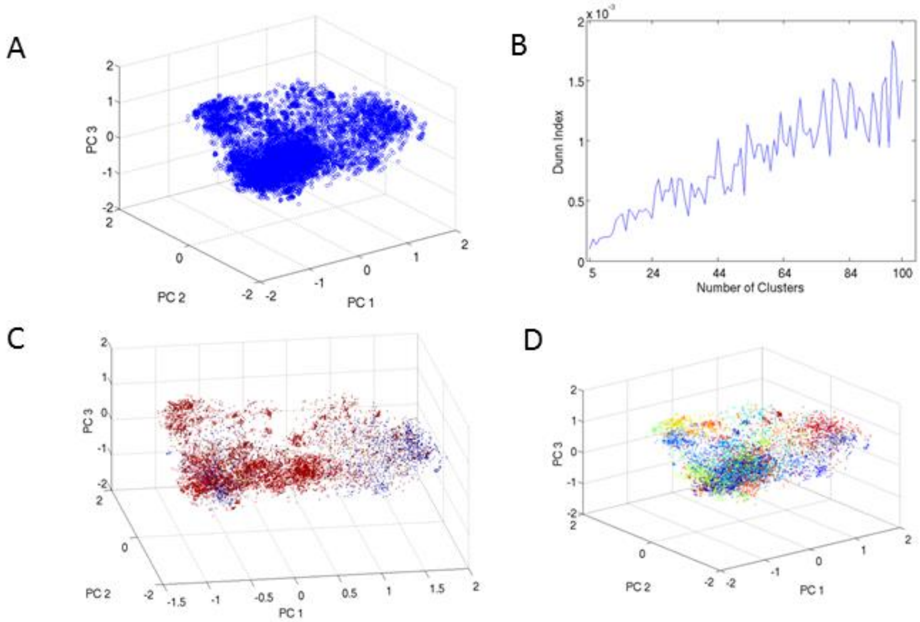

Figure 1. Antivirus label clustering. **A.** 3-D scatter plot of the AV label Malicia dataset projected into the space of the first three principal components. **B.** Dunn index as a function of the number of clusters in a k-means clustering of the Malicia dataset in the AV label space. **C.** 3-D scatter plot of Malicia dataset in principal components space with colorization based on Malicia family labels. **D.** 3-D scatter plot in PC space with colorization based on 54 cluster k-means results.

