CCC Dataset 2010 によるマルウェア配布元 IP アドレス評価に関する一考察

須藤　年章

インターネットマルチフィード株式会社
〒100-0004 東京都千代田区大手町 1-5-1 大手町ファーストスクウェアイーストタワー3F
sudo@mfeed.ad.jp

あらまし　マルウェアに感染させられるユーザは、スパムメール内のリンクへのアクセスやインジェクションされたサイトの閲覧など、いろいろな手法により気づかないうちにマルウェア配布元へと誘導される。マルウェア配布元は、専用に構築されたサイトだけではなく、無料のホスティングサービスやオンラインストレージなど様々なインターネットサービスが利用されている。また、海外のプロバイダやホスティング事業社の中には、提供しているサービスがマルウェア配布元や詐欺行為、スパム送信などに利用されることを容認しているものが多数存在する。本論文では，CCC DATAset 2010 の攻撃元データを用いて，攻撃元のIPアドレスについて解析し悪性度評価とその利用方法を検討する。

Consideration concerning malware distribution former Internet Protocol address evaluation by CCC Dataset 2010

Toshiaki Sudoh

Internet Mulifeed Corporation
OTEMACHI 1st.SQUARE EAST TOWER ,3F 1-5-1,Otemachi,Chiyoda-ku,Tokyo 100-0004,Japan
sudo@mfeed.ad.jp

Abstract　A user infected of the malware is induced to the malware distribution origin before it notices by various techniques to inspect the site where the access and the injection were done to the link in the spam mail. As for the malware distribution origin, various internet services like not only the site constructed for exclusive use but also a free hosting service and the online storage, etc. are used. Moreover, a lot of things to allow the provided service to be used for the malware distribution origin, fraudulence, and the spam transmission, etc. exist in an overseas provider and the hosting business company. In this thesis, Internet Protocol address in the attack origin is analyzed by using attack former data of CCC DATAset 2010 and the malignancy evaluation and the use are examined.

1.　はじめに

マルウェア感染は、OS の脆弱性をネットワーク越しに利用して感染させる手法から、スパムメールに添付したファイルや、メール文中に書き込んだリンクによりマルウェア配布サイトへ誘導する手法や、改ざんされ不正コードを書き込まれた一般のホームページを閲覧することにより気づかないうちにマルウェア配布サイトへと誘導される手法などが利用される例が増えている。マルウェア感染のトリガとなる手法が様々に工夫されているとともに、これらの手法により誘導されるマルウェア配布サイトの構築にも工夫が凝らされている。マルウェア配布サイトは、専用に構築されたサイトだけではなく、無料のホスティングサービスやオンラインストレージなど様々なインターネットサービスが利用されており海外の

プロバイダやホスティング事業社の中には、提供しているサービスがマルウェア配布元や詐欺行為、スパム送信などに利用されることを容認しているものが多数存在する。またマルウェア配布専用のネットサービスが構築されている例もある。本論文では, CCC DATAset 2010 の攻撃元データを用いて, 攻撃元の IP アドレスについて悪性度評価方法とその利用方法を検討する。

2. マルウェア配布元

マルウェア配布に利用されるサービスの例を表 1 にまとめる。

表 1 マルウェア配布に利用されるサービス

マルウェア配布に利用されるサービス	例
ホスティングサービス	ネームベースバーチャルホスティング
	IP ベースバーチャルホスティング
	ホームページ作成サービス
	画像/動画ホスティング
	オンラインストレージ
	物理サーバー貸
クラウドサービス	
ボット	一般ユーザの端末
	ボットを利用したホスティングサービス
一般サーバの乗っ取り	インジェクションによる
	アカウントハックによる
専用サーバ	専用に構築

このような配布元が利用される大きな理由は、追跡やフィルタリングなどのセキュリティ対策をうけにくくするためであり大きく二通りのアプローチがある。

(1) IP アドレス、ドメインを頻繁に変えることで、追跡およびフィルタリング対策を妨害する
(2) 一般サービスを利用することで、IP アドレス、ドメインを一般サービスと共通化し対策を困難にする

