Computer Security Symposium 2014
22 - 24 October 2014

DNS の名前解決に着目した悪性ドメインの利用状況に関する調査

水谷 正慶 †

† 日本アイ・ビー・エム 東京基礎研究所
135-8511 東京都江東区豊洲 5-6-52 NBF 豊洲キャナルフロント
masa@jp.ibm.com

あらまし 情報技術の発展および普及に伴い，インターネット上での悪意ある人物によるセキュリティ侵害が深刻な課題となっている．悪意ある人物がインターネット上でボットネットなどを利用して攻撃してくるなかで，ドメイン名は被害者と攻撃者，あるいは悪意のあるプログラムと攻撃者を結びつける上で重要な役割を担っている．本稿では悪意のある人物が利用していると考えられるドメイン名約 42 万件を 4ヶ月にわたって監視を続け，どのような傾向が見られたのかについて報告するとともに，悪意あるドメイン名のリストを用いたセキュリティ対策の有効性について議論する．

Analysis of Malicious Domain Name Usage focused on DNS Name Resolution

Masayoshi Mizutani†

†IBM Japan, Tokyo Research
NBF Toyosu Canal Front Building 5-6-52 Toyosu,Koto-ku Tokyo, 135-8511 Japan
masa@jp.ibm.com

Abstract The Domain Name System (DNS) plays important role in the Internet for not only innocent purposes but also malicious activities. The knowledge of DNS activities by malicious users is required to monitor and protect end-user systems. In this paper, we investigated about 420 thousands malicious domain names for 4 months, and analyzed name resolution data. Our experiments show characteristics of malicious domain activities, and we discussed how a black list of malicious domain name could be used.

1 はじめに

Domain Name Service(DNS) はインターネットにおけるサービスを示すドメイン名から，IP アドレスなどサービス利用に必要となる情報を DNS サーバから取得するためのプロトコルである．今日のインターネットでは World Wide Web(WWW) や電子メール配信などの目的で広く利用されているが，これらは正当な目的に限らず，悪意ある活動においても同様に多くの場面で利用されている．例としてエンドユーザを誘導するために正規のサービスと誤認しやすいドメイン名をもつ攻撃サイトを用いたり，悪意あるソフトウェアの総称であるマルウェアが Command & Control サーバ (C&C サーバ) へ接続するために DNS の機能を悪用するなどの事例が挙げられる．このような攻撃からエンドユーザを守るため悪意あるユーザが利用しているドメイン名を収集，ブラックリストとして管

理し，システム管理者が管理対象のネットワーク内でブラックリストに含まれるドメイン名の問い合わせが発生していないかをチェックすることで，悪意ある活動に関わってしまっているエンドユーザがいないかを監視することができる．また，ブラックリスト中のドメイン名の問い合わせに対し，管理対象のキャッシュサーバで応答を遮断する，あるいは意図的に誤った応答をすることによって，悪意あるサーバへの通信を水際で防ぐことができる．

ドメイン名のブラックリストによる悪意ある活動の監視，遮断はドメイン名の問合せを制御できれば効果的だが，ドメイン名問合せを伴わない通信や監視，遮断の方法によっては有効に機能しない場合もある．例えば運用しているサーバに対して外部からの接続が怪しい接続元かどうかを判断したいと考えた時，多くの場合悪意ある目的で利用されるドメイン名は逆引きが正しく設定されていないため，ブラックリストに存在するドメイン名に紐付いた IP アドレスなのかどうかを判断することができない．また，ファイアウォールや Intrusion Detection System(IDS)，Intrusion Prevention System(IPS) なども，DNS の問い合わせに関わらず通信の監視をしているため，監視対象の通信がブラックリストに含まれるドメイン名を利用しているのかどうかの判断が難しい．さらにマルウェアなどで通信対象のサーバのアドレスがハードコーディングされており，ドメイン名問合せが発生しない場合もある．これは，ブラックリストに含まれるドメイン名のアドレスをあらかじめ取得してキャッシュし，検査対象の IP アドレスを効果的に発見できる可能性があるが，一方で対象ドメイン名に登録された IP アドレスが頻繁に変わることもありえるため，悪意のある人物がドメイン名をどのように利用しているのかを俯瞰的に把握する必要があると考えられる．