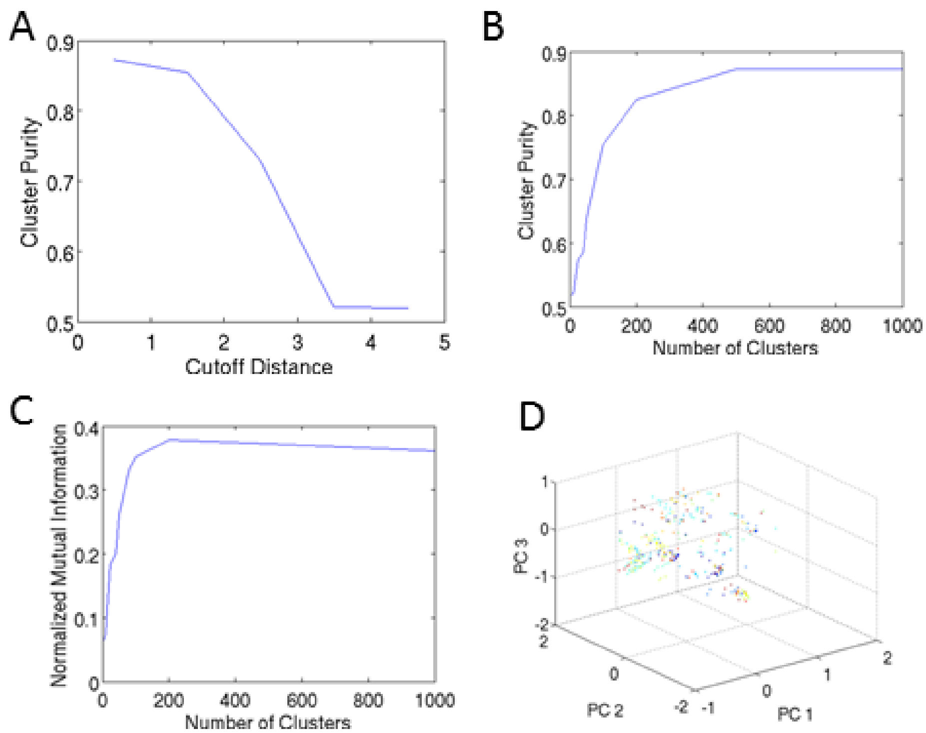

Figure 2. Import feature clustering. **A.** Cluster purity as a function of linkage tree cutoff distance. **B.** Cluster purity as a function of number of clusters. **C.** Normalized mutual information as a function of number of clusters. **D.** Projection of data into space of the first three principal components of the import data with colorization based on 200 cluster results.

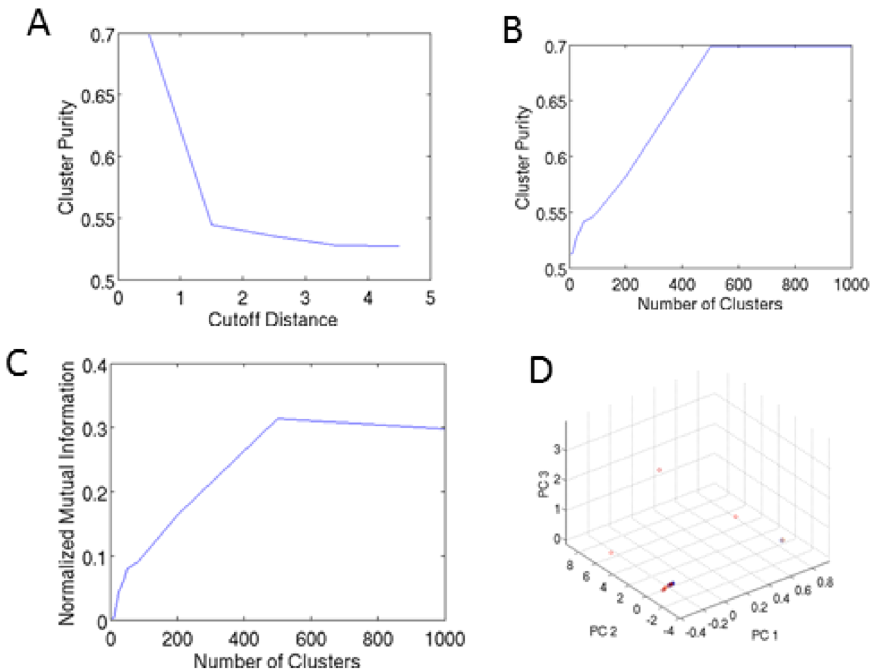

Figure 3. Export feature clustering. **A.** Purity as a function of linkage tree cutoff distance. **B.** Purity as a function of number of clusters. **C.** Normalized mutual information as a function of number of clusters. **D.** Data projected into the space of the first three principal components with colorization based on results of separating data into 200 clusters.

