Detektor anomálií DNS provozu

Statistická metoda

Karel Slaný • karel.slany@nic.cz • 30.11.2013

CZ.NIC | SPRÁVCE DOMÉNY CZ

Obsah

- Stručný popis funkce
- Ukázky nalezených anomálií

Metoda založena na

G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, K. Cho, *Extracting hidden anomalies using sketch and non Gaussian multiresolution statistical detection procedures*, 2007

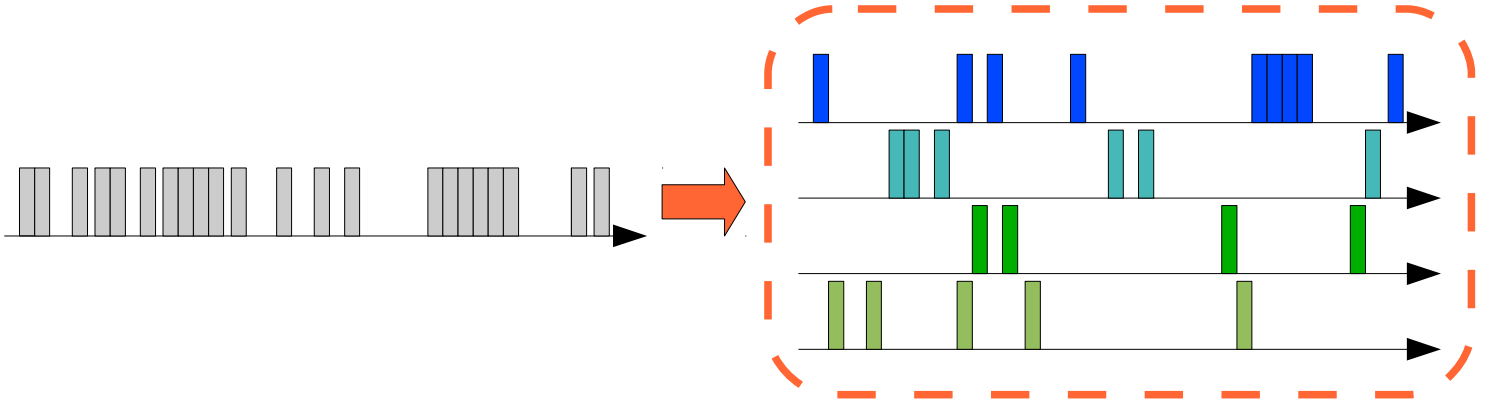
- aplikovatelné na velké soubory dat
- detekce krátce i déletrvajících anomálií
- nízké výpočetní nároky

Metoda

- Analyzuje provoz v rámci klouzajícího okna.
- Algoritmus iteruje přes tyto kroky:
  - náhodné rozdělení do skupin
  - agregace dat ve skupinách
  - konstrukce gama rozdělení a odhad parametrů
  - zjištění referenčních hodnot
  - výpočet vzdálenosti
  - sloučení dat ze skupin a identifikace anomálií

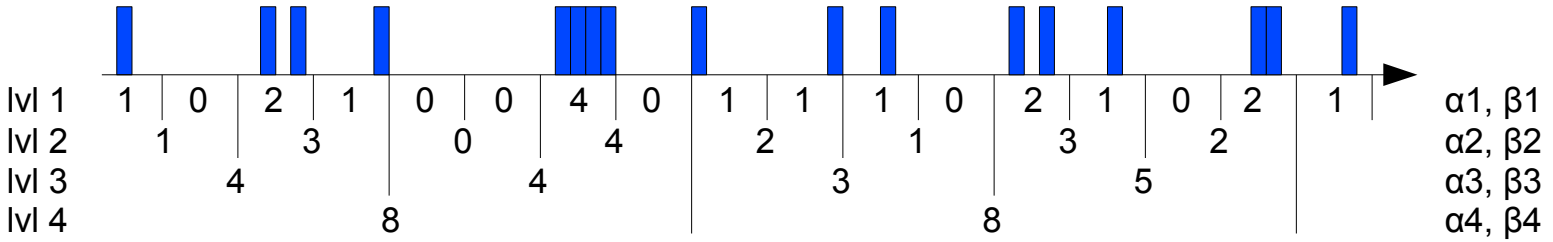
Náhodné rozdělení

- Data z klouzajícího okna jsou rozdělena do skupin (sketches) za pomocí univerzální hašovací funkce.
- Zvolený atribut analyzovaného paketu slouží jako hašovací klíč.
- Velikost hašovací tabulky je omezena.



