マルウェアの機能推定と亜種分類について

甲斐　博 （愛媛大学大学院理工学研究科）

2021/5/21 仕繰セミナー

私の研究テーマ ①

- 数式処理システムによる計算（代数計算）

$$\mathrm{GCD}\left(\left(x+\frac{1}{3}\right)(x+1),\left(x+\frac{1}{3}\right)(x+2)\right)=x+\frac{1}{3}$$

$$\mathrm{GCD}(x+0.333,x+1/3)=1$$

- 数値数式融合計算（ハイブリッド計算）

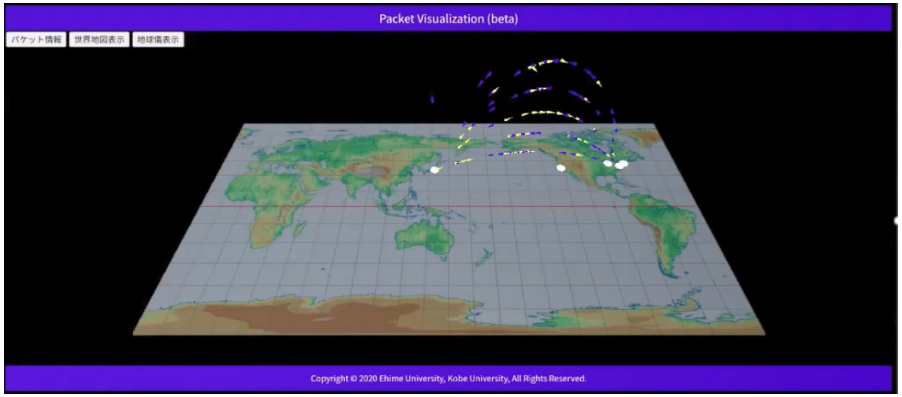
$$\mathrm{ApproximateGCD}\left(x+0.333,x+\frac{1}{3},acc=0.001\right)=x+0.333$$

- 例：　不定積分、因数分解、グレブナ基底、・・・

私の研究テーマ ②

- 暗号・プロトコルの応用
  - SAS(Simple And Secure authentication protocol) 認証方式と応用
  - 秘密分散法

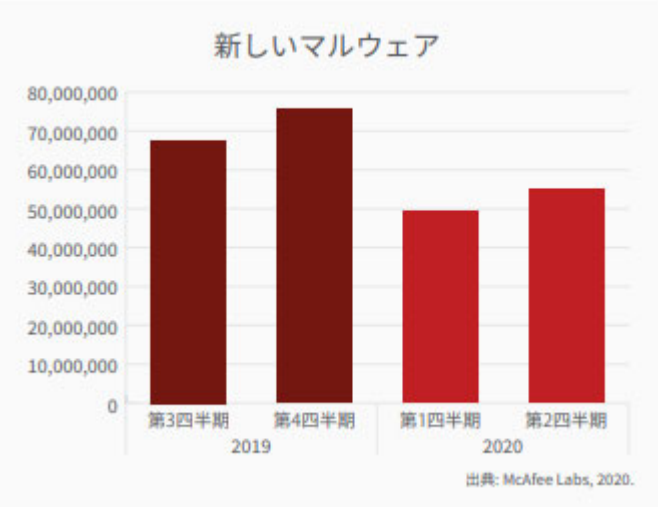
- サイバー攻撃の対策技術
  - ネットワークパケットの可視化
  - **マルウェアの亜種分類、機能推定**

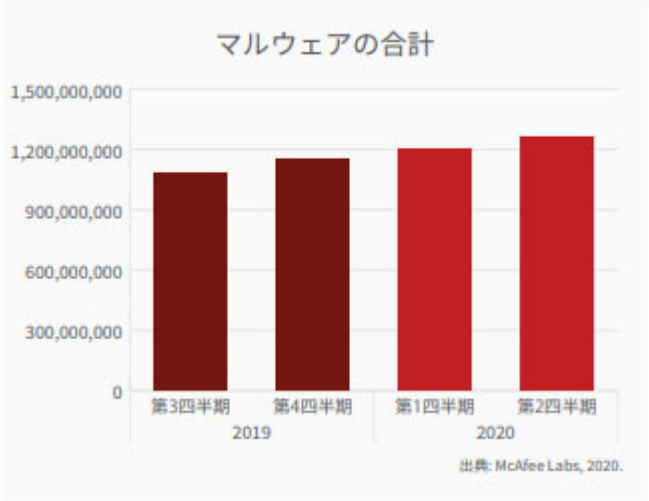


研究背景

近年，多種多様なマルウェアが増加傾向

新種のマルウェア以外に亜種のマルウェアが多く観測





[1] McAfee: McAfee Labs 脅威レポート 2020 年11月
https://www.mcafee.com/enterprise/ja-jp/assets/reports/rp-quarterly-threats-nov-2020.pdf

マルウェアとは

- 定義
  - 悪意のある（malicious）ソフトウェア（software）の総称
- 種類
  - トロイの木馬、ワーム、コンピュータウィルス、ボットなど
- 現状
  - 日々、新しいマルウェアが作成されている
  - 完全な新種のマルウェアは少ない
  - 既存のマルウェアを変更したマルウェア（マルウェアの亜種）が多い

マルウェア解析に関する研究

- マルウェア解析の目的
  - マルウェアに機能を明らかにする
  - マルウェアの亜種かどうかの判定

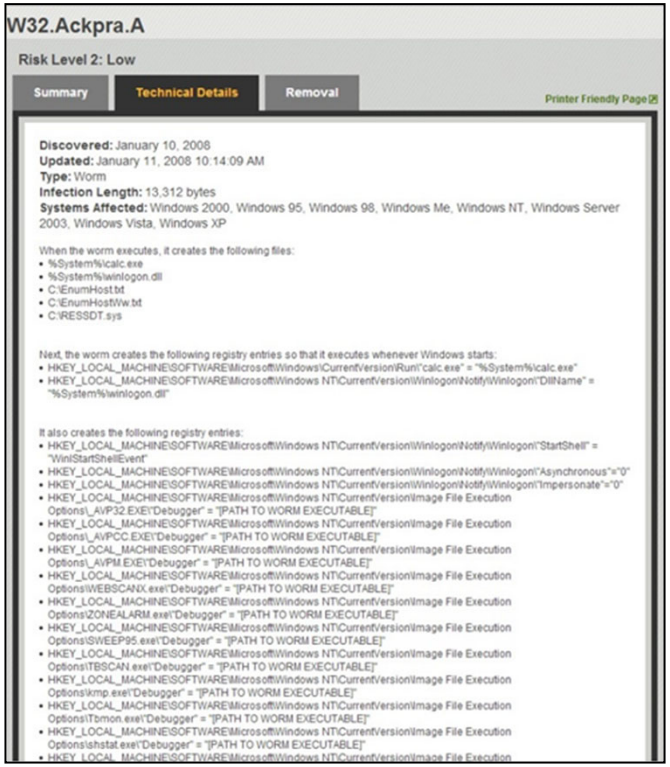
- 研究内容
  - マルウェアの亜種は機能的に類似するか　⇒　可視化
  - マルウェア解析結果を使って、亜種分類が可能か　⇒　動的解析を使った分類
  - マルウェア亜種の機能の推定ができるか　⇒　動的解析を使った機能推定

