ウイルスの時間的な関連性に注目した見える化

小櫻 文彦　津田 宏　鳥居 悟

株式会社富士通研究所　ソフトウェア＆ソリューション研究所

〒211-8588 神奈川県川崎市中原区上小田中 4-1-1

E-mail:　{kozac, htsuda, torii.satoru}@jp.fujitsu.com

あらまし　マルウェア研究用データセットCCC DATAset 2009の攻撃元データの分析において、同一ハニーポットが一定時間内に連続してダウンロードするウイルスの関連に注目し、それを連鎖感染とみなし見える化した。これにより、年や月単位でのウイルスの連鎖感染の傾向が明らかになった。

キーワード　マルウェア，ボット，ハニーポット，ログ，連鎖感染

Visualization of malware attacks based on their sequences

Fumihiko KOZAKURA, Hiroshi TSUDA, and Satoru TORII

Software and Solution Laboratories, Fujitsu Laboratories LTD.

4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, 211-8588 Japan

E-mail:　{kozac, htsuda, torii.satoru}@jp.fujitsu.com

Abstract　From the attack host log of CCC DATAset 2009, we investigate the sequential relations among malwares and visualize the relations to clarify the trends of chain infections.

Keyword　malware，bot，honeypot, log, infection chain

1. はじめに

インターネット上での不正アクセス活動は活性化するとともに、マルウェアの挙動も巧妙になっている。ダウンローダーを介して複数のマルウェアを連鎖的にダウンロードされるものもある[1]。こうした多様な不正アクセス活動に対して、その動向を知り、それに基づいた対策が求められている。

本稿では、サイバークリーンセンター(https://www.ccc.go.jp/) が収集した研究用データセットCCC DATAset 2009[2] の攻撃元データ(以降、CCC2009攻撃元データ)において、ウイルスの時間的な関連性に着目した分析結果を報告する。ハニーポットにウイルスがダウンロードされ一定時間内に同一ハニーポットにダウンロードされたウイルスに連鎖感染の関連があると仮定し、そのウイルス間の関連を見える化することで攻撃や対策の動向を分析する。

2. 関連研究

松木[3]は研究用データセット CCC DATAset2008 の 2 日間の攻撃通信データおよび攻撃元データから、連鎖感染と考えられる関連性の可視化を行い、その傾向を論じている。

3. ウイルスの時間的な関連性の分析

今回、CCC2009 攻撃元データには、ハニーポットを識別できる識別子が含まれ 1 ハニーポットにおけるウイルスのダウンロード状態の把握が可能となったことを受け、我々はウイルスの時間的な関連性に着目した分析を行った。

ここで時間的な関連性として、あるウイルスがダウンロードされてから一定期間内にダウンロードされたウイルス(複

数の場合もある)は関連性があるとする。時間的関連性は全て連鎖感染とは限らないが、大量の関連の中から全体の傾向を見える化するため、多少のランダムな誤りはあったとしても問題ない。その見える化の特徴を利用してウイルス間の関連を把握できるようになれば、ウイルスの特徴が把握できるのではないかと考えた。

3.1. CCC2009 攻撃元データの概要と特徴

CCC2009 攻撃元データの総件数は2,470,766件でありその特徴について以下に記載する。

(1) 月別攻撃回数

2008 年 5 月	287,954 件
2008 年 6 月	298,220 件
2008 年 7 月	283,305 件
2008 年 8 月	258,284 件
2008 年 9 月	224,711 件
2008 年 10 月	223,775 件
2008 年 11 月	167,812 件
2008 年 12 月	151,471 件
2009 年 1 月	189,168 件
2009 年 2 月	182,319 件
2009 年 3 月	116,742 件
2009 年 4 月	87,005 件

