Centro de
Investigación en
Computación

Repositorio de Malware

Introducción

Revisión de conceptos

Laboratorios de malware

Course

Análisis y Detección de Malware

Instructor

Acosta Bermejo Raúl

Lecture notes

2024-B
4 de septiembre del 2024

Instituto
Politécnico
Nacional

Table of contents (outline)

Tabla de contenido

1. Introducción
2. Casos de Estudio (5)
    1. VirusSign
    2. VirusShare
    3. VX-Heaven
    4. Kaggle
    5. theZoo
    6. Contagio
3. Conclusiones

Introduction

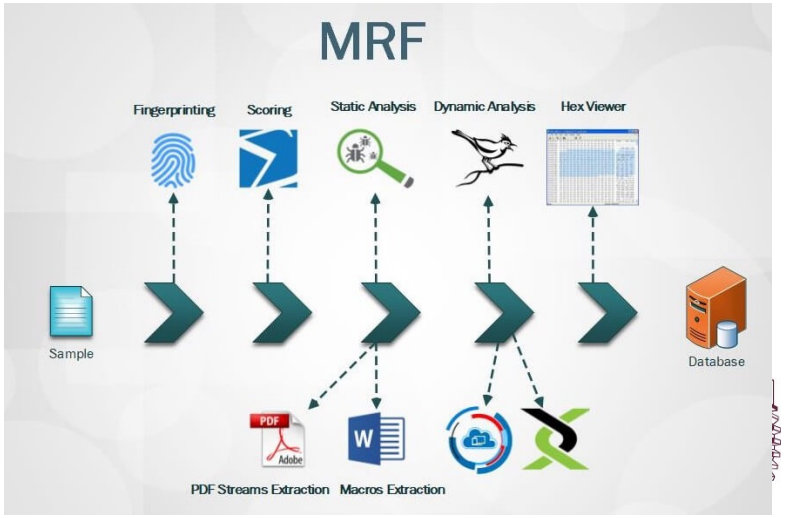
Introducción

Motivación

Introduction

Conceptos básicos

1. La idea es no sólo almacenar las muestras sino ofrecer diferentes funcionalidades de análisis (como el Lab de Malware).
2. La plataforma puede definir un “framework”, es decir, un marco de trabajo que incluya: i) un portal, ii) scripts, y iii) herramientas.
3. Se usan diferentes nombres como.
   - Malware DB.
   - Malware Repository Framework (MRF)
4. Links
   - https://github.com/Tigzy/malware-repo
     2017, junio.
     Aplicación Web con PHP.



VirusSign

Muestras recientes

BD privada

Pero ...

VirusSign

Descripción

Es una base de datos comercial con las siguientes características:

- El volumen de muestras que tiene es:
  - **600 millones** de muestras **300TB**, 4 de sept. 2024.
- Costos: Por mes
  - Paquete básico $50 USD mensuales / $1000 mil pesos (tomando $20)
  - Otros paquetes hay que contactarlos.
- Servicios:
  - Ancho de banda (bps) según cuenta: 2M, 200M, 1G.
    Automated Malware Analysis System (AMAS).
- https://www.virussign.com/

VirusSign

Descripción

Free

$0/mo

Sign up

- Up to 100 records/day
- Up to 2Mbps/s
- Windows only

Basic

$49.99/mo

Subscribe

- 10k records/day (Avg.)
- Up to 200Mbps/s
- Windows | Linux | Unix

Professional / Business

-

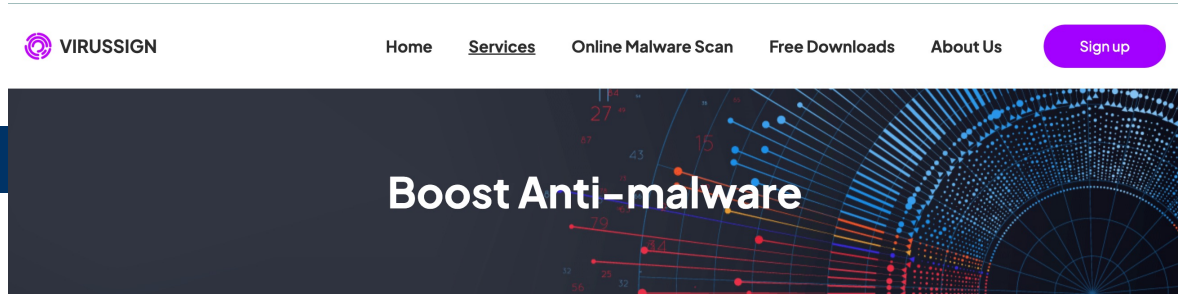
Contact us →

- 200k-500k records/day
- 1Gbps/s
- Windows | Linux | Unix
- Android | iOS | macOS

If you're interested in specific datasets, such as Android, macOS, iOS, Linux, metadata-only, or more customized plans, or if you wish to obtain our collection of over 300TB and approximately 600 million past malware samples, please contact us.

VirusSign

Descripción



Malware database

Daily malware feeds keep your anti-malware detection and defense capabilities up-to-date.

Data plans

Malware analysis

Our AI-powered AMAS rapidly and independently identifies new malware.

Learn more

Biggest repository

Over 300TB and approx 600 million nonredundant samples, it is the most valuable resource to empower your AV, EDR, SIEM, DLP, CTI, Firewall, and more.

Online search

We offer free online malware scanning and search. You can quickly confirm if the files or hashes have been identified as malware in our database.

Malware analysis

Speed up your response with our Automated Malware Analysis (AMAS), which rapidly identifies new malware no need third-party scanners.

VirusShare

La BD pública más grande

BD pública

VirusShare

Descripción

Es una base de datos pública con las siguientes características:

- Identificación de las muestras mediante hashes (MD5, SHA1, SHA256).
- Contiene muchas muestras:
  - Del periodo: 2012-06-15 a 2024-09-4, 12 años.
  - Al 21 de noviembre del 2023 hay 86 millones de muestras.
  - Cada semestre descargamos las nuevas.
- Requiere cuenta pero es gratuita.
  - https://virusshare.com/

86,722,860

VirusShare

Descripción

VirusShare.com - Because Sharing is Caring

Home • Hashes • Torrents • Research • About

Search ?

