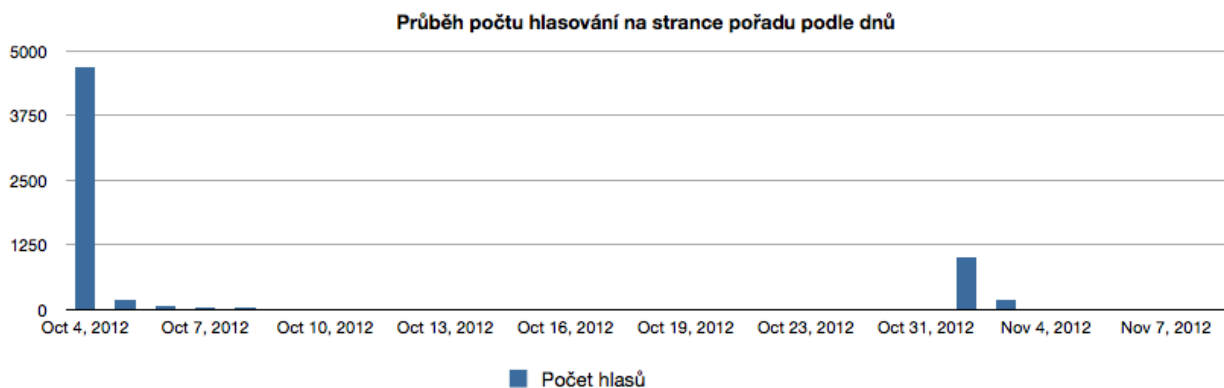


## Hlasování k pořadu Hyde Park ČT24 ze 4.10.2012 (host: Jaromír Drábek)

### Předběžný rozbor anomálií spojených s hlasováním o výkonu hosta v tomto pořadu

#### Cíle

Cílem rozboru byla identifikace případných podezřelých hlasů během samotného pořadu, případně detekce další manipulace po pořadu. Podkladem byl XLS soubor s logy hlasování poskytnutý Českou televizí.



V denní agregaci jednotlivých hlasů jsou patrné dvě špičky. Jedna v den vysílání 4. října a druhá 1. listopadu 2012. Celkem bylo hlasováno 3 190 hlasy negativně a 3 137 pozitivně. Při první špičce bylo zasláno 4 693 hlasů, a to v den vysílání. Při druhé špičce bylo přidáno 1 014 nových hlasů.

Během 4. října to bylo 3 032 kladných hlasů a 1 661 záporných. Při druhé špičce bylo všech 1 014 hlasů záporných. Ani jeden tento poměr neodpovídá poměrům počtu hlasů v drtivé většině dalších dnů, kdy byly počty hlasů sice výrazně menší, ale většinou v poměru 3:1 mezi počtem záporných a kladných hodnocení.

Podle interních pravidel ČT by nemělo dojít k odeslání více než 3 hlasů z jedné IP adresy. Přesto bylo takto odesláno 105 hlasů, jedná se však o zanedbatelný počet a proto budeme tento incident nadále ignorovat.

### Podíl spambotů na celkovém počtu hlasování

V této části analýzy dochází k porovnání IP adres hlasujících vůči seznamu IP adres registrovaných jako zneužívaných spamboty. API stránek Stop Forum Spam (<http://stopforumspam.com/>) vyhodnotilo celkem 1 444 IP adres, z nichž se hlasovalo jako IP adresy podezřelé. Z 1 444 hlasů bylo 1 153 hlasů negativních, 291 pozitivních. Jejich distribuce v čase však byla zřetelně jednostranná. Nejvíce podezřelých IP adres hlasovalo 1. listopadu, a to 960. Následovalo hlasování 4. října, kdy se jedno o 307 hlasů. Třetí den bohatý na podezřelé hlasy pak byl 2. listopad se 172 hlasy.

