

1. Collecting PE File- Honey Client , Amun, VX Heaven
2. Dynamic Analysis- CW Sandbox, Anusbis
3. Human Based Behavior Analysis- is it benign type or malware ? Knowledge store – malware Profile + feature – might have different result on different malware analysis tool
4. Statistical Report- Given by cuckoo sandbox
5. PE file future Extraction- using Perfile for feature Extraction.
6. Feature Selection SPARK -MLlib was used as first stage for feature   
   selection
7. Future reduction:- " ChiSqSelector was applied getting 10% of the most relevant features, Relief,   
   Chi-squared, F-Statistics, PCI etc . It was possible to select the best features runningthe Random Forest library {randomForest} on R- Cran{R studio}{ ranking by decrese on accuracy and ranking on decrease on node impurtiy -see importantace on randomForest package. } In this process sixstep of reduction were developed running for each step the algorithms for classification. First it was reduce to 1000 to then 500 then 300m,finally 9.
8. Feature Vector - Convert to Feature Vector-using online tool :- jubatus- <http://jubat.us/en/>- Disrtibuted online machine learning framework
9. Train calc and update each FV weight based on its frequency and label it
10. Classification :- Learning during training using Genetic algorithm , Deep learning , Decision Tree
11. Performance analysis :- using Deep learning or Decision Tree