マルウェアの可視化に関する研究

- 目的
  - マルウェアの類似性や差異を直感的に確認できその分類を行う
- 方法
  - 機能特性に基づくマルウェアの可視化
- 結果
  - 可視化による類似性の直観的判定

マルウェアの解析結果

- セキュリティベンダによるマルウェアの解析結果の公開
  - Symantec Security Response
- 内容
  - マルウェア名
  - 種類
  - マルウェアの機能
  - 対策・駆除方法



W32.Ackpra.A

Risk Level 2: Low

Summary | Technical Details | Removal | Printer Friendly Page

Discovered: January 10, 2008
Updated: January 11, 2008 10:14:09 AM
Type: Worm
Infection Length: 13,312 bytes
Systems Affected: Windows 2000, Windows 95, Windows 98, Windows Me, Windows NT, Windows Server 2003, Windows Vista, Windows XP

When the worm executes, it creates the following files:
- %System%\calc.exe
- %System%\winlogon.dll
- C:\EnumHost.txt
- C:\EnumHostWw.txt
- C:\RESSDT.sys

Next, the worm creates the following registry entries so that it executes whenever Windows starts:
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\"calc.exe" = "%System%\calc.exe"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winlogon\"DllName" = "%System%\winlogon.dll"

It also creates the following registry entries:
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winlogon\"StartShell" = "WinStartShellEvent"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winlogon\"Asynchronous"="0"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winlogon\"Impersonate"="0"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options_AVP32.EXE\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options_AVPCC.EXE\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options_AVPM.EXE\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\WEBSCANX.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\ZONEALARM.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\SWEEP95.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\ITBSCAN.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\kmp.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\Tbmon.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\shstat.exe\"Debugger" = "[PATH TO WORM EXECUTABLE]"
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution

マルウェアの可視化方法

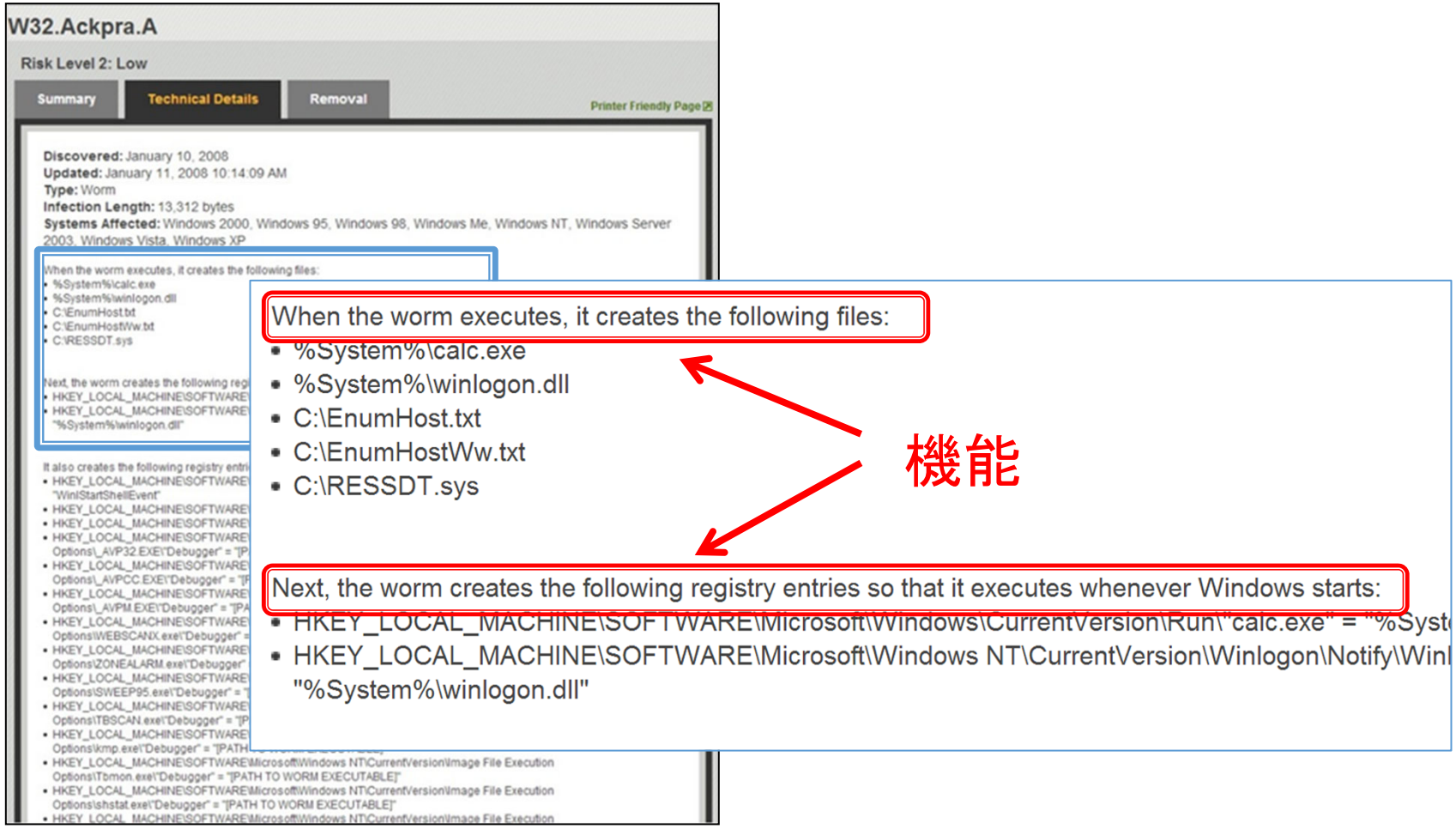
1. セキュリティレスポンスから情報を取得
   - **種類**：マルウェアの種類
     - 例）ワーム，トロイの木馬，…
   - **機能**：マルウェアの動作
     - 例）ファイルを作成する，ファイルを削除する，…
   - **機能対象の数**：動作の対象となるオブジェクトの数
     - 例）作成するファイルの数，削除するファイルの数，…
2. 機能を3Dモデルで表現