System currently contains 41,221,083 malware samples.

Report for a sample recently added to the system:
51d7e0c6df18d0cb7044a8c05d5fecbdba80b6a02937900755a976b9c6728d03

VirusShare info last updated 2021-10-05 00:00:04 UTC

MD5	e4b403f069a0c3adf3b207f40db41f79
SHA1	0bdf770f69132769779a9caca48e66425c8d3a3c
SHA256	51d7e0c6df18d0cb7044a8c05d5fecbdba80b6a02937900755a976b9c6728d03
SSDeep	768:H52M/1jHu0Xs/WhOmebwXRD8chY5sqpQ4ua:H5V/Q/WecachYGqpQ4n
Size	37,325 bytes
File Type	HTML document, UTF-8 Unicode text, with very long lines, with CRLF, LF line terminators
Mime Type	text/html
Extension	html
TrID	HyperText Markup Language with DOCTYPE (80.6%) HyperText Markup Language (19.3%)
Detections (1/10)	ClamAV: PUA.Win.Tool.Packed-177 *VirusShare Scanned 2021-10-05 00:00:04 UTC*
ExIF Data	Clckd: 7b3cdf8224a8e0e0fdda865475d406ce Description: Young Non Nude Models Galleries Finding Girls For Sex In Addis Ababa, Ethiopia Guys Bodacious photo session of nude guy and a g XXX Dessert Nude pics of season hubley Geri Burgess Nude Delight Zishy / Hotty Stop July | 2011 | My Asian Gfs young girls non nude art, earlie teens girls non nude, […] FileSize: 36 KiB FileType: HTML FileTypeExtension: html Generator: WordPress 5.4.7 Google: nositelinkssearchbox GoogleSiteVerificat…

MD5	
SHA1	8bf213c163a1be25527876769e50a07f68f6fa44
SHA256	3b7b47d2f4bcaea7d2024d048fc2eafcc097b196579d8438a64defc667665474
SSDeep	3072:SF48emwdydxyfkMY+BES09JXAnyrZaII+YQ:Se8emwdyd0sMYod+X3oI+YQ
Size	147,416 bytes
File Type	HTML document, UTF-8 Unicode text, with very long lines, with CRLF line terminators
Detections	Ad-Aware = Trojan.HTML.Ramnit.A ALYac = Trojan.HTML.Ramnit.A Antiy-AVL = Trojan[Dropper]/VBS.Ramnit.a

VirusShare

Descripción

Varias listas con muestras específicas

VirusShare_Android_APK_2022-3	203.83 GB	2023-01-06 21:47:26
All Android APK files collected in 2022 Part 3 of 3. 20,206 samples.		
VirusShare_ELF_20200405	2.40 GB	2020-04-05 21:50:52
All new ELF binaries collected since the previous release in 2019.		

Descargas

Es posible descargar todas las muestras (cuenta gratis) usando torrents.

VirusShare.com - Because Sharing is Caring

Home • Hashes • Torrents • Research • About • Swag Shop

Need more than a few samples? Want to download a large number of malware samples? Want to download almost the entire corpus? Here's how:

VirusShare has a torrent tracker serving up packages of malware packaged in the order they were received and indexed on the system. Each is a password-protected zip file with the usual 'infected' password. Torrents 0 through 148 contain 131,072 samples each. Torrents 149 and later contain 65,536 samples each.

If you are going to download, please continue the sharing and leave your torrent client seeding as long as possible so that other people like you can get their samples too.

File	Size	Added
VirusShare_00000.zip	13.56 GB	2012-06-15 00:39:38
VirusShare_00001.zip	85.33 GB	2012-06-15 22:12:42
VirusShare_00002.zip	46.62 GB	2012-06-16 09:01:01

VirusShare_00476.zip	30.58 GB	2023-07-14 18:31:51
VirusShare_00477.zip	19.78 GB	2023-08-07 13:12:07
VirusShare_00478.zip	19.46 GB	2023-08-13 23:01:39
VirusShare_00479.zip	17.82 GB	2023-08-20 11:14:24
VirusShare_00480.zip	19.56 GB	2023-09-09 13:41:21
VirusShare_00481.zip	24.11 GB	2023-09-22 13:51:57
VirusShare_00482.zip	28.97 GB	2023-10-21 13:04:27
VirusShare_00483.zip	36.96 GB	2023-11-04 12:10:06
VirusShare_00484.zip	24.15 GB	2023-11-10 09:16:07
VirusShare_00485.zip	46.76 GB	2024-02-07 17:37:12
VirusShare_00486.zip	20.29 GB	2024-02-14 16:29:00

VirusShare

Descripción

La mayoría de las muestras de los repositorios no permiten (al menos la versión gratuita) saber o elegir las muestras según:

1. El tipo de malware (15)

   Adware, Backdoor, Botnet, Ransomware, Rootkit, Trojan, Spyware.

2. El tipo de archivo (70)

   Ejecutable (PE, ELF, etc.), Script (Bash, JavaScript, Perl, PHP), Office y PDF, Imagen (jpg, png, bmp, etc.), etc.

3. El tipo de arquitectura

   Intel (AMD), M1, ARM, Atom, etc. / Número de bits (16, 32, 64)

4. Periodo

   Evolución: virus metamórficos a polimórficos, gusanos.