**V praxi tak bylo minimálně 960 hlasů z 1 044 v den 1. listopadu podezřelých. Podotýkám, že se jednalo výhradně o hlasy negativní, podíl hlasů poslaných z podezřelých IP adres byl 92 procent hlasů. V případě hlasů z 2. listopadu se jednalo o poměr ještě větší: 172 podezřelých hlasů ze 176.**

**Pokud jde o hlasy ze 4. října, tak z podezřelých IP adres bylo 307 z celkově 4 693 odevzdaných. Jednalo se tedy o necelých 7 procent hlasů. Z pohledu hlasování to pak bylo 289 hlasů pozitivních a 18 negativních.**

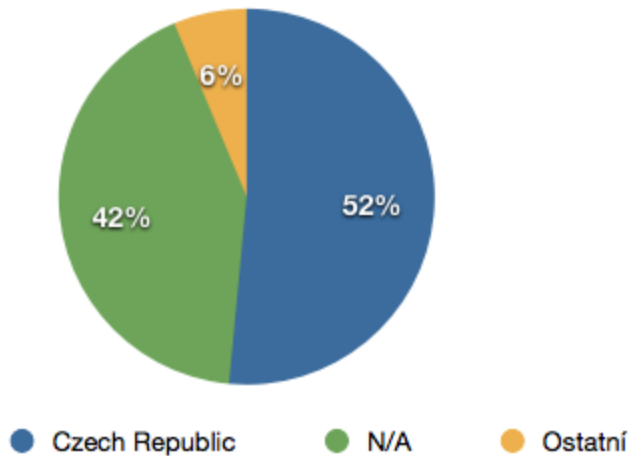
V případě podezřelých hlasů ze 4. října je sice jejich procento vyšší a poměr neodpovídá celkovému rozložení hlasů, ale i pokud by tyto hlasy byly odečteny, nedošlo by k žádné zásadní změně v celkovém poměru. Přesto, i pokud jde o hlasování 4. 10., patří tento podíl spolu s radikálně odlišnou distribucí poměrů hlasů k podezřelým faktorům.

Z analýzy hlasů z podezřelých IP adres lze učinit první předběžný závěr: Hlasy z 1. listopadu a 2. listopadů jsou z velkou pravděpodobností zmanipulované. Svědčí o tom jak radikální vychýlení v prospěch jedné volby, tak obrovské zastoupení hlasů z podezřelých IP adres. Dále se mu tedy nebudeme věnovat a soustředíme se na hlasy pouze ze dne vysílání.

### Distribuce hlasů podle zemí

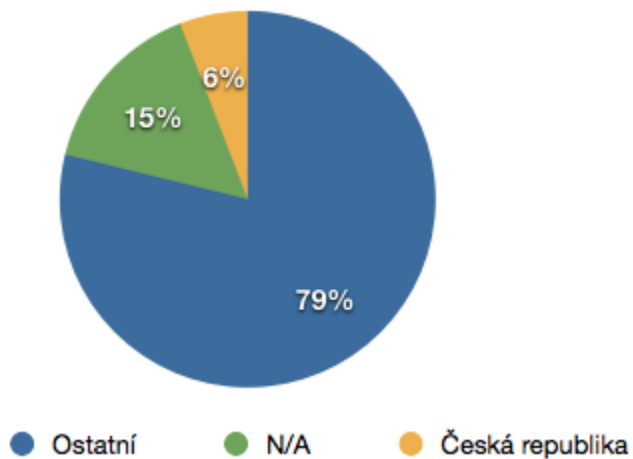
Během hlasování 4. října přišly hlasy z celkem 26 strojově identifikovatelných lokací. Celkem 1 980 hlasů přišlo z lokací strojově přímo neurčitelných.

**Podíl zemí podle IP adres**



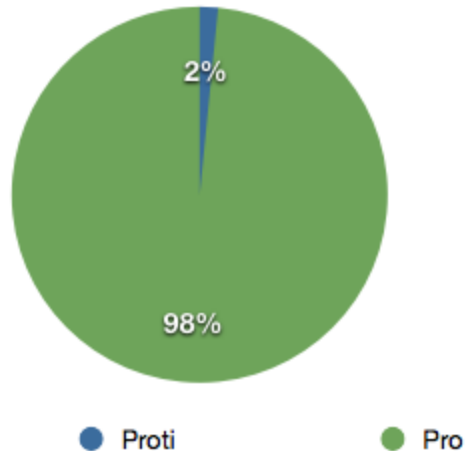
Drtivá většina IP adres definovaných jako podezřelé (viz předchozí část) přišla z právě z lokací v cizině.