本論文では悪意のある人物が利用していると考えられるドメイン名，悪性ドメイン名約 42 万件を 2014 年 4 月 2 日から同年 8 月 12 日までの 4ヶ月にわたって監視を続け，その結果から悪性ドメイン名が長期的にどのように利用されているのかを調査，分析した．対象となる悪性ドメイン名について定期的に問合せを実施し，応答に含まれる IP アドレスの情報を蓄積し続けた．この結果から悪性ドメイン名の生存時間の分布，悪性ドメイン名が持つ IP アドレスの変化，さらに悪性ドメイン名間の関係についての分析を実施し，以下の知見を得た．

- 監視期間中に一度でも IP アドレスを応答した悪性ドメイン名は合計で 307,929 件であったのに対し，監視終了時に IP アドレスを応答したのは合計数の約 52.78%である 162,530 件であった．ただし，全体の約 27.42%にあたる 84,448 件の悪性ドメイン名が 2014 年 6 月 12 日に一斉にシャットダウンしたとみられる事象も観測されており，これを除外すると約 72.73%の悪性ドメイン名が 4ヶ月間活動を続けていると言える．
- 監視期間中，約 63%の悪性ドメイン名が 1 種類のみの IP アドレスを応答しており，約 86%のドメイン名が 10 種類以下の IP アドレスを応答している．
- 取得した IP アドレスは 4ヶ月後でも約 76%が有効である．しかし，観測される IP アドレスの総数は約 150%になっており，増加分の約 50%については False Negative が発生しうる．
- 悪性ドメイン名が持つ IP アドレスを高い網羅率で維持したい場合，1 時間未満の周期でドメイン名問合せを繰り返すのが望ましい。

本稿では第 2 節で悪性ドメイン名情報の取得方法について述べ、第 3 節で分析結果を示す。その後、第 4 節で考察を述べ、第 5 節でまとめを示す。

2 実験環境

悪性ドメイン名として hpHosts[1] が提供するドメイン名リストを利用した。hpHosts はマルウェア配布サイトやフィッシングサイト、攻撃コードの配布を実施しているサイトのドメイン名を管理している。また hpHosts は悪意

ある目的で利用されているとみられるドメイン名の問合せに対して、意図的に誤った応答をすることで悪意ある目的で利用されているとみられるサーバへの接続を水際で防ぐための設定ファイルを `http://hosts-file.net/download/hosts.txt` にて配布している。本稿ではこの設定ファイルに含まれるドメイン名を悪性ドメイン名として定義する。実験では 2014 年 4 月 2 日時点での最新版の hosts.txt を利用し、これ含まれていた 421,805 件のドメイン名について調査した。

調査は日本時間の 2014 年 4 月 2 日午後 3 時から開始して同年 8 月 12 日午前 10 時まで、1 時間毎に全ドメイン名に対して A レコードの問い合わせを DNS キャッシュサーバ経由で実施した。調査期間中は悪性ドメイン名のリストは更新せず、継続的に同じリストに対して調査を継続した。

DNS 問合せの送信には Node.js[2] を利用し、421,805 件のドメイン名を並列して問合せるツールを実装した。問合せには Node.js で提供されている dns.resolve4 関数を使用し、問合せは DNS キャッシュサーバを経由して実施した。負荷軽減のため問い合わせ時刻をずらしながらクエリを送信しており、また応答受信までの遅延も発生しているため、実際の問い合わせ間隔には誤差がある。問い合わせが失敗して応答が無かった場合などに問い合わせは再送せず、応答なしと記録した。