(2) ウイルス出現回数上位 10

1:PE_VIRUT.AV	222,207 件
2:未知検体群	139,310 件
3:PE_BOBAX.AK	94,325 件
4:TSPY_KOLABC.CH	93,005 件
5:PE_VIRUT.D-1	70,618 件
6:BKDR_VANBOT.AHH	70,028 件
7:BKDR_SDBOT.BU	50,059 件
8:PE_VIRUT.D-4	45,615 件
9:BKDR_SCRYPT.ZHB	40,117 件
10: WORM_SWTYMLAI.CD	38,768 件

月別攻撃数の変化を見ると、ハニーポットへの攻撃が減少していることが伺える。全体としては減少傾向であるものの、PE_VIRUT.AV のように毎月目立った動きをするウイルスや WORM_SWTYMLAI.CD のように限られた一定期間しか生存が確認できていないウイルスが混在している。

3.2. CCC2009 攻撃通信データの分析

CCC2009 攻撃元データとは別に、期間が 2009 年 3 月の 2 日間と限られるがより詳細な情報が CCC2009 攻撃通信データとして公開されている。そこで連鎖感染の関連の確認を CCC2009 攻撃通信データで行う。CCC2009 攻撃通信データのビュアーとして富士通研究所で開発したツール[4]を使用した。

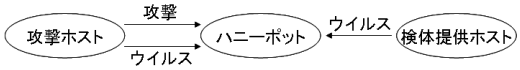


図 1　ハニーポットとホストの関係

図 1 はハニーポットとホストとの関係を表したものである。ホストにはハニーポットに攻撃をしてウイルスを提供する攻撃ホストとウイルスだけを提供する検体提供ホストがある。そこで CCC2009 攻撃通信データで出現しているウイルスについて攻撃ホストから提供されているウイルスなのか検体提供ホストから提供されているウイルスかの以下のように分類した。

(1) 攻撃ホストからのみダウンロード
- BKDR_RBOT.ASA
- PE_VIRUT.AV
- BKDR_POEBOT.GN
- BKDR_MYBOT.AH
- PE_BOBAX.AK
- その他

(2) 検体提供ホストからのみダウンロード
- WORM_SWTYMLAI.CD
- TROJ_BUZUS.AGB
- TROJ_DLOADR.CBK
- TROJ_AGENT.ANDF
- その他

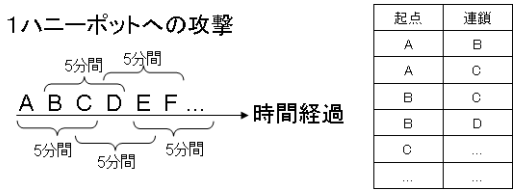
この分類を連鎖感染の視点で見ると(1)は起点になる可能性が大きく、(2)は連鎖である可能性が大きい。そこで(2)のウイルスより前に、(1)のウイルスが存在するか手動で調べた。

その結果、WORM_SWTYMLAI.CD や TROJ_BUZUS.AGB の前には、

PE_VIRUT.AV が出現しており、TROJ_DLOADR.CBK や TROJ_AGENT.ANDF の前には PE_BOBAX.AK が起点となっていることが分かった。また、(2)のウイルスで 1 つを除き連鎖感染における起点と連鎖の関連を確認できた。

3.3. ウイルスの時間的な関連性の定義

一定期間内に出現したウイルスは連鎖感染の関連があると仮定したものの、あまり長い期間を取ると、複数の連鎖が絡む可能性も高くなり、またハニーポットが定期的にリセットされるため長期間の監視には向かない。そこで今回は時間的な関連として、図 2 のようにあるウイルスを起点とし 5 分以内に出現したウイルス(複数の場合もある)を連鎖として定義する。なお、5 分以内とは、3.2 の攻撃通信データの分析でも確認した期間であるが、汎用的に意味のある数字ではないことを断っておく。