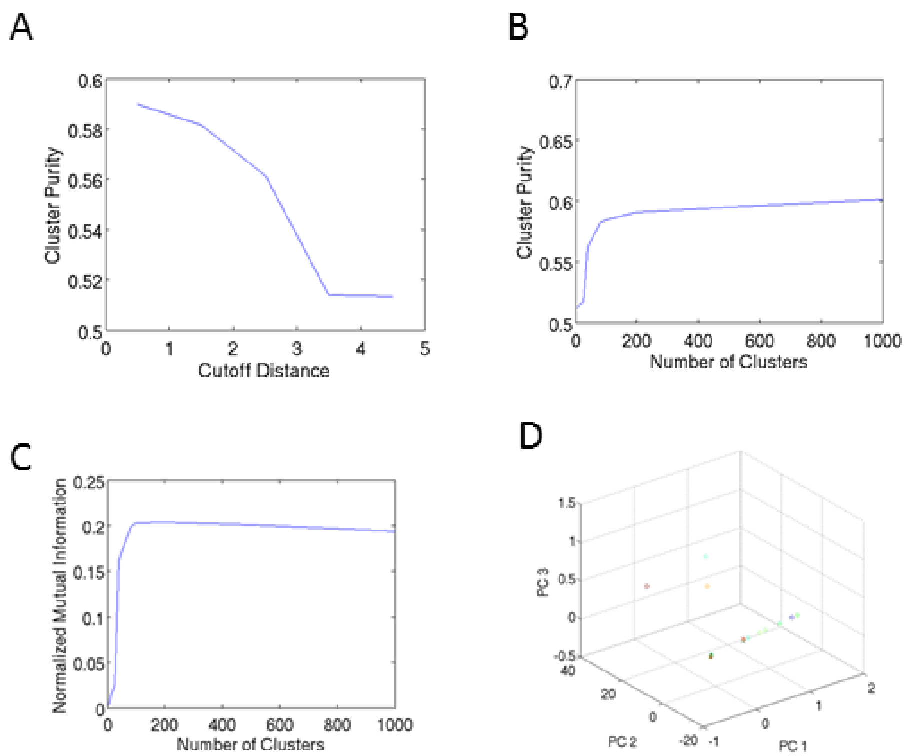

Figure 4. DNS request feature clustering. **A.** Purity as a function of linkage tree cutoff distance. **B.** Purity as a function of number of clusters. **C.** Normalized mutual information as a function of number of clusters. **D.** Data projected into the space of the first three principal components with colorization based on results of separating data into 200 clusters.

