{研究用データセット：攻撃元データ編}

ナレッジマネジメントツールによるマルウェア挙動の見える化

小櫻文彦　津田　宏　鳥居　悟（富士通研究所）

CCC DATAset 2008/2009[1] の攻撃元データに対して，筆者らがナレッジマネジメントのツールとして開発した見える化ツール，ビジネス情報ナビゲータ[2] を適用し，複数のマルウェアの時間・空間的な関連性を見える化した事例[3]を紹介する．

見える化とは

「見える化」とは，現場のデータ分析から問題点を明らかにし，業務改善を図るなど，ナレッジマネジメントの文脈でしばしば用いられる言葉である．ここで大切なのは，単に何でも手持ちのデータを視覚化ツールに入れれば何か有益な結果が出るわけではないことである．利用シーンや解決したい問題点を明らかにして，データやツールの特徴を活かした視覚化を行い，それを解釈して改善につなげることが課題である．実際には，データベースやテキスト文書など異なった種類の情報を組み合わせることも少なくない．現場のデータは個々のシステムごとに独立で作られていることも多く，ばらばらなデータの関連付けといったデータの統合も課題となる．

ナレッジマネジメントツール

ナレッジマネジメントツール，ビジネス情報ナビゲータは，富士通研究所が開発した，相関情報の見える化技術である．組織内には用途別に独自に設計されたデータベースや，テキストのような非構造化情報といった異種情報源が数多く存在し，統合活用できていないことが多い．ビジネス情報ナビゲータは，それらの情報源から関係情報をメタデータとして抽出し，統合メタデータベースを介して，検索・マイニングを行い，隠れた関連性を相関マップとして見える化する．

1つの応用例[4] として，組織内での特定のスキルを持った人や人脈を検索する人物情報検索システム（KnowWho）がある．一般に業務上の情報の大半は人から得ると言われていて，スキルのある人を知っていることは重要である．そこで特定のスキルを持った人を検索する場合，結果がリストだけでは情報の把握が難しい．そこで人脈を含め相関マップとして見える化することで，起点となっている人が分かるなど隠れた情報を簡単に得ることができる．さらに人が複数の業務を行うことで発生する分散された情報から，人のメタデータとして自動抽出し統合することで，組織情報からは得られにくい実務から発生した最新の人脈を含め見える化できる．

また，滋賀銀行の適用事例[☆1] では，銀行が所有するさまざまなシステムのデータから，お客様企業のお金や商品の流れなどの関係を抽出・統合し，お客様と取引先の関係など，ビジネス状況全体の相関関係をさまざまな相関マップとして見える化している．これら相関マップにより，営業担当者がお客様の置かれた企業環境と地域のビジネス環境を容易に把握でき，ビジネスマッチングなどの業務支援になっている．

マルウェア分析への適用

図-1 は今回のマルウェア分析にビジネス情報ナビゲータを適用した構成図である．これにより，ビジネス情報ナビゲータの動きを説明しよう．

入り口となる CCC DATAset 2009 攻撃元データは，一旦 RDF（Resource Description Framework）という形式に変換される．RDF は，セマンティック Web の中心となる規格であり，主語（subject）述語（predicate）目的語（object）の三つ組みで，「A さん（主語）の上司（述語）は C さん（目的語）である」というような関係情報を記述する．

ここで主語，述語，目的語として記述される語彙は「オントロジー」という形式であらかじめ定義しておく必要がある．オントロジーとは，メタデータを記述する際の語彙定義であり，クラスの階層や，クラスの取り得る属性，属性間の階層関係などにより定義する（セマンティック Web ではオントロジーを記述する OWL（Web オントロジー言語）という言語がある）．