   → 可視化による直感的な分類

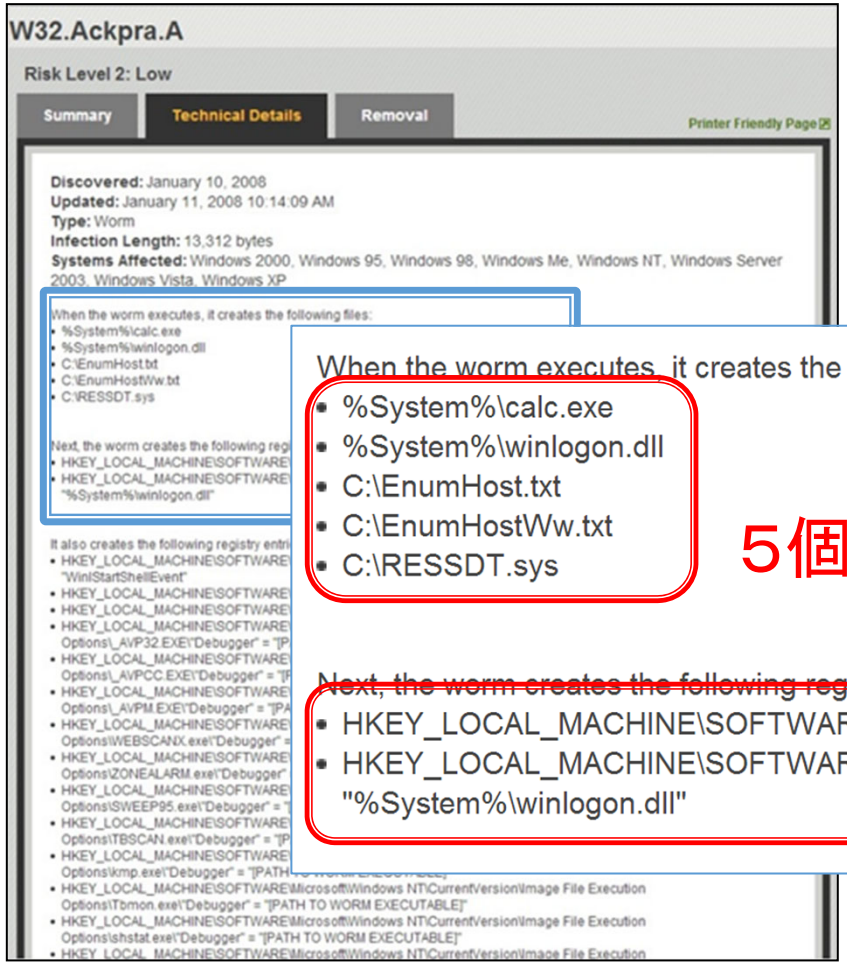
種類の抽出


W32.Ackpra.A
Risk Level 2: Low
Summary
Technical Details
Removal
Printer Friendly Page
Discovered: January 10, 2008
Updated: January 11, 2008 10:14:09 AM
Type: Worm
Infection Length: 13,312 bytes
Systems Affected: Windows 2000, Windows 95, Windows 98, Windows Me, Windows NT, Windows Server 2003, Windows Vista, Windows XP
When the worm executes, it creates the following files:
• %System%\calc.exe
• %System%\winlogon
• C:\EnumHost.txt
• C:\EnumHostWw.txt
• C:\RESSDT.sys
Next, the worm creates
• HKEY_LOCAL_MACH
• HKEY_LOCAL_MACH
"%System%\winlogo
It also creates the follo
• HKEY_LOCAL_MACH
"WinlStartShellEvent"
• HKEY_LOCAL_MACH
• HKEY_LOCAL_MACH
• HKEY_LOCAL_MACH
Options_AVP32.EXE
• HKEY_LOCAL_MACH
Options_AVPCC.EXE
• HKEY_LOCAL_MACH
Options_AVPM.EXE\"
• HKEY_LOCAL_MACH
Options\WEBSCANX
• HKEY_LOCAL_MACH
Options\ZONEALARM
• HKEY_LOCAL_MACH
Options\SWEEP95.e
• HKEY_LOCAL_MACH
Options\TBSCAN.exe
• HKEY_LOCAL_MACH
Options\kmp.exe\"De
• HKEY_LOCAL_MACH
Options\Tbmon.exe\"
• HKEY_LOCAL_MACH
Options\shstat.exe\"D
• HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution
Discovered: January 10, 2008
Updated: January 11, 2008 10:14:09 AM
Type: Worm
種類
Infection Length: 13,312 bytes
Systems Affected: Windows 2000, Windows 2003, Windows Vista, Windows XP

機能の抽出


W32.Ackpra.A
Risk Level 2: Low
Summary
Technical Details
Removal
Printer Friendly Page
Discovered: January 10, 2008
Updated: January 11, 2008 10:14:09 AM
Type: Worm
Infection Length: 13,312 bytes
Systems Affected: Windows 2000, Windows 95, Windows 98, Windows Me, Windows NT, Windows Server 2003, Windows Vista, Windows XP
When the worm executes, it creates the following files:
%System%\calc.exe
%System%\winlogon.dll
C:\EnumHost.txt
C:\EnumHostWw.txt
C:\RESSDT.sys
When the worm executes, it creates the following files:
%System%\calc.exe
%System%\winlogon.dll
C:\EnumHost.txt
C:\EnumHostWw.txt
C:\RESSDT.sys
機能
Next, the worm creates the following registry entries so that it executes whenever Windows starts:
HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\"calc.exe" = "%Syst
HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winl
"%System%\winlogon.dll"

機能対象の抽出



When the worm executes, it creates the following files:

- %System%\calc.exe
- %System%\winlogon.dll
- C:\EnumHost.txt
- C:\EnumHostWw.txt
- C:\RESSDT.sys

5個の機能対象

Next, the worm creates the following registry entries so that it executes whenever Windows starts:

- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\"calc.exe" = "%Syst
- HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon\Notify\Winl "%System%\winlogon.dll"

2個の機能対象

機械的に情報を抽出

- 方法

1. 種類の抽出
   - 「Type:」の右の文字列
2. 機能の抽出
   - 機能と英単語の対応表を作成
   - 対応表で定義した英単語を含む文を，その単語と対応する機能に変換
3. 機能対象の数の抽出
   - 機能と判定されない文は機能対象とみなす
   - 直前に現れる機能の機能対象として数を記録

