ATHENA Research & Innovation Information Technologies

Πλατφόρμα ανάλυσης και κατηγοριοποίησης συμπεριφοράς κακόβουλου λογισμικού με τη χρήση μηχανικής μάθησης

Κυριάκος Στεφανίδης, Βασίλης Πίκουλης

Ερευνητής Β', Συνεργαζόμενος Ερευνητής

Ινστιτούτο Βιομηχανικών Συστημάτων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝΔΥΣΕΩΝ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΔΙΑΡΘΡΩΤΙΚΩΝ ΠΡΟΓΡΑΜΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΑνΕΚ

ΕΠΑνΕΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εισαγωγή

Η απειλή των κυβερνοεπιθέσεων και του κακόβουλο λογισμικό

- Ως το 2023, περί τα **5.2 τρις δολάρια**[1] περιουσιακών στοιχείων θα βρίσκονται υπό απειλή
- **28%** του συνόλου των κυβερνοεπιθέσεων αφορά τη χρήση κακόβουλου λογισμικού

Η ανίχνευση κακόβουλου λογισμικού παραμένει ανοιχτό και δύσκολο πρόβλημα

- Απειλή που συνεχώς εξελίσσεται
- Η έλευση της Βαθιάς Μηχανικής Μάθησης έφερε πληθώρα νέων τεχνικών
- Έλλειψη πρότυπων συνόλων εκπαίδευσης και ελέγχου των μεθόδων

[1]Kelly Bissel, et. al., *Ninth Annual Cost of Cybercrime Study*. The Cost of Cybercrime, Ponemon Institue LLC, Accenture plc, 2019

Αποτελέσματα Επιγραμματικά

Δυο συστήματα ανίχνευσης κακόβουλου λογισμικού δημιουργήθηκαν στο πλαίσιο του έργου

- Συστήματα που βασίζονται στη **στατική** ανάλυση του λογισμικού
- Το πρώτο σύστημα χρησιμοποιεί το σύνολο χαρακτηριστικών EMBER και **random forests**
- Το δεύτερο βασίζεται στη χρήση **graph kernels** και Support Vector Machines

Τα συστήματα ενσωματώθηκαν στην πλατφόρμα ανάλυσης κακόβουλου λογισμικού Sisyfos

- Υποστηρίζεται χειροκίνητη και αυτόματη ανάλυση όπως και μαζική ανάλυση αρχείων
  - Μέσω web interface
  - Μέσω REST API
- Συμπεριλαμβάνει διάφορα ενσωματωμένα modules στατικής και δυναμικής ανάλυσης

Προσπάθεια διασύνδεσης της πλατφόρμας με την πλατφόρμα κυβερνοασφάλειας αυτόνομων οχημάτων του ευρωπαϊκού έργου nIoVe

https://www.niove.eu/

Ανίχνευση κακόβουλου λογισμικού με χρήση χαρακτηριστικών του EMBER και random forests

Το σύνολο δεδομένων EMBER

- **600.000** δείγματα για εκπαίδευση, **200.000** για επαλήθευση/έλεγχο
- Δείγματα από καλόβουλο & κακόβουλο λογισμικό
- **2351** χαρακτηριστικά για κάθε δείγμα (strings, κεφαλίδα, εντροπία, κ.α.)
- Εργαλείο ανοικτού κώδικα για την εξαγωγή των χαρακτηριστικών από νέα δεδομένα

Ταξινόμηση με χρήση random forests

- 1.000 δυαδικά δέντρα αποφάσεων
- 48 χαρακτηριστικά ανά δέντρο
- Η απόφαση λαμβάνεται από την πλειοψηφία των επιμέρους ταξινομήσεων

Υλοποίηση σε Python

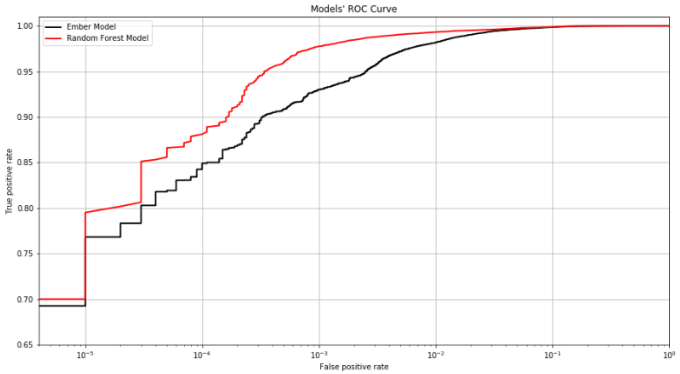
- Χρήση των βιβλιοθηκών scikit-learn και numpy

Ανίχνευση κακόβουλου λογισμικού με χρήση χαρακτηριστικών του EMBER και random forests

