RS∧°Conference2015

Singapore | 22-24 July | Marina Bay Sands

SESSION ID: TTA-F04

High dimensional visualization of malware families



Matt Wolff

Chief Data Scientist

Cylance, Inc. @cylanceinc

Glenn Chisholm

Chief Technology Officer

Cylance, Inc. @cylanceinc





Visualization of Malware Families

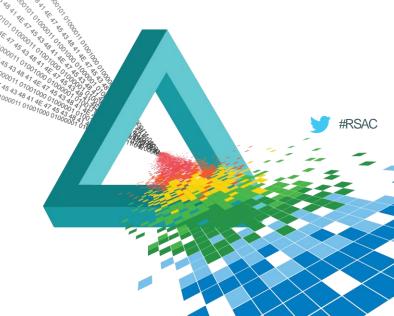
- Having an understanding of the relationships between individual pieces of malware provides insight to the analyst.
- This can be acomplished through detailed reverse engineering or reliance on industry naming and identification.
- We seek to utilize automated feature extraction to develop a feature set and capability to visualize relationships between sample sets of malware.
- Ultimately we represent a sub set of malware using Chernoff Faces and explore the relationships



RS∧°Conference2015

Singapore | 22-24 July | Marina Bay Sands

Naming and Identification of Malware





Malware Families

- Malware is commonly named to allow easy identification of function and purpose (Also for marketing value).
- Traditional Anti Virus companies have used one of:
 - Generic designations such as Trojan, Password Stealer or the ever helpful Generic.
 - Specific familial names to identify a specific branch of malware, this is common with known attack kits such as Zeus.
 - An attack specific name where a group of tools used in an attack are named together despite not having any relationship or in some cases being malicious.





Malware Families

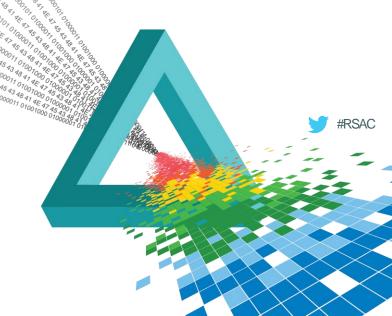
- Understanding the reliability of these names is important.
- An incorrectly named sample may lead to poor decisions by a security team.
 - Not every organisation has access to reverse engineers
 - If you don't truly know its purpose, responding is impossible
- Can we trust the names we get for a set of samples?
- Can we begin to see subsets of samples?



RS∧°Conference2015

Singapore | 22-24 July | Marina Bay Sands

The 65,000 Dridex Samples





Dridex

- We were presented with this set by some financial sector researchers who called these out as a set of Dridex samples.
- Dridex is a set of malware samples that targets end users to steal financial information.
- Once resident it monitors for specific financial institutions identified through a configuration file.
- Interesting in its wide distribution and well structured campaigns.





```
CRYPT4.BBZ
WIN32/DRIDEXH WIN32/MGON.OV
WORM:WN32/DORKBOTI
TROJAN-DOWNLOADERWIN32.DALEXIS
SUSPICIOUS.CLOUD.5
TROJANDOWNLOADERWIN32/DYNAMERIAC
TROJANSWIN32/INJECTOR.BRDO GENERIC.SEIO
WIN32/RIVETOR.BRDO GENERIC.SEIO
WIN32/RIVETOR.BRDO GENERIC.SEIO
WIN32/RIVETOR.BRDO GENERIC.SEIO
WIN32/RIVETOR.BRDO GENERIC.SEIO
WIN32/RIVETOR.BRDO GENERIC.SEIO
WIN32/RARAGANY
PACKED.GENERIC.459
TROJAN.GEN.2 TROJ.AGENT_054753.TOMB GENERIC_S.EHX TROJAN.WIN32.SENCPK.ZEA W32/VIRRNSM-E
MALWONTON -AN UDS:DANGEROUSOBJECT.MULTI.GENERIC
MALWONTON -AN UN32/ROPESTJ
TROJAN.WIN32/ROPESTJ
TROJAN.WIN32/ROPESTJ
TROJAN.WIN32/ROPESTJ
TROJAN.WIN32/ROPESTJ
TROJAN.WIN32.SHIPUP.BOH

MAL/ENCPK.-MP
GENERIC_S.EIC
MAL_ARKAM-4
GENERIC_S.EIL
TROJAN.WIN32.SHIPUPBOG
TROJAN.WIN32.SHIPUPBOG
```



Literally hundreds of names for a group of samples, can we start to understand which samples are closely related?