Escenarios

Ransomware	Rootkit	Trojan
Windows	Linux	Web
	ARM (webcam)	Apache

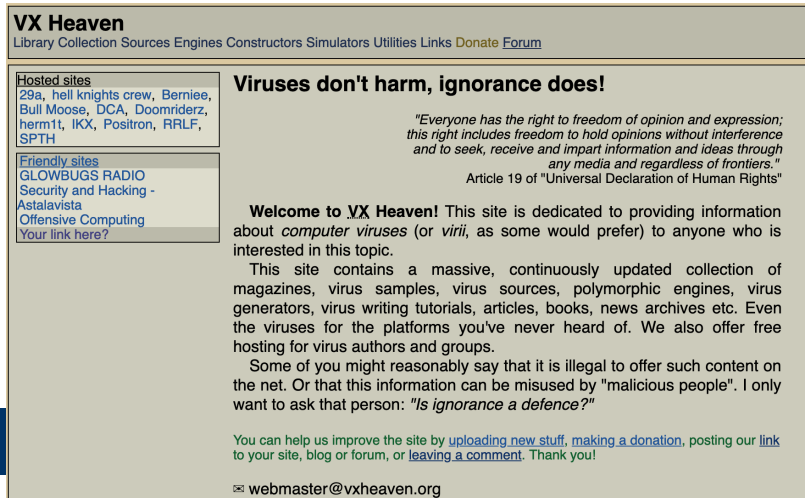
VX-Heaven

Una de las más citadas en los primeros artículos

BD públicas

VX-Heaven

Introducción



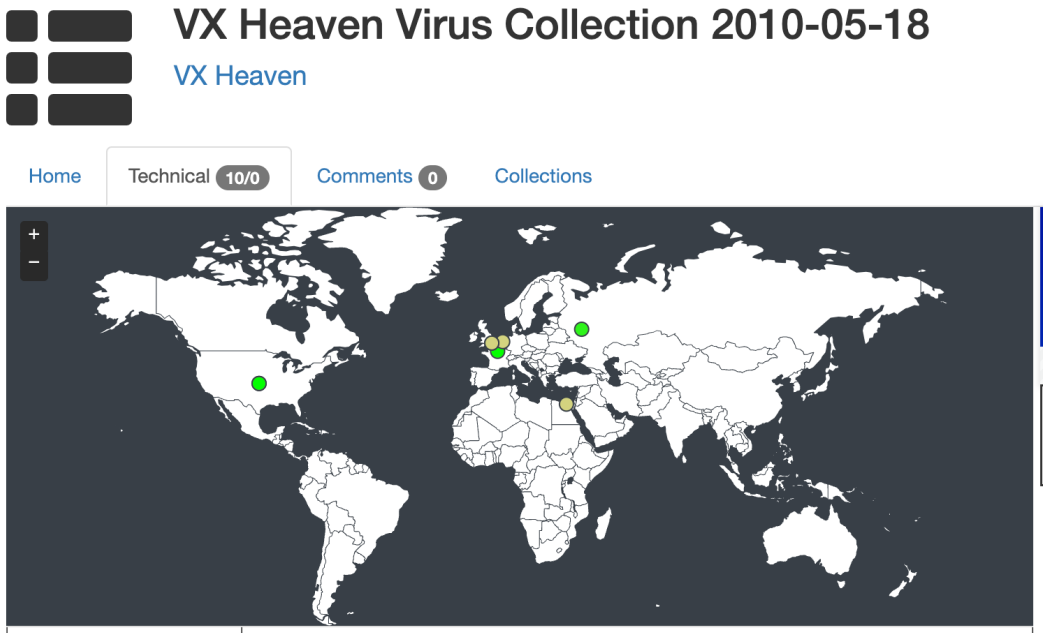
Es una base de datos pública con las siguientes características:

- Es posible descargar de forma directa las muestras.
  - 83.133.184.251/virensimulation.org/index.html
  - Deja de funcionar con frecuencia. Cuando esta fuera de línea usar:
  - http://academictorrents.com/details/34ebe49a48aa532deb9c0dd08a08a017aa04d810/tech&dllist=1.
- Identificación de las muestras (MD5, SHA1, SHA256).
- Contiene varias muestras de malware viejas y algunas recientes. En total se tienen:
  - 270,898 muestras.

VX-Heaven

Introducción

Muestras 2010 => Registro 2014



VX Heaven Virus Collection 2010-05-18

VX Heaven

Home | Technical 10/0 | Comments 0 | Collections

Info hash	34ebe49a48aa532deb9c0dd08a08a017aa04d810
Last mirror activity	4:24 ago
Size	47.88GB (47,880,726,923 bytes)
Added	2014-07-02 18:50:10
Views	3126
Hits	8224
ID	2831
Downloaded	3050 time(s)

VX-Heaven

Introducción

Las muestras están organizadas por familias de malware, protocolos, etc. En particular de la siguiente forma:

1.	Backdoor	51,459
2.	Constructor	933
3.	DoS	257
4.	Email	3,511
5.	Exploit	1,938
6.	Flooder	267
7.	HackTool	775
8.	Hoax	1,216
9.	IM	513
10.	IRC	545
11.	Net	1,938
12.	not-virus	148
13.	P2P	573
14.	Packed	379
15.	Rootkit	3,210
16.	SMS	57
17.	SpamTool	64
18.	Spoofer	26
19.	Trojan	169,392
20.	VirTool	604
21.	Virus	26,941
22.	Worm	6,152

VX-Heaven

Introducción

Similar a VirusShare

¿De que tipos de Archivos son?

Los datos de esta página son el resultado del cálculo que se obtiene de la información en la base de datos.

No	Tipo	Cantidad
1	PE	219,113
2	EXE	25,146
3	Unknown	14,344
4	ASCII text	5,578
5	CDF Microsoft	3,740
6	Boot Sector	591
7	HTML	496
8	ELF	410
9	ISO-8859 text	353
10	PDF	296
11	DOS floppy	261
12	DOS Batch	113
13	OMF Microsoft	111
14	Java class	100
15	Autorun	69
16	PHP	40

Kaggle

Plataforma de ML con concursos / Uno de malware

BD públicas

Kaggle

Descripción

Es un sitio web donde se pueden realizar proyectos científicos del área de Ciencia de Datos. Se caracteriza por

- Concursos con premios
  - https://www.kaggle.com/c/malware-classification
- Datos disponibles (Datasets)
- El procesamiento se hace mediante el concepto de Kernels.