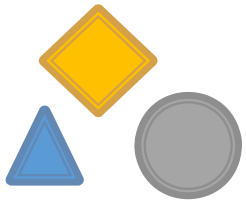
機能と英単語の対応表

機能群	機能	単語
情報収集	特定ファイルの調査	search
	ユーザ情報の収集	collect
	特定箇所の文字列収集	password
感染行動	ファイル・レジストリ・DLLの作成	create, copy
	ファイル・レジストリ・DLLの書き換え	modify, add
	ファイル・レジストリ・DLLの削除	deleate
破壊活動	バックドアの作成	back, door
	システムの破壊	end
外部への動作	サイトへのアクセス	connect
	Eメールの送信	mail
	サーバへのログイン	server
	ファイルのダウンロード	download
	ネットワーク共有	network

可視化モデル

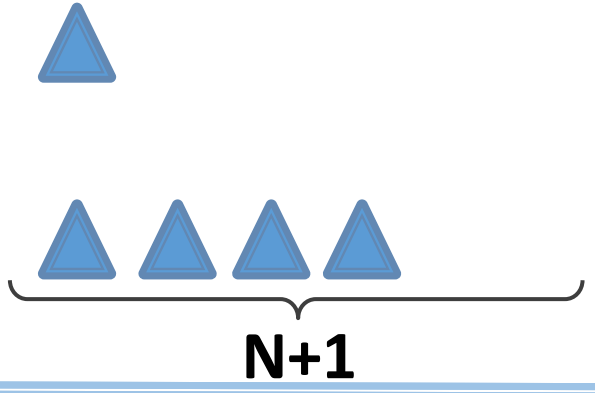
- 種類　　種類モデル

- 機能　　機能モデル



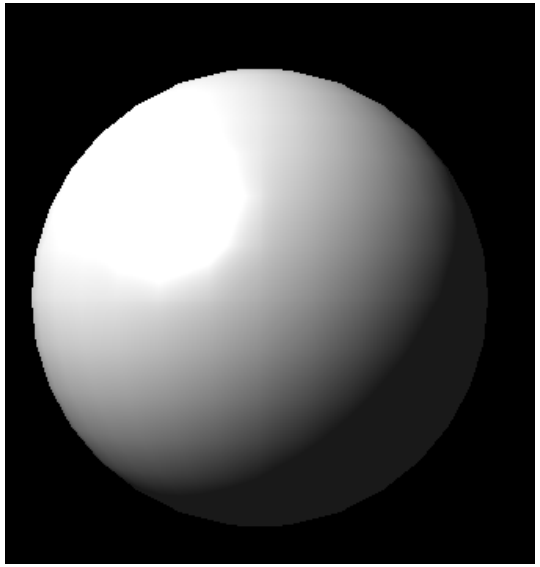
  - 機能モデルの数は「機能対象の数＋１」とする
    - ある「機能A」の機能対象がない（0個）
      - 「機能モデルA」1個
    - ある「機能A」の機能対象がN個ある
      - 「機能モデルA」N＋1個

…

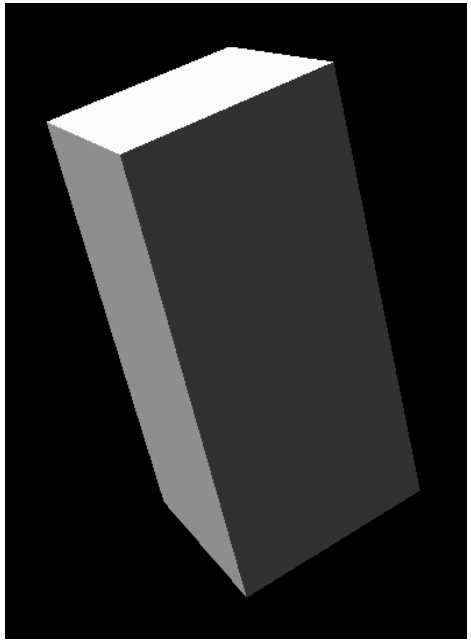


種類モデルの定義

- 種類ごとに種類モデルを定義



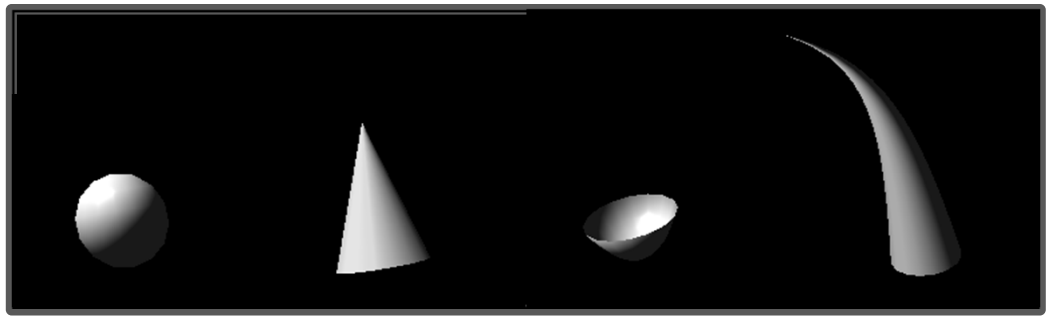
ワーム



トロイの木馬

機能モデルの定義

- 機能ごとに異なる機能モデルを定義
  - 機能群
    - 情報収集:楕円体
    - 感染行動:円錐
    - 破壊活動:お椀型
    - 外部への動作:つの型
  - 機能群に含まれる機能
    - 色分け



情報収集　感染行動　破壊活動　外部への動作

例：マルウェアの可視化

- 検体名
  - **Trojan.Desktophijack.C（トロイの木馬）**
  - **W32.Beagle.CX@mm　（ワーム）**
  - **W32.Beagle.DA@mm　（ワーム）**

} 亜種

抽出した機能特性(1/2)

種類		
トロイの木馬		
機能群	機能	機能対象数
感染行動	ファイル・レジストリ・DLLの作成	3
感染行動	ファイル・レジストリ・DLLの作成	3
感染行動	ファイル・レジストリ・DLLの書き換え	5
感染行動	ファイル・レジストリ・DLLの作成	3
感染行動	ファイル・レジストリ・DLLの書き換え	4
感染行動	ファイル・レジストリ・DLLの書き換え	1
感染行動	ファイル・レジストリ・DLLの書き換え	4
感染行動	ファイル・レジストリ・DLLの書き換え	5
感染行動	ファイル・レジストリ・DLLの書き換え	1
感染行動	ファイル・レジストリ・DLLの書き換え	5
感染行動	ファイル・レジストリ・DLLの書き換え	15
外部への動作	ファイルのダウンロード	2
感染行動	ファイル・レジストリ・DLLの作成	1