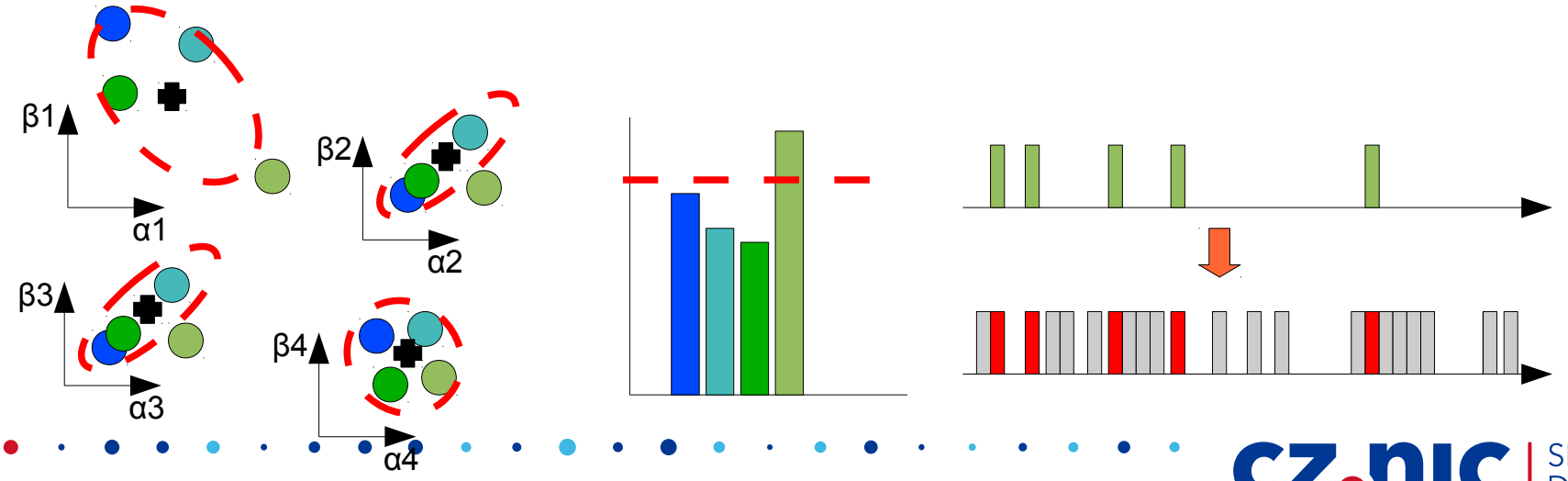
Agregace a Gama rozdělení

- Data ve skupinách jsou společně agregována za použití různých agregačních stupňů.
  - Je vytvořena posloupnost obsahující množství zaznamenaných paketů v průběhu zvolených intervalů, které odpovídají použitým agregačním stupňům.
  - Agregační stupeň mění rozlišení časové osy.
- Agregovaná data jsou modelována Gama rozložením.
  - Pro každé rozložení jsou vypočítány parametry pro tvar (α) a měřítko (β).



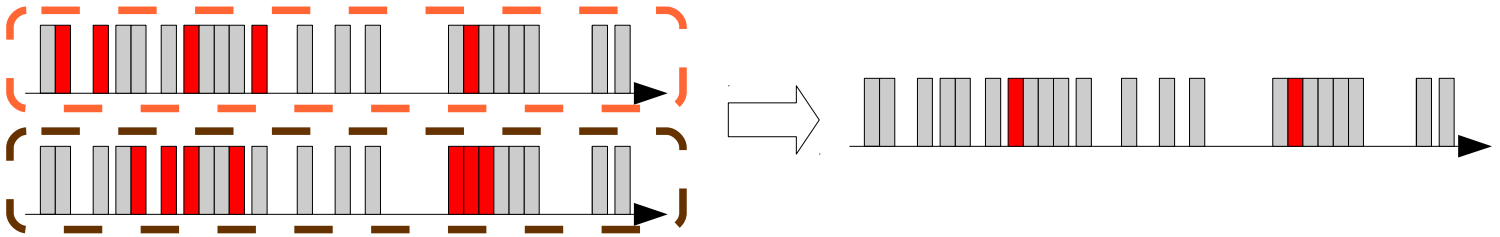
Identifikace anomálií

- Pro každý odpovídající agregační stupeň ve všech skupinách je vypočítána střední hodnota a rozptyl vypočtených parametrů Gama distribuce.
- Pro každou skupinu je spočtena Mahalanobisova vzdálenost ke středu.
- Skupiny s průměrnou vzdáleností větší, než je uvedený práh, jsou označeny jako anomální.



Zpřesnění identifikace

- Všechny atributy (hašovací klíče), které se nacházejí v označené skupině, jsou pokládány za podezřelé.
- Použitím další hašovací funkce se docílí jiného rozdělení do skupin a tím i odlišných podezřelých atributů.
- Seznam anomálních atributů se získá průnikem anomálních skupin.