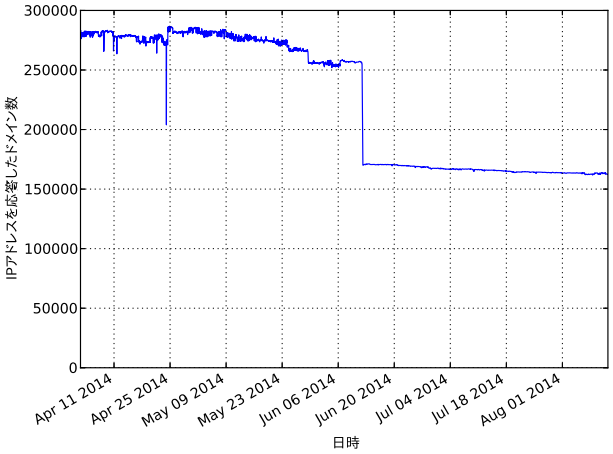

図 1: DNS 問い合わせに対して IP アドレスを含む応答を返したドメイン名数の推移

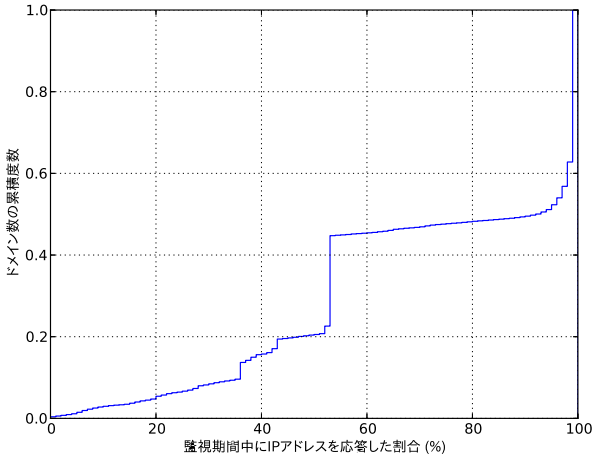

図 2: 観測期間中に DNS 問い合わせに対して IP アドレスを含む応答を返した割合の累積度数分布

3 実験結果

調査対象となった悪性ドメインのうち、一度でも A レコードを含む応答があったのは 307,929 件であった。これは調査対象の全ドメイン数、421,805 件の約 73.00%にあたる。本稿ではこの結果に対してリストに含まれる悪性ドメイン名が機能している期間、悪性ドメイン名から取得した IP アドレスの利用可能期間や IP アドレスを取得すべき周期について分析した。

3.1 ドメインリストの有効性に関する調査

hpHosts で管理されている悪性ドメインのリストが時間経過とともに有効性がどのように変化したかを調査するため、観測期間中に IP アドレスを含む応答を返したドメイン名の総数を、折れ線グラフによって図 1 で示す。Y 軸が IP アドレスを含む応答を返したドメイン名の数、X 軸が時系列となる。グラフ途中にある急激なドメイン数の落ち込みは 2014 年 6 月 12 日に大量のサブドメインを持ついくつかのドメイン名が IP アドレスを含む応答をしなくなったためである。表 1 に示すドメイン名およびそのサブド

表 1: 2014 年 6 月 12 日にシャットダウンしたと見られるドメイン名とサブドメイン数（TLD を含む末尾 4 文字は*に置き換え）

ドメイン名	リストに含まれるサブドメイン数
dont64worrycas*.***	8165
allforyou64loan*.***	9838
fast64loan*.***	9937
999uploans6*.***	10596
64luckypaybil*.***	11040
64moneyforbil*.***	11486
1000paydaycash6*.***	11518
64happy1000loan*.***	11868

メイン名、合計 8 個のドメイン名が合計 84,448 件のサブドメインを hosts.txt 内に保持しており，これらが一斉に応答しなくなったため，急激なドメイン数減少が起きている。本稿執筆時点で各ドメイン名の WHOIS 情報を確認したところ、登録日時が調査開始以前であり、更新期限が調査終了後であったため、何かしらの人為的な操作によってシャットダウンさせられたのではないかと推測される。監視終了時に IP アドレスを応答したドメイン名は 162,530 件であったが、この 6 月 12 日にシャットダウンした悪性ドメインのケースを例外として捉えると、応答があった総数 307,929 件から 84,448 件を除いた 223,481 件に対して約 72.73%が 4ヶ月間活動を続けていると言える．

図 2 は観測期間中に各悪性ドメインが IP アドレスを含む応答を返した割合を累積度数分布で表している。ドメイン名に対する問い合わせの応答は不連続に発生、消滅するため、応答は連続して返ってきたものに限らず、問い合わせ数合計に対して応答があった割合を用いている。図 1 で示したように、6 月 12 日にシャットダウンしたと思われる悪性ドメインが応答率約 54%の箇所に現れている。また、これ以外にも約 10%ほどのドメイン名が合計で 50 日以下しか有効になっていないという点も見て取れる。しかし一方で、半分以上のドメイン名は応答率が 90%、合計約 118 日以上活動しており、半分以上のドメインについては長期的にリストが有効であることを示している。