表 1 において(1)に当てはまるものはボットを利用する手法であり、その他の手法は(2)に当てはまる。これらのサービスを利用するとすべてのユーザの IP アドレスやドメインは共通になるため、単純にフィルタリングすることによる一般ユーザへの影響が大きく、対策が困難になる。一部のクラウド系サービスにおいては、IP アドレスが動的に変化するものもあるため、(1)(2)の機能を同時に満たす場合もある。

3. 分析

3.1. 攻撃元 IP アドレスの分類

CCC DATAset 2010 の攻撃元データから攻撃元の IP アドレスを抽出し評価する。実際の対処には、攻撃元 IP アドレスそのものだけではなく、攻撃元 IP アドレスを含む大小のアドレスブロックや経路情報などの単位でのフィルタリングやレピュテーションが行われる。これは隣接するアドレスブロックは、共通の攻撃に利用される可能性が高いと考えられ、未検出の攻撃元をまとめて対処することや、将来的な攻撃を未然に防ぐことを目的とするためである。例えば下記のようならアドレスブロックを定義が考えられる。

(1) Host（/32）
攻撃元 IP アドレスそのもの
(2) Net（/24）
攻撃元 IP アドレスを含む/24 のネットワーク。
x.x.x.0～x.x.x.255 の 256 個の IP アドレス。
(3) Net（/16）
攻撃元 IP アドレスを含む/16 のネットワーク。
x.x.0.0～x.x.255.255 の 65536 個の IP アドレス。
(4) Route
攻撃元 IP アドレスを含む route 情報
(5) AS
攻撃元 IP アドレスの所属する AS 単位

この定義は一例であり、スパムへの対応、マルウェア配布をブロックするなど目的によって効果が異なるため最適なアドレスブロックを検討する必要がある。

3.2. IP アドレスの推移

図 1 に 2007 年 11 月から 2010 年 4 月の期間で一日毎に観測されたユニークな攻撃元 IP の数を前述の 5 つのブロック単位で集計した結果を示す。図 1 から得られる特性をまとめると次のようになる。

- Host(/32)を見ると 2007 年末および 2008 年初めをピークに減少しており、2010 年 4 月にはピーク時の 20%にまで減少している。
- Host(/32)と Net(24)の件数はすべての期間においてほぼ重なっており、一つの/24 のアドレスレンジに複数の攻撃元 IP アドレスが含まれる割合は低いことを示す。
- Net(/16)までアドレスレンジを広げると、2007 年末～2008 年初めの期間では一つの/16 のアドレスレンジに平均して 4 個の攻撃元 IP アドレスが含まれ。2009 年以降は一つの/16 のアドレスレンジでも 2 個の攻撃元 IP アドレスしか含まなくなっている。
- Net(/16)と route の二つのアドレスブロックも近い結果になっている。攻撃元 IP アドレスに該当する route 情報の prefix 長別分布は図 2 のようになっており、/16 の route 情報が最も多く、また/21 の route 情報に該当する攻撃元 IP アドレスも突出して多くな

