3Z-1

DNSクエリデータの解析によるクエリパターンのクラス分類

風戸　雄太†　　福田　健介‡　　菅原俊治†
早稲田大学†　　国立情報学研究所‡

1　はじめに

近年，インターネット利用者の増加によりDNSトラフィック量も増加しており，DNSの状況と挙動を把握することはネットワークの健全な運用のために不可欠である．

このようなDNSの計測手法には，権威DNSサーバで計測をする方法とDNSキャッシュリゾルバで計測する方法がある．権威DNSサーバで計測をした研究として，文献[1]では権威DNSサーバの最上位の1つであるFルートDNSサーバに到着するクエリを解析し，ルートサーバに到着する問い合わせクエリのほとんどは無意味な繰り返しクエリであることを示した．また文献[2]では，.comと.netの権威DNSサーバのデータを解析し，悪質な動作をするドメインと正常なドメインをクラスタリングによる分類を行っている．一方DNSキャッシュリゾルバで計測した研究として，文献[3]ではWeimerが提案したパッシブ計測手法[4]を用いて，大学構内のローカルキャッシュリゾルバから送信されるDNSクエリを収集・解析し，スパム対策システムが存在しないドメインへのクエリを大量に送信していることを明らかにした．しかし，国内ではこのような調査は充分されていない．

そこで本研究では，パッシブ計測手法を用いてインターネットバックボーンで収集したDNSクエリパケットを解析し，統計的なデータの解析およびDNSの回答がエラーであったクエリパターンの分類を行った．その結果，ネットワーク上の一部のDNSキャッシュリゾルバの異常検出と，エラークエリの90.7 %について分類し，それらの原因が，ウイルス・スパム対策，正しくないTLDの使用，サーバの設定ミスなどであることが明らかになった．

2　DNSクエリ解析手法

2.1　データ測定

解析に使用するDNSクエリデータとして，学術情報ネットワークSINET内のバックボーンリンクの1つに設置された測定用マシン上で，UDP port 53を含むパケットをtcpdumpコマンドを用いて収集した．測定期間は2012年7月25日14時-7月27日14時と10月16日14時-10月23日14時である．この期間に保存した全UDPパケット数は5,459,975,555であった．

2.2　統計データ解析

(a) 測定環境のネットワーク内のキャッシュリゾルバとネットワーク外の権威DNSサーバ間，(b) 測定環境のネットワーク外のキャッシュリゾルバとネットワーク内の権威DNSサーバ間のDNSクエリについて，それぞれ次の4項目の解析を行う．

1. 問い合わせクエリ・応答クエリのトラフィック変動
2. 問い合わせクエリのクエリタイプ
3. 応答クエリの回答率・回答エラーとクエリタイプ
4. DNSサーバ・キャッシュリゾルバのクエリ・IPアドレス

2.3　クエリパターン分類

クエリパターン分類の対象としたクエリは，順引き問い合わせの回答成功クエリとName Errorの回答エラークエリである．対象であるクエリの問い合わせ名を1日分ごとにデータとしてまとめた．クエリパターン分類規