Big Game Hunting

Peculiarities In Nation State Malware Research
Marion Marschalek
Senior Malware Researcher
Cyphort Inc.
The centrifuges that spin nuclear material at Iran’s enrichment facilities.
Offense
Going
Commercial
AV 2.0

... where the customer is the product
How Anti-Virus went Threat-Intel

Malware.. ‘watching’
Actor tracking
Publicity
APT numbering, logos & names
Haystack Processing

\[ \sim 70,000 - 300,000 \] new samples/day
(Depending which report you trust)

Sample trading

Automated processing
Needle Processing

Threat Intelligence
Telemetry Data
Leaked Documents
Infected Machines
Gossip
Endpoint Wars

Endpoint agents
Threat indicators
Mitigation tactics
Silent data exchange
Agent

Threat detection & mitigation

Signature hits

Timestamps

Hit frequencies

Binaries

Q&A Data

Threat Indicators

Endpoint Wars
Endpoint Wars backstage

Signature generation & testing
Silent signatures
Binaries
Telemetry
‘Free’ security products
FRENEMIES & THE FUNGUS AMONGUS

Or: When Malware Became Intellectual Property
[REDACTED] "Where did you find this malware?"

Me: "It was sent to me by targeted activists."

[REDACTED] "That's Cheating."
Alberto Nisman
YO SOY NISMAN
Todo parece indicar que Nisman fue engañado. A su teléfono Motorola xt626 llegó un archivo con el título "estrictamente secreto y confidencial.pdf.jar". Acaso creyendo que se trataba de un documento importante, lo abrió sin advertir la extensión ".jar". Allí estaba el virus.
Attribution IS tricky?

Network based indicators point to Argentina and Uruguay

Also use of hosting services in the US, Germany, and Sweden
Quand les Canadiens partent en chasse de « Babar »

Le Monde | 21.03.2014 à 12h26 • Mis à jour le 19.05.2014 à 14h13 | Par Jacques Follorou et Martin Untersinger

ntrass.exe
– DLL Loader uploaded to a victim as part of tasking seen in collection
– Internal Name: Babar
– Developer username: titi

Babar is a popular French children’s television show
Titi is a French diminutive for Thiery, or a colloquial term for a small person

C’est une véritable traque qu’ont menée les services secrets techniques canadiens du Centre de la sécurité des télécommunications du Canada (CSEC). Elle est relatée dans le document fourni au Monde par Edward Snowden, dans lequel ils présentent leurs trouvailles. Avaré en détails, ce document permet néanmoins de retracer l’enquête qui a permis de pointer la France du doigt.

Comme dans une partie de chasse, ce sont des empreintes qui attirent en premier lieu l’attention des services canadiens. La note interne indique en effet que le CSEC collecte quotidiennement et automatiquement un certain nombre de
MISERY BUSINESS

Who wrote the malware?
Who controlled the malware?
Who were the victims?
What was the aim of the operation?
MISERY BUSINESS

BINARY ➔ CONTEXT

BINARY ➔ BINAR A CO WIN!
SH* Academics say

Source code authorship attribution

Automatic detection of stylistic features in binary code

Problems?
Datafication of RE results

Different domains & lots of attributes

Any attribute can be faked or random

Assumption: Impossible that all vary in all cases
Goal: Even out individual human / compiler influence
STRING CONSTANTS
Error messages
String formatting style
English grammar mistakes
C&C commands
Timestamp formatting

IMPLEMENTATION TRAITS
Memory allocation habits
Use of global variables
Multi-threading model
Software architecture and design
Constructor design
Dynamic API loading technique
Exception handling
Usage of public source code
Programming language and compiler
Compilation time stamps and time zones

CUSTOM FEATURES
Obfuscation techniques
Stealth and evasion techniques
Use of encryption and compression algorithms
Encryption keys
Re-used source code
Malware specific features
System infiltration
Propagation mechanisms
Artifact naming schemes / algorithms
Data exfiltration techniques
System / OS version determination technique
C&C command parsing implementation