Dridex

- Are any of these samples infact Dridex?
- If so what are the names for the actual Dridex samples?
- The vast majority of these names are of no actual value, yet people rely upon them.
 - Some imply these are ransomware
 - Some are banking trojans (Maybe Dridex?)
- Can we begin to find populations of samples?

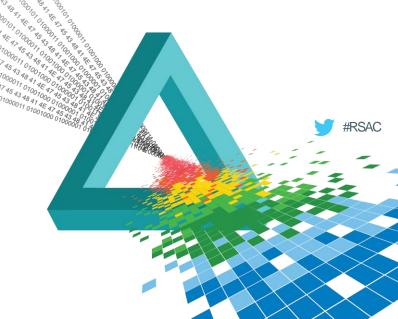




RS∧°Conference2015

Singapore | 22-24 July | Marina Bay Sands

Clustering the Samples





Fetaure Extraction

- Extracting a large number of features from the samples:
 - Structural
 - Headers
 - Sections
 - Content:
 - Strings
 - Resources
 - Code
- Can we establish relationships between samples?



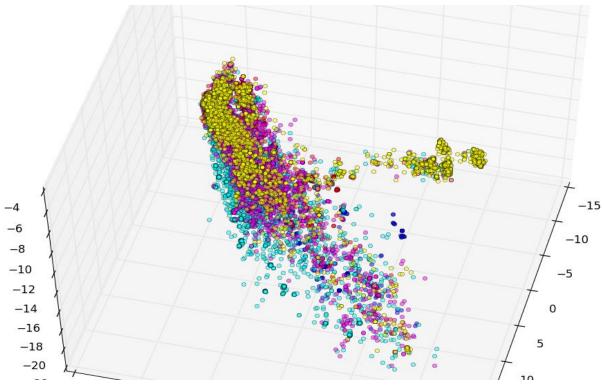


3 Dimensional visualization

- With the extracted features, we then use techniques to reduce the dimensionality down to 3
 - SVD
 - T-SNE
- What does the data look like in 3 dimensions?



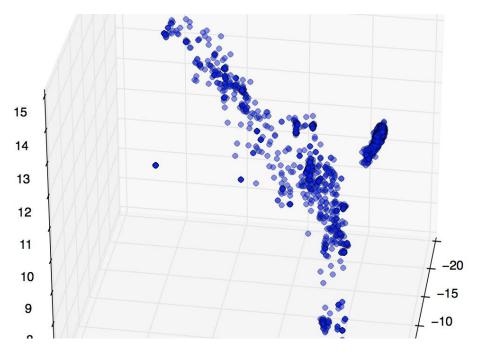




General malware. Reduced using SVD.







Clustering in 3 Dimensions

"Dridex" samples. Reduced using SVD.





Notes

- Visually, it is difficult to differentiate between different malware groups in 3 dimensions.
- Even the Dridex samples do not appear as a single cluster, but as multiple points in space.
- WordCloud method on Dridex clusters
 - Group Dridex into small clusters of close samples
 - Visualize the different names applied to samples in the cluster