Použití na DNS provoz

- Metoda byla navržena pro TCP/IP provoz.
  - Jako vstupní data slouží IP adresy.
- Metoda byla aplikována na DNS provoz.
  - Je možné také používat dotazované doménové jméno popřípadě jiné atributy DNS zprávy.

Software

- Dříve samostatná aplikace, dnes existuje a je udržována jako modul dns-collectoru.
- zlepšení:
  - Možnost zpracovávat data přenášena přes TCP a data z fragmentovaných paketů.
  - Snížena paměťová náročnost při zpracování archivovaného provozu.
  - Přidána možnost vyloučit zvolené identifikátory z analýzy.
  - Zadává se horní hranice počtu skupin. Skutečný počet je určen za běhu.

Použití

- Vstupem je zachycený síťový provoz ve formátu podporovaném libpcap.
- příklad výstupu:

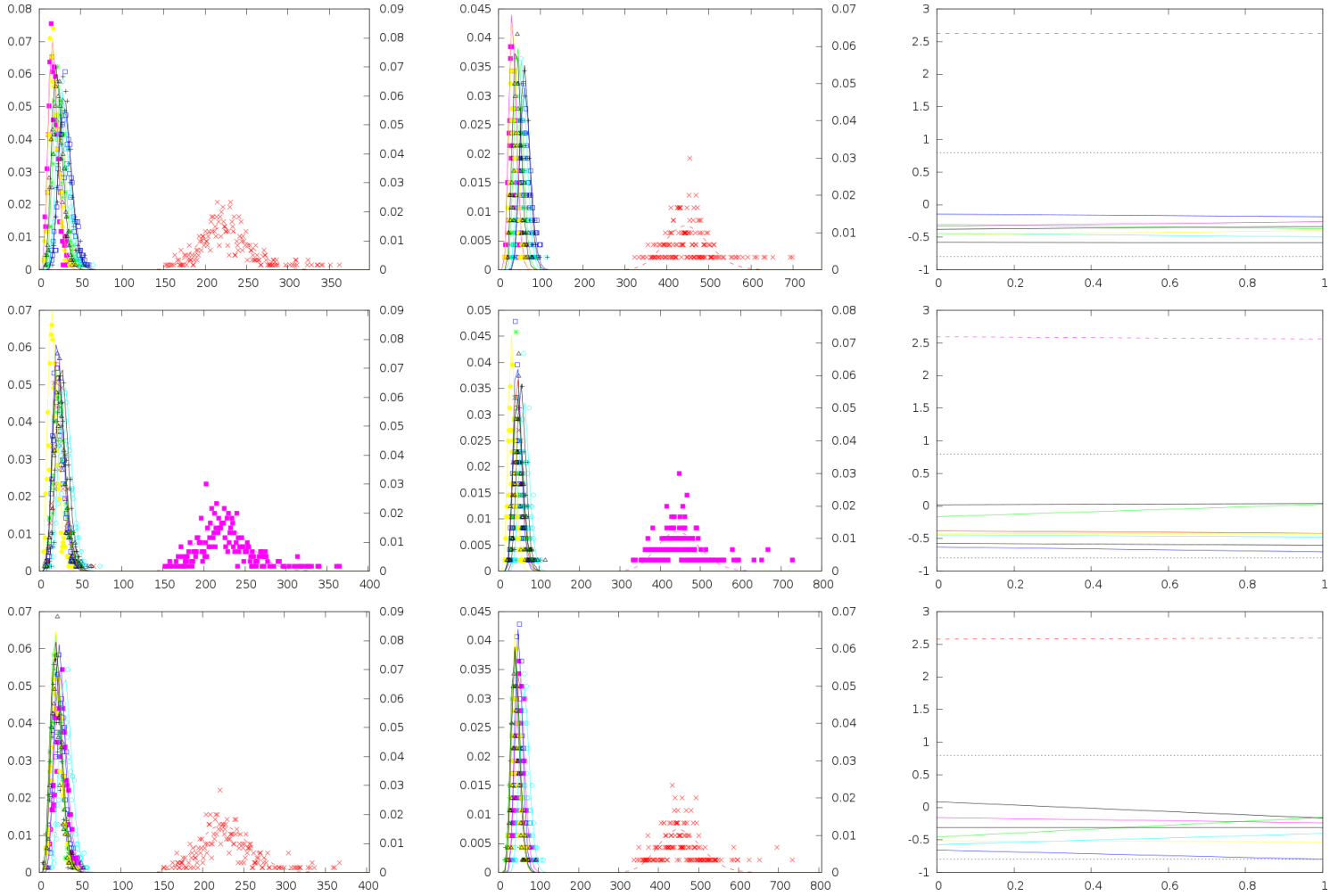
  From: Fri Nov 1 01:00:00 2013

  To: Fri Nov 1 01:09:59 2013

  found anomalies (4 / 154041) : dns.technorail.com., dns2.technorail.com., dns3.arubadns.net., dns4.arubadns.cz.
- Je možné generovat vizualizace průběhu detekce a výsledných detekovaných anomálií ve formátu pro gnuplot.