Ακρίβεια μοντέλου: **99.2%**

Βελτιώνει το μοντέλο αναφοράς του EMBER

Μεγαλύτερη πιθανότητα ανίχνευσης για το ίδιο ποσοστό εσφαλμένου συναγερμού



Πιθανότητα εσφαλμένου συναγερμού	Πιθανότητα ορθής ανίχνευσης	
	EMBER	**Random Forests**
1%	98.5 %	**99.0 %**
0.1 %	96.4 %	**98.8 %**

ATHENA Research & Innovation Information Technologies

Ανίχνευση κακόβουλου λογισμικού με χρήση χαρακτηριστικών του EMBER και random forests

Τα σύνολα εκπαίδευσης και ελέγχου του EMBER:

- Έχουν προέλθει από την ίδια διαδικασία συλλογής και από τον ίδιο οργανισμό
- Συλλέχθηκαν με απόσταση μόλις δύο μηνών μεταξύ τους

Έλεγχος απόδοσης σε **νέο σετ** δεδομένων

- **89.000** κακόβουλα εκτελέσιμα αρχεία από το αποθετήριο VirusShare
- Δεδομένα που έχουν συλλεχθεί με διαφορετικό τρόπο χρόνο από το EMBER
- Από το κάθε δείγμα εξάγεται το ίδιο σετ των 2351 χαρακτηριστικών που χρησιμοποιήθηκαν για την εκπαίδευση

Συγκριτικά αποτελέσματα:

Δεδομένα ελέγχου	Μοντέλο ανίχνευσης	
	EMBER	**Random Forests**
EMBER	98.9 %	**99.0 %**
VirusShare	84.2 %	**87.7 %**

Ανίχνευση κακόβουλου λογισμικού με χρήση χαρακτηριστικών του EMBER και random forests

Μελέτη ευαισθησίας ως προς τα χρησιμοποιούμενα χαρακτηριστικά:

- Χρήση **μικρού** υποσυνόλου **σημαντικών** χαρακτηριστικών
- Εξαγωγή σημαντικότερων χαρακτηριστικών με βάση το **κέρδος πληροφορίας**

Σημαντικότερα χαρακτηριστικά

- Εικονικό μέγεθος τμημάτων στην μνήμη
- Εισαγόμενες & εξαγόμενες συναρτήσεις
- Εντροπία τμημάτων

Αποτελέσματα μοντέλου με χρήση των **10 σημαντικότερων** χαρακτηριστικών:

Αριθμός χαρακτηριστικών	Ακρίβεια	Χρόνος εκπαίδευσης (λεπτά)
2351	99.2 %	2195
10	**96.1 %**	**23**

Ανίχνευση κακόβουλου λογισμικού με χρήση χαρακτηριστικών του EMBER και random forests

Επίδραση του μεγέθους του συνόλου εκπαίδευσης:

- Αναμενόμενα μη ικανοποιητική απόδοση για πολύ μικρά μεγέθη συνόλων (≤ 1000)
- Απώλεια ≤ 1% της απόδοση σε μεσαίου μεγέθους σύνολα (≈ 50.000),
- Η αύξηση στην απόδοση γίνεται ολοένα και «ακριβότερη» για ≥ 100.000

Πλήθος δειγμάτων εκπαίδευσης	Ακρίβεια	Χρόνος εκπαίδευσης (λεπτά)
1.000	92.6 %	0.2
5.000	97.0 %	2.1
10.000	97.6 %	8.8
30.000	98.2 %	21.6
100.000	98.6 %	132
600.000	**99.2 %**	**2195**

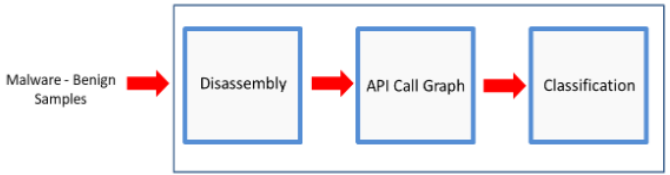
Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Το σύστημα συνδυάζει τη στατική ανάλυση με τη μοντελοποίηση γράφων

Ταξινόμηση με χρήση Support Vector Machines

Βασικά βήματα αλγορίθμου:

1. Αποσυναρμολόγηση (**disassembly**) του δείγματος
2. Γράφος Ελέγχου Ροής ➔ Γράφος Κλήσεων API ➔ *Αφηρημένος Γράφος Κλήσεων*
3. Χρήση **πυρήνα** βασισμένου στα random walks για συγκρίσεις μεταξύ γράφων.
4. Ταξινόμηση με χρήση γραμμικού **SVM**.



ATHENA Research & Innovation Information Technologies

Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Σύνολα δεδομένων:

- 997 δείγματα **κακόβουλων** εκτελέσιμων από τα VirusShare, VirusTotal
- 557 **καλόβουλα** εκτελέσιμα από έμπιστες πηγές (Windows, Git, Cygwin, κ.α.)