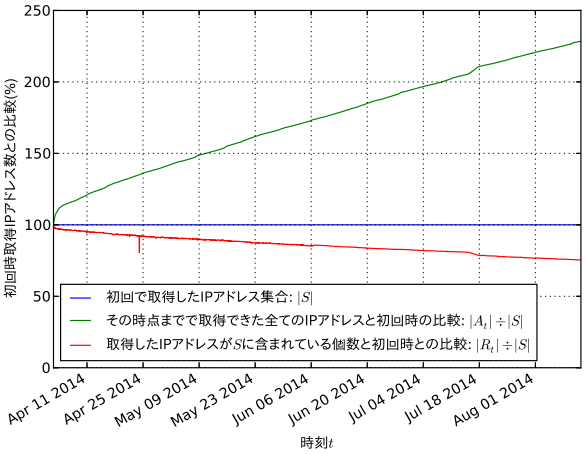

図 3: 取得した IP アドレスの時間経過によるロス率と新しく出現した IP アドレス数の割合

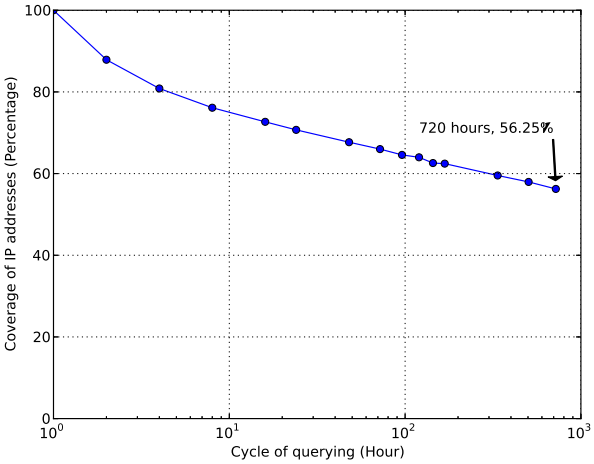

図 4: 問い合わせ周期による IP アドレスの網羅率 (1 時間周期で問い合わせをした際に得られた IP アドレス数を 100%とする)

3.2 ドメイン名から取得した IP アドレスの有効性に関する調査

次にドメイン名から取得した IP アドレスの有効期間を調査するため、最初に悪性ドメインのリストから取得した IP アドレスとその後に出現した IP アドレスに関する分析を実施した結果を図 3 に示す。最初の問い合わせ時に取得できた IP アドレスの集合 (以下、S とする) に対して、その後の時刻 t に取得された IP アドレ

スで S に含まれているものの集合を R_t、時刻 t までに得られた IP アドレスの総数を A_t とする。図にはそれぞれ初回に取得した IP アドレスの個数と時刻 t までに得られたアドレス総数の比較 $\frac{|A_t|}{|S|}$ と、時刻 t に取得した IP アドレスに S に含まれるものがどれだけ残っていたかという比較 $\frac{|R_t|}{|S|}$ を表している。最初に取得できた IP アドレスの個数、$|S|$ は 41,827 件であり、調査終了時刻 e でも観測された S に含まれる IP アドレス $|R_e|$ は 31,586 件、観測された全ての IP アドレス数 $|A_e|$ は 95,570 件であった。IP アドレスの観測数については、調査期間中に急激な変化は発生していない。図 1 に出現した 2014 年 6 月 12 日のドメイン名減少においても、応答しなくなったドメイン名 84,448 件は全て 1 の IP アドレスのみを返答していたため、図 3 にはその影響はでていない。

図 3 では、調査終了時点の IP アドレスと初期に取得された IP アドレスの比較 $\frac{|R_e|}{|S|}$ が約 0.727 となっている。これはリストに含まれる悪性ドメインのために利用する IP アドレスが変更、あるいは削除されているということであり、初回に取得した S をもとに監視、遮断を試みた場合には 4ヶ月経過した後も False Positive が約 28%になると考えることができる。しかし、一方で $\frac{|A_e|}{|S|}$ が約 2.284 となっており、時間経過とともに初回取得時を上回る個数の新たな IP アドレスが悪性ドメインのために使われていることになる。False Negative は約 56%以上で発生すると考えられる。本実験では 1 時間周期でアドレスを取得しているが、より短い Time To Live(TTL) パラメータを設定される可能性もあるため、実運用では False Negative の割合はより高くなるのではないかと予想される。監視や遮断の目的によって許容される False Positive, False Negative の割合は異なるが、両者をなるべく低くしたい場合は定期的に悪性ドメインから取得される IP アドレスを更新しなければならないと言える。