Vizualizace průběhu detekce

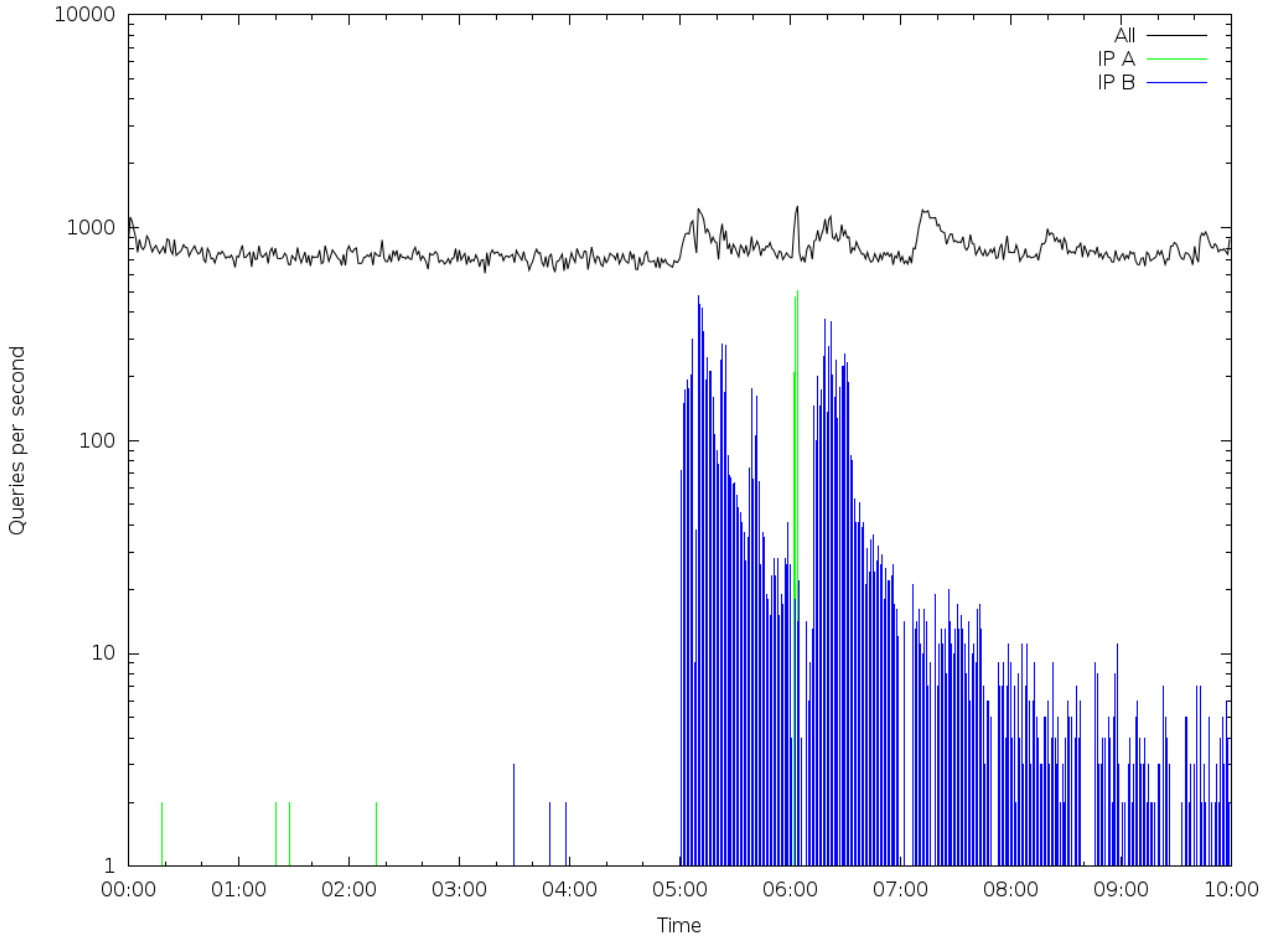
aggregation 2, hash 3, shetch 8, threshold 0.8



