亞洲大學|資電學院
Asia University| College of Information and Electrical Engineering
資訊工程學系
Department of Computer Science and Information Engineering




<勒索病毒預測>

指導老師：林俊淵老師　　**組員：賴星翰、蔡嘉妮**

摘要：
此專題使用機器學習的方式針對勒索病毒加密前api調用行為特徵禁行分析，目的是希望能找出勒索軟體家族間有沒有特定的加密行為模式，期望找到的重要行為模式能實際應用，提前偵測到電腦中的加密病毒，降低勒索風險，引用的數據集是倫敦帝國學院（RISS）研究小組整理的勒索軟件特徵。該資料集包含了十幾個勒索軟件家族的應用程式介面（API）數據，並且還包含了多種良性軟件數據。且家族有11個類別為此我們需要透過對應策略來提高準確率。確保找到的特徵行為是加密前的行為特徵，對於預防病毒才有實質幫助。

實作方法：
針對582比勒索事件分析其家族的特徵模式，因家族類別有11個，多類別且極度不平衡的情況下會導致準確率非常低，對應策略採用one-vs-rast 加上balance對類別進行處裡，機器學習模型採用決策數，因決策樹最方便能看出特徵重要性，且準確率非常高，查看不同家族間的top20特徵彼此間有相似的情形，最終查看排除特徵發現加密特徵是一連串的行為，流程是:找到系統目標路徑、產生虛擬空間、將勒索病毒檔案從資料庫抓取到虛擬空間、最後將病毒映射到目標路徑中，這些行為是我們研究找到的加密前重要特徵，期望畢業前以此模式維依據開發防毒軟體。

家族	決策數	knn	svm	logision	random	gb	Voting
2	0.8803419	0.8803419	0.8803419	0.8461538	0.9145299	0.6324786	0.9145299
2pca	0.8803419	0.8974359	0.8547009	0.7948718	0.9059829	0.6068376	0.8888889
6	0.8461538	0.8547009	0.8547009	0.8376068	0.8888889	0.8547009	0.8717949
6pca	0.8119658	0.8632479	0.7692308	0.7692308	0.8888889	0.6837607	0.8547009
9	0.9316239	0.9316239	0.9230769	0.9230769	0.9401709	0.6068376	0.9316239
9pca	0.9316239	0.9316239	0.9487179	0.9059829	0.9487179	0.8034188	0.9401709
5	0.974359	0.9487179	0.9230769	0.9487179	0.965812	0.7777778	0.965812
5pca	0.9316239	0.957265	0.9230769	0.9316239	0.9487179	0.8461538	0.965812
7	0.9230769	0.9145299	0.8888889	0.8888889	0.9401709	0.7350427	0.923077
7pca	0.8888889	0.9230769	0.8803419	0.7948718	0.9316239	0.6581197	0.9230769

	1896event	5287event	3668event	760event	2545event
pca1	0.104361	0.104401	0.104817	0.104933	0.105648
pca2	0.120896	0.122992	0.124899	0.128004	0.132446
pca3	0.079735	0.080355	0.085136	0.090844	0.109731

	DecisionTreeClassifier	KNeighborsClassifier	svm	LogisticRegression	RandomForestClassifier	MultinomialNB
original	0.9672131147540983	0.940983606557377	0.9442622950819672	0.9475409836065574	0.9573770491803278	85.57377049180329
PCA	0.9311475409836065	0.9377049180327869	0.9475409836065574	0.9508196721311475	0.9540983606557377	85.57377049180329
PCA-balance	0.9475409836065574	0.9377049180327869	0.921311475409836	0.9475409836065574	0.9540983606557377	85.57377049180329