INFRASTRUCTURE
C&C servers
Countries / languages used for domain hosting and naming
Beaconing style
Communication protocol and port
Communication intervals
<table>
<thead>
<tr>
<th></th>
<th>NBOT/TFC</th>
<th>Bunny</th>
<th>Babar</th>
</tr>
</thead>
<tbody>
<tr>
<td>String constants</td>
<td>Error / status messages: No</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>String formatting style: All</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>English grammar mistakes: No</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Timestamp formatting: Time</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td>Implementation traits</td>
<td>Memory allocation habits:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Use of global variables: Few</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Multi-threading model:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Software architecture and</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>design: MSVC++ default</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Dynamic API loading technique</td>
<td>Present, subset of APIs only, per API, API name identifie</td>
<td>Present, subset of APIs only, per API, API name identifie</td>
</tr>
<tr>
<td></td>
<td>Exception handling: C++ EH and unhandled exception filter: ExitThread()</td>
<td>C++ EH and unhandled exception filter: ExitThread()</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Usage of public source code:</td>
<td>None (known)</td>
<td>Keylogger from codproject.com, OpenscoreAMR, libunwind</td>
</tr>
<tr>
<td></td>
<td>Programming language and</td>
<td>C++ / MSVC++ 8.0</td>
<td>C++ / MSVC++ 8.0</td>
</tr>
<tr>
<td></td>
<td>compiler: C++ / MSVC++ 8.0</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Compilation time stamps and</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>time zones:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td>Custom features</td>
<td>Obfuscation techniques:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Stealth and evasion techniques:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Use of encryption and</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>compression algorithms: API</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>name obfuscation custom</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>algorithm: API name</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Obfuscation of subset of</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>APIs: Emulator check:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>XOR key 0x40C777</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Re-used source code in</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>general: API name hash</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>generation, API name</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>hashing and loading:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Malware specific features:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>System infrastructure:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Malware configuration:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>C&amp;C servers:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Countries / languages used</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>for domain hosting and</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>nominating: US/English</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>User agent / beaconing style:</td>
<td>Many</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>Communication protocol / port</td>
<td>HTTP/80</td>
<td>HTTP/80</td>
</tr>
<tr>
<td></td>
<td>Communication intervals:</td>
<td>Many</td>
<td>Many</td>
</tr>
</tbody>
</table>

**Science, Yo!**
Stylometry in Attribution

BABAR linked to French government

BUNNY spearphishing with 0-days

NBOT Denial-of-Service

DINO spying in Iran

CASPER active in Syria in 2014
What It’s Not

No authorship attribution
Manual work
Not feasible for automation / machine learning
Interpretation in the eye of the analyst
Soft Attribution vs Hard Attribution
SECRET MALWARE IN EUROPEAN UNION ATTACK LINKED TO U.S. AND BRITISH INTELLIGENCE

BY MORGAN MARQUIS-BOIRE, CLAUDIO GUARNIERI, AND RYAN GALLAGHER

11/24/2014
USERDNSDOMAIN = BGC.NET
USERDOMAIN = BELGACOM
USERNAME = id051897a
USERPROFILE = C: \Users \id051897a
A cyber attack on Belgacom raised considerable attention last week. Documents leaked by Edward Snowden and seen by SPIEGEL indicate that Britain's GCHQ intelligence agency was responsible for the attack.
"Looking at the code closely, we conclude that the "QWERTY" malware is identical in functionality to the Regin 50251 plugin."
"BLIND FREDDY COULD SEE E_QWERTY IS A REGIN PLUGIN"
Curiouser and Curiouser!
- cried Alice
SSOOOUUU...
String obfuscation with XOR 9Bh
Checking for running
security processes (and dummyyy.exe)
Control component talking to a device driver `\\.\asr2892`
Sending IOCTLs 220004 & 220008

Orchestrator component executing binaries from disk
Drops ‘msrun.exe’ from .rsrc section
Redirects standard handles of spawned process, piping output back to launcher

2002
Prepared to run on _old_ Windows versions
Using APIs deprecated after Win95/98/ME
Function to check for the MZ value,
the PE value and the NE value
Implementation traits and user agent string indicate Win NT 4.0 as target platform

Persists as shell extension for the icon handler
Wants to run in the context of the ‘Progman’ window
Implant to monitor terminal server sessions

Global hook to filter for WM_KEYFIRST, WM_SYSKEYDOWN, WM_CHAR, WM_SYSCHAR

Loads msob4k32.dll and 6 exports by ordinal

2007-2009
String obfuscation using XOR 9Bh
Evasive when network sniffer products are running

Super stealthy network communication:
Versatile communication method
9+ C&C servers, infrequent intervals
Communication done through injected standard browser instance

2007-2009
Fine tuned to paddle around Kaspersky security products

2011
“We’re All Mad Here”

MORGAN
@headhntr

MARION
@pinkflawd

#FREECLAUDIO

@botherderder