Detekované anomálie – SrcIP

andraste, 1.11.2013 UTC 00:00

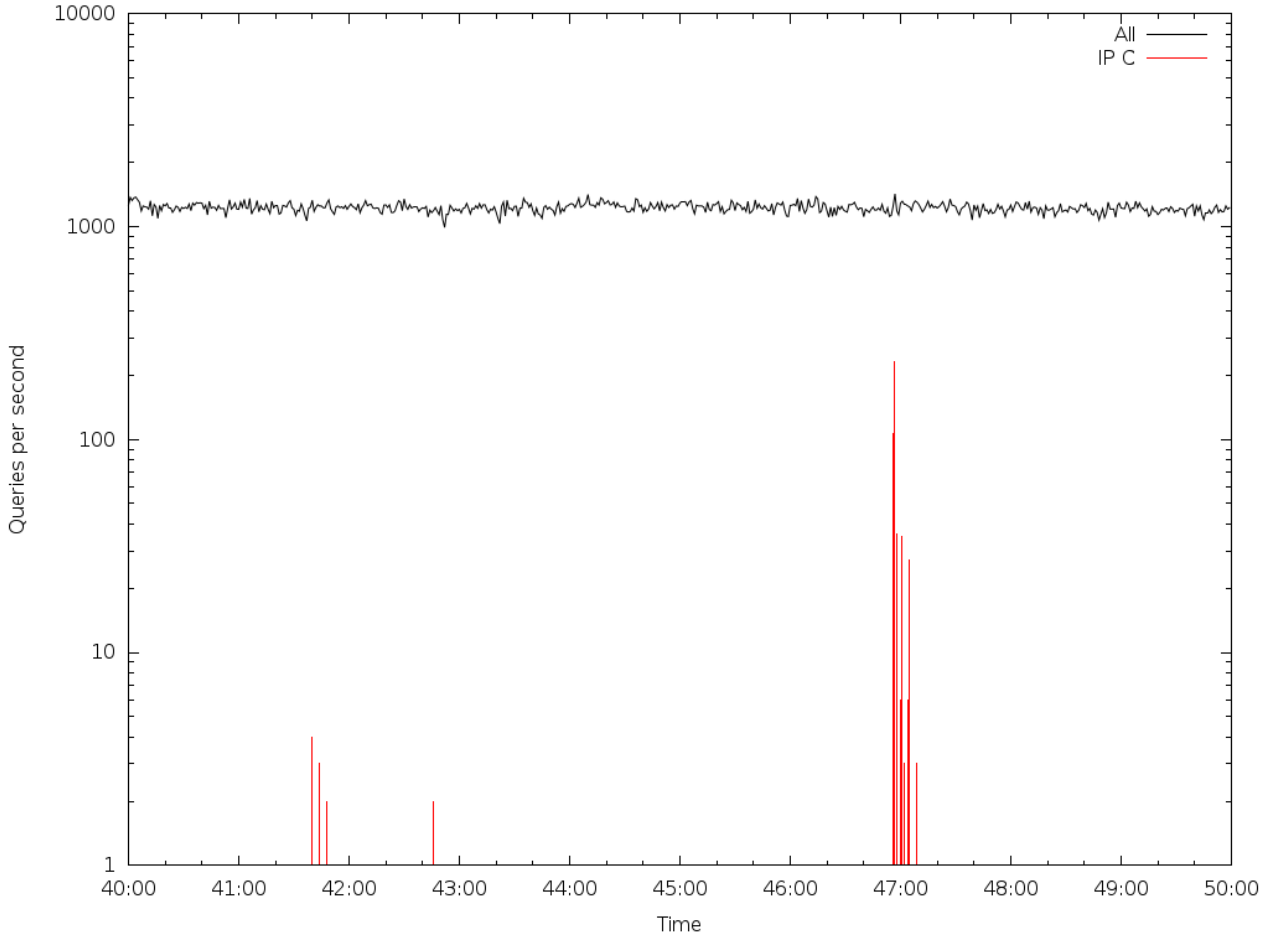
dotazy na A záznamy webserverů,
jména většinou seřazena.
(zdroj US)
Zřejmě webcrawlery.