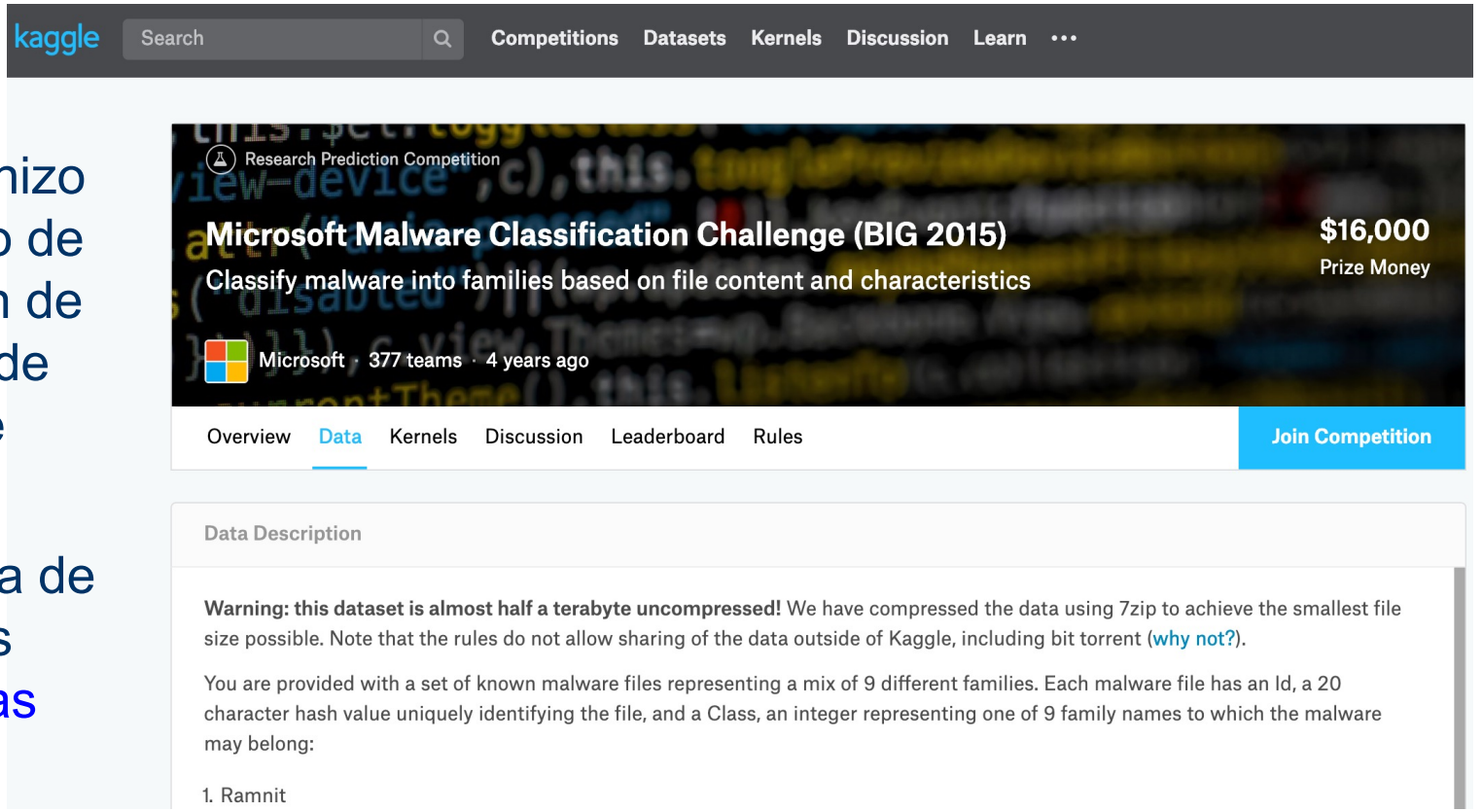
  Permiten explorar y ejecutar código de Maquinas de Aprendizaje (*machine learning*), en un ambiente computacional de nube que permite realizar análisis colaborativos y reproducibles.

Kaggle

Descripción

En el 2015 hizo un concurso de clasificación de muestras de malware

Sólo muestra de Windows sanitizadas



theZoo

Código fuente de malware

BD públicas

theZoo

Descripción

Es una colección de muestras de malware:

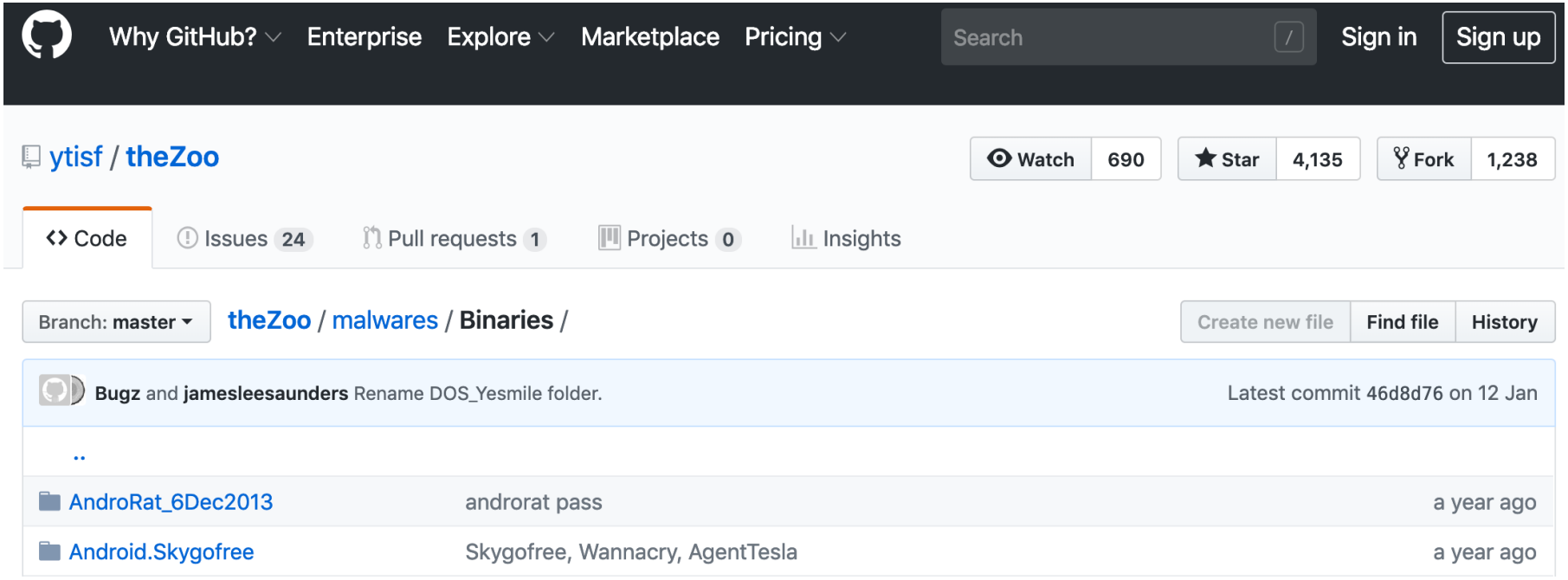
- Contiene pocas muestras (214) gratuitas pero:
  - Contiene los archivos binarios.
  - Los códigos fuente de varias de las muestras.