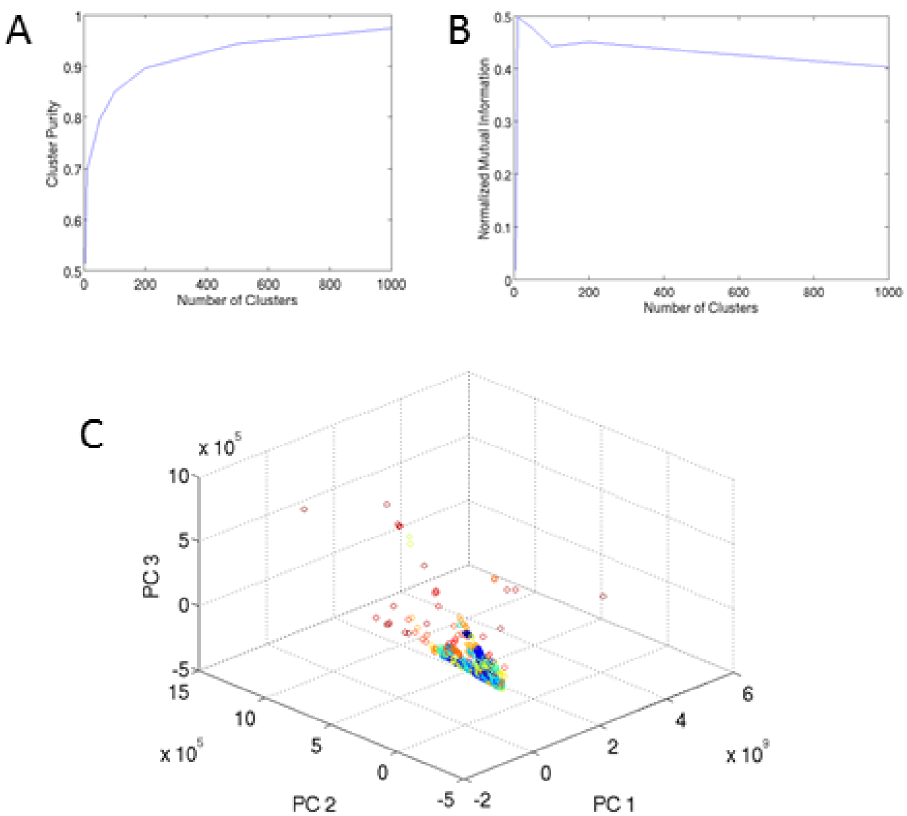

Figure 5. Size information clustering. **A.** Purity as a function of number of clusters. **B.** Normalized mutual information as a function of number of clusters. **C.** Data projected into the space of the first three principal components of the size data. Colorization is related to the results of a 200 cluster analysis.

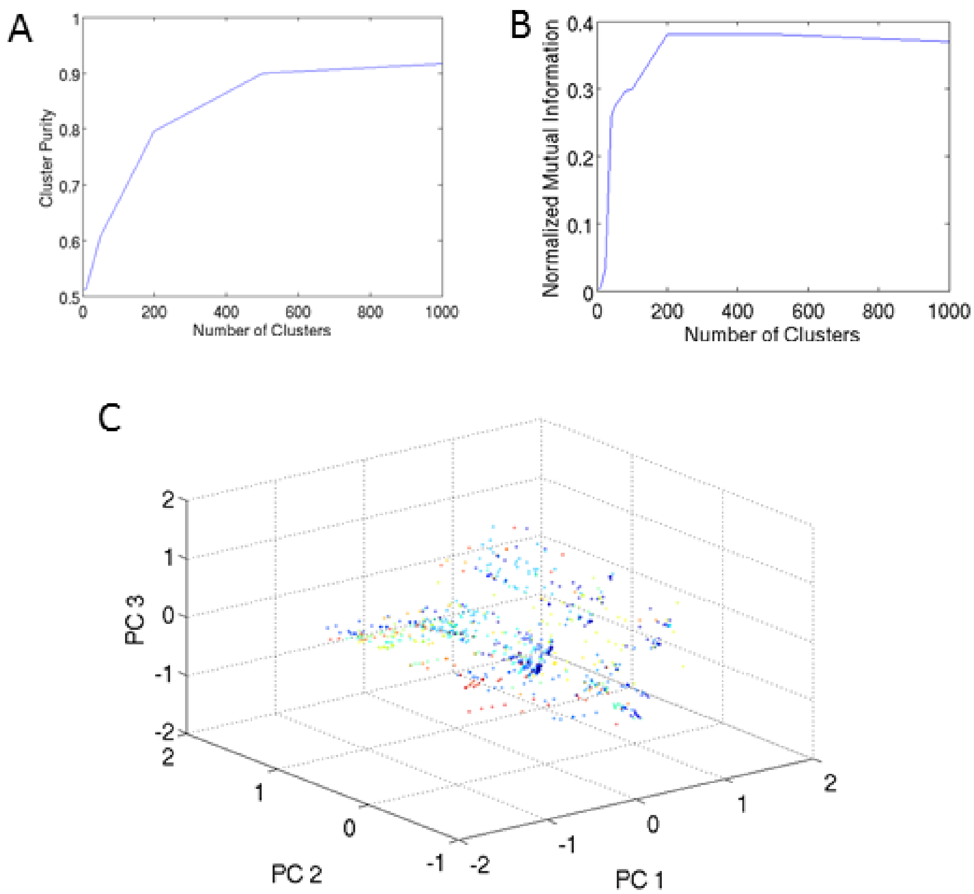

Figure 6. Import, export, and DNS request clustering. **A.** Purity as a function of number of clusters. **B.** Normalized mutual information as a function of number of clusters. **C.** Data projected into the principal components space and colorized based on a 200 cluster breakdown.