Detekované anomálie – SrcIP

andraste, 1.11.2013 UTC 14:40

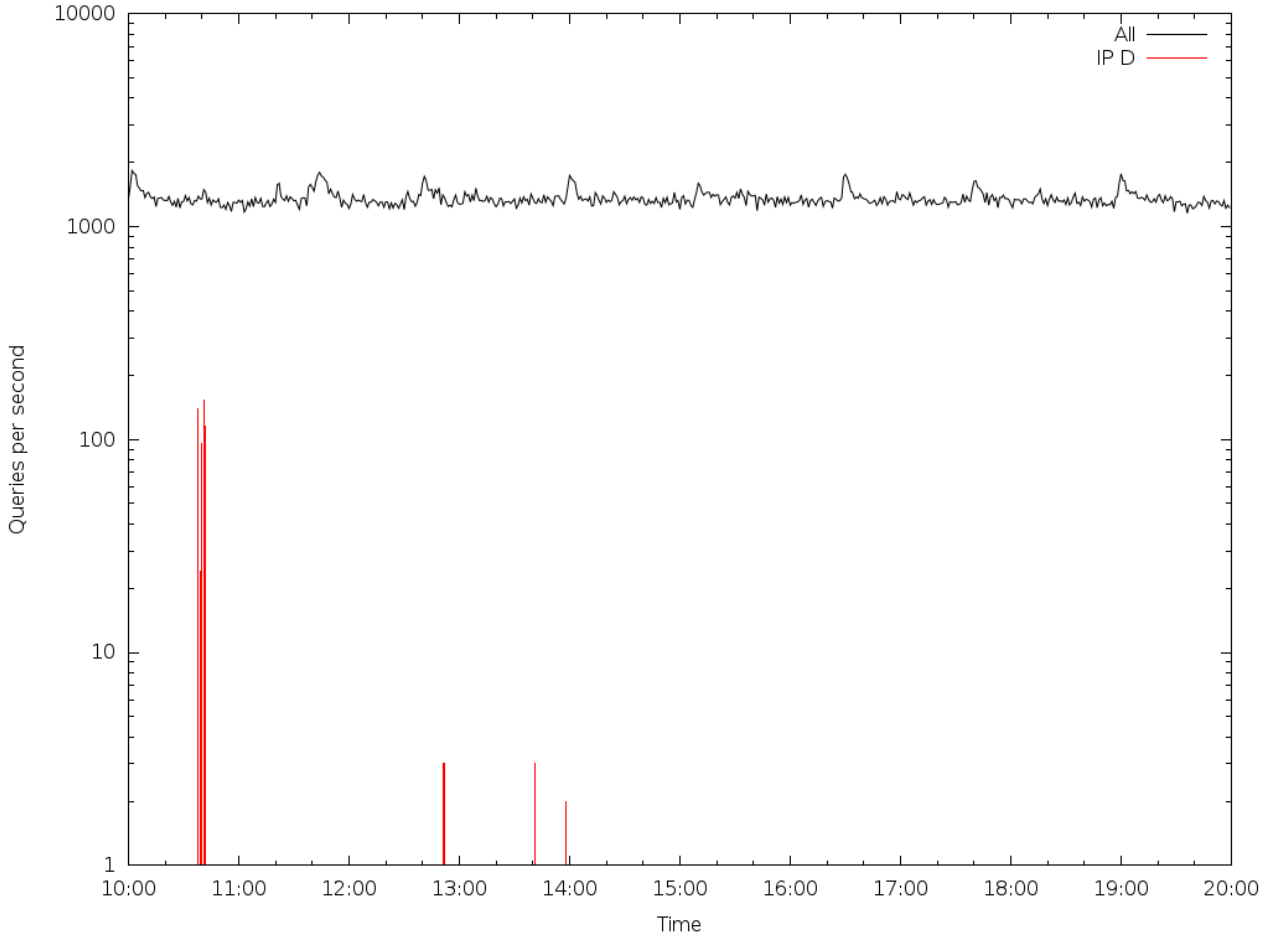
dotazy na MX a A
mail*.*.cz, mx*. , ms*. ,
smtp*. , antispam*. , gw*.,
nagios*. v české zóně.
(zdroj AU)



