozložení výpočtu na lokální kontrolu podmínek při změně hran.
 - Minimální počet replikací
 - Vyvážení zátěže





Cíl

- Systém na monitorování sítí
- Založeno na měření běžných služeb na klientech (např. bittorrent)

Řešené problémy

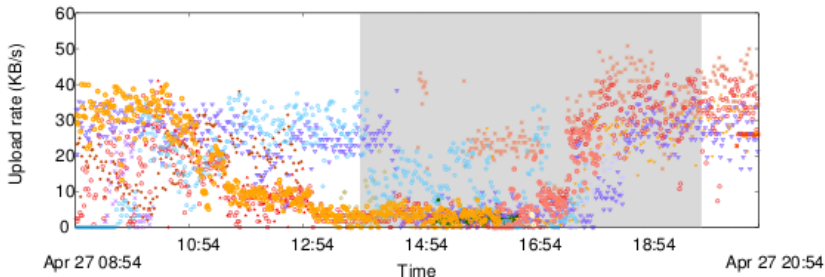
- Jak a co měřit.
 - Rychlost přenosu
 - Odezvy
 - ...
- Jak motivovat uživatele, aby si instalovali agenta.
 - Uživatel musí mít ze systému profit, jsou mu poskytována monitorovací data, takže on sám má představu, zda není nějaký problém.
- Jak zachovat soukromí. Uživatelů.
 - Používání generovaných ID a BGP prefixů

Implementace

- Detekce lokálního stavu
 - Rychlost přenášených dat
 - Odezvy
 - Ztrátovost
- Distribuované uložení
- Uzly detekují které další uzly leží ve stejné síti dle BGP prefixů
- Uzly korelují získané informace
- Plugin do Vuze (bittorent)

Výpadek v síti British Telecom

- 14:54 GMT - BT informuje o možných problémech v síti.
- 20:50 GMT - Problém je označen jako vyřešený.



Vlastnosti NEWS

- Distribuovaný monitorovací systém
- Čím více agentů, tím důvěryhodnější a přesnější detekce
- Možnost detekce problému nad rámec jedné sítě
- GPL licence

Dosažené výsledky

ISP	Počet uživatelů	Detekovaných událostí
Deutsche Tel.	6760	69
HTP	3652	112
HanseNet	3216	17
Neuf Cegetel	2821	108
Arcor	2245	29
Cableuropa	1999	245
Proxad/Free	1769	176
France Tel.	1688	31
Tel. Italia	1651	20
Telefonica	1337	27

O co v článku jde

- Analyzovat algoritmy detekce útoků Sybily
- Srovnat jejich vzájemné schopnosti
- Potenciálně najít další možnosti pro detekci těchto útoků

Útok Sybily

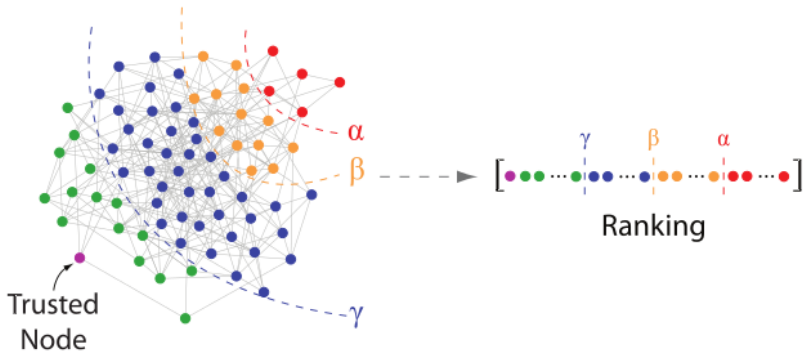
- Vytvoření většího množství “nepravých” identit jedním uživatelem
- Cílem je ovlivnit chování sociální sítě
 - Získání privilegií na P2P sítích
 - Různá hlasování
 - atd.

Princip algoritmů

- Založené na procházení hran od důvěryhodného uzlu
 - SybilGuard
 - SybilLimit
 - SybilInfer
- Založené na “volení” důvěryhodných uzlů
 - SumUp

Souhrn

- Až na malé rozdíly pracují algoritmy podobně
- Základní předpoklad: Falešné identity nevytváří příliš velké množství vazeb na reálné uzly.
- Vždy se vybere důvěryhodný uzel
- Postupuje se po hranách od důvěryhodného uzlu a hodnotí se počet “hopů” k měřenému uzlu.
- Dalším kritériem je kolik cest k danému uzlu z důvěryhodného uzlu vede.

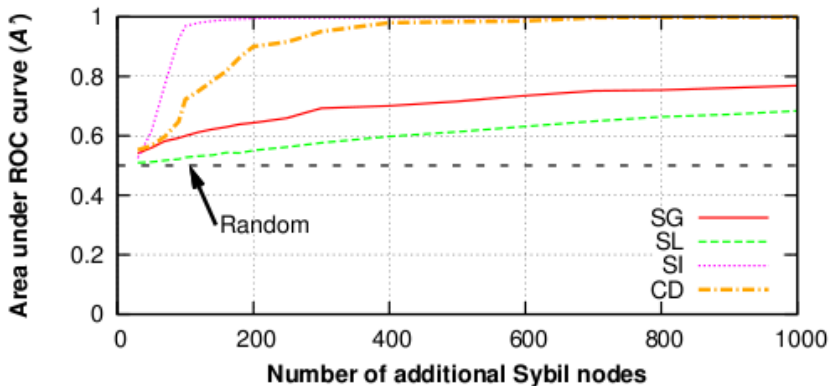


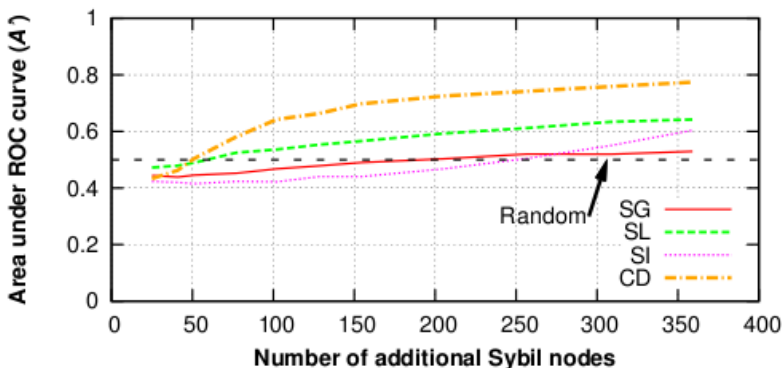
Vlastnosti algoritmů

- Algoritmy označují za legitimní uzly zejména uzly v komunitě s důvěryhodným uzlem.
- Na uzly mimo komunitu se algoritmy dívají předujatě.
- Algoritmy vlastně detekují komunity.
- Uzly jsou označený číselně, čím vyšší číslo, tím pravděpodobnější Sybilin uzel.

Srovnání algoritmů

- Jaká je pravděpodobnost, že libovolný Sybilin uzel získá nižší ohodnocení, než libovolný regulární uzel.
- Na syntetických sítích některé algoritmy fungují dobře. (SybilInfer a SybilLimit)
- Na reálných sítích nepracuje žádný z algoritmů uspokojivě.





Závěr

- Are Sybils necessarily bad?
- Předpoklad, že Sybily tvoří vlastní komunitu nemusí být pravdivý.
- Čím více se útočící uzly napojí na normální uživatele, tím je pravděpodobnost jejich detekce menší