- https://github.com/ytisf/theZoo
- 2014, diciembre.

theZoo

Descripción

A Live Malware Repository



theZoo

Descripción

```
Dbot v3.0 (price: 100 usd):
- stable irc bot
- multicommand topic parsing
- multicommand chat parsing
- irc connection timeout
- unlimited number of irc servers
- xor encoded strings (antivirus anti-heuristic)
```

DBot v3.1 - March 2007

DBotv3.1_March2007.zip

```
3. Bot commands
---------------

Standard non-protected commands (assuming prefix = !):
() == needed parameter, [] == optional parameter

!v                      bot version
!d                      disconnect
!r                      reconnect to same server
!q                      reconnect to next server
!n [nick]               change nick [to nick]
!restart                restarts the bot (this is not reconnect)
!scanstop               stop scanning
!patch                  patch win xp sp2 tcpip.sys (report to techchannel)
!total                  total transfers via bots ftpd
!vnc                    report own vnc server status (to techchannel)
!getcftp                get current cftp settings
!j (#chan)              join #chan
!p (#chan)              part #chan
!bk (x)                 start botkiller, repeat every x minutes
                        (if x == off, bk is turned off)
!setcftp (host) (port) (user) (pass) (file)     change cftp settings

Scanning command:
!scan (threads) (rand/seq) (ip/b/y) (vnc mode) (transfer mode) [lanbots range]

threads = number of threads (recommended 64, not go over 96)
rand/seq = 1 means random, 0 means sequential scanning
ip/b/y = put in IP to start scan from or b for C&D IP range scanning (own)
         or y for B random IP range scanning (own)
vnc mode = 0 - no vnc scan,
           1 - with vnc scan,
           2 - with vnc scan and rooting,
           3 - vnc scan only
transfer mode = 0 - all bots use ftp
                1 - lan bots use cftp, wan bots use ftp
                2 - all use cftp
```

- d3des.c
- d3des.h
- dbot.dsp
- dbot.dsw
- dbot.ncb
- dbot.opt
- dbot.plg
- downloader.cpp
- downloader.h
- features.txt
- ftpd.cpp
- ftpd.h
- include.h
- main.cpp
- main.h
- manuals.txt
- md5.cpp
- md5.h
- MD5ChecksumTest.exe
- misc.cpp
- misc.h
- netapi.cpp
- netapi.h
- PackFiles.sh
- patcher.cpp
- patcher.h

Contagio

Blog

BD públicas

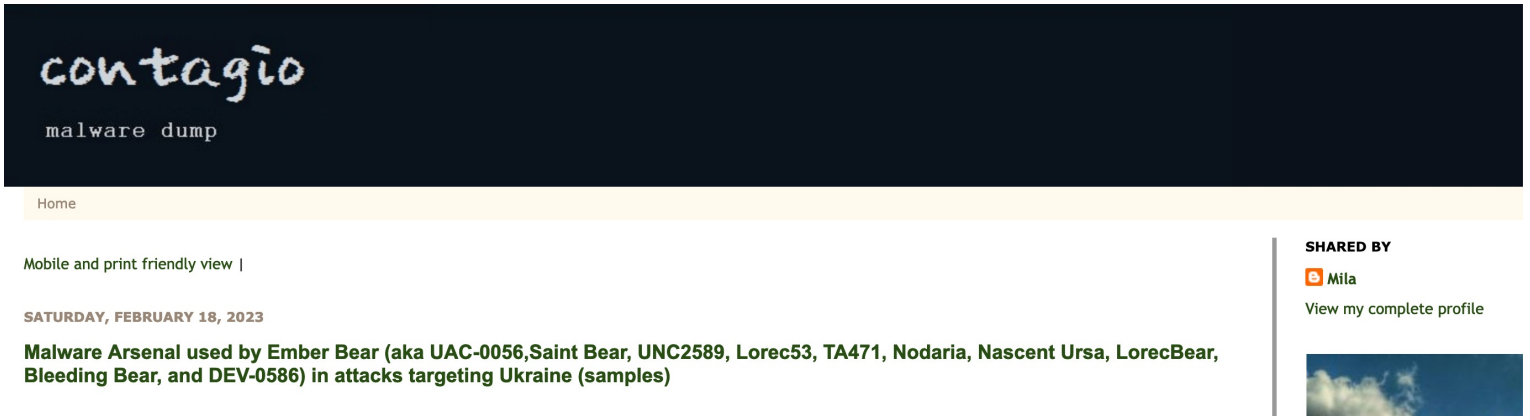
Contagio

Descripción

Se define como:

> *Contagio is a collection of the historic malware samples, threats, observations, and analyses.*

- Son gratuitas
  - Todas comprimidas con contraseña: infected.
  - Varias tienen una explicación de su funcionamiento.
  - Periodo: **2008-2024, 16 años**.
- https://contagiodump.blogspot.com/



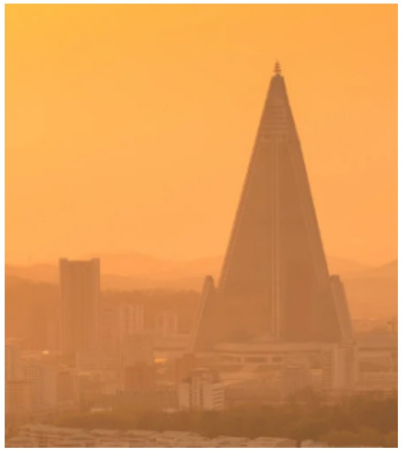
Contagio

Descripción

La búsqueda se limita a las etiquetas que usan en el blog y por fecha. Ejemplo del malware el más reciente:

MONDAY, SEPTEMBER 2, 2024

2022-2024 North Korea Citrine Sleet /Lazarus FUDMODULE (BYOVD) Rootkit Samples



2024-08-30 Microsoft: North Korean threat actor Citrine Sleet exploiting Chromium zero-day

2024-03-01 Lazarus group operations—A deep dive into FudModule Rootkit by Lucas Mancilha

2024-02-28 Avast: Lazarus and the FudModule Rootkit: Beyond BYOVD with anAdmin-to-Kernel Z

2024 Blackhat Asia Speakers: Luigino Camastra, Igor Morgenstern Video
Slides

2024-04-18 Avast: From BYOVD to a 0-day: Unveiling Advanced Exploits inCyber Recruiting Scam

2022-09-30 ESET: Lazarus & BYOVD: evil to the Windows core

2022-09-22 Ahnlab: Lazarus Group's Rootkit Attack Using BYOVD

- ► 2020 (2)
- ► 2019 (5)
- ► 2018 (1)
- ► 2017 (4)
- ► 2016 (4)
- ► 2015 (5)
- ► 2014 (5)
- ► 2013 (17)
- ► 2012 (59)
- ► 2011 (77)
- ► 2010 (191)
- ► 2009 (55)
- ► 2008 (1)

SHORTCUTS

- Mobile Malware mini-dum
  leave a sample.

CATEGORIES - SORT OF

- - HTA files (1)
- - JAVA (16)
- - MOBILE MALWARE (6)
- - OSX (10)
- alienspy (1)
- APT1 (1)
- Aurora (2)
- Autocad (1)
- Backdoor.Olyx (1)
- Backdoor.Wirenet (1)

Repositorio de Ciseg

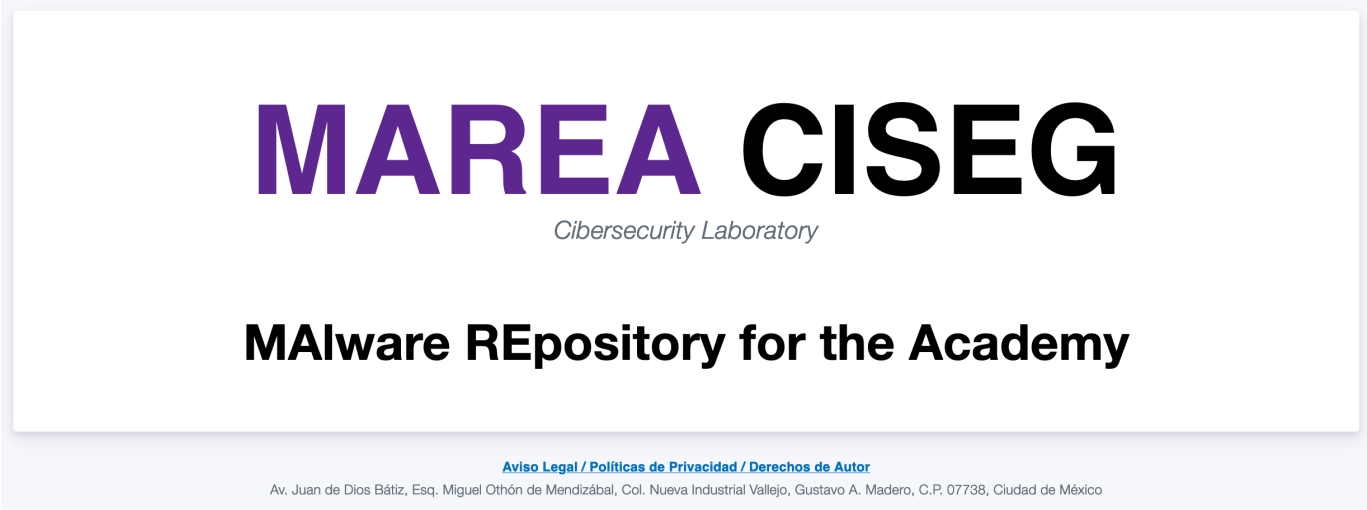
Compilación de varias Tesis, sabático, etc.

BD pública y privada

Ciseg Lab

Introducción

- Se han compilado y organizado las bases de datos anteriores.
- La organización empieza con el tipo de archivo y sistema operativo.
- Para trabajos de tesis de varios niveles: NS, Maestría y Doctorado.