**Rozložení pravděpodobných spambot účtů podle zemí**



Pokud jde o hlasy z lokací strojově automaticky neurčitelných (dle záznamu v db WHOIS), pak na první pohled překvapí, že se jednalo výhradně o hlasy podporující jednu stranu.

### Hlasy z neurčitelné země



Při bližší analýze se ukázalo, že se jedná především o hlasy z IP rozsahu u společnosti Telefonica, zbytek tvořily především anonymizéry TOR a další adresy. V případě rozsahu Telefonica se však jednalo minimálně o 1 765 hlasů, které byly v drtivé většině pozitivní, což je distribuce krajně nepravděpodobná. Více už v dalším kroku.

### Distribuce podle adresních rozsahů

Z analýzy IP adres hlasujících vyplývá, že mezi IP adresami, které se účastnily hlasování, dominují silně dva rozsahy. První je podle záznamu WHOIS veden u společnosti Telefonica. Během vysílání z něj přišlo 1 765 hlasů, z nichž bylo 1750 pozitivních a 15 negativních.

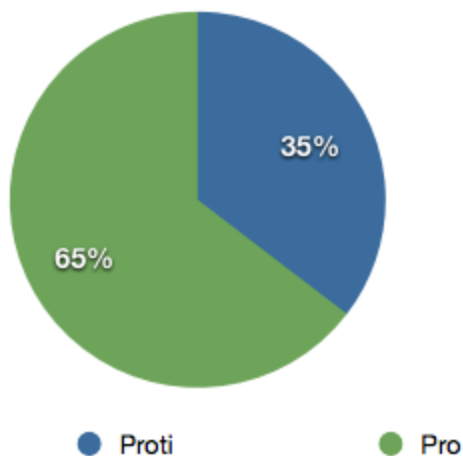
Druhý výrazný rozsah IP adres je registrován na společnost Vodafone. Prostřednictvím těchto IP adres bylo zaznamenáno celkem 414 hlasů, z nichž 407 bylo pozitivních a 7 negativních.

Je vysoce nepravděpodobné, aby došlo k přirozenému hlasování z jednoho rozsahu v takové míře. I tyto hlasy lze považovat za zmanipulované.

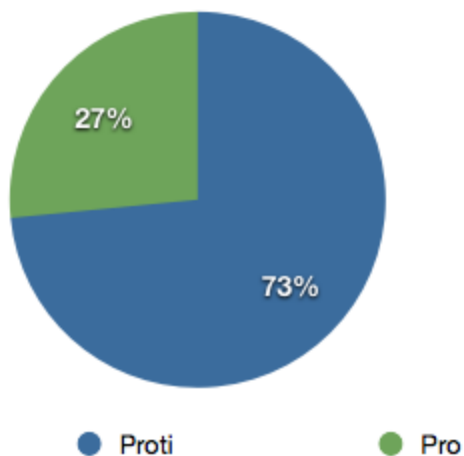
### Závěr a korekce výsledků

Na základě předběžné analýzy můžeme říci, že je více než pravděpodobné, že s hlasováním k pořadu Hyde Park 4. října bylo manipulováno jak v době vysílání pořadu, tak později. K manipulaci byly zřejmě použity jak hlasy podezřelé z účasti ve spamovací síti, tak automatizované hlasy z nakoupených IP rozsahů. Proti oběma těmito metodám se Česká televize neměla v době vysílání možnost efektivně bránit.

**Původní výsledek hlasování z 4.10.**



**Výsledek po korekci podezřelých hlasů**



Nyní můžeme přejít ke korekci hlasování ze 4. října a odečíst hlasy spambotů (Proti 18, Pro 289), hlasy z podezřelého rozsahu Telefonica (Proti 15, Pro 1 750) i rozsahu Vodafone (Proti 7, Pro 407). Po korekci dostaneme následující výsledky.

<b>Původní hlasování ze 4. října</b>	
Proti	1,661
Pro	3,032
<b>Výsledek po odečtení podezřelých hlasů</b>	
Proti	1,621
Pro	586