Η πλατφόρμα *Ghidra* της NSA

- Εργαλείο ανοικτού κώδικα για την ανάλυση λογισμικού (*disassembly*)
- Εύρεση δομικών μπλοκ κώδικα, και των κλήσεων API που αυτά πραγματοποιούν

Αφηρημένος Γράφος Κλήσεων Συστήματος

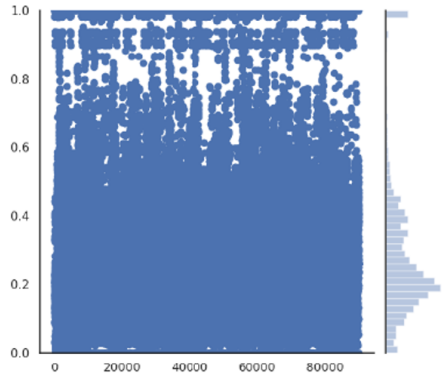
1. Γράφος Ελέγχου Ροής (διασύνδεση μεταξύ μπλοκ κώδικα)
2. Γράφος Κλήσεων API (αντικατάσταση κόμβων/μπλοκ με κλήσεις API)
3. Αφηρημένος Γράφος (συγχώνευση κόμβων, εξάλειψη σειράς εκτέλεσης)

Random Walk Graph Kernel

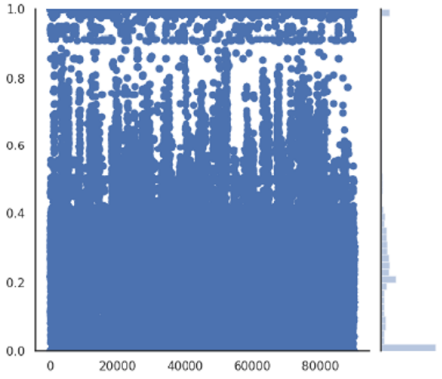
- Εύρεση «εσωτερικού γινομένου» μεταξύ γράφων (ποσοτικοποίηση **ομοιότητας** γράφων)

Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

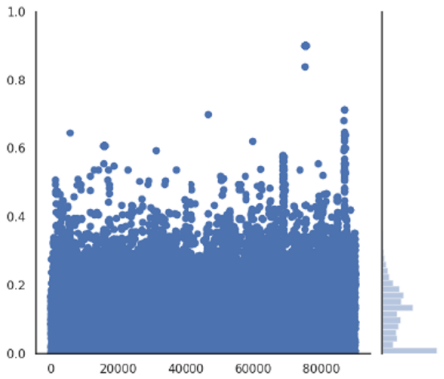
Τιμές πυρήνα (ομοιότητας) για ζεύγη δειγμάτων του συνόλου εκπαίδευσης



καλόβουλο/καλόβουλο



κακόβουλο/κακόβουλο



καλόβουλο/κακόβουλο

Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Εκπαίδευση SVM

- Χωρίζουμε το διαθέσιμο σύνολο σε δείγματα εκπαίδευσης/δείγματα ελέγχου
- Μέσω πειραματισμού καθορίστηκε ως βέλτιστη η αναλογία 90% - 10%
- 100 διαφορετικά ζεύγη συνόλων σύνολα με τυχαία επιλογή

Μετρικές

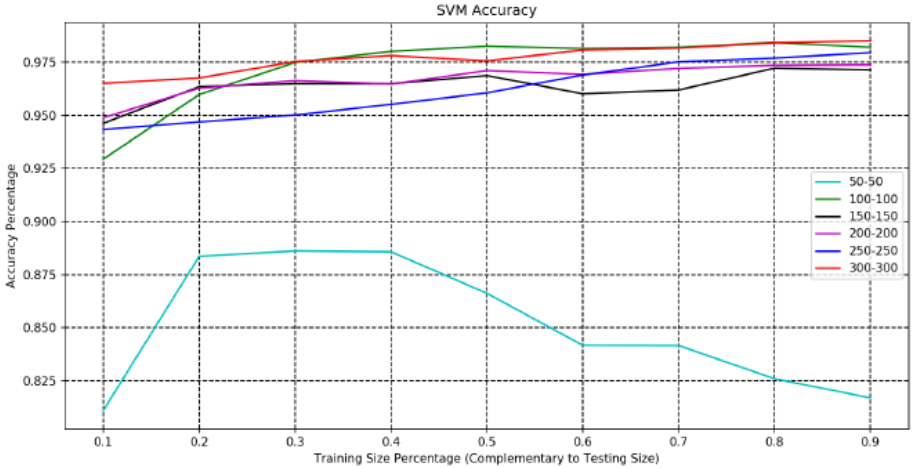
- Ακρίβεια (ποσοστό σωστών προβλέψεων)
- ROC AUC (εμβαδό κάτω από την καμπύλη ROC)
- Precision (ποσοστό σωστών ανιχνεύσεων)
- Recall (ποσοστό κακόβολου λογισμικού που ανιχνέυθηκε)

