DETECTION AND PREVENTION OF ADVANCED PERSISTENT THREATS (APT) ACTIVITIES IN HETEROGENOUS NETWORKS USING SIEM AND DEEP LEARNING

Code developed as a part of IBM funded project
Members : Balaji Bharatwaj M, Aditya Reddy M
Research Supervisor – Senthil Kumar T
In support with : Sulakshan Vajipayajula
Architect CTO Office, IMB Security Bangalore – svajipay@in.ibm.com

Problem statement

To design a model that detects Advanced Persistent Threats in heterogenous networks using Hidden Markov Models

Methodology

1. Use Wireshark to track the network and obtain the .pcap file
2. Export the .pcap file into CSV (Comma separated value) files. Then convert that into a variable dataframe for easy manipulation
3. Extract the required features from the variable data frame
4. Apply Hidden Markov Model to find the outliers from the dataset
5. Results can be visualized as a Dendrogram and Packet analysis diagram

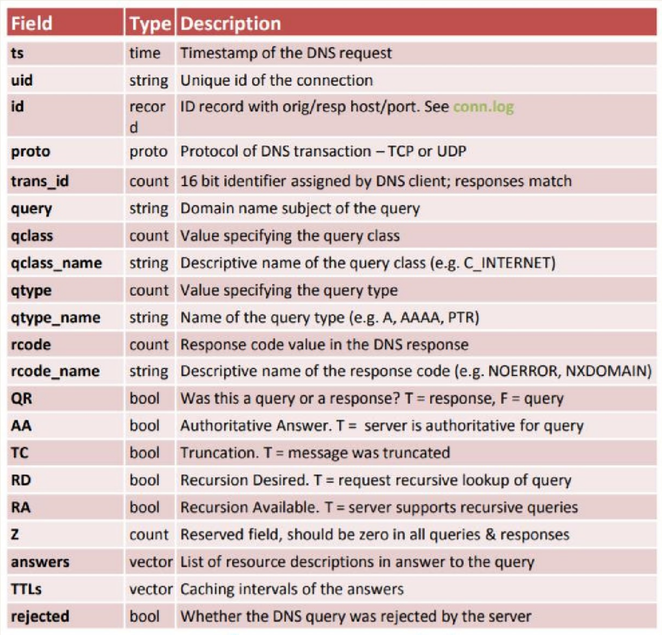
Current Status

Completed the implementation using Hidden Markov Model and output is visualized using a Dendrogram and Packet Analysis diagram

Dataset Information

DNS log dataset is used for our implementation.

Source of the dataset – secrepo.com – Samples of security related data



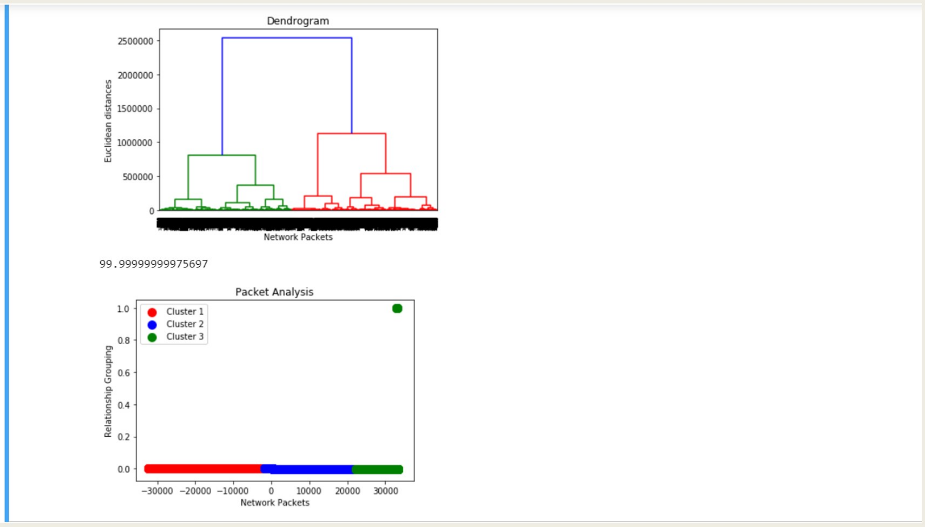
Field	Type	Description
ts	time	Timestamp of the DNS request
uid	string	Unique id of the connection
id	record	ID record with orig/resp host/port. See conn.log
proto	proto	Protocol of DNS transaction – TCP or UDP
trans_id	count	16 bit identifier assigned by DNS client; responses match
query	string	Domain name subject of the query
qclass	count	Value specifying the query class
qclass_name	string	Descriptive name of the query class (e.g. C_INTERNET)
qtype	count	Value specifying the query type
qtype_name	string	Name of the query type (e.g. A, AAAA, PTR)
rcode	count	Response code value in the DNS response
rcode_name	string	Descriptive name of the response code (e.g. NOERROR, NXDOMAIN)
QR	bool	Was this a query or a response? T = response, F = query
AA	bool	Authoritative Answer. T = server is authoritative for query
TC	bool	Truncation. T = message was truncated
RD	bool	Recursion Desired. T = request recursive lookup of query
RA	bool	Recursion Available. T = server supports recursive queries
Z	count	Reserved field, should be zero in all queries & responses
answers	vector	List of resource descriptions in answer to the query
TTLs	vector	Caching intervals of the answers
rejected	bool	Whether the DNS query was rejected by the server

Dataset Features – Source: secrepo.com

Algorithms used

- For Comparison – KNN, K-Means
- Implementation - HMM

Results - Output



```
99.99999999975697
```