Detekované anomálie – SrcIP

andraste, 1.11.2013 UTC 16:10

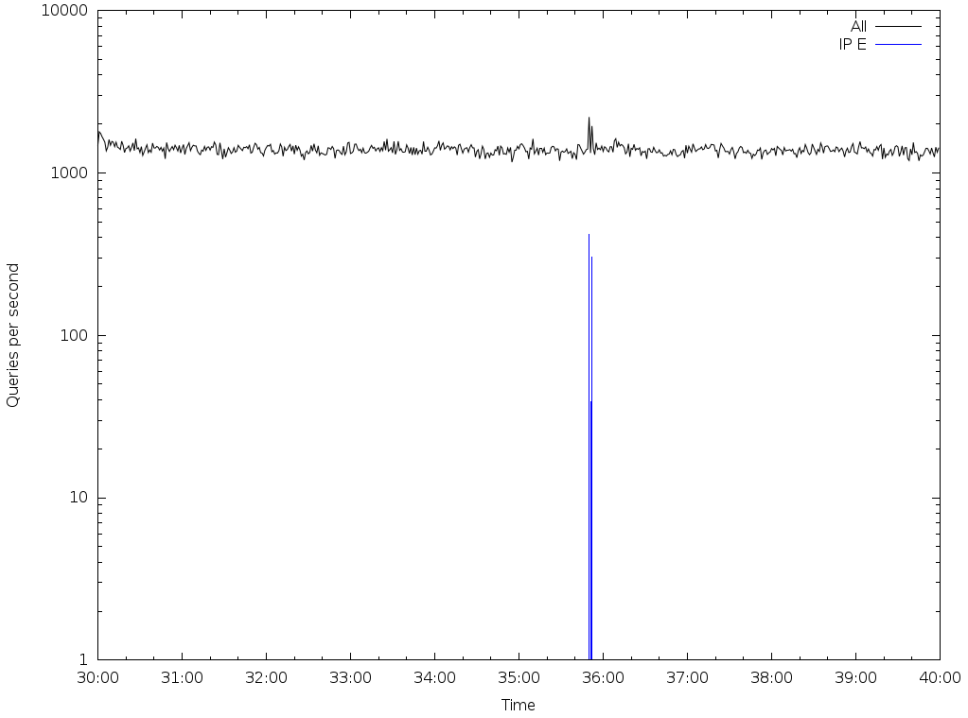
dotazy na A záznamy *.mp3-zdarma-stazeni.cz, kde * je ve formátu jméno-interpreta-název-skladby.
(zdroj RU)
odpovědi jsou NXDomain

