Analisis tren dan perbandingan performa metode forecasting regresi linier berganda dan regresi dengan ARIMA error berdasarkan data aktivitas malware C&C di Indonesia = Trend analysis and performance comparison of forecasting method multiple liniear regression and regression with ARIMA errors based on C&C malware activity data in Indonesia

Ahmad Shoheh Dwi Ristono, author

Abstrak

ABSTRAK
Penelitian ini membahas tentang analisis data time series berdasarkan data laporan trafik DNS Sinkhole dari ID-SIRTII. Analisis ini terbagi menjadi dua yaitu analisis tren dan perbandingan performa dua metode forecasting berdasarkan jumlah aktivitas malware berbasis command and control C C selama tiga bulan awal tahun 2016 di Indonesia. Analisis tren dilakukan dengan mengelompokkan malware berdasarkan keluarga dan varian teraktif. Tren analisis menunjukan total aktivitas malware sebanyak 1452585872 yang terdiri dari 12 keluarga malware berbeda. Keluarga malware yang paling aktif yaitu berasal dari keluarga B85 Dorkbot dengan persentase 40,49 dari total keseluruhan aktivitas malware dalam tiga bulan tersebut. Varian teraktifnya yaitu B85-R2V dengan persentase 37 dari total keseluruhan aktivitas malware. Perbandingan performa forecasting menggunakan dua metode yang akan dibandingkan yaitu regresi linier berganda dan regresi dengan ARIMA error. Berdasarkan perbandingan nilai MAPE dalam prediksi jangka waktu 1 minggu, kedua metode hampir memiliki kemampuan yang sama dalam memprediksi 10 varian malware C C terbanyak. Sedangkan dalam jangka waktu 2 minggu, metode regresi dengan ARIMA error memberikan kemampuan prediksi yang lebih baik dibandingkan dengan metode regresi linier berganda.
<hr>
<i>ABSTRACT
This bachelor thesis discuss about time series data analysis of DNS Sinkhole traffics from ID SIRTII. There are two main sections of this analysis, trend analysis and performance comparison of two forecasting methods based on C C malware activities in the first three months of 2016 in Indonesia. Trend analysis is performed by grouping malware based on family and the 10 most active variants. It shows 1452585872 total activity of malware consist of 12 different malware families. The most active malware is come from B85 Dorkbot family with 40.49 of total malware activities in those three months. Its most active variant is B85 R2V with 37 of total malware activities. Performace comparison of forecasting methods use multiple linear regression and regression with ARIMA errors. Based on the comparison of MAPE values in one week prediction period, both methods almost have the same ability to predict the top 10 C C malware variants. While within 2 weeks prediction period, regression method with ARIMA errors gives better prediction ability compared with multiple linear regression method.