っており prefix 長が/16 よりも長い route 情報が利用された割合が全体の 66%を占めるためである。

- AS 数は 200～300 個程度で全期間を通して大幅な変化はない。

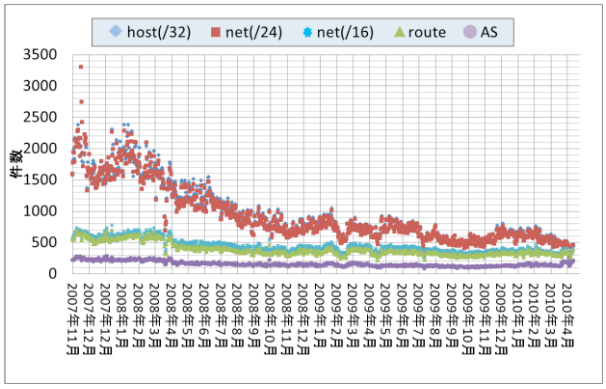

図 1　利用された IP アドレス数

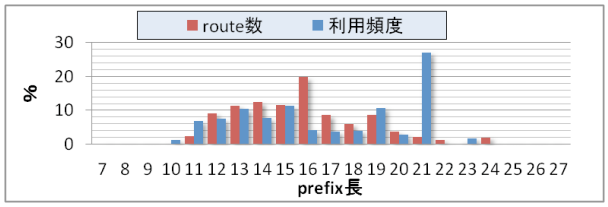

図 2 prefix 長

3.3.　攻撃元 IP アドレスの利用頻度を元にした評価

図 3 に攻撃元 IP アドレスが攻撃に利用された頻度を分析した結果を示す。

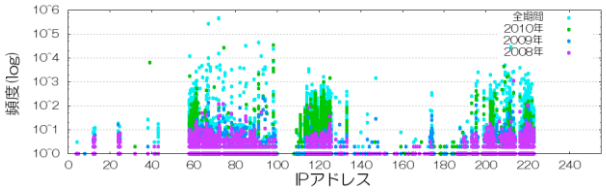

図 3　攻撃元 IP アドレスの利用頻度

IP アドレス毎に攻撃元としての利用頻度に差異が見られるが、単純にブロードバンドユーザに割り当てられているアドレスブロックほど頻度が高くなっていたり、58～62、105～112、175～185 近辺のアドレスブロックは2008 年にはインターネット上では全く、もしくは部分的にしか利用されていなかったアドレスブロックであるため 2009 年以降、2010 年以降しか観測されていなかったり2010 年以降の利用頻度が高くなるなどの特性に影響を与えている。ただし実際に対策を行うための評価としては、その時期において実際に攻撃が多いIP アドレス、IP アドレスブロックという単純な判断で利用することになるので問題ないと考えられる。

3.3.1.　/24 内の攻撃元 IP アドレス数を元にした評価

図 4 に/24 のアドレスブロック内の攻撃元 IP アドレス数について分析した結果を示す。

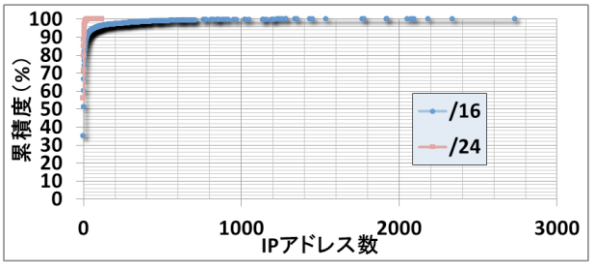

図 4　アドレスブロック中の攻撃元IP 数の累積度布

最大は 118 個で一つの/24 のアドレスブロックの IP アドレス数256 個の46%が攻撃元IP アドレスに該当するものであったことになる。ただし、これは接続の都度 IP アドレスが動的にアサインされるアクセスラインサービスを利用している単独もしくは少数の端末である可能性が高いため、単純にこの/24 のアドレスブロック中に大量の攻撃元が存在することはできないが、アドレスブロック的には悪性度が高いと判断することができる。/24 のアドレスブロックに 1 個しか攻撃元が含まれなかったものが 55.8%を占め、10 個以下で 96.7%を占めている。したがって悪性 IP アドレスを検出した場合に、その IP アドレスを含む/24 のアドレスブロックをまとめてブラックリストに登録するような運用を行った場合、/24 のアドレスブロックの 96%が無関係なアドレスであり、実運用上は問題を生む可能性が高い。

3.3.2.　/16 内の攻撃元 IP アドレス数を元にした評価

/16 にアドレスブロックを広げ同様の分析を行った結果を図 4 に示す。最大は 2734 個で/16 の 65536 個の 4.1%が攻撃元 IP アドレスに該当するものである。/16 のアドレスブロックに 1 個の攻撃元しか含まなかったものが 34.8%、100 個以下で 95.7%を占める。したがって/16 のアドレスブロック単位でも評価は可能だが、96%以上の無関係な IP アドレスまで含めてしまうため、/24 よりも影響範囲が広がり実運用上問題がある。