Trojan.Desktophijack.C

抽出した機能特性(2/2)

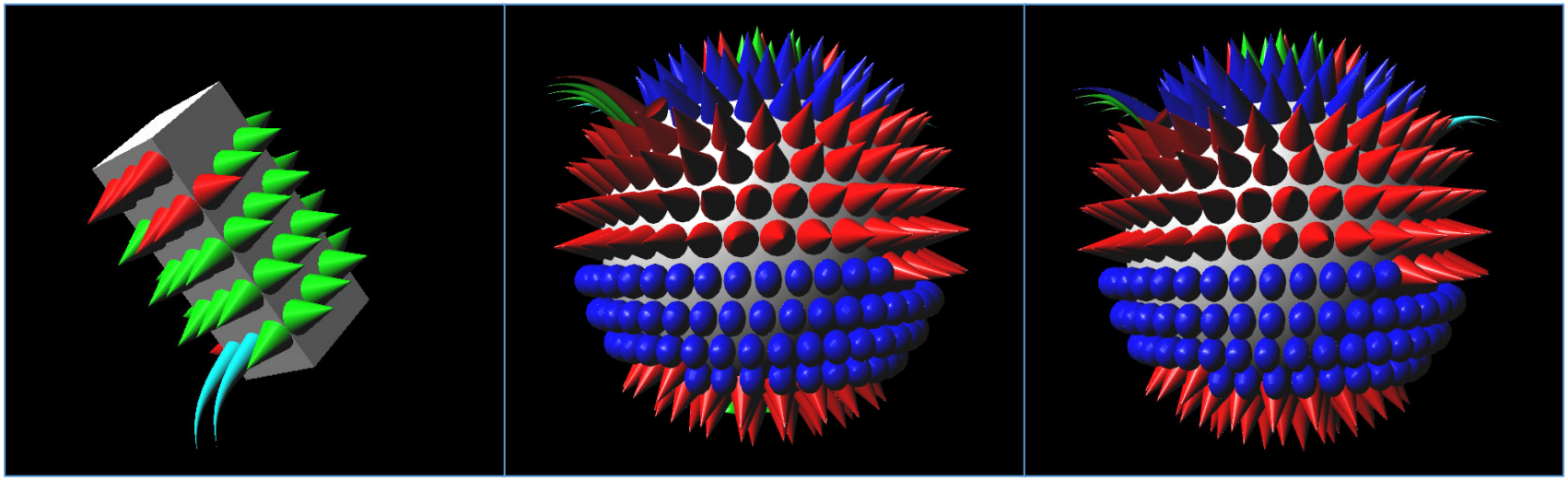
種類		
ワーム		
機能群	機能	機能対象数
感染行動	ファイル・レジストリ・DLLの作成	1
感染行動	ファイル・レジストリ・DLLの書き換え	6
感染行動	ファイル・レジストリ・DLLの作成	8
感染行動	ファイル・レジストリ・DLLの削除	40
感染行動	ファイル・レジストリ・DLLの削除	3
破壊活動	バックドアの作成	1
外部への動作	サーバへのログイン	1
外部への動作	Eメールの送信	3
外部への動作	ファイルのダウンロード	15
感染行動	ファイル・レジストリ・DLLの作成	130
情報収集	特定箇所の文字列収集	1
情報収集	特定箇所の文字列収集	121
感染行動	ファイル・レジストリ・DLLの作成	1
感染行動	ファイル・レジストリ・DLLの作成	46
破壊活動	システムの破壊	3

W32.BeagleCX.@mm

種類		
ワーム		
機能群	機能	機能対象数
感染行動	ファイル・レジストリ・DLLの作成	1
感染行動	ファイル・レジストリ・DLLの書き換え	5
感染行動	ファイル・レジストリ・DLLの作成	8
感染行動	ファイル・レジストリ・DLLの削除	41
感染行動	ファイル・レジストリ・DLLの削除	2
破壊活動	バックドアの作成	1
外部への動作	サイトへのアクセス	1
外部への動作	Eメールの送信	3
外部への動作	ファイルのダウンロード	14
感染行動	ファイル・レジストリ・DLLの作成	131
情報収集	特定箇所の文字列収集	1
情報収集	特定箇所の文字列収集	121
感染行動	ファイル・レジストリ・DLLの作成	1
感染行動	ファイル・レジストリ・DLLの作成	46
破壊活動	システムの破壊	3

W32.BeagleDA.@mm

可視化モデル



Trojan.Desktophijack.C　　W32.Beagle.CX@mm　　W32.Beagle.DA@mm

研究目的・提案手法

研究目的

解析結果により、マルウェアの機能を推定できるか

提案手法

検体間の類似度に基づいた機能推定

マルウェアの類似度導出

類似度を用いることでマルウェアの相関関係を数値化

マルウェアの機能推定

類似度に基づき機能毎にポイントを付与し機能推定

マルウェア解析手法

- 表層解析

  解析対象のプログラムを実行せず、表層的な特徴をもとに過去の解析データなどを調査する方法

- 静的解析

  マルウェアのバイナリファイルを逆アセンブリし、マルウェアの特徴を解析する方法

- 動的解析

  マルウェアを実際に動作させ、その挙動を監視して解析する方法

先行研究

関連する研究

大久保らの研究[2]

静的解析結果（マルウェア検体からバイトコードを抽出）から，
LCSやN-gramを用いて類似度導出・機能推定を行う

本研究

マルウェアの動的解析結果を用いて，
大久保らが提案した機能推定方法の有効性を検討する

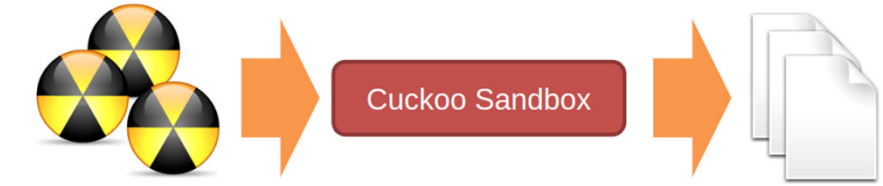
[2] 大久保諒，伊沢亮一，森井昌克，井上大介，中尾康二: マルウェアの類似度に基づく機能推定，CSS2013，2013.