Dridex Clustering Analysis

- Applying clustering techniques we extract 236 separate clusters from the initial set of 64,971 samples.
- The majority of clusters are smaller than 10 samples.
- 1661 of the samples did not fall into any cluster.
- However one cluster has 45,426 of the samples within it.
- Starting with this single cluster our goal will be to randomly select out samples for detailed analysis and determine the efficacy of the technique.





```
NOT-A-VIKUS:PSWTOOL.WIN32.NETPASSCIP
                                                                  TROI/ZBOT-EHY
           WIN32/CRYPTOR
                                                            S.REPUTATION.1
          TROJAN.WIN32.SHIPUPBOG
                ILCRYPT
                                                              GENERICOBFUSCATEDG
                                TSPY_SIMDA BK22031D.TOMC
            PWS:WN32/SIMDA
                                                             SUSPICIOUS.CLOUD.5
                             TROJAN-DOWNLOADERWIN32.DALEXIS
            GENERIC S.EHV
                             A VARIANT OF WN32/KRYPTIK.AFEW
                                                               MAL/ENCPK-OI
                             TROJANDOWNLOADER:WIN32/OGIMANT.A
       RDN/GENERICDX!DKG
                                                                 TROI DALEXIS.SMH
                              HEUR:TROJAN .WIN32 .GENERIC
        TROJAN:WIN32/ROPEST.J
                                                               TROI GEN.R047C0DCI15
                             VIRUS.WIN32.NABUCUR.C(V)
           GENERIC S.EHX
                                                                   TROI_KRYPTK.SML3
                           UDS:DANGEROUSOBJECT.MULTI.GENERIC
        TROIAN.GEN.2
Null :→
       W32/VIRRNSM-E
                                                                 TROI/SIMDA-BA
                                  TROIAN .WIN32 .GENERIC!BT
                                                                     GENERIC32NDT
          VIRUS:WIN32/NABUCURD
                                                               GENERIC_S.EHZ
                             TROJANDOWNLOADER:WIN32/DALEXIS.A
       WORM:WIN32/DORKBOT.
                              TROJANDOWNLOADER:WIN32/ZEMOT.A
                                                               GENERIC_S.EIA
            WIN32/AGENT.UNQ
                               VIRTOOL:WIN32/CEEINJECTGEN!KK
                                                              CRYPT4.BBZ
        TROJAN.WIN32.SHIPUPBOE
                                                             RANSOM:WIN32/CROWTI.A
                                 TROJAN.WIN32.ENCPKZEA (V)
                                                              RDN/GENERICDX!DKN
              MAL/WONTON-AN
                                TROJANDROPPERWIN32/GEPYSA
                                                              SHEUR4.ACIK
                                    TROJAN.WIN32.CRIDEXC (V)
        VIRUS.WIN32.POLYRANSOMF
                             TROIANDROPPERWIN32/ROVNIX P
             TROI/AGENT_AIRO
                                                            W33 DVNCUMI UCKVUINES
```

AV Names for Cluster 1

Picking the biggest cluster what does AV think of these samples





```
A VARIANT OF MSIL/AUTORUN.SPY.AGENT.BT
                                                                 GENERIC SEHU
         W32/VIRRANSOMB
          TROJ/AGENT-AIRO
                                                                 PAK GENERIC002
          TROJAN.WIN32.SHIPUPBOG
                                                           VARIANT OF WIN32/DRIDEXH
                              A VARIANT OF WIN32/INJECTOR.BRDO
                                                             SHEUR4CEOO
              GENERIC36ATUZ
                              TSPY SIMDA BK22031D.TOMC
          GENERIC MALWARE, MS
                                                             GENERIC_S.EHZ
                           A VARIANT OF WIN32/KRYPTIK .AFEW
       INFOSTEALER.SHIZ
                                                                TROI KRYPTK.SML3
                            TROIANDOWNLOADER:WIN32/OGIMANT.A
          GENERIC S.EHV
                                                                 PWS:WIN32/SIMDA
                             HEUR:TROJAN .WIN32 .GENERIC
        TROJAN:WIN32/ROPEST.J
                                                              TROJ_GEN.R047C0DCJ15
                             VIRUS.WIN32.NABUCUR.C(V)
       TROI GEN.R047C0RCC15
                                                                   CRYPT3.BWKU
                           UDS:DANGEROUSOBJECT.MULTI.GENERIC
                                                                  MAL/ENCPK-OJ
             LUHE.FIHA.A
Null :→
        SHEUR4.ACIK
                                                                ROI/SIMDA-BA
                                TROIAN.WIN32.GENERIC!B7
                                                                     GENERIC32NDT
          VIRUS:WIN32/NABUCURD
                             TROJANDOWNLOADER:WIN32/DALEXIS.A
                                                               GENERIC S.EHX
           WIN32/AGENT.UNO
                                                              RANSOM:WIN32/CROWTI
                               VIRTOOL:WIN32/CEEINJECTGEN!KK
        TROJAN.WIN32.SHIPUPBOE
                                                            RANSOM:WIN32/CROWTI.A
                                 TROIAN.WIN32.ENCPKZEA (V)
                                                              PACKEDGENERIC459
          SUSPICIOUSCLOUD5
                                TROJANDROPPERWIN32/GEPYSA
                                                              MAL/WONTON - AN
       A VARIANT OF WIN32/WIGON.OV
                                  TROJAN.WN32.CRIDEX.C(V)
             GENERIC S.EIA
                             TROJANDROPPERWIN32/ROVNIX.P
```

AV Names for Cluster 80

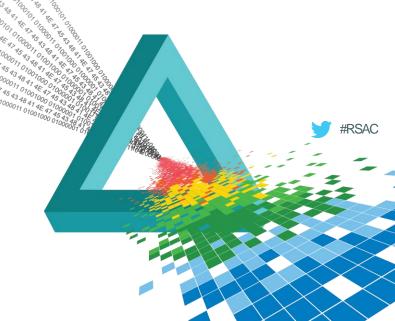
These samples show a strong trend to being ransomware. However some are called CRIDEX or Infostealer