3.3.3.　アドレスブロックの考え

今回のデータを対象にした場合、悪性な IP アドレスを含む/24 のアドレスブロックをそのまま悪性と判断しても効果は非常に低く、逆に無関係な IP アドレスをファイルタリングするなどの悪影響が想定される。ただし、隣接の IP アドレスは、同じホスティングサービスの提供が行われていたり、ラウンドロビンされる IP アドレスの一つであったり、また隣接への感染なども考えられるため検出漏れへの対応や今後の可能性として悪性度のポイントを加算することは検討の余地がある。例えば/24 の 256 個のアドレスがすべて悪性と確証を得られた場合に

悪性度 100%になるようなポイントの割り振りが考えられる。アドレスブロックを/16 に広げる、あるいは経路情報までひろげてポイントを加算する場合も同様である。

3.4. 攻撃元 IP アドレスが利用された期間の評価

次に IP アドレスが攻撃に利用される時期および期間の違いに着目して分析する。

図 5 は2009 年 1 月に利用された攻撃元 IP アドレスについてそのアドレスがそれ以前及びそれ以後どの期間で観測されたかを表している。

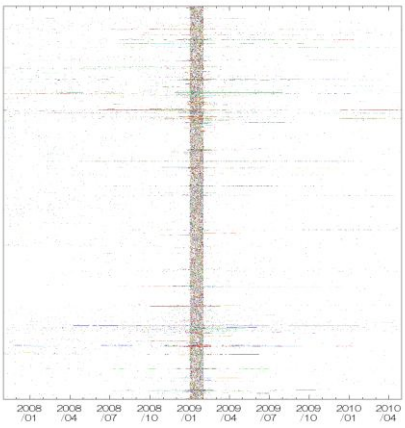

図 5　2009 年 1 月の攻撃元 IP の攻撃利用期間

X 軸が時期を表し、Y 軸が IP アドレスの違いを表している。したがって X 軸に平行な線が現れる場合、同一の IP アドレスが長期にわたり利用されていたことを表す。一年以上継続して利用されている IP アドレスや一度観測されていたアドレスが長期間あけて再度観測されるようなものもある。2009 年 1 月に集中してプロットが現れているが、分析対象とした攻撃元 IP アドレスが 2009 年 1 月に観測されたものであることから、短期間にしか利用されていないアドレスが大量にあることを表している。これらの IP アドレスは短期間にしか攻撃利用されないため、悪性と判断し対策を実施したとしても、直後には改善され攻撃とは無関係になってしまう。このような無実の IP アドレスをフィルタリングしつづけてしまうような問題を避けるためには、改善度評価に関しても必要な条件を定義し実施する必要がある。

図 6 はそれぞれの攻撃元IP アドレスからの攻撃が観測され続けた時間の累積度分布を表す。X 軸は攻撃元 IP アドレスが観測されつづけた時間を秒で表しており、観測期間の違いによる特性の違いを見るために全期間と2008 年 1 月、2009 年 1 月、2010 年 1 月の攻撃元IP アドレスについて解析を行った。ここで見られる特性をまとめると次のようになる。

- 一回しか観測されなかった攻撃元IP アドレスが、全体では 49%、2008 年 1 月では 42%、2009 年 1 月のデータでは44%、2010 年 1 月のデータでは 52%ある。
- 攻撃が観測された期間が 1 時間以下であった攻撃元 IP アドレスは全体では76%、サンプルで抽出した 3 つに期間のデータでは 70～73%であった。短時間の攻撃は攻撃元IP アドレスを変えながら行われる連続攻撃などその観測時期毎の攻撃の流行の影響うけていると想定される。そのためサンプル期間毎に量にばらつきが出ている。
- 攻撃が観測された時間が 24 時間以下であった攻撃元 IP アドレスは全体の 92%を占めた。サンプルで抽出した 3 つの期間の攻撃元IP アドレスでは 86%となっており、90%程度の IP アドレスが 24 時間以内で利用が終了している。