データセット

FFRI Dataset

株式会社FFRIが独自に収集したマルウェアの動的解析結果

Cuckoo Sandbox上でマルウェアを実行



引用[3]

MWS Dataset 2020[3]にてFFRI Dataset 2013 ~ 2020が提供

➡ 本研究ではFFRI Dataset 2016, 2017を使用

[3] 寺田真敏，他: マルウェア対策のための研究用データセット MWS Datasets ～コミュニティへの貢献とその課題～，情報処理学会，Vol.2020-IFAT-139 No.8，2020 年 7 月.

データセット

実験に使用する検体

各ファミリから解析対象検体5検体，解析済み検体30検体

合計解析対象検体45検体，解析済み検体270検体

ファミリ名
TROJ_FORUCON.BMC
BKDR_FYNLOS.SMM
PE_PARITE.A
BKDR_BLADABI.SMC
TROJ_AGENT_006376.TOMB
PE_URSNIF.B-O
TROJ_KRYPTK.SMS
BKDR_VAWTRAK_GB280010.UVPM
RANSOM_CRYPNAN_GA250444.UVPM

類似度導出までの手順

1. 検体情報からAPIコールを抽出
2. 以下の2つの方法で解析対象検体と解析済み検体との類似度（Jaccard係数）を導出
    a. APIコール列のN-gram（ N連接）の集合
    b. APIコール列のLCS（Longest Common Subsequence）の長さ

検体情報


```
{
    “virustotal”:{
        “scans”:{
            “TrendMicro”:{
                “result”:”TROJ_GEN.R00JC0DBL17”,
            }
        }
        “sha1”:”1e9c8dfc…”
    }
    “behavior”:{
        “processes”:{
            {…}
            {
                “calls”:[
                    { “api”:”CoInitializeEx” },
                    { “api”:”CreateDirectoryW”},
                    …
                ]
            }
        “summary”:{
            “directory_created”:[],
            “dll_loaded”:[]
        }
    }
    …
}
```


TrendMicroによる命名

マルウェアのSHA1ハッシュ値

マルウェアが呼び出すAPIコール

マルウェアの機能概要

実験結果

ファミリ	SHA1	N-gram:FP	N-gram:TP	LCS:FP	LCS:TP
TROJ¥_ FORUCON.BMC	0d5d903*	0.0714	1.0000	0.0714	1.0000
	e709a42*	0.2143	1.0000	0.2143	1.0000
	e6e0f81*	0.0714	0.8667	0.0667	0.9333
	f27c0f3*	0.1053	0.7727	0.1500	0.7727
	bb27b1f*	0.0000	0.7368	0.0000	0.7368
BKDR¥_ FYNLOS.SMM	e8f04e2*	0.4286	1.0000	0.4286	1.0000
	8f985c3*	0.2143	0.9167	0.2143	0.9167
	933a537*	0.2857	0.9091	0.3333	0.9091
	d52f74e*	0.0714	1.0000	0.0714	1.0000
	958969f*	0.2143	1.0000	0.2143	1.0000
PE¥_ PARITE.A	96c7bff*	0.0000	1.0000	0.0000	1.0000
	78c0c01*	0.0000	1.0000	0.0000	1.0000
	cbddedf*	0.0000	1.0000	0.0000	1.0000
	ed60e47*	0.0000	1.0000	0.0000	1.0000
	c54a317*	0.0000	1.0000	0.0000	1.0000
BKDR¥_ BLADABI.SMC	a851047*	0.0526	1.0000	0.0526	1.0000
	97bd321*	0.0526	1.0000	0.0526	1.0000
	6d80bff*	0.1579	1.0000	0.1579	1.0000
	e8b328b*	0.1053	1.0000	0.1053	1.0000
	4867967*	0.1053	1.0000	0.1053	1.0000
TROJ¥_ AGENT¥_ 006376.TOMB	276c1d2*	0.0000	1.0000	0.0000	1.0000
	488c3af*	0.0000	1.0000	0.0000	1.0000
	800e5f9*	0.0000	1.0000	0.0000	1.0000
	c8918c9*	0.0000	1.0000	0.0000	1.0000
	b5148f3*	0.0000	1.0000	0.0000	1.0000

実験結果

ファミリ	SHA1	N-gram:FP	N-gram:TP	LCS:FP	LCS:TP
PE¥_ URSNIF.B-O	9deb785*	0.0000	0.9375	0.0000	0.9375
	0878a0f*	0.0667	0.9333	0.0667	0.9333
	8e85723*	0.0000	1.0000	0.0000	1.0000
	3ad8b9b*	0.0000	1.0000	0.0000	1.0000
	c21563f*	0.0000	1.0000	0.0000	1.0000
TROJ¥_ KRYPTK.SMS	9cc9c76*	0.0000	1.0000	0.0000	1.0000
	4036ffe*	0.0000	1.0000	0.0000	1.0000
	d39ebc5*	0.0000	1.0000	0.0000	1.0000
	2c7996e*	0.0000	1.0000	0.0000	1.0000
	e09f4ca*	0.0000	1.0000	0.0000	1.0000
BKDR¥_ VAWTRAK¥_ GB280010.UVPM	cf3185a*	0.0000	1.0000	0.0000	1.0000
	b7bdb31*	0.0000	1.0000	0.0455	1.0000
	b6b1a89*	0.0000	1.0000	0.0455	1.0000
	e21998a*	0.2381	1.0000	0.2727	1.0000
	65e216e*	0.2381	1.0000	0.2381	1.0000
RANSOM¥_ CRYPNAN¥_ GA250444.UVPM	a5f56b9*	0.0000	1.0000	0.0000	1.0000
	2f9f5dd*	0.0000	1.0000	0.0000	1.0000
	6750ebe*	0.0000	1.0000	0.0000	1.0000
	18261c5*	0.0000	1.0000	0.0000	1.0000
	771982f*	0.0000	1.0000	0.0000	1.0000
AVERAGE	/	0.0598	0.9794	0.0646	0.9809

動的解析を使った分類に関する研究

- 研究目的
  - マルウェアがどのマルウェアファミリに属するかを分類
- 方法
  - 動的解析結果から得られるAPIコール列＋LZW圧縮アルゴリズムの辞書
  - SVMによる二値分類
- 結果
  - 分類精度の平均値が83.80%

データセット

FFRI Dataset

FFRI社が収集したマルウェアの動的解析結果

[5]

使用するデータセット

- FFRI Dataset 2013 ~ 2016 (JSON形式)
- Kasperskyの命名規則に則り、APIコールが取得されている検体
- 亜種の数が100検体以上あるファミリ