起点	連鎖
A	B
A	C
B	C
B	D
C	…
…	…

図 2 ウイルスの時間的関連

3.4. 月別データによる見える化

CCC2009 攻撃元データの 3 月分のデータにおけるウイルスの関連を見てみよう。可視化ツールとしては、富士通研究所で開発した関係メータデータの分析システムであるビジネス情報ナビゲーター[5][6]を使用した。ビジネス情報ナビゲーターは、RDF(Resource Description Framework)[7]形式で記述されたグラフ関係によるメタデータの視覚化が可能で、人脈検索や顧客関係の分析に用いられている。

図 3 は CCC2009 攻撃元データの 3 月分のデータから関連を求め上位 15 関連を見える化したグラフである。青色のノードが連鎖感染の起点、緑色のノードが連鎖感染の連鎖と色分けし、起点から連鎖の方向には矢印を付けている。また、1 つのウイルスが起点と連鎖の両方に現れる場合もある。

ここから CCC2009 攻撃通信データの分析でも確認したように、PE_VIRUT.AV を起点として PE_VIRUT.AV → WORM_AUTORUN.CZU → WORM_SWTYMLAI.CD という連鎖や、さらに TROJ_BUZUS.AGB への連鎖が確認できる。また、PE_BOBAX.AK → TROJ_DLOADR.CBK という連鎖も確認できる。

また、CCC のホームページで公開されている「2009 年 03 月度 サイバークリーンセンター 活動実績」[8]の上位 5 検体の動向観測のグラフからは、2 位の WORM_MAINBOT.AH が増えると 4 位の WORM_AUTORUN.CZU も増えていることが読み取れる。この 2 つのウイルスの関連も、WORM_MAINBOT.AH → WORM_AUTORUN.CZU という連鎖を見ることができる。このように検体の検出数からは見えてこない関連を得られた。

図 4 は図 3 と同様に CCC2009 攻撃元データの 4 月分のデータから関連を求め上位 15 関連を見える化したグラフである。3 月と比べると、WORM_AUTORUN.CZU には、大きく 2 つ(PE_VIRUT.AV および WORM_MAINBOT.AH)から連鎖しているのは同じであるが、その先 WORM_SWTYMLAI.CD がなくなり未知検体(数値ノード)になっているのが大きく異なる。

この図はまた、「2009 年 04 月度 サイバークリーンセンター 活動実績」[9]の上位 5 検体の動向観測の報告で WORM_AUTORUN.CZU と未知検体が同一サイトから大量配布されている、とあるのと一致する。

3.5. 全期間による見える化

図 5 は全期間の攻撃元データから関連を求め上位 15 ウイルスの関連を見える化したものである。データ量が多くなるため月々のデータの見える化より単純で分かり易い関連を見ることができる。

ここから大別して 4 つの連鎖グループ

に分類されているだけではなく、TSPY_KOLABC.CH を連鎖感染の起点として WORM_SWTYMLAI.CD と BKDR_POEBOT.GN に連鎖の関連があることがわかる。この関連はトレンドマイクのホームページの TSPY_KOLABC.CH の詳細情報 [10] に、WORM_SWTYMLAI.CD と BKDR_POEBOT.GN を生成すると書かれている通り現実を表現している。

3.6. ウイルスに注目した見える化

ウイルス出現回数上位 10（2 位の未知検体群を除く）のウイルスについて、連鎖感染において、起点と連鎖のどちらになりやすいかも見える化から区別ができる。起点としてダウンロードされるウイルスは図 6 の PE_VIRUT.AV のように、PE_VIRUT.AV を中心として放射線上に連鎖のノード群が広がる形になる。逆に連鎖としてダウンロードされるウイルスは連鎖感染の連鎖であるウイルスを中心として放射線上に起点のノードが広がる。また、その両方を持っているウイルスはそれらをマージしたようなグラフとなる。それを基にウイルス出現回数上位 10 分類した結果を以下に記載する。