悪性ドメインの IP アドレスを周期的に取得する場合、どの程度の周期が適切なのかを調査するために周期の変化によって IP アドレスの網羅率がどのように変化するかを分析し図 4 に示

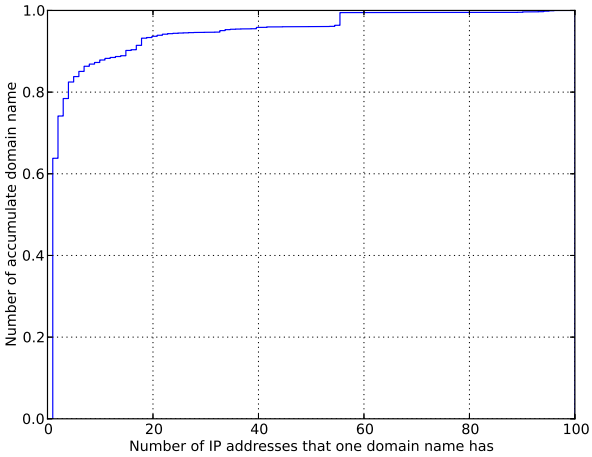

図 5: 観測期間中に各ドメイン名の応答に含まれたユニークな IP アドレス数の累積度数分布

した。それぞれ 2、4、8、16、24、48、72、96、120、144、168、336、504、720 時間周期で IP アドレスを取得した場合に、1 時間周期で IP アドレスを取得した場合とくらべてどの程度の割合が取得できるかを分析した。図中の Y 軸が 1 時間周期で IP アドレスを取得した際の数を 100%としたとき、異なる周期で取得した IP アドレス数の割合を示している。X 軸は取得周期で単位は 1 時間である。10^0 の始点が 100%を示しているが、次の 2 時間周期の点ですでに網羅率が 90%以下となっており、さらに 4 時間周期で約 80%になっている。その後の変化は比較的なだらかであるため、2 日間 (72 時間) 周期で 70%以上、一週間周期で 60%以上を維持できる。今回の実験では 1 時間周期での問い合わせが最短の周期だったが、この時点ですでに多くの IP アドレスを取りこぼしている可能性があるため、悪性ドメインから得られる IP アドレスを完全に把握するのは困難であると予想される。

3.3 悪性ドメイン名の利用方法に関する調査

悪性ドメインのリストから IP アドレスを全て網羅するためには高頻度で IP アドレスの問い合わせをする必要があるが、全体の傾向を把握することでより効果的な IP アドレスの取得を試みるため、悪性ドメインと IP アドレスの関