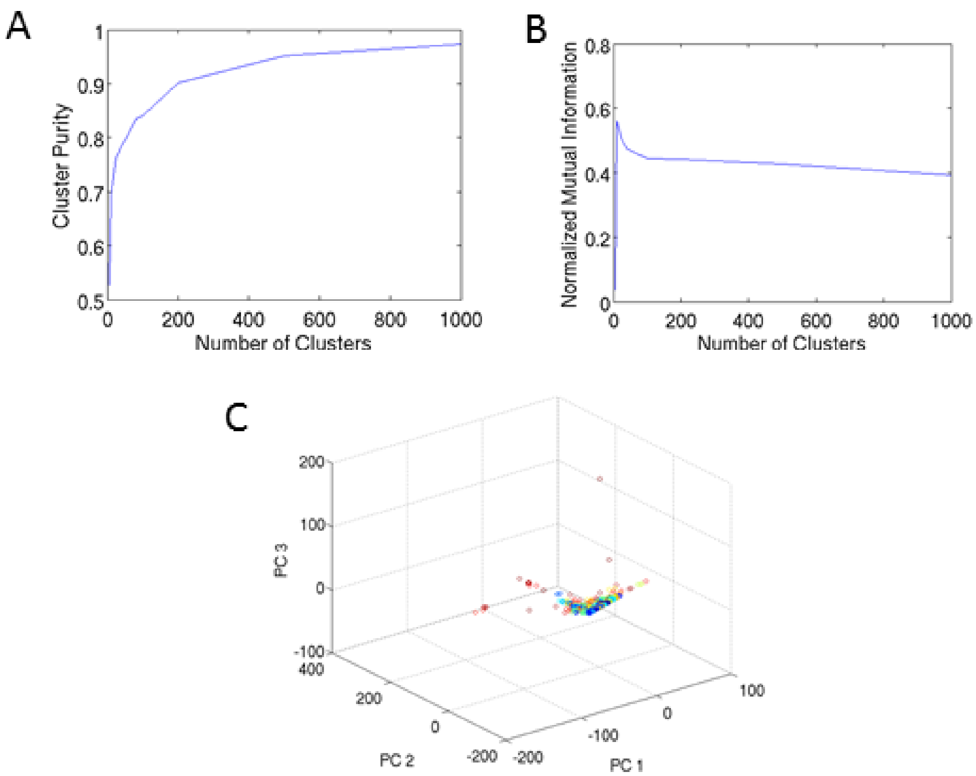

Figure 7. Import, export, DNS request, and size information clustering. **A.** Purity as a function of number of clusters. **B.** Normalized mutual information as a function of number of clusters. **C.** Data projected into the principal components space and then colorized based on the results of a 200 cluster breakdown.

Family	Precision	Recall	Precision – Null Excluded	Recall – Null Excluded
Winwebsec	96%	89%	97%	90%
Zbot	61%	95%	73%	97%
ZeroAccess	47%	96%	56%	97%
Security Shield	11%	100%	No convergence	No convergence

Table 1. Import Support Vector Machine Analysis.

Family	Precision	Recall	Precision – No Null	Recall – No Null
Winwebsec	61%	100%	69%	81%
Zbot	No convergence	No convergence	26%	100%
ZeroAccess	98%	95%	100%	96%
SecurityShield	2%	91%	No convergence	No convergence

Table 2. Export Support Vector Machine Analysis.

Family	Precision	Recall	Precision – No Null	Recall – No Null
Winwebsec	57%	100%	64%	100%
Zbot	24%	100%	25%	100%
Zeroaccess	73%	40%	100%	39%
SecurityShield	2%	100%	No convergence	No convergence

Table 3. DNS Request Support Vector Machine Analysis.

Family	Precision	Recall	Precision – No Null	Recall – No Null
Winwebsec	93%	99.7%	95%	99.5%
Zbot	42%	93%	56%	95%
Zeroaccess	26%	99.6%	35%	100%
SecurityShield	4%	80%	2%	69%

Table 4. Size Information Support Vector Machine Analysis.