1:PE_VIRUT.AV	起点
2:未知検体群	—
3:PE_BOBAX.AK	起点
4:TSPY_KOLABC.CH	連鎖
5:PE_VIRUT.D-1	起点
6:BKDR_VANBOT.AHH	起点
7:BKDR_SDBOT.BU	連鎖
8:PE_VIRUT.D-4	起点
9:BKDR_SCRYPT.ZHB	起点/連鎖
10: WORM_SWTYMLAI.CD	連鎖

この分類によりウイルスの連鎖感染における特徴を得ることができる。起点は PE ファイル感染型が多く松木[3]と同じ結果である。また、両方に分類されるウイルスも存在し時期により分類が変化することがある。

4. まとめ

本稿では、時間的な関連性に注目してウイルスの連鎖感染の見える化を行った。実際の関連の情報がないにも関わらず、複数の視点から見える化することで以下の動向を得られることがわかった。

(1) ウイルスのグループ化と連鎖感染の流れ
(2) 連鎖感染における、起点/連鎖の判定
(3) 全期間/月別の連鎖動向の違い

攻撃元データに連鎖感染の起点判定のために攻撃をされた形跡の情報があるとより正確な時系列の分析ができると考えている。今後、有用な情報が増えることを期待する。

参考文献

[1] “情報セキュリティ白書 2008”, 独立行政法人情報処理推進機構,2008.6.

[2] 畑田,他, “マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有”,MWS2009,2009.10

[3] 松木, “時系列分析による連鎖感染の可視化と検体種別の推測”, MWS2008, 2008.10

[4] 東角,鳥居, “TCP セッションの特徴に基づくボット制御通信の検知方式の検討”, MWS2009,2009.10

[5] 松井,津田,片山, ナレッジマネジメントツール:ビジネス情報ナビゲーター, FUJITSU, pp.325-330, Vol.57. No.3, 2006.5

[6] 小櫻,津田,鳥居,”ウイルスのライフサイクルに着目した攻撃挙動の見える化”, MWS2008, 2008.10

[7] RDF (Resource Description Framework), http://www.w3.org/RDF/

[8] 2009 年 03 月度 サイバークリーンセンター活動実績, https://www.ccc.go.jp/report/200903/0903monthly.html

[9] 2009 年 04 月度 サイバークリーンセンター活動実績, https://www.ccc.go.jp/report/200904/0904monthly.html

[10] トレンドマイクロ社,TSPY_KOLABC.CH 詳細情報, http://www.trendmicro.co.jp/vinfo/graywar e/vegraywareDetails.asp?GNAME=TSPY%5FKOLABC%2ECH&VSect=Td

