Sumeat tiivisteet

Soveltaminen haittaohjelma-analyysissa

Sumea tiivistys

- Tunnistettavan syötteistä samankaltaisuuksia
- Ominaisuuksien valinta
- Tiivistetty ominaisuusjono
  - Syötteestä valitaan ominaisuuksia tiivistettäväksi
- Tavujono
  - Syöte tietyn kokoisiksi lohkoiksi
  - Lohkoista rakennetaan tiiviste

Paloittain määritelty tiiviste

- Koko syöte lohkotaan
- Ominaisuudet syötteen lohkoja
- Ominaisuuksien tiivistys
- Tiivisteen muodostus
- Lohkokoko
  - Vakio
  - Kontekstista riippuva
- Suorituu heikosti ohjelmistojen analysoinnissa

Epätodennäköiset ominaisuudet

- Samankaltaisissa syötteissä samoja epätyypillisyyksiä
- Ominaisuudet muusta syötteestä poikkeavimpia
- Vertailussa etsitään toisesta syötteestä näitä
- Sopii katkelmien löytämiseen
  - Esim. tietyn koodinpätkän sisältävän ohjelman tunnistus
- Vajaa syötekattavuus
  - Ei hyvä kokonaisten tiedostojen vertailussa
  - Mahdollistaa tietynlaisia hyökkäyksiä
- Sdhash

Enemmistö

- Enemmistö ei muutu pienistä muutoksista
- Jaetaan syöte lohkoihin
- Arvo alkioiden enemmistön perusteella
- Ei mahdollista suurta määrää hyökkäystapoja
- Mvhash-b

Ohjelmistot

- Ssdeep
  - Palottain määritelty, kontekstiriippuvainen
  - Tunnistaa heikoimmin haitallista sisältöä
  - Asteikko 0-100
- Sdhash
  - Epätyypilliset ominaisuudet
  - Ominaisuudet 64 tavun merkkijonoja
  - Soveltuu haittaohjelma-analyysiin
  - Asteikko 0-100
- Mvhash-b
  - Lohkon alkioiden enemmistö skaalattuna arvoon 0 tai 255
  - Suorituskykyinen
  - Asteikko 0-100, käänteinen

Haittaohjelma-analyysi sumein tiivistein

- Käytössä staattisessa analyysissa
  - Kryptografiset tiivisteet soveltuvat heikosti
- Ohjelmisto
  - Verrataan tiivisteindeksiin
  - Osuma?
- Katkelma
  - Etsitään tiettyjä piirteitä sisältävä ohjelmisto
  - Tietty versio, kirjasto, koodinpätkä ym.

Harhauttaminen

- Haittaohjelman naamiointi
  - Lähdekoodin muuttaminen
  - Käännös ja linkitys
- Aiheettomat osumat
  - Muutetaan merkityksetöntä dataa
  - Tiiviste samankaltainen kuin haittaohjelman
  - Kuluttaa tutkintaresursseja
- Ohjelmistot reagoivat eri tavoin
- Osa alttiimpia tietynlaisille hyökkäyksille

Esimerkki naamioidusta roskapostista (Oliver, Forman ja Cheng, 2014)

Haasteet

- Semanttista merkitystä ei tulkita
- Käyttö ja tulkinta haastavaa
- Johdettavissa harhaan
- Dataa on helppo naamioida
- Pakattua dataa ei voi vertailla
- Ohjelmistojen erot