RS∧°Conference2015

Singapore | 22-24 July | Marina Bay Sands

Higher Dimensional Visualization





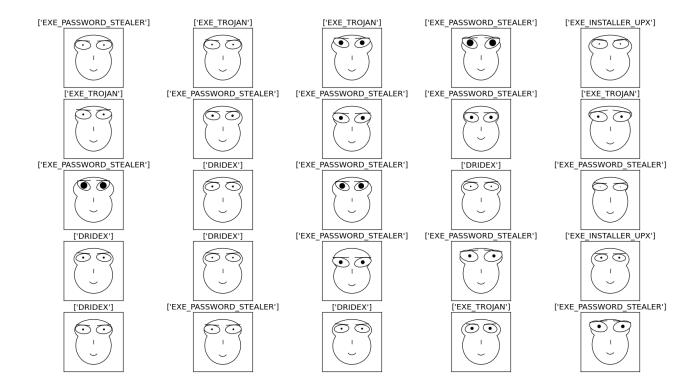
Chernoff Face Visualization

- Human faces provide a variety of characteristics that can be easily interpreted.
- The shape, size, placement and orientation of face features such as eyes, ears, mouth, etc. represent values of variables
- A visualization technique, known as Chernoff Faces, leverages this data to generate high dimensional visualizations of data
- First, we take a look at the similarity of various types of malware samples using 17 dimensions





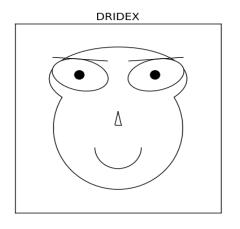
Chernoff Face Visualization: General Malware

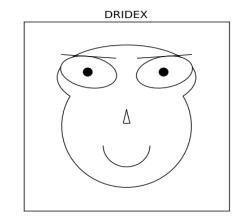


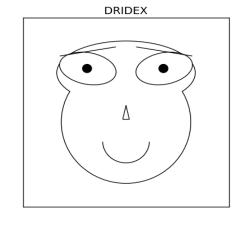


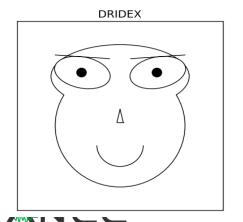


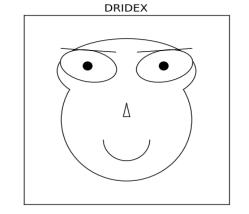
Chernoff Face Visualization: DRIDEX

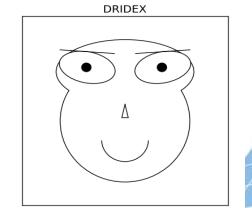














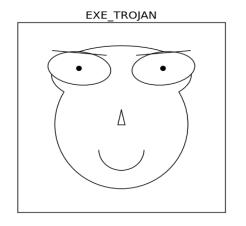
Chernoff Face Visualization: DRIDEX

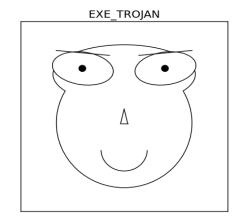
- At a higher dimension, many of the DRIDEX clusters look similar.
- This is a strong indication that many of these samples are in fact related, and the DRIDEX name is applied to samples that are similar

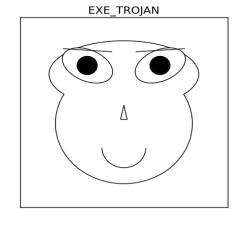


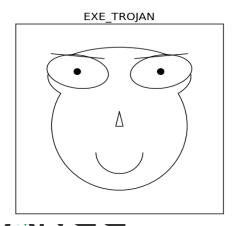


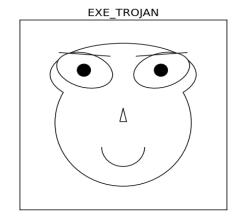
Chernoff Face Visualization: TROJAN

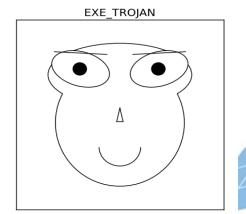














Chernoff Face Visualization: TROJAN

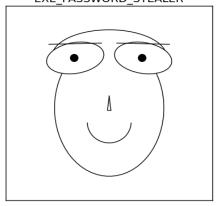
- With the more generic name Trojan, we are starting to see some variety in the dataset
- With these clusters, significant differences in the eyes indiciate that the variance between trojan clusters is more noticable than with DRIDEX