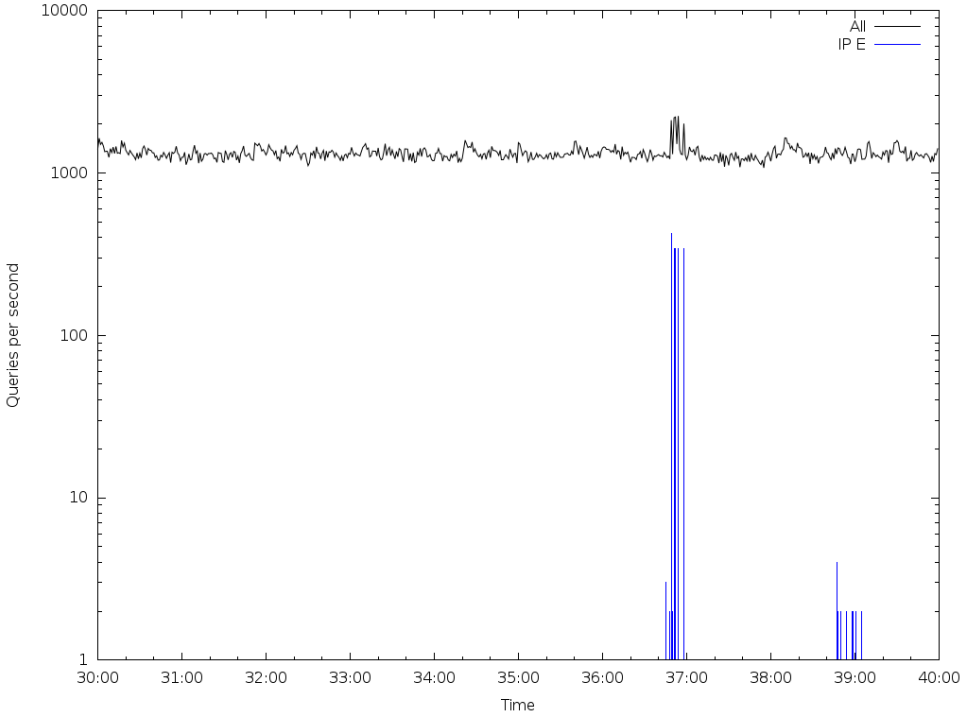


Detekované anomálie – SrcIP

dns-s-01, 1.11.2013 UTC 13:30 a 15:30

~700 dotazů na A a.ns.secunet.cz a b.ns.secunet.cz, (zdroj US), zřejmě skript

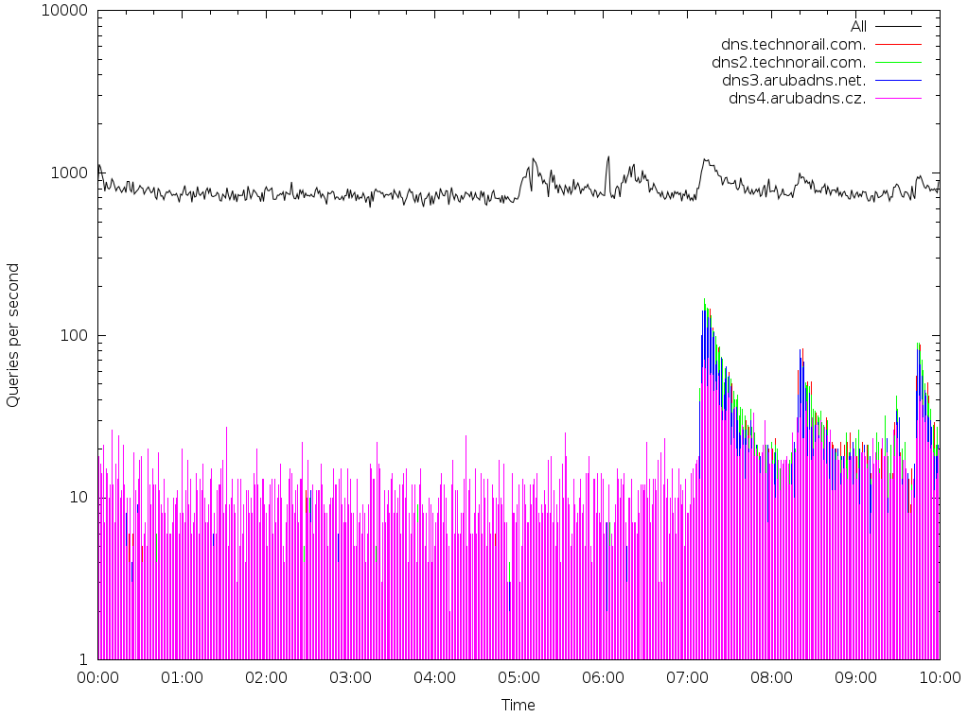


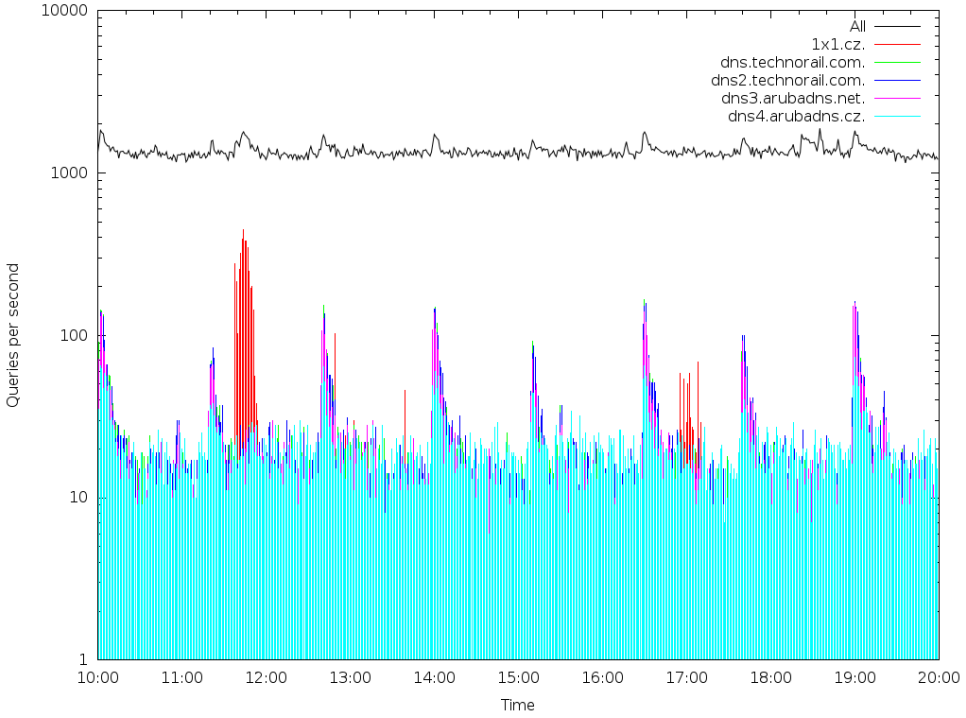


Detekované anomálie – QName

andraste, 1.11.2013 UTC 00:00 a 16:10

Dotazy na A záznamy. (IP z celého světa) Pozorováno v průběhu celého dne. Pokus o amplification attack?





Shrnutí

- Velkou část detekovaných anomálií lze rozdělit do několika skupin:
  - legitimní provoz – resolvery, web crawlery
  - výčet zóny – slepý, se slovníkem nebo se znalostí obsahu zóny
  - podezřelý – špatně fungující resolver, špatně nakonfigurovaná zóna, testovací skripty
- TODO:
  - detekce podle ASN/GeoIP, spolupráce s DSC(ng)

Děkuji za pozornost

Karel Slaný • karel.slany@nic.cz