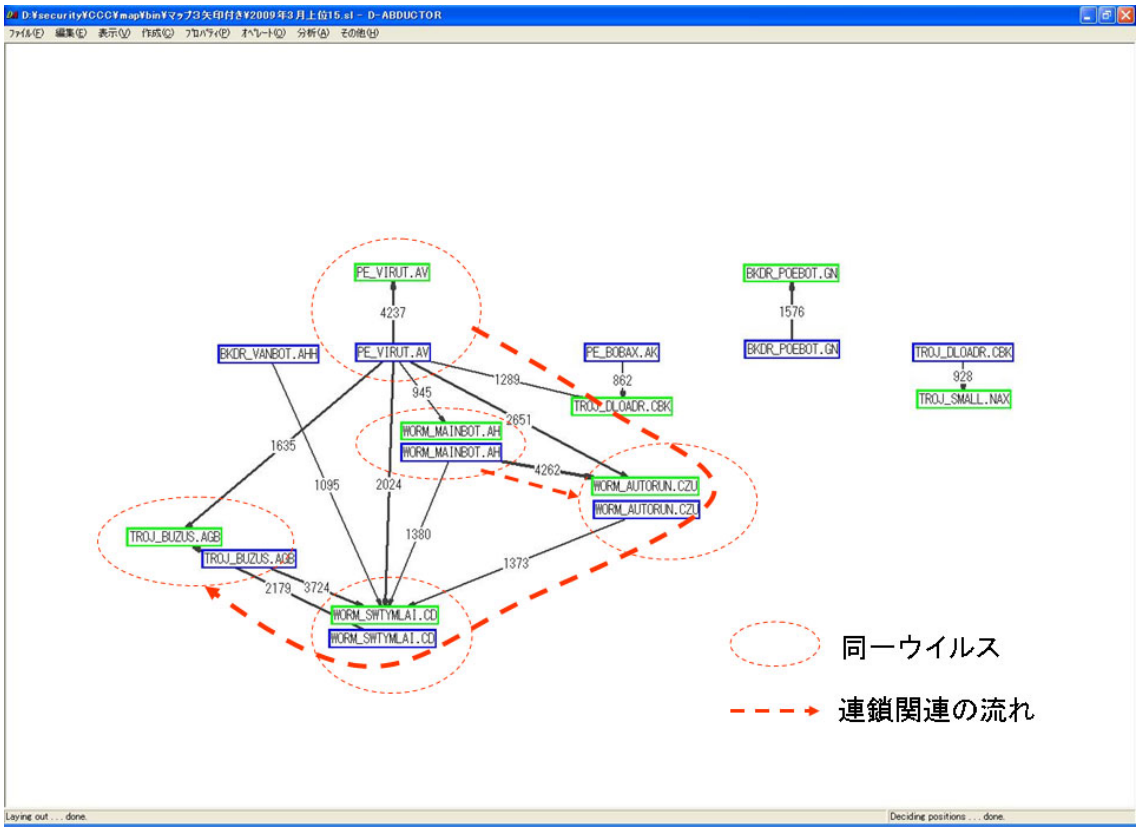


図３　攻撃元データの３月分の上位 15 関連

