Web-palveluihin kohdistuvien hyökkäysten analyysi

Jyväskylän yliopisto

Joel Lehtonen
Kristian Siljander

Yleistä

- Osa ISSM-projektia
- Kirjoitamme gradua aiheesta
  - WWW-palvelun arkkitehtuuri
  - Web 2.0:n tietoturvariskit
  - Palveluihin kohdistuvat hyökkäykset
  - Anomalioiden tunnistaminen

Mitä meillä on?

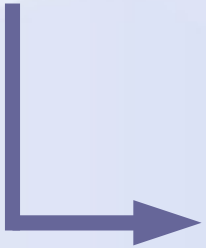
- Data saatu Ixonos Oyj:n tuotannosta poistetuilta palvelimilta
- 24 GiB pakattua Apache-lokia
- Noin miljoona? sivupyyntöä
- Uniikkeja IP-osoitteita n. 10 000 kpl
- Kerätty n. 6kk ajanjaksolta

Esikäsittely

- Lokit luetaan MySQL-tietokantaan käyttäen PhasefulSplitter-sovellusta
- MySQL:n käyttö helpottaa suuren tietomassan käsittelyä.
- Sivupyynnön sisältö pilkotaan useisiin sarakkeisiin.

Esikäsittely

130.234.49.2 - - [10/May/2009:15:53:01 +0300] "GET /images/icn.gif HTTP/1.1" 200 2680 "http://www.jyu.fi/a.html" "Mozilla/5.0 (SymbianOS/9.2;...)"



Sarake	Esimerkki
IP	130.234.49.2
Date	2009-05-10 12:53:01
Server	kotka
Service	mobi
Request	GET /images/icn.gif
Response	200
Bytes	2680
Referer	http://www.jyu.fi/a.html
Browser	Mozilla/5.0 (SymbianOS/9.2;...)

Klusterointi

- Käyttämämme anomalia-analyysi pohjautuu diffuusiokuvauksiin
- Datan tulee olla luokka-asteikollista
- Numeerinen data on klusteroitava
  - esim. bytes
- Osa sarakkeista on jo luokka-asteikollisia
  - esim. response, browser

Lisätiedon kerääminen

- Joistakin sarakkeista on saatavilla lisää tietoa yhdistämällä eri lähteitä
  - GeoIP-tietokanta
  - Selainten tunnistetietojen luokittelu

Esimerkki: IP-osoite

130.234.169.73

Sarake	Esimerkki
Country Code	FI
Region	15
Region Name	Western Finland
City	Jyväskylä
Latitude	62.2333
Longitude	25.7332
ISP	University of Jyvaskyla network
Organization	University of Jyvaskyla

Avoimia kysymyksiä

- Tiedon kategorisoinnin menetelmät
- Anomalioiden tunnistaminen
- Menetelmien toteutettavuus käytännössä