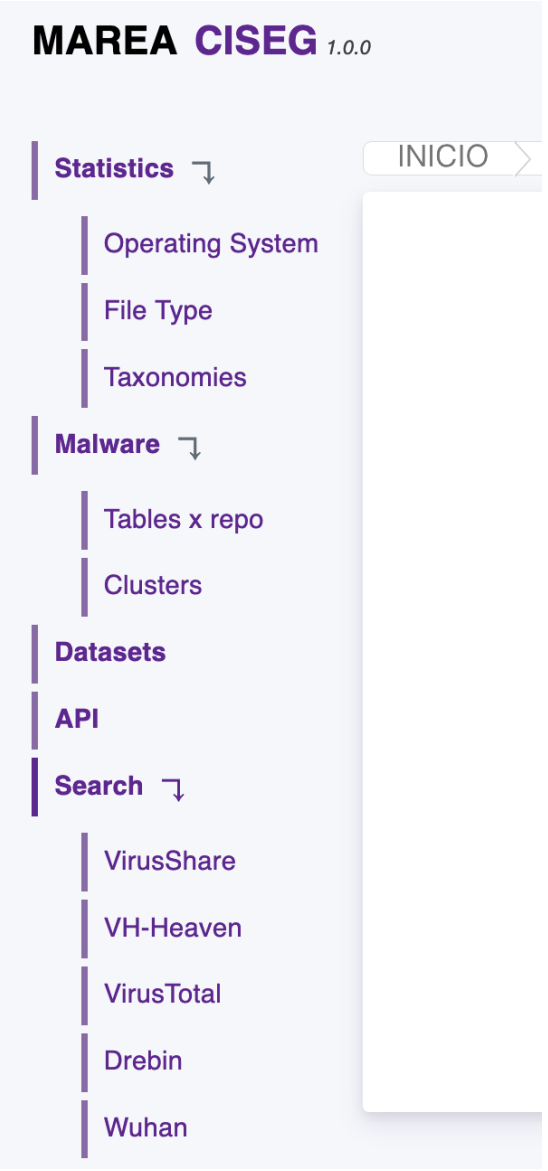


Ciseg Lab

Introducción

- Está en constante desarrollo y ya tiene varias funcionalidades como:
  1. Búsqueda por valor hash
  2. Agregar muestras. Solo el administrador.
  3. Crear datasets.
  4. Acceso mediante API/REST.
- Se requiere cuenta en el sistema y hay diferentes roles
  1. Administrador
  2. Soporte
  3. Analista

Ejemplo del menu

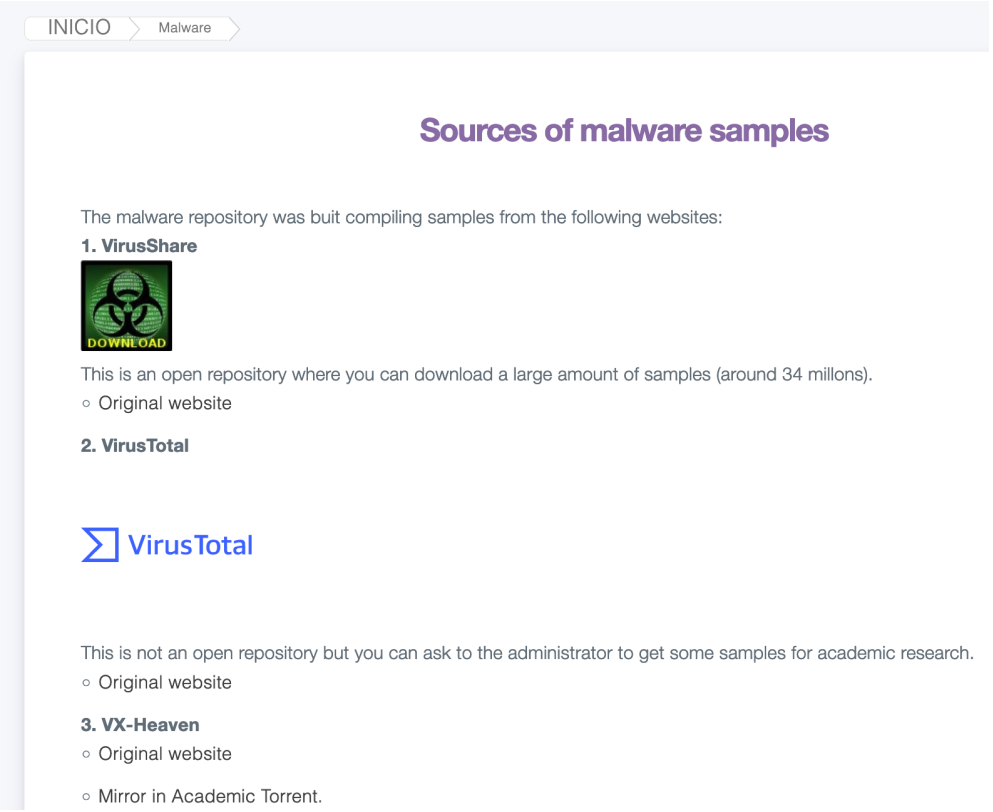


Ciseg Lab

Introducción

- Veremos el sistema y algunas estadísticas:
  http://148.204.64.218/marea/
- Implementación
  1. Aplicación web
  2. Almacenamiento
     - Muestras en NAS
     - Archivos del procesamiento en local (MV).

NAS vs SAN



API rest

Explicación, ejemplos

VT y otros sitios

API rest

Introducción

1. Definido en el año 2000 por el Dr. Roy Fielding en su tesis doctoral.
2. Es una arquitectura pero no un standard o protocolo.