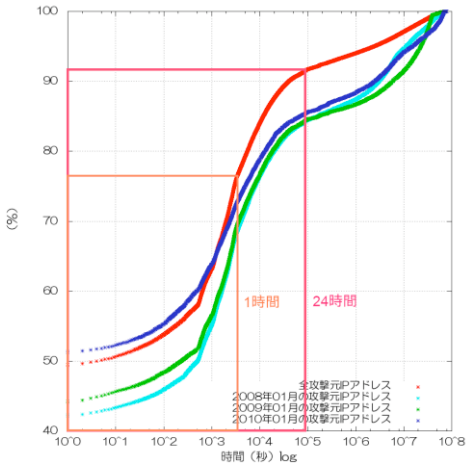

図 6 攻撃元 IP アドレスの利用期間の累積度分布

3.4.1. 攻撃期間の考え方

したがってこれらの情報から悪性度評価に利用できる要素としては次のようなものが考えられる。

- 攻撃継続時間
- 過去に攻撃利用された履歴
- 最後に攻撃が観測されてからの経過時間

攻撃継続時間の長短は、実運用上は長くても短くてもたとえばフィルタリング対象とすべき悪性ポイントを割り振る必要があるが、前述のアドレスブロックなどの要素への悪性ポイントの加算などの検討要素としても利用することができる。過去の履歴は、過去に攻撃利用されたことがある場合は加点を大きくするなどの要素として使う。また最後に攻撃が観測されてからの経過時間は逆に悪性度を下げるための要素であり、一度悪性判断された IP アドレスからの攻撃が観測されなくなってからの時間経過にあわせてポイントを減算していく。ただし、過去

の攻撃履歴の要素による再発の可能性としてのポイントを残すために 0 になる期間はそれぞれの運用状況にあわせて長期間に設定するなどの検討が必要である。

3.5. 攻撃元 IP アドレスとハッシュの関係を利用した評価

図 7 に攻撃元 IP アドレスと、その IP アドレスが関連したユニークハッシュ数の関係を示す。

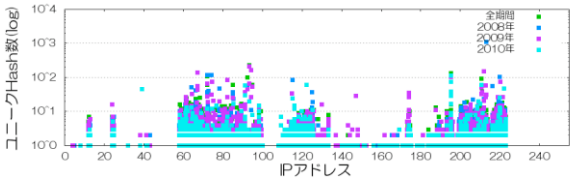

図 7 攻撃元 IP アドレス (/32)とハッシュ数の関係

アドレス毎、時期毎に関連するユニークハッシュ数に差異があることがわかる。さらに図 8 にそれぞれの攻撃元 IP が関連したユニークハッシュ数の累積度分布を示す。最大では1118 種類のハッシュに関連した IP アドレスが存在したが、1 個のハッシュにしか関連しなかったIP アドレスが全体の 81.6%を、3 個以下のハッシュにしか関連しないIP アドレスが 98.9%を占める。関連するユニークハッシュ数に応じてその IP アドレスの悪性度を評価することができるが、1 個のハッシュにしか関連しない攻撃元 IP アドレスがほとんどを占めるため、一つの IP アドレスへの対策を行ったとしても僅かなマルウェアへの対策にしかならず対策の効率性の向上は期待できない。すべてのマルウェアに対応するには、大量の IP アドレスに対して網羅的に対策を行う必要がある。