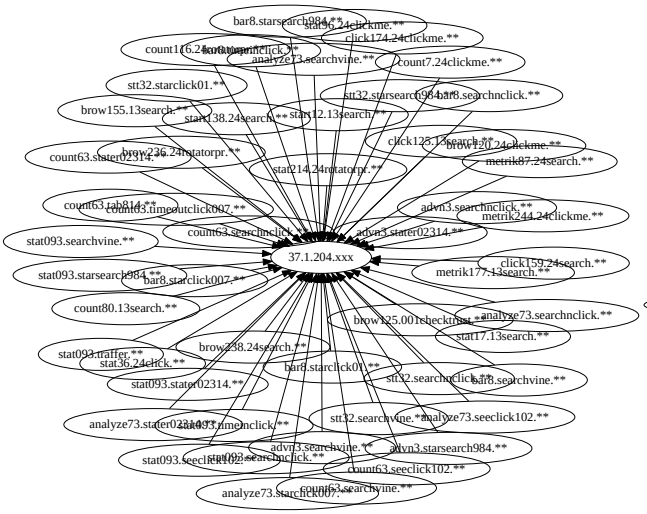

図 6: ドメイン名と IP アドレスをグラフ構造にして抽出した例 1: 1 つの IP アドレスを複数のドメインが共有している状態

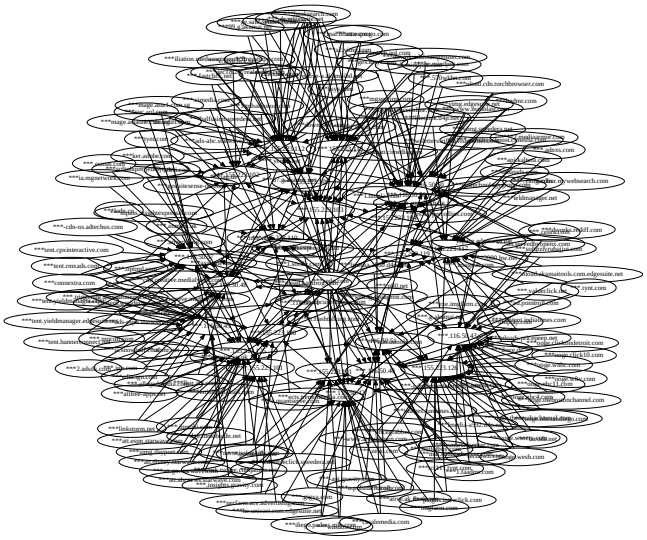

図 7: ドメイン名と IP アドレスをグラフ構造にして抽出した例 2: ドメイン名と IP アドレスが複数あり複雑に関係を持っている状態

係について調査した。図 5 は各悪性ドメイン名が問い合わせに対していくつの IP アドレスを応答したのか数え上げ、累積度数分布にして表示したものになる。表示範囲が限られていたため，応答 IP アドレスが 100 件未満のものしかグラフに描画されていないが，応答 IP アドレスが 100 件を超えるドメイン名は全体で 1,426 件．応答があった 307,929 件に対して約 0.46%であったため、ここではひとまず議論の対象から除く。図中では 60%以上のドメイン名が調査期間中に 1 つだけの IP アドレスしか応答していないことを示している。さらに 90%以上のドメイン名が 4ヶ月間に 20 件以下の IP アドレスしか応答していないことが分かる。ここで示している応答 IP アドレス数はドメイン毎にいくつ応答したかを示しているため、重複している IP アドレスも多く含まれている。別途、数え上げた結果、この図中に示されている 307,929 件のドメイン名は 65,307 件の IP アドレスしか使用しておらず、残りの 1,426 件のドメイン名によって残り 30,263 件の IP アドレスが利用されていることがわかった。

そこでドメイン名と IP アドレスの関係の全体像を把握するため、ドメイン名と IP アドレスをそれぞれノードとし問い合わせ元になったドメイン名と結果として得られた IP アドレスをエッジで接続したグラフ構造を作成して、ドメイン名と IP アドレスのグループ分けを実施した。これにより、調査期間中に応答を返した全てのドメイン名、得られた全ての IP アドレスと、その問い合わせ結果から 31,312 個のグラフが作成された。そのうちの 2 例を図 6 と図 7 に示す。それぞれの図ではドメイン名、IP アドレスの一部をマスキングしている。図 6 では 1 つの IP アドレスを複数のドメイン名が共有して利用している状態を示している。図 5 で示した 1 つしか IP アドレスを返さなかったドメイン名は多くがこの形態に属している。一方、図 7 は複数のドメイン名が複数の IP アドレスを利用しており、利用している IP アドレスも様々である。

これらのグラフとして分離したドメイン名および IP アドレスのグループの分布を調査するため、図 8 においてグラフに含まれるドメイン名数と IP アドレス数によって散布図として示した。図中において IP アドレス数が 10^2 未満のものについては、図 5 中に示されていることになる。図 5 で示した結果と同様、ほとんどのグラフが IP アドレス 10^2 未満に収まっているが、4 つほどのグラフが IP アドレスを約 1,000 件以上含んでいるのが分かる。特に上部の 2 つは 10^4 を超えている。このようなグラフは IP アドレス、あるいはドメイン名の数は多いものの、全体のグラフ数から見ると少ない存在であ