図-1 に今回筆者らが定義したオントロジーの一部を

☆1 http://pr.fujitsu.com/jp/news/2008/05/7.html

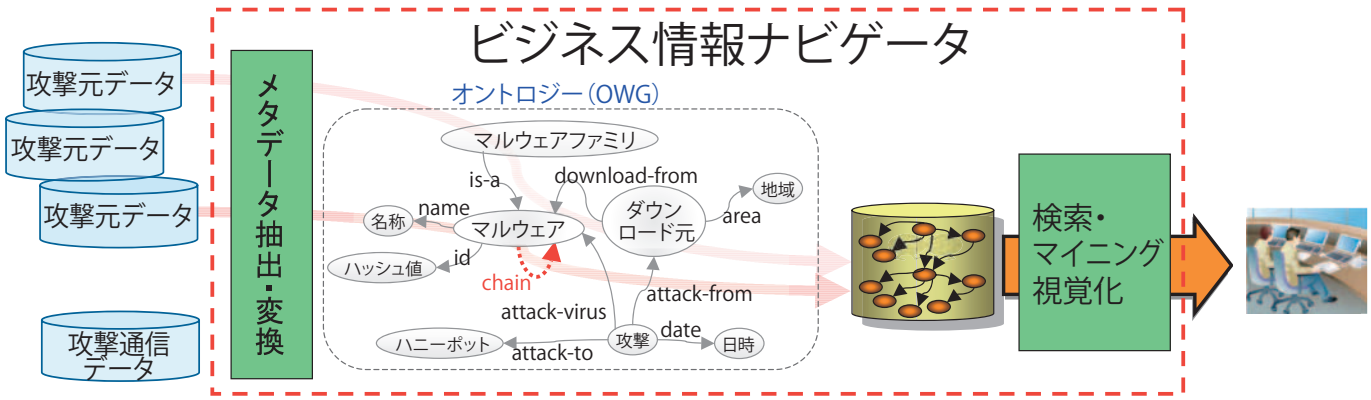

図 -1　CCC DATAset の分析の流れ

示す．楕円で示される「マルウェア」や「攻撃」がクラスで，クラスの間には属性や関係情報が定義されている．たとえば，「マルウェア」の name 属性が「名称」とか，「マルウェア」は「マルウェアファミリ」の一種である（is-a）などである．このように語彙を厳密に定義しておくことで，入り口のデータはばらばらでも RDF のレベルで情報を合わせて分析することができる．

一旦攻撃元データを RDF に変換することで，多様な視点から相関マップを作ることが可能になる．分析者がマルウェアと何の関係を見たらどうなるかなど試行錯誤する場合にも，オントロジーがあれば見通しは良くなる．具体的な例を以降で順次紹介しよう．

攻撃元と地域の見える化のための RDF 変換

マルウェアファミリが，世界のどの地域にどのように感染しているかの動きを見てみよう．図 -1 のオントロジーでは，「マルウェアファミリ」の下位概念（is-a）に「マルウェア」があり，それには「ダウンロード元」（download-from）と，ダウンロード元の area 属性に「地域」が紐付いている．こうした関係をたどれば，マルウェアファミリと地域の関係が見える化できそうである．

ただし，これらの関係がそのまま攻撃元データに含まれているわけではない．見える化をするには，まず元データを加工しながら RDF に自動変換することが必要になる．

マルウェアとマルウェアファミリの is-a 関係を，筆者らは，攻撃元データのマルウェア名称を利用して算出した．マルウェア名称は，たとえば PE_VIRUT.AV というように，マルウェアファミリ名（PE_VIRUT）を元に作られる．そこで文字列の関係から PE_VIRUT.AV というマルウェアは，PE_VIRUT というマルウェアファミリの下位概念（is-a）であると見なした．また，ダウンロード元の地域については，ダウンロード元の IP アドレスから whois サーバの情報を利用して獲得した．

初期の攻撃におけるマルウェアファミリと地域の相関マップ

図 -2 は，CCC DATAset 2008 の攻撃元データから，ダウンロード時には未知検体だったマルウェアに対し，そのマルウェアファミリがどの地域からよくダウンロードされるかの関連を見える化したものである．なおマルウェアの初出の攻撃では，ダウンロード時点で名称は未知（UNKNOWN）であるが，後でアンチウイルスベンダによりマルウェアファミリや名称が決まる．