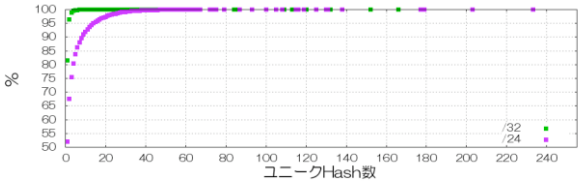

図 8 ユニークハッシュ数の累積度分布

3.5.1. /24 のアドレスブロックとハッシュ数の関係を元にした評価

図 9 に攻撃元 IP アドレスが含まれる/24 のアドレスブロックと、その IP アドレスが関連したユニークハッシュ数の関係を示す。前述の攻撃頻度と同様に/24 にまとめても大き差異は見られない。さらに詳細に分析するため図 8 にそれぞれの攻撃元 IP が関連したユニークハッシュ数の累積度分布を示す。

最大で 1132 種類のハッシュに関連したIP アドレスが存在したが、/32 で解析した結果とくらべても僅かな差しかない。また24 にアドレスブロックを広げても 1 個のハッシュにしか関連しなかったアドレスブロックが全体の51.4%あり、また 30 個以下のハッシュにするアドレスブロックで 99.2%を占める。したがって/24 単位で対策を行ったとして も大量のマルウェアに対する対策の効率が向上するわけではない。

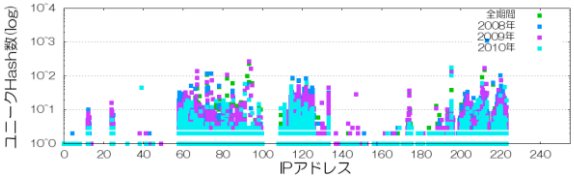

図 9 Net(/24)とハッシュ数の関係

3.5.2. ハッシュ数に関する考え方

単純に関連するハッシュ数に比例して悪性度判定を行うことができるが、今回のデータでは 1 個のハッシュにしか関連性を持たないIP アドレスが圧倒的に多く大きな差を出しにくい結果になった。さらにアドレスブロックを広げることで対策の効率を向上させることが考えられるが、/24 のアドレスブロックに広げた場合でも関連するハッシュが一つしかないものの割合が非常に大きく、これも大きな違いを出すことはできなかった。しかし一般的な悪性度評価としては、関連するハッシュ数の大小による評価は有効であると考えられる。

3.6. AS に関する評価

各 AS に含まれる攻撃元 IP アドレスの数の大小は各 AS のもつ IP アドレスの全体数に比例すると想定されるため、単純にその AS の悪性度を表すものではない。ただし前述の通り悪性サイトを積極的に取り込んでいるプロバイダや、問題があったとしても特に何の対処もしないプロバイダや事業者が存在する。そのため対策を行わなければならない側の立場では、その AS に所属する攻撃元 IP アドレスの数の大小や攻撃頻度を元に AS 単位で悪性度評価を行うことも必要にある。図 10、図 11 は、AS 単位での攻撃頻度とユニークハッシュ数を分析した結果示す。図 10 は 2007 年 11 月から 2010 年 4 月までの全期間のデータを分析し、位 20AS をまとめたものである。

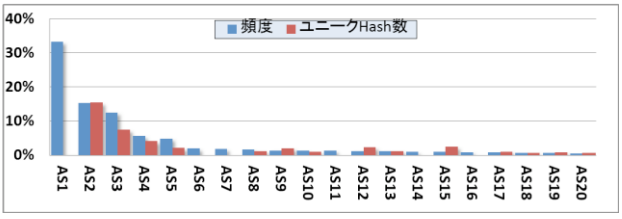

図 10 AS とハッシュ数の関係（全期間）

一位の AS は攻撃の 33%を占めているが関連するハッシュ数は 488 個で全体の 0.2%しか利用しておらず、攻撃頻度は非常に高いが、少数の攻撃にのみにしか関連していない状況を表している。二位の AS は攻撃の 15%を占めているとともにユニークハッシュ数は、26359 個で全体の 15%を占めるということで攻撃頻度も大きくたくさんの攻撃に関連している。図 11 は同様の解析を 2010 年のデータに限定して解析したものである。関連する AS が大きく変わっているが、一位の AS はやはり攻撃頻度は高いがユニークハッシュ数が少なく全体の 0.2%しかないなど、AS 毎に特徴があらわれている。