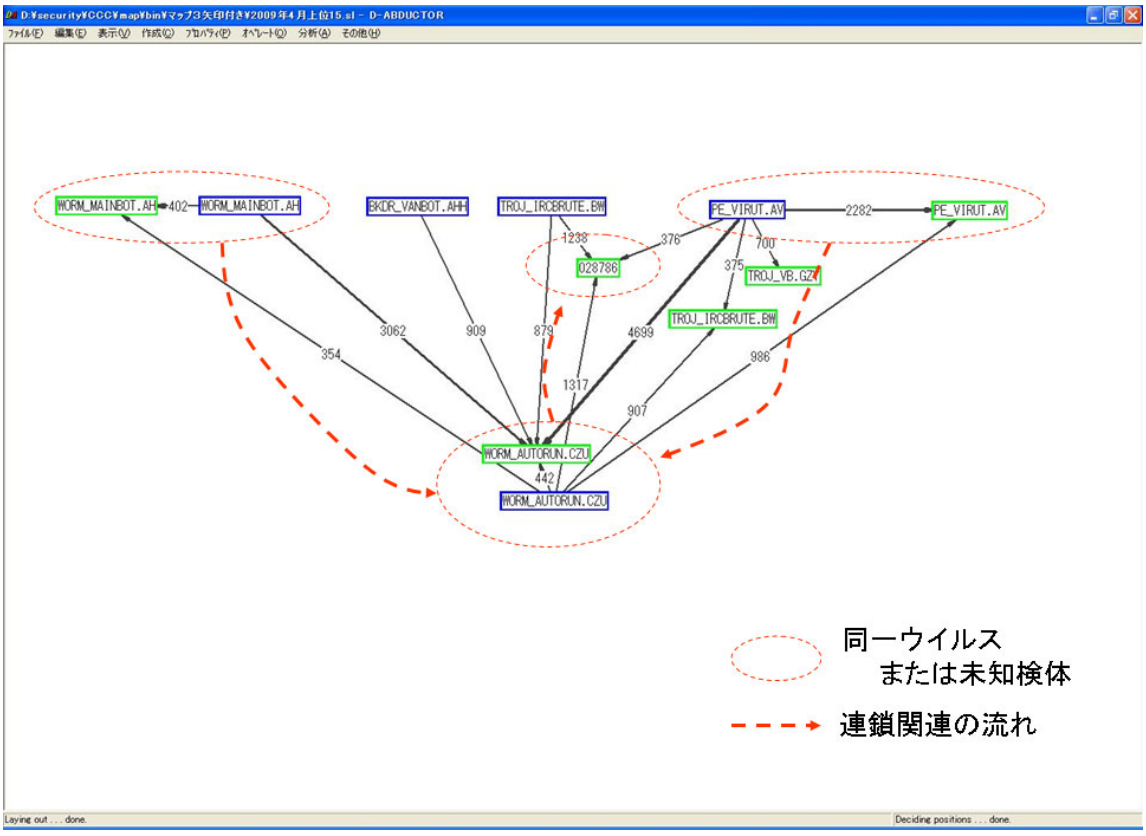


図４　攻撃元データの４月分の上位 15 関連

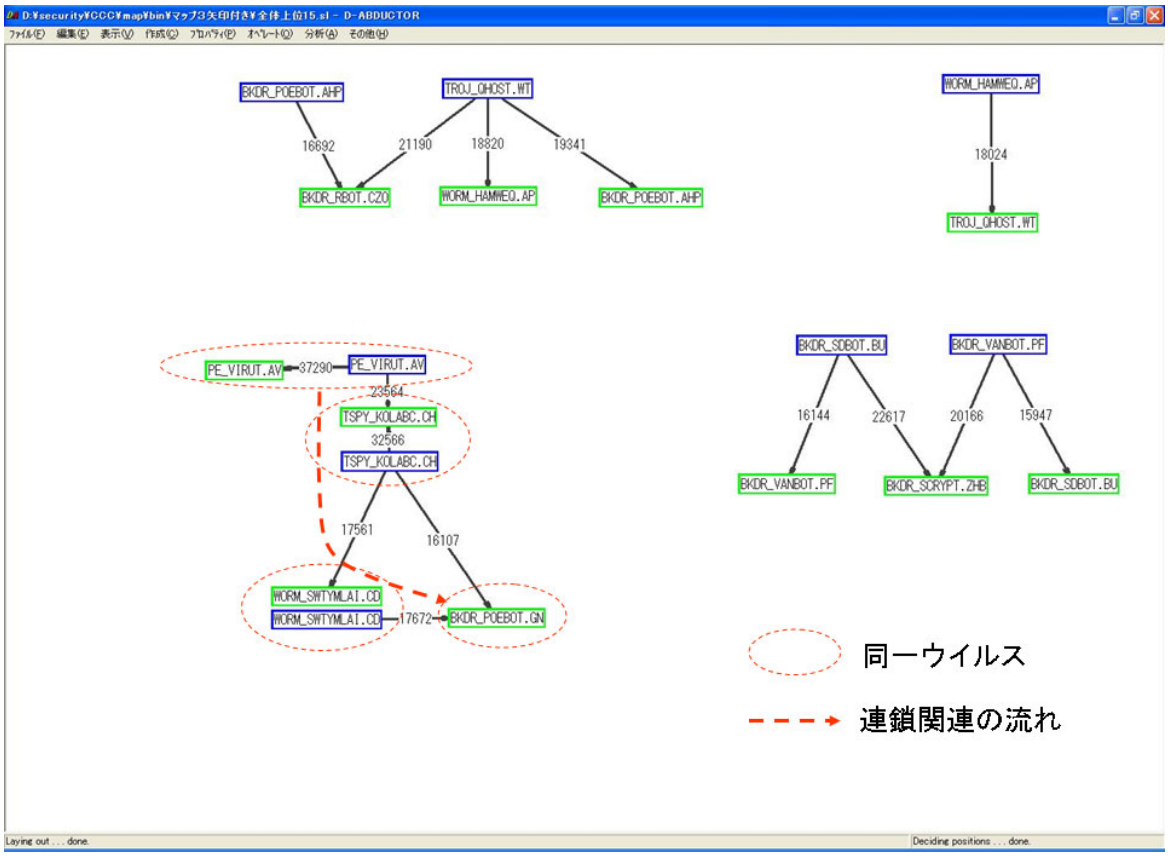


