Analisis dan implementasi knowledge discovery menggunakan algoritma multidensity dbscan untuk mengidentifikasi tren malware pada database honeypot dionaea = Analysis and implementation of knowledge discovery using dbscan multidensity algorithm to identify malware trend in database of dionaea honeypot

Martin Dominikus Tjandra, author

Abstrak

ABSTRAK
Dionaea adalah honeypot yang memiliki tujuan utama mendapatkan salinan dari malware. Setelah mendapatkan salinannya, proses knowledge discovery dilakukan untuk mendapatkan informasi dari database Dionaea. Dua alasan utama penggunaan knowledge discovery adalah data terlalu banyak namun informasinya sedikit, dan untuk mengekstrak informasi yang berguna dan menafsirkannya. Proses knowledge discovery memiliki beberapa fase, yaitu pembersihan data, seleksi data, transformasi data, prekalkulasi, data mining, evaluasi pola, dan penyajian informasi. Proses data mining menggunakan variasi algoritma DBSCAN, yaitu multidensity DBSCAN. Analisis dibagi menjadi dua, yaitu analisis cluster dan dataset. Analisis cluster menjelaskan hubungan antara lokasi negara penyerang berdasarkan daerah Internet Registry-nya dan persentase deteksi malware berdasarkan beberapa vendor antivirus. Dari analisis dataset, didapatkan informasi bahwa malware yang paling sering digunakan penyerang atau tren malware, berjenis Downadup, yaitu sebesar 71.1%. Negara yang paling sering menyerang adalah Rusia dan beberapa negara Eropa. Sebagai pembanding, laporan tahunan yang dipublikasi Microsoft, ENISA, dan F-Secure pada akhir 2014 menunjukkan tren malware yang sama, yaitu berjenis Downadup.
<hr>
ABSTRACT
The main purpose of implementation of Dionaea is to get copy of malwares. After that, knowledge discovery is applied to get information from Dionaea?s database. Two main reasons to use data mining method are data is too large but only contain few informations, and to extract useful informations and interpret them. Knowledge discovery process have several steps, they are data cleaning, data selection, data transformation, precalculation, data mining, pattern evaluation, and knowledge representation. Data mining process uses multidensity DBSCAN. There are two main sections of analysis, cluster analysis and dataset analysis. Cluster analysis show the relation between attackers? country location which is based on their Regional Internet Registry and malware detection rate from several antivirus vendor. Dataset analysis shows the most frequent country whose attacker is Conficker variant, 71.1% of all dataset is Conficker worm incident and the mode of attacker country is Russia and severals Europe countries. This outputs show similarity about threat landscape and malware in Asia, compared to annual report by Microsoft, Enisa, and F-Secure which was published at the end of 2014, which stated Downadup as most frequent malware.