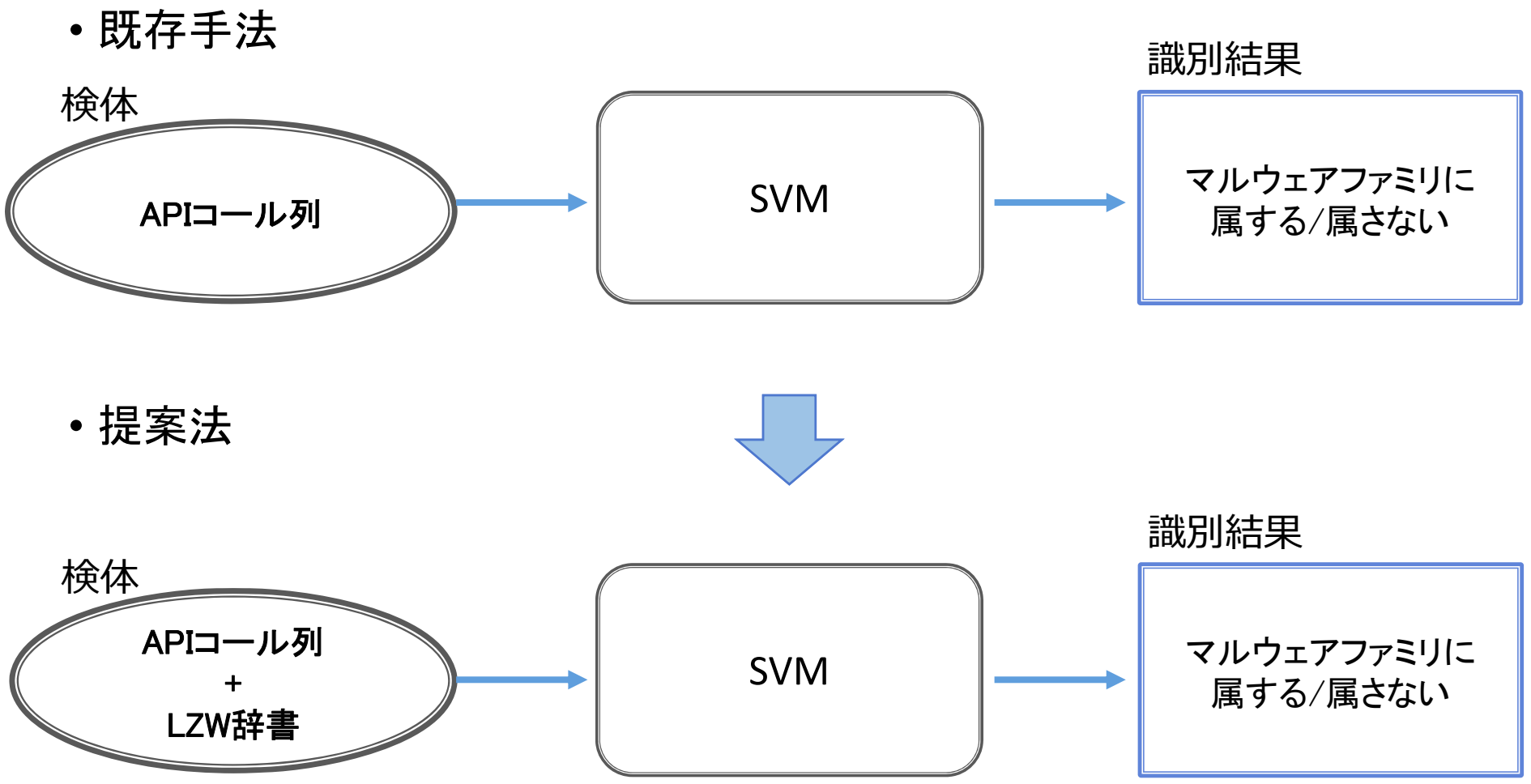
[5]寺田真敏，秋山満昭，松木隆宏ほか：マルウェア対策のための研究用データセットMWS Datasets ～コミュニティへの貢献とその課題～，情報処理学会，Vol.2020-IFAT-139 No.8 (2020).

条件を満たすマルウェアファミリ

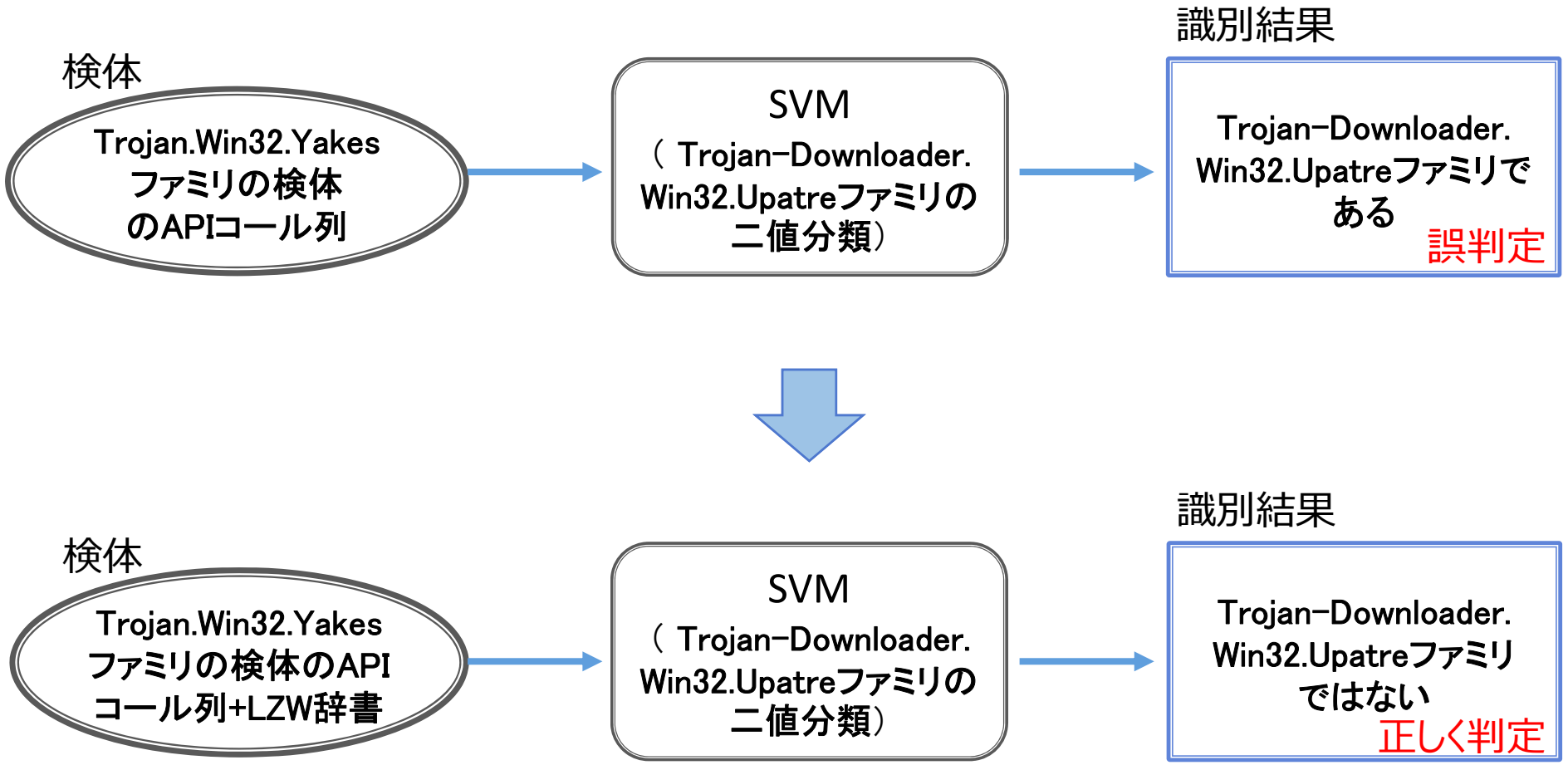
No.	ファミリ名
1	Trojan.Win32.Waldek
2	Trojan.Win32.Agent
3	Trojan-Ransom.Win32.Foreign
4	Trojan-Dropper.Win32.Injector
5	Trojan.Win32.Yakes
6	Trojan.Win32.Llac
7	Backdoor.Win32.Matsnu
8	Trojan-PSW.Win32.Tepfer
9	Trojan.Win32.Jorik
10	Backdoor.Win32.Androm
11	Worm.Win32.WBNA