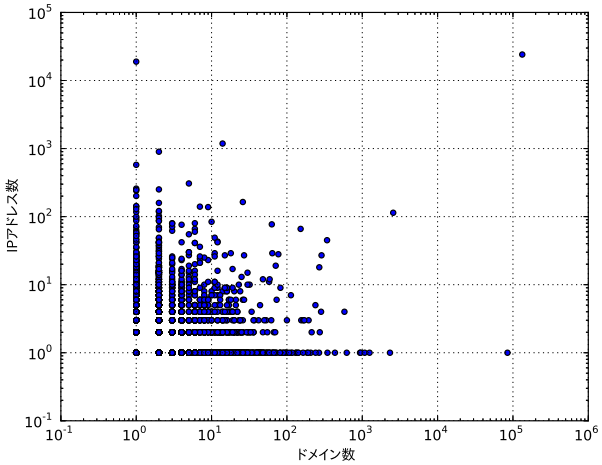

図 8: グラフ構造によって分割した各グラフが持つドメイン名数と IP アドレス数の分布

ると言える。そのため、このようなグループを例外的に処理することによって、取得した IP アドレスの監視、遮断の False Negative を低減させる手法を提案できるのではないかと考えられる。また、同一のグループに所属していることから、ドメイン名などに一定の規則性を持つと見られるため、DNS トラフィックの特徴から悪意あるドメイン名を発見する手法 [4] との連携も可能性として考えられる。

4 考察

本稿でとりあげた悪性ドメイン名は例外的に大量の IP アドレスやドメイン名を持つものが含まれており、かつドメイン名の利用停止のような状況が調査期間中に発生していたが、これらの事象を除けば安定して使われているドメイン名が多いと言える。Fast Flux[3] に代表されるドメイン名の悪用方法などでは IP アドレスが急激に変化したり、大量のドメイン名を短期間で乗り換えていくような悪意あるユーザの活動が考えられたが、本稿での調査結果ではそのような活動をしているドメイン名は一部であることが明らかになった。これは Fast-Flux のような手法を用いることで防御側からの追跡性を下げることができる一方、大量の悪性ドメイン名やボットネットを運用するのにもコストがかかるため、攻撃者側もそのようなバランスをとった結果なのではないかと考えられる。

今後の課題として、取得する情報や問い合わせ頻度の精査が必要であると考えられる。取得する情報は IP アドレスだけではなく、TTL や Name Server(NS) レコード、ゾーン情報なども分析の対象となりえる。また、WHOIS 情報もあわせて定期的に確認することによって、ドメイン名間の明確な関係を調査できるようになると期待される。一方、取得する IP アドレスの種類についても注意を払う必要がある。本稿での調査でも取得した IP アドレスには大手のクラウドサービスのものが多く含まれていた。通常のクラウドサービスでは悪意のないユーザの利用も多く考えられ、IP アドレスの再利用によりそれらのホストの IP アドレスと混同してしまう可能性がある。さらに、大手クラウドサービスでは広い IP アドレスレンジを所持している場合が多く、特に Contents Delivery Network(CDN) サービスを利用されていた場合、アドレスが頻繁に変わる原因になりうる。そのため、対象となる IP アドレスやネットワークの性質も考慮する必要がある。また、本稿では 1 箇所から取得したブラックリストを利用していたため、今後はより広い範囲のリストに着目する必要がある。

5 まとめ

本稿では、悪意あるユーザが利用していると見れるドメイン名を調査し、実際に機能していた 307,929 件について 4ヶ月間情報を取得し続け、名前解決によって得られた IP アドレスがどのように活用できるかという観点から分析を実施した。その結果、例外的にドメインが一斉に使用されなくなるケースや多数のドメイン名、IP アドレスを有するケースがみられたが、半数以上のドメインは長期的に安定して長期的に運用されていた。また、これらは限られた IP アドレスをもちいて運用されていたため、名前解決によって得られた IP アドレスを利用することで、IP アドレスベースの監視や遮断に一定の効果が見られるのではないかと期待される。今後はより広い範囲や多様な情報を用いて、悪意あるユーザによる活動を抑制するための知見を

得られるよう、調査、分析を継続する。

参考文献

[1] I.T. Mate, hpHosts, 2005, http://www.hosts-file.net/

[2] Node.js, 4 Joyent, Inc, http://nodejs.org/

[3] William Salusky, Robert Danford, “Know Your Enemy: Fast-Flux Service Networks”, July, 2007 http://www.honeynet.org/papers/ff/

[4] Leyla Bilge, Engin Kirda, Christopher Kruegel, Marco Balduzzi, Sophia Antipolis, “EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis”, ACM Trans. Inf. Syst. Secur., Apr. 2014.