三角形の頂点に北米（左下），欧州（上），アジア（右下）の 3 つの地域を固定し，関連するマルウェアファミリ（緑枠四角形）を互いに引っ張り合う．こうすると，3 つの地域から攻撃するファミリは真ん中へんに，2 つの地域からのファミリは三角形の辺上に，また地域固有のものは三角形の頂点付近に，それぞれ自動的に配置される．なお，マルウェアファミリのうち黄色く塗られているものは，ダウンロード数が上位のものである．このような相関マップにより，マルウェアファミリごとに初出の攻撃で地域性があることが分かる．

連鎖関係のマイニング

ビジネス情報ナビゲータでは，攻撃元データから，隠れた関係をマイニングして見える化することもできる．ここでは，マルウェアの連鎖感染[5]のモデル化を考える．連鎖感染とは，あるマルウェアに感染すると，それが他のマルウェアのダウンローダとしての機能を持ち，次に別のマルウェアにも感染することである．CCC DATAset 2009 の攻撃通信データにも，あるマルウェア A をダウンロードした後に別のマルウェア B をダウンロ

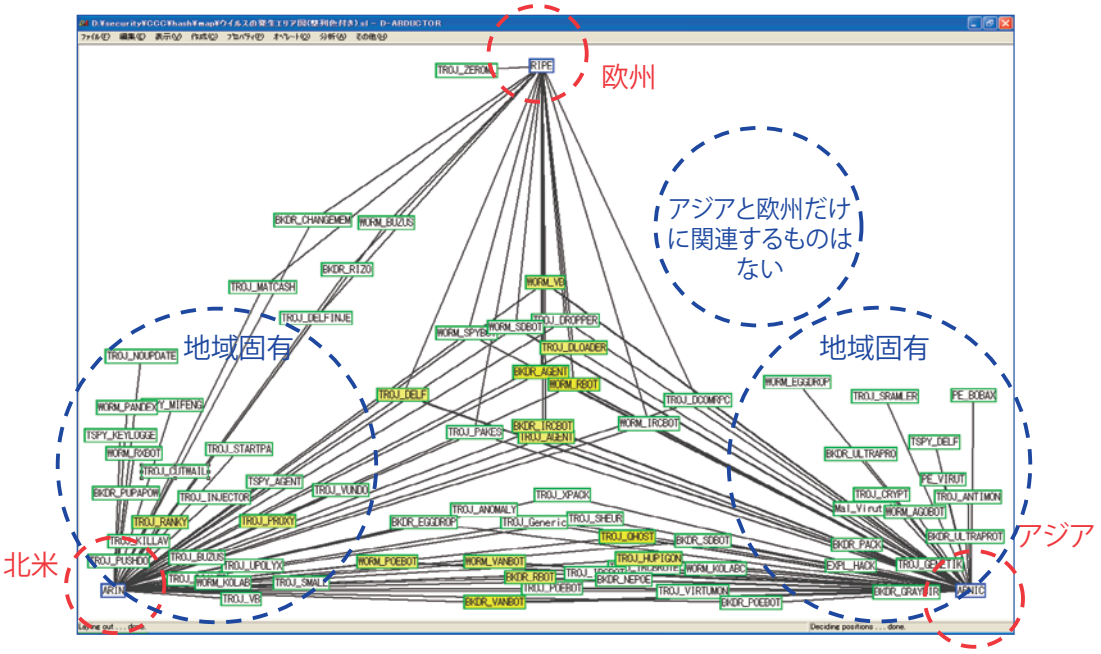

図 -2　マルウェアファミリによる地域分布

ードするという連鎖が散見された．

このようなマルウェア間の連鎖関係は，図 -1 のオントロジーでは，chain という関係で定義している．これは，連鎖感染をモデル化した関係であり，次のように攻撃の情報からマイニングして得られるものである．