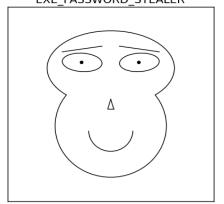


Chernoff Face Visualization: STEALER

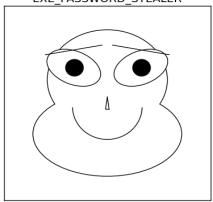




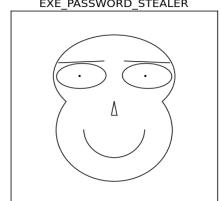
EXE_PASSWORD_STEALER



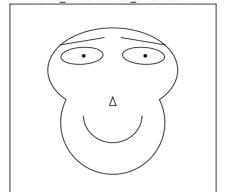
EXE PASSWORD STEALER



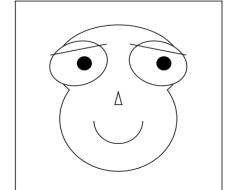
EXE_PASSWORD_STEALER



EXE PASSWORD STEALER



EXE PASSWORD STEALER





ence2015



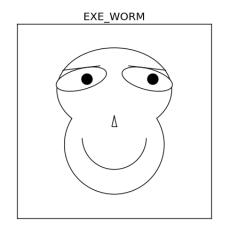
Chernoff Face Visualization: STEALER

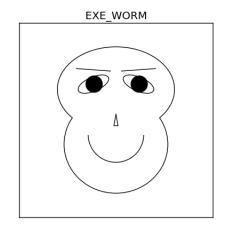
- Another "generic" name, STEALER
- Here we see a large variance in the different clusters. It would be reasonable to state at this point that many of these clusters are in fact not related at all in terms of their features.
- While the malware in this group may steal information, the methods used and functionality contained in the samples varies greatly.

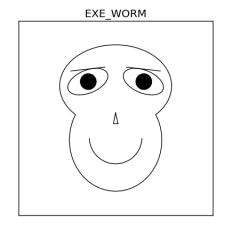


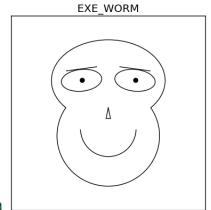


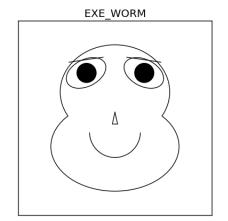
Chernoff Face Visualization: WORM

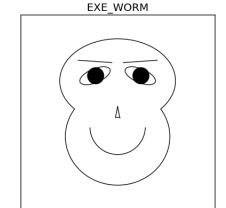
















Chernoff Face Visualization: WORM

 Similar to what we saw with STEALER, large variance in what is considered WORM as well

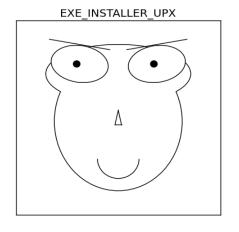


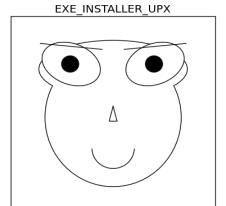


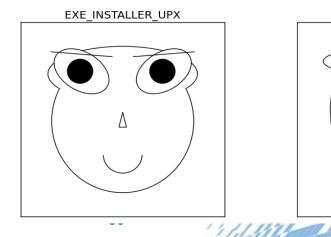
Chernoff Face Visualization: INSTALLER

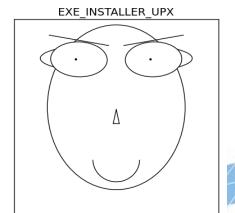
















Chernoff Face Visualization: INSTALLER

- Collection of UPX packed malware samples that work as droppers
- While there is still a good amount of variance in the clusters, not as significant as we saw with STEALER and WORK
- The reduction in variance compared to previous groups can be attributed to the use of packing





Analysis

- Higher dimensional visualization helped to identify similarities and variance in different malware naming groups
- Sample like Dridex, and probably other well known malware campaigns, often have high similarity between samples
- Other malware names, such as worm or password stealer often have weak similarity between samples





Apply Slide

- Higher dimensional visualization can be an effective tool to when looking for similarities between or in groups
- Is some cases, malware names may have minimal correlation between samples of the same name
- A single sample of malware can have a wide range of different names, making it difficult to identify what the sample is





Additional info

- Questions?
- Contact us: datascience@cylance.com
- Special thanks to Xuan Zhou for her contributions