No.	ファミリ名
12	Trojan-PSW.Win32.Fareit
13	Trojan-Downloader.Win32.Upatre
14	Backdoor.Win32.DarkKomet
15	Trojan.Win32.Scar
16	Packed.Win32.Tpyn
17	Trojan-Spy.Win32.Zbot
18	Worm.Win32.Vobfus
19	Hoax.Win32.ArchSMS
20	Trojan.Win32.Kovter
21	Downloader.Win32.LMN
22	Trojan.Win32.Inject

既存手法と提案手法


• 既存手法
検体
APIコール列
SVM
識別結果
マルウェアファミリに
属する/属さない
• 提案法
検体
APIコール列
+
LZW辞書
SVM
識別結果
マルウェアファミリに
属する/属さない

亜種推定が誤判定をする場合



Trojan-Downloader.Win32の検体の例

■ Trojan.Win32.Yakes ファミリの検体のAPIコール列

0 1 0 0 0 0 15 15 15 1 16 16 10 10 10 10 10 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 11 5 11 5 5 10 10 11 17 18 19 20 5 10 20 5 10 20 5 10 2 20 0 1 0 6 21 22 9 23 23 23 23 10 23 2 2 2 2 2 17 10 23 23

Trojan.Win32.Yakes.olac

■ Trojan-Downloader.Win32.Upatre ファミリの検体の APIコール列

0 1 0 2 2 2 2 2 2 2 2 2 2 2 0 1 3 4 5 6 7 8 9 10 10 11 5 5 12 13 14

Trojan-Downloader.Win32.Upatre.fopg

良い影響を与える検体間の比較

◘ APIコール+LZW辞書情報

0 1 0 0 0 0 15 15 15 1 16 16 10 10 10 10 10 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 2 2 2 2 2 17 10 11 5 11 5 5 10 10 11 17 18 19 20 5 10 20 5 10 20 5 10 2 20 0 1 0 6 21 22 9 23 23 23 23 10 23 2 2 2 2 2 17 10 23 23 0 1 1 0 0 0 0 0 0 0 15 15 15 15 15 1 1 16 16 16 16 10 10 10 10 10 10 10 10 10 2 2 2 2 2 2 2 2 17 17 10 10 2 2 2 2 2 2 2 17 10 10 2 2 2 2 2 2 2 17 17 10 2 2 2 2 2 2 2 17 17 10 2 2 2 2 2 2 17 10 10 2 2 2 2 2 2 17 17 10 2 2 2 2 2 2 17 10 10 11 11 5 5 11 11 5 5 5 10 10 10 11 11 17 17 18 18 19 19 20 20 5 5 10 20 20 5 10 10 20 20 5 10 2 2 20 20 0 0 1 0 0 6 6 21 21 22 22 9 9 23 23 23 23 23 23 23 10 10 23 23 2 2 2 2 2 2 17 17 10 23

Trojan.Win32.Yakes.olac

0 1 0 2 2 2 2 2 2 2 2 2 2 2 0 1 3 4 5 6 7 8 9 10 10 11 5 5 12 13 14 0 1 1 0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 1 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 10 10 10 11 11 5 5 5 5 12 12 13 13 14

Trojan-Downloader.Win32.Upatre.fopg

亜種推定の結果 (Accuracy)

No.	既存手法	提案手法	No.	既存手法	提案手法
1	91.59%	**93.68%**	13	76.79%	**80.01%**
2	72.02%	**73.30%**	14	81.25%	**82.18%**
3	86.31%	83.08%	15	75.00%	72.49%
4	70.24%	**83.33%**	16	79.76%	**84.93%**
5	81.25%	79.16%	17	74.80%	**78.16%**
6	76.19%	**81.55%**	18	94.05%	91.06%
7	None	**85.76%**	19	94.84%	**95.21%**
8	75.60%	**76.06%**	20	90.48%	**90.66%**
9	90.48%	**91.07%**	21	89.29%	86.25%
10	79.88%	75.91%	22	69.64%	**88.24%**
11	91.87%	90.92%	avg.	**81.94%**	**83.80%**
12	79.37%	**80.56%**	max.	**94.84%**	**95.21%**
			min.	**69.64%**	**72.49%**

まとめ

- マルウェアの可視化

  ⇒　亜種の機能的類似性の直感的理解

- マルウェアの動的解析結果を使った機能推定

  ⇒　機能推定のTP値0.98, FP値0.06

- マルウェアの動的解析結果を使った亜種分類

  ⇒　亜種分類精度の平均値83.80%