攻撃には（日時，ハニーポット，ダウンロード元，マルウェア）の情報が含まれている．各攻撃のインスタンスから，あるマルウェアがダウンロードされてから一定時間内によくダウンロードされるマルウェアを連鎖関係と定義する．

具体的には，起点マルウェアから 5 分以内にダウンロードしたマルウェアを連鎖の候補として取り出す．こうした起点マルウェアと連鎖候補の組合せのうち頻度が上位のマルウェアを，連鎖関係とする．なお，単純に時間的に連続した攻撃を連鎖関係としないのは，ハニーポットにて異なる連鎖感染が同時に発生する場合があるからである．

連鎖関係にあるマルウェアの見える化

図 -3 は，攻撃元データから得られた連鎖関係の一部である．連鎖関係は，起点マルウェア（青枠）から連鎖先マルウェア（緑枠）の矢印で示される．なお矢印上の数値は出現回数である．同じマルウェア同士が矢印で結ばれている場合があるが（TSPY_KOLABC.CH），これは同じマルウェア同士の連鎖感染というよりは，同じマルウェアの感染が，複数の異なる攻撃元からたまたま同じタイミングに発生したと解釈すべきであろう．

図 -3 には TSPY_KOLABC.CH を起点として WORM_SWTYMLAI.CD と BKDR_POEBOT.GN への感染の流れが見て取れる．実際に，トレンドマイクロ社の情報[6]にも，TSPY_KOLABC.CH が WORM_SWTYMLAI.CD と BKDR_POEBOT.GN を生成するという説明がある．マイニングで得られた連鎖関係が，実際の連鎖感染と一致した例になっている．

未知検体への連鎖感染

図 -4 は，2009 年 4 月におけるマルウェアの連鎖関係を見える化したものである．注目したいのが中央部に位置している 028786 という数値だけのマルウェアである．これはこの時点では未知検体で，3 つのマルウェアから連鎖先になっていることが見て取れる．

同月度のサイバークリーンセンター活動実績[7]の上位 5 検体の動向観測の報告では，WORM_AUTORUN.CZU と未知検体の流布について注意勧告がされており，それを裏付ける相関マップと言える．

多くの連鎖感染をもたらすマルウェア

連鎖関係の連鎖元に注目すると，数多くの連鎖をもたらすマルウェアを見つけ出すこともできる．図 -5 は CCC DATAset 2009 攻撃元データの中でダウンロード数最多の PE_VIRUT.AV の連鎖関係を見える化したものである．