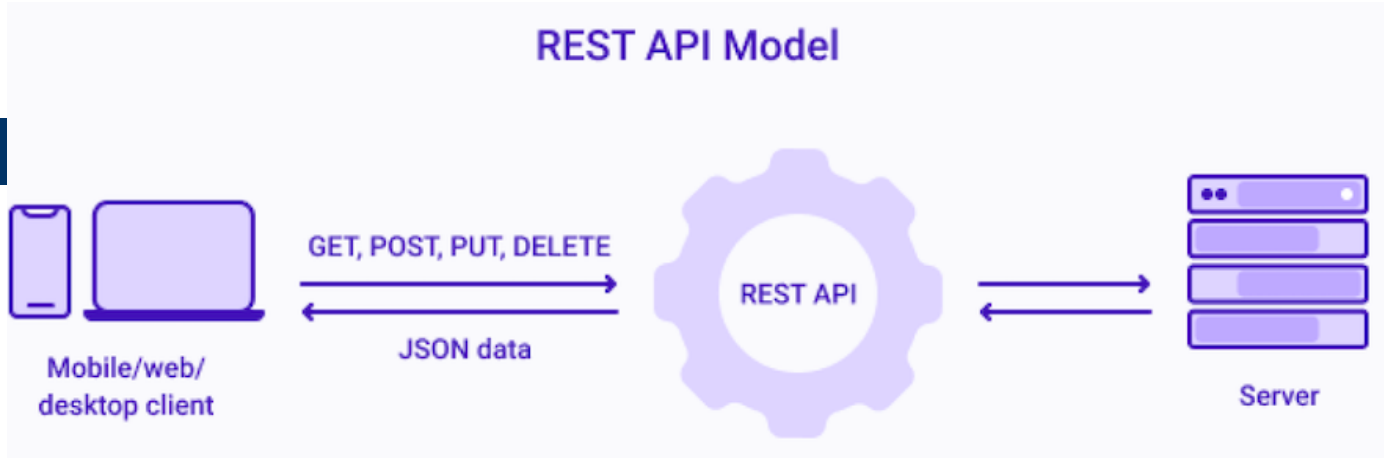
   Características o criterios:

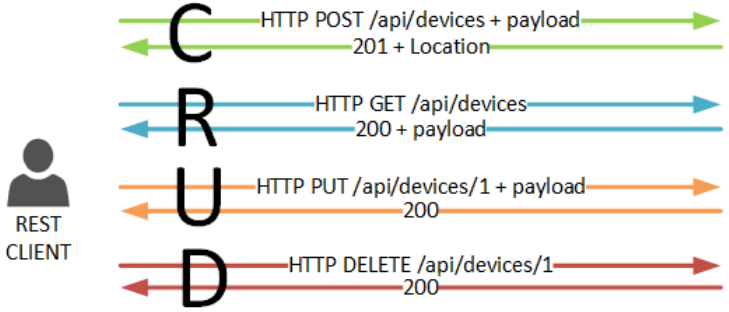
   - Modelo Cliente-Servidor.
   - Sin estado. Stateless.
   - Uso de memoria cache: mejorar el rendimiento.
   - Diseño en capas para ocultar intermediarios.
   - Uso de URL o URI como identificador único de los recursos.
   - Puede ser RESTful (usa los 5 criterios del REST) o RESTless.
   - Orientado a recursos, usa las operaciones de los verbos HTTP.
3. Implementación:
   - Protocolos: HTTP, SOAP (Simple Object Access Protocol).
   - Formato: XML, Json

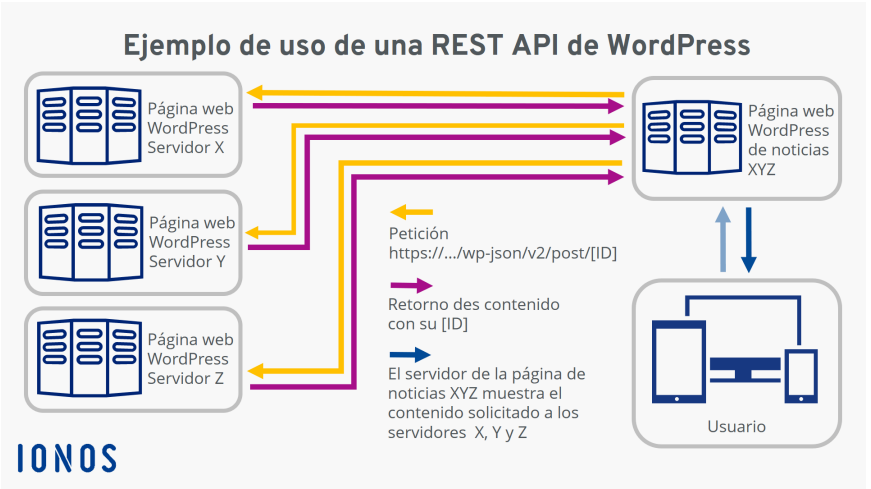
API rest

Intro

Arquitectura







API rest

Introducción

Ejemplo de API

- GET / tickets **R**ecupera una lista de tickets
- GET / tickets / 12 **R**ecupera un ticket específico
- POST / tickets **C**rea un nuevo ticket
- PUT / tickets / 12 **U** Actualiza ticket # 12
- PATCH / tickets / 12 **U** Actualiza parcialmente el ticket # 12
- DELETE / tickets / 12 – **D** Elimina el ticket # 12

API rest

Introducción

Ejemplos y referencias

1. VirusTotal
   - General: https://developers.virustotal.com/reference/overview.
   - Ejemplo de uso:
     https://developers.virustotal.com/reference/file-info
2. URLs
   - https://jsonapi.org/
   - https://github.com/public-apis/public-apis
     https://ipapi.co/#api, https://ipapi.co/148.204.64.218/json
   - Desarrollo de APIs: https://www.postman.com/
     https://documenter.getpostman.com/view/8854915/Szf7znEe
3. Temas avanzados

   Como se realiza la autenticación?

   Password, certificado o API key. Oauth 2.0

Conclusiones

Cosas que recordar

To remember

A recordar

Lo más importante es:

- Como usar el repositorio
  - Descargas manuales
  - Descargas mediante un API Rest y una API key.
- Las muestras están protegidas
  - Comprimidas con contraseña (*infected*)
  - Sanitizadas (a veces y no se pueden ejecutar/dañar).
- Se pueden seleccionar por tipo de:
  - Sistema Operativo, Archivo, etc.
  - Familia de malware.
  - *Dataset*: muestras con un criterio específico.

The end

Contacto

Raúl Acosta Bermejo

http:www.cic.ipn.mx
http://www.ciseg.cic.ipn.mx/

racostab@ipn.mx
racosta@cic.ipn.mx

57-29-60-00
Ext. 56652