



Autors: Labi Raksta ID= 324

## Par captchas izmantošanu cīņā ar spam-botiem



Jau pusotru gadu WAP blogiem "uzmācas" boti, kas piepludina komentārus ar nederīgu informāciju. Ņemot vērā ierobežojumu rakstīt komentārus ne biežāk kā reizi 25 sekundēs, dienā tiek ģenerēti ap 200 komentāri. Tādā veidā, atstājot šo vietni bez pieskatīšanas, mēneša laikā tiek radīta tāda "miskaste", ka nekas cits neatliek kā dzēst nederīgos komentārus. Pietam, ja tas jādara katru dienu, tad drīz vien var nojūgties. Vairākas reizes gadījās tā, ka komentāri pārsniedza datubāzes datu apjoma kvotu, līdz ar to blogs "apstājās", jo nevarēja pievienot ne komentārus, ne jaunus rakstus.

Standarta aizsardzības metodes, IP un User-Agent (UA) bloķēšana, šajā gadījumā ir bezspēcīgas, jo spam-botam ikreiz mainās IP un dažreiz arī UA. Tāpēc nekas cits neatlika, kā ieviest komentāru pārbaudi ar drošības kodu jeb angļu valodā captcha ("*Completely Automated Public Turing test to tell Computers and Humans Apart*" ["*pilnīgi automātisks publisks Tjūringa tests, lai atšķirtu datorus no cilvēkiem*"]) vairāk var palasīt (WEB)<http://lv.wikipedia.org/wiki/CAPTCHA> vai angļiski (WEB)<http://en.wikipedia.org/wiki/CAPTCHA>. *Captcha's* galvenā doma ir atšķirt datoru skriptus (botus) no cilvēka. Laikam, mūsdienu internetā bez tā nevar iztikt, ja mājaslapai "*tiel tas gods*" paciest nemitīgu botu "uzmākšanos". Sākotnēji ar drošības kodu aizsargātiem pakalpojumiem bija 99.9% garantija, ka neviens algoritms nevarēs to "nolasīt". Taču līdz ar mākslīgās inteliģences attīstību, *Captcha* pietiekami efektīvi vairs nesniedz vēlamu rezultātu.

Kā tiek nolasīts teksts no attēla? - Eksistē vairāki optiskās rakstzīmju atpazīšanas mehānismi OCR (Optical Character Recognition), kas spēj algoritmiski noteikt tekstu, kas "slēpjas" attēlā, piemēram,



(WEB) <http://www.free-ocr.com/> un (WEB) <http://www.onlineocr.net> ir bezmaksas OCR servisi, kas spēj atpazīt pat latviešu valodu. Mūsdienās pastāv divas koderu frontes - vieni ir tie, kas mēģina pilnveidot drošības kodu nolases algoritmus, bet otri ir tie, kas mēģina ģenerēt tādus attēlus, kurus spēj nolasīt tikai cilvēks, papildinot kodu ar grafiskiem fona trokšņiem, laužtām līnijām, saplūstošām krāsām un burtu savērpšanu. Šobrīd grūti pateikt, kuri no viņiem ir pārāki, tomēr jāsaprot, ka neviena *Captcha* nav 100% droša.

Attēlā redzamais WAP blogu drošības koda mehānisms ir pavisam vienkāršs - attēlā parādās unikāla 5 simbolu kombinācija, kas var būt attēlota ar vienu no 2 dažādiem šriftiem, papildus teksts tiek sašķiepts pēc gadījuma rakstura izvēlēta leņķa. Pārbaudot paštaisīto drošības attēlu iepriekš minētajos OCR servisos, izrādījās, ka otrais (<http://www.onlineocr.net>) veiksmīgi spēj noteikt attēla saturu, ja tas nesatur ciparu 4, taču pirmais (<http://www.free-ocr.com/>) nespēja pareizi noteikt nevienu no piedāvātajiem drošības kodiem. Tomēr domāju, ka WAP blogu vajadzībām esošā *captcha* būs pietiekami laba.

Nobeigumā gribēju piedāvāt lasītājiem notestēt jaunizveidoto drošības kodu un komentāros padalīties ar iespaidiem, pamanītajām kļūdām un nepilnībām, norādot versiju (WAP/WEB).

Labots: 20:28 25-03-13 Uzrakstīts:20:27 25-03-13