Network Security Using Graph Theory Techniques

ISCX
Information Security Centre *of* Excellence

Ali Shiravi, Ali A. Ghorbani

Filtration

Filtering State Machines are used to block and let through certain entities related to a particular context. These machines can be combined to create functional means of dealing with selection and generation of data or information. The amount of state held between independent input entities varies. However, it is generally required to be very limited, due to the fact that these machines should perform very fast with limited resources. By State Machines we intend to indicate a lack of semantic understanding toward what is being accepted, denied, and converted.

ComGraph

The extracted flows and their corresponding metadata is fed into the ComGraph module in order to extract nodes of interest and the interactions among them. The ComGraph Module is responsible for translating flow info into nodes and edges for placement within the communication graph. It is composed of many graphs for each host, a group of hosts, subnets, and also for multiple services of interest. It also contains the snapshots of the these graphs at regular intervals.

Flow Extraction

In the implemented framework, a sophisticated Flow extractor is implemented to initially extract the flows by precisely following the TCP State Machine on both ends of the communication. It is designed to be placed at the network gateway and monitor all passing communication. In cases where the TCP flow is incomplete, due to dropping of packets, or is anomalous, several heuristics are utlized to infer the state of the flow. I will also mark incomplete and suspicious flows for later use in the Analytics.

Time Series

Several graph metrics, both static and dynamic (time-window based) are extracted from the corresponding graphs. This is done for each host, a group of hosts, different subnets, and for all required services. These extracted features are then represented in the form of time-series. To reduce the dimensionality of the time series data, we apply simple but effective approximation and sampling methods such as PAA.

Behaviour Analysis

Our current process of identifying behavioural trends is to take the time series information and feed it into various clustering algorithms. The clustering that is currently taking place is done at three stages. The first stage is to find particular types of behaviour of various hosts using the time series raw data (multiple metrics) and see whether a particular recurring types of behaviour is identified. Second, the clustering is applied to the time series as whole on the combination of hosts and metrics that we find appropriate. At the third stage we apply clustering to the clusters of the previous stage to find super-clusters. The final output are profiles for each host, a group of related hosts and whole subnets.

Situation Awareness

Our focus here is to move from the traditional data fusion to developing processes and related algorithms that can accurately and effectively abstract this collaborative behaviour and present it to the decision makers. This is driven by the high volume of traffic moving over the network and the complex relationships that form between hosts. This provides decision makers with reliable and insightful view of their network environment. In formal terms, what we are doing here with analyzing Communication Graphs is situation assessment in order to arrive at a state of knowledge regarding the network.

Monitoring

Through the precise monitoring of the various behavioural profiles and their corresponding members, we are able to identify whether a particular host or a service used by that host is behaving abnormally. These are subsequently flagged for analysis by the system admin.