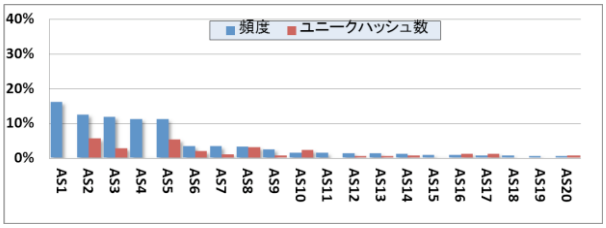

図 11　AS とハッシュ数の関係（2010）

一般的には攻撃頻度の大きさとユニークハッシュ数の大小は比例するが、攻撃頻度が大きいが、関連するハッシュ数が少ない AS が存在する。このような AS は専門的にマルウェア配布用のサイトを設置しているサーバやホスティング事業社、またはそれらの事業社を抱えた AS である場合が多いため、このような特性をもつ AS に関しては悪性度評価を高めるような対策を行うことも可能だと想定される。

3.6.1.　AS に関する考え方

AS 単位での分析することにより下記のような評価対象を利用できる。

- 評価アドレスブロックの拡大
- 運営主体自体の評価
- 経路情報と組み合わせた対策

主に運営主体自体の評価を要素として加えることができ、これはフィルタリングなどの実施だけではなく、運営主体そのものへの直接的な改善依頼や、インターネットコネクティビや AS 単位でのトラフィック流通の抑止などより強硬な対処へと発展するものに利用できる。実運用上のフィルタリングとしては、小規模な事業社を対象とする場合はいいが、ISP を象にする場合は、無関係な IP アドレスが圧倒的に多いため一般的には利用が難しいが、企業やセキュリティサービスなどでは不必要な AS との通信としてフィルタリングする運用も可能である

4.　その他の評価要素

近年、一つの攻撃が継続する期間が非常に短くなっており、検出後対策を検討する時点では、攻撃元からマルウェアがダウンロードできなくなっていることが増えている。ただし一度攻撃に利用された IP アドレスは時間をおいて再度別の攻撃に利用される可能性があるため、それを踏まえた悪性度評価を行っておく必要がある。したがってマルウェア配布が確認されない攻撃元IPアドレスや下記のような攻撃元IPアドレスも悪性度評価に加えることで対象の拡大と精度の向上を図る必要がある。

- スキャン攻撃元
- 改竄などの不正アクセスの疑いのある IP アドレス

4.1.　再評価周期

様々な指標を元に悪性と判断するだけでは、改善され問題のなくなった IP アドレスを悪性と評価しつづけ誤った対策に利用されることになるため、再評価を繰り返し、悪性度の加点、減点を繰り返し行う必要がある。本データでは、利用期間が 24 時間以内の IP アドレスが 90%、1 時間以内のものが 76%を占めるため最低でも 24 時間以下の周期で再評価を行う必要がある。

5.　おわりに

マルウェアを検出して、それを解析して対策するという手法は確実であり効果も高いが、大量のマルウェアや攻撃に対してより迅速にあるいは未然に対策を実施する手法として、IP アドレスやサービスなどの評価を継続的に行いその情報を元にしてインフラそのものがもつアーキテクチャとしてフィルタリングなどを実施する対策も有効な手法だと考えられる。本研究の結果はあくまでも CCC で観測された情報のみを元にしているデータであるため、インターネット上で発生しているすべての攻撃を網羅しているものではなく、また、実運用上は誤検出の要素も多分に含むため、複数のレピューテーションデータベースの並行利用と丁寧な運用により精度を高める必要がある。

謝辞

本研究の一部は Telecom-ISAC Japan の支援を受け実施している．本研究を進めるにあたり，有益な助言と協力を頂いた Telecom-ISAC Japan の関係者各位に深く感謝致します．

6.　参考文献

1)　畑田充弘，他：マルウェア対策のための研究用データセット ～MWS 2010 Datasets～，MWS2010（2010 年 10 月）