Αποτελέσματα ταξινόμησης

- Όλες οι χρησιμοποιούμενες μετρικές οδηγούν σε επιδόσεις άνω του **98%**,
- Πολύ ικανοποιητική απόδοση δεδομένου του σχετικά **μικρού συνόλου** εκπαίδευσης
- Καταδεικνύει την **αποτελεσματικότητα** των τεχνικών που χρησιμοποιήθηκαν

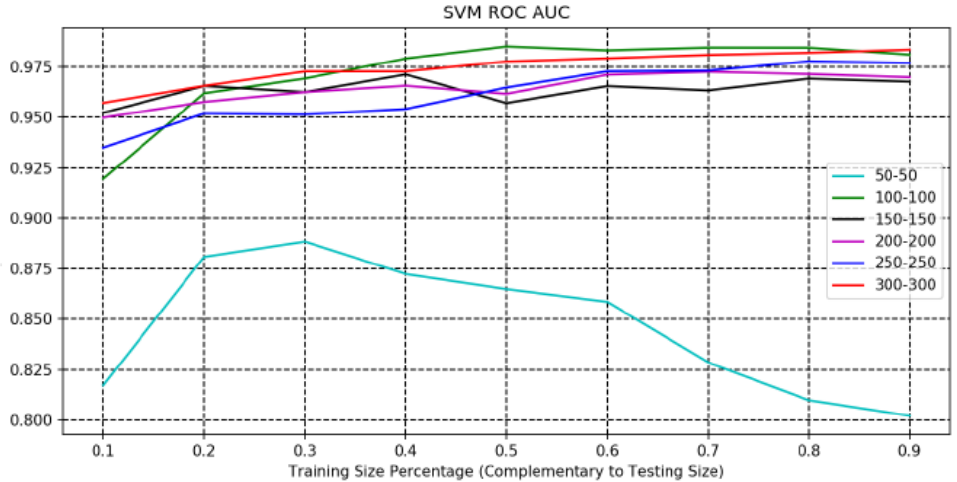
Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Ακρίβεια μοντέλου για διάφορα μεγέθη και αναλογίες συνόλων εκπαίδευσης



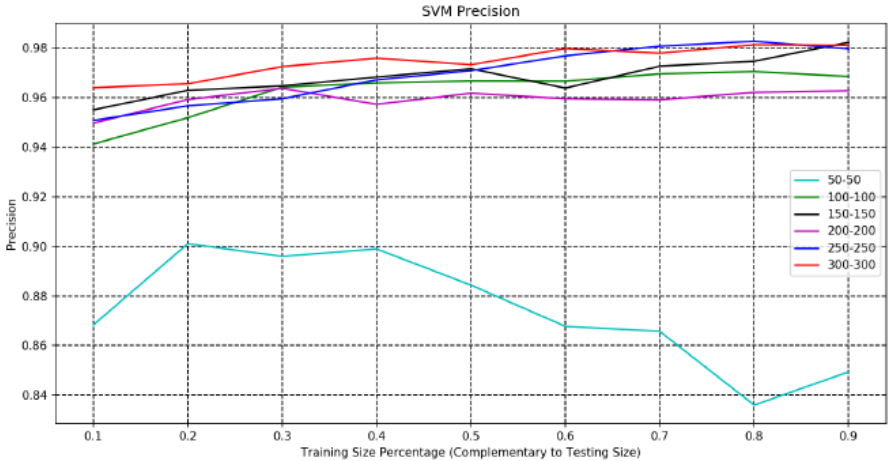
Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Μετρήσεις **ROC - AUC** για διάφορα μεγέθη και αναλογίες συνόλων εκπαίδευσης



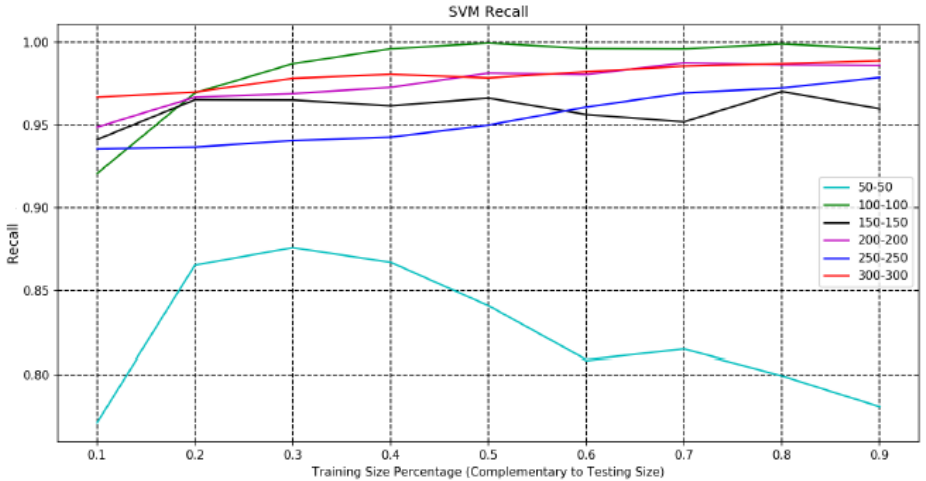
Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Μετρήσεις του **precision** για διάφορα μεγέθη και αναλογίες συνόλων εκπαίδευσης