図５　全データ上位 15 関連

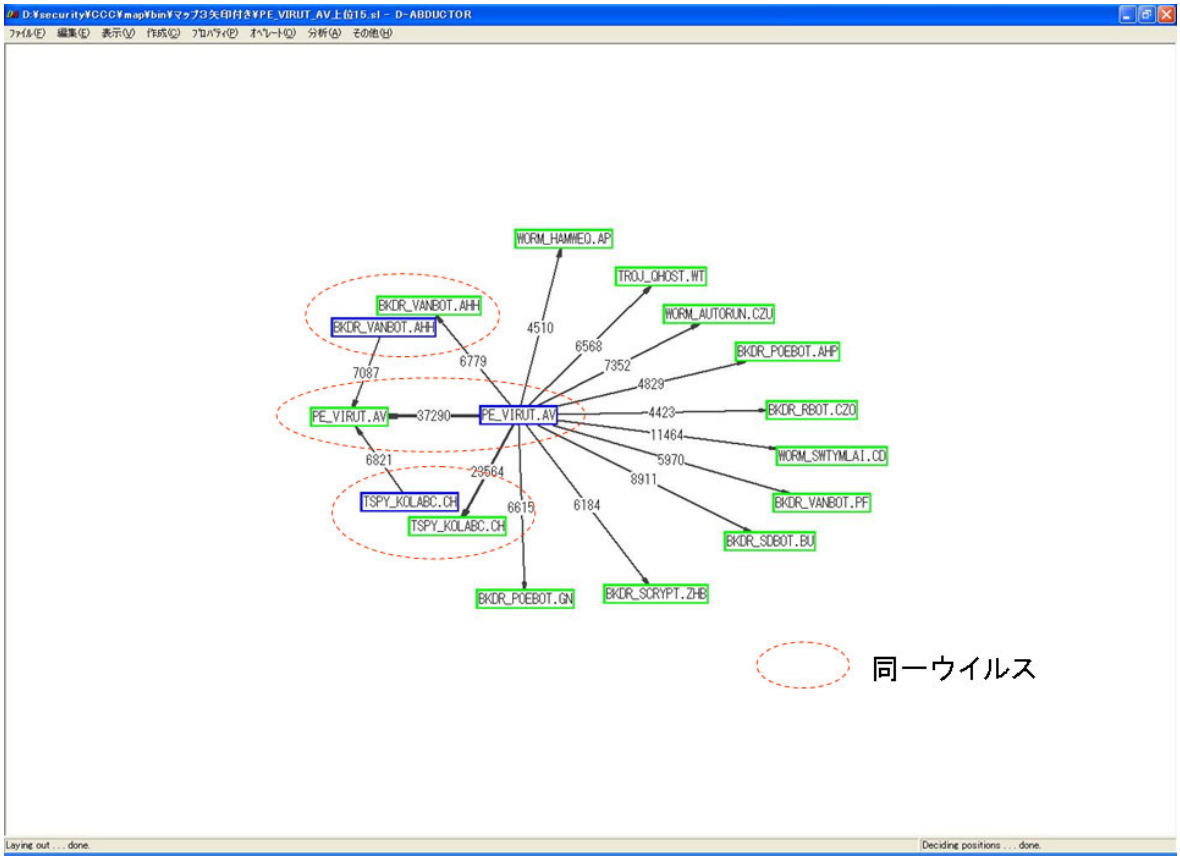


図６　PE_VIRUT.AV 上位 15 関連