PE_VIRUT.AV は連鎖先より起点になることが多く，多様な連鎖の起点になっていることが一目で分かる．

このように，あるマルウェアに注目して見える化することで，そのマルウェアが連鎖を起こしやすいかどうか，また連鎖の役割が，起点なのか連鎖先なのかそれとも両

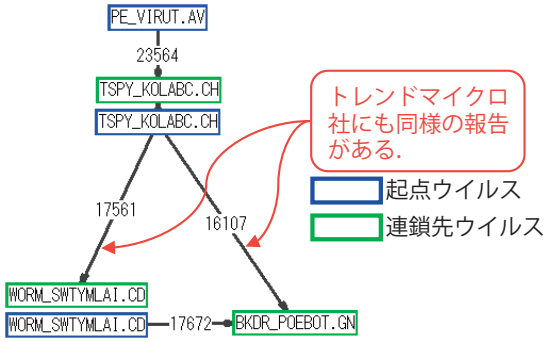

図 -3　見える化した連鎖の一例

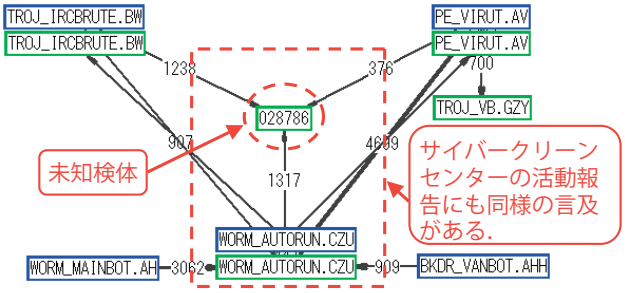

図 -4　見える化した未知検体への連鎖の一例

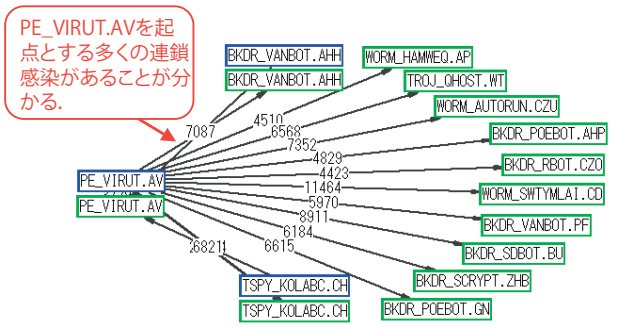

図 -5　見える化した PE_VIRUT.AV を連鎖感染の起点とするマルウェア群

方の要素を持っているかといった性質を確認できる．

まとめ

本稿では，ビジネス情報ナビゲータの出力する相関マップにより，CCC DATAset 2008/2009 の攻撃元データにおける，マルウェアの挙動をマクロに見える化した事例を紹介した．RDF による関係情報に変換することで，表などでは表現が難しいマルウェア間の複雑な関係が表現でき，マルウェアの脅威を容易に把握することが可能となる．

ビジネス情報ナビゲータは，元々はナレッジマネジメントツールとして開発したものであるが，オントロジーを活用し試行錯誤を繰り返すことで，攻撃ログ中の情報のつながりや，連鎖関係など，マルウェアの挙動の見える化にも応用できることを示した．セキュリティの分野において，本手法以外にもまだ，さまざまな分析アプローチが可能であることを示唆していよう．

今回は 1 種類のハニーポットのログを対象としたため，ビジネス情報ナビゲータの特徴である異種情報源の統合という機能は使っていない．多種類のハニーポットログや，セキュリティのニュースサイトの情報など，形式の異なる情報を意味的に統合して組み合わせると，新たな知見につながる分析結果も得られることであろう．

ダウンローダやトロイの木馬と呼ばれる複数のマルウェアに感染してしまう連鎖感染は，悪意ある者が意図的に仕組んだものと考えられており，このような手口が一般的になってきている．今後は，起点マルウェアを検知したら連鎖先マルウェアをブロックするといった，連鎖感染の遷移に着目したセキュリティ対策の研究に取り組みたい．また，新たなデータを使用した分析にも挑戦したい．

参考文献

1）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 (Oct. 2009).
2）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，FUJITSU, Vol.57. No.3, pp.325-330 (May 2006).
3）小櫻，津田，鳥居：ウイルスの時間的な関連性に注目した見える化，MWS2009 (Oct. 2009).
4）井形，小櫻，片山，津田：セマンティクグループウェア：RDF を用いた KnowWho の実現，SIG-SWO-A303-05 (Mar. 2004).
5）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 (Oct. 2008).
6）トレンドマイクロ社，TSPY_KOLABC.CH 詳細情報，http://www.trendmicro.co.jp/vinfo/grayware/ve_graywareDetails.asp?GNAME=TSPY%5FKOLABC%2ECH&VSect=Td
7）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

（平成 21 年 12 月 30 日受付）

小櫻文彦
kozac@jp.fujitsu.com
（株）富士通研究所 ソフトウェア＆ソリューション研究所 セキュアコンピューティング研究部 研究員．データベースや情報漏洩対策セキュリティに興味を持つ．

津田　宏（正会員）
htsuda@jp.fujitsu.com
（株）富士通研究所 ソフトウェア＆ソリューション研究所 主管研究員．セマンティック Web や情報漏洩対策セキュリティに興味を持つ．博士（理学）．

鳥居　悟（正会員）
pro104@labs.fujitsu.com
（株）富士通研究所 ソフトウェア＆ソリューション研究所 主管研究員．ネットワークによるセキュリティ対策に興味を持つ．