Ανίχνευση κακόβουλου λογισμικού με βάση τον Γράφο Κλήσεων API

Μετρήσεις του **recall** για διάφορα μεγέθη και αναλογίες συνόλων εκπαίδευσης



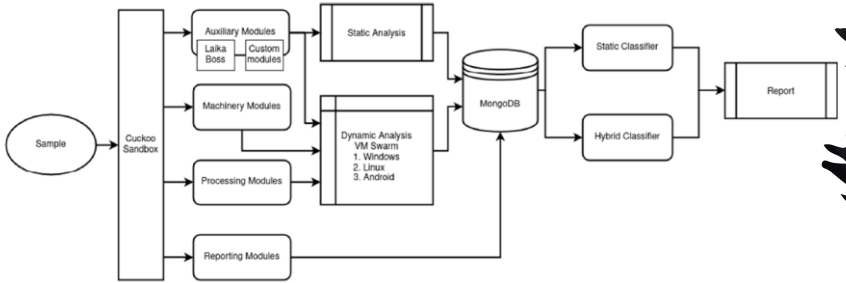
Ολοκλήρωση με την πλατφόρμα sisyfos

Παρέχει ένα ολοκληρωμένο περιβάλλον ανάλυσης κακόβουλου λογισμικού

- Βασισμένο στο cuckoo sandbox

Τρία βασικά υποσυστήματα

- Orchestrator
- Στατική ανάλυση
- Δυναμική ανάλυση
- Ταξινόμηση



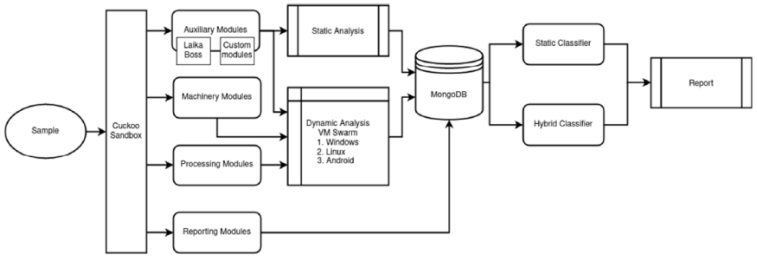
Orchestrator

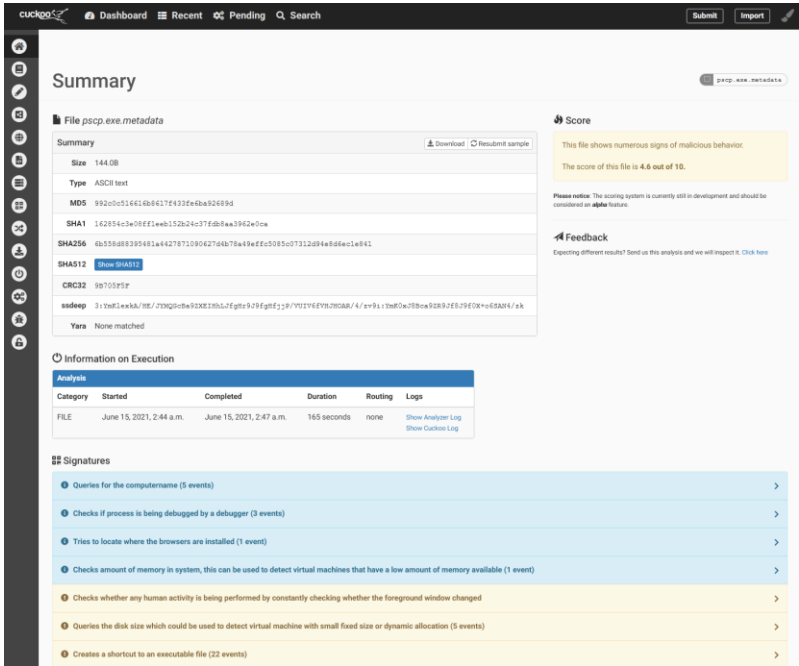
Python Scripts για την μαζική φόρτωση και ανάλυση δειγμάτων

sample -> cuckoo and modules -> database -> classifier

Web interface μέσω του cuckoo sandbox

Web service για την μαζική μεταφόρτωση





Στατική Ανάλυση

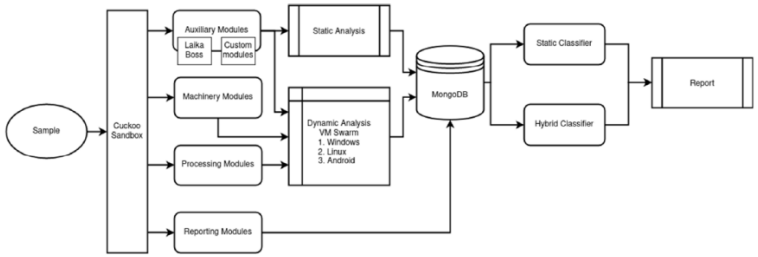
Εύκολη εξαγωγή χαρακτηριστικών

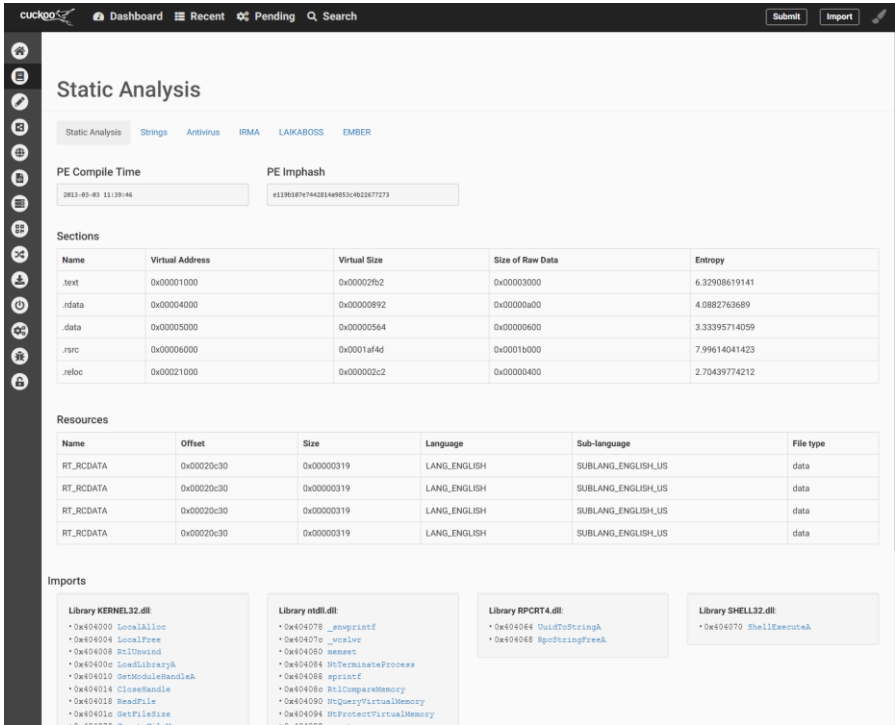
Βασικά εργαλεία στατικής ανάλυσης (cuckoo)

- file metadata
- resources
- basic binary info
- printable strings
- Virus scanning services

Άλλα εργαλεία

- entropy measurement
- Ember feature extraction
- Binary disassembly (υπο ολοκλήρωση)
- Laika Boss
  - object unpacking and scanning





Δυναμική Ανάλυση

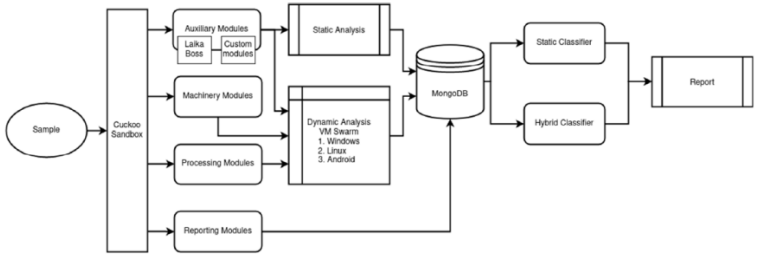
Sandboxes (windows VMs)

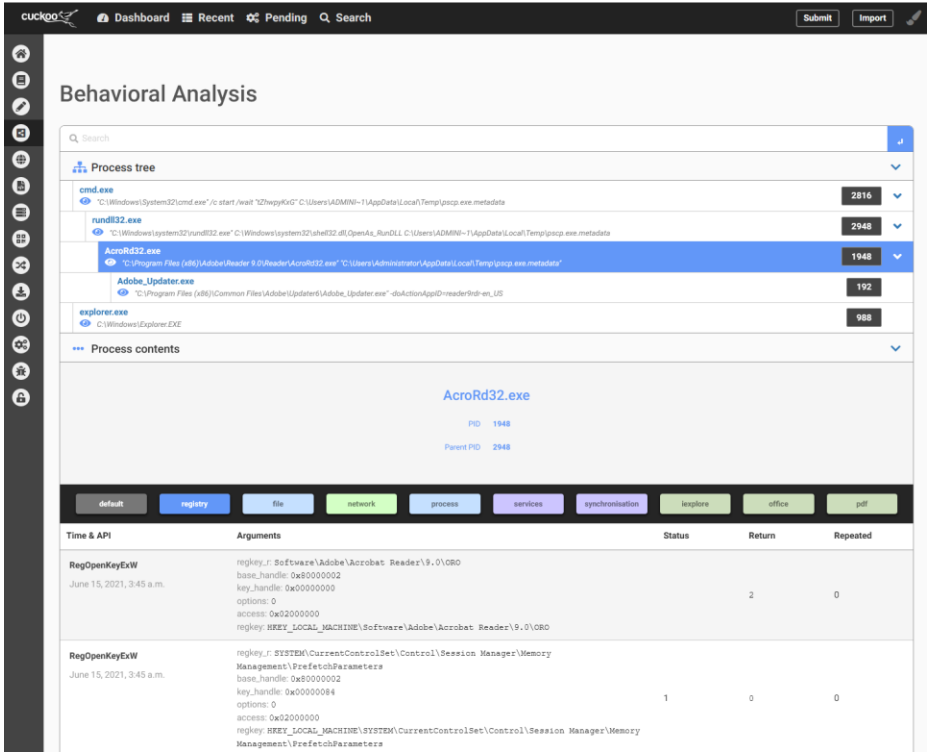
Cuckoo/VirtualBox για την διαχείρηση των VMs

- 6-~50 παράλληλα VMs δυνατότητα
- VMCloak για απόκρυψη του sandbox
  - Εγκατάσταση κοινών εφαρμογών

Χαρακτηριστικά

- file system
- registry
- network
- dynamic signatures
  - unpacking
  - code injection in child processes

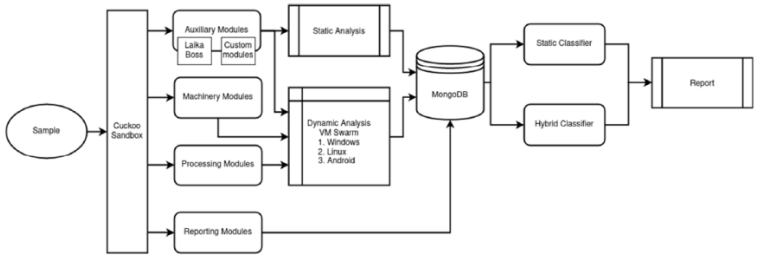


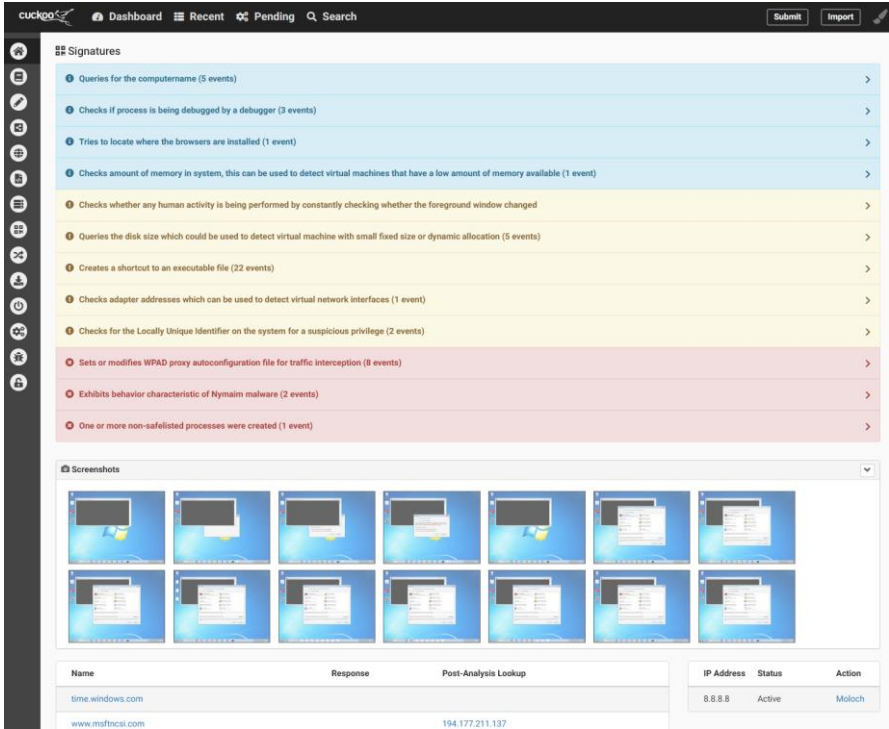


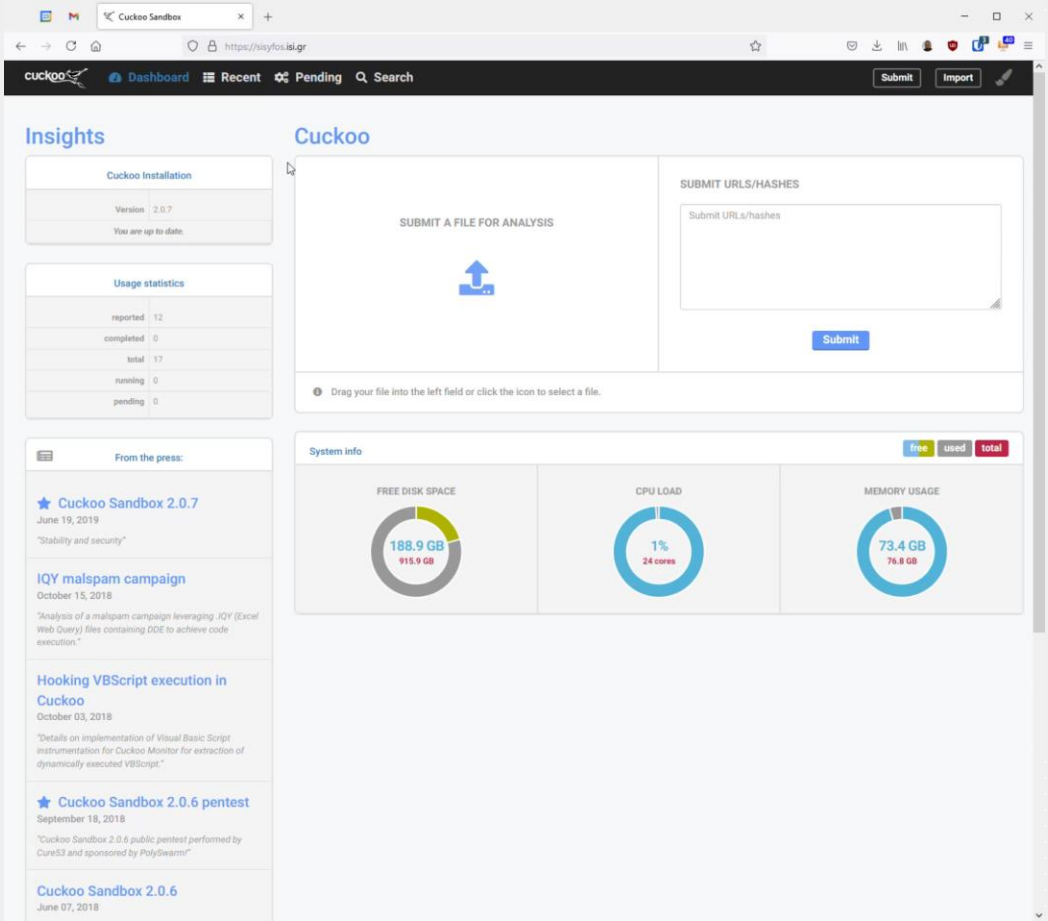
Ταξινόμηση

Τρείς ταξινομητές

- Ενσωματωμένος ταξινομητής του cuckoo (hybrid)
- random forests (static)
- API call graph (static)






Cuckoo Sandbox
https://sisyfos.isi.gr
cuckoo
Dashboard
Recent
Pending
Search
Submit
Import
Insights
Cuckoo Installation
Version 2.0.7
You are up to date.
Usage statistics
reported 12
completed 0
total 17
running 0
pending 0
From the press:
Cuckoo Sandbox 2.0.7
June 19, 2019
"Stability and security"
IQY malspam campaign
October 15, 2018
"Analysis of a malspam campaign leveraging .IQY (Excel Web Query) files containing DDE to achieve code execution."
Hooking VBScript execution in Cuckoo
October 03, 2018
"Details on implementation of Visual Basic Script instrumentation for Cuckoo Monitor for extraction of dynamically executed VBScript."
Cuckoo Sandbox 2.0.6 pentest
September 18, 2018
"Cuckoo Sandbox 2.0.6 public pentest performed by CureS3 and sponsored by PolySwarm!"
Cuckoo Sandbox 2.0.6
June 07, 2018
Cuckoo
SUBMIT A FILE FOR ANALYSIS
SUBMIT URLS/HASHES
Submit URLs/hashes
Submit
Drag your file into the left field or click the icon to select a file.
System info
free
used
total
FREE DISK SPACE
188.9 GB
915.9 GB
CPU LOAD
1%
24 cores
MEMORY USAGE
73.4 GB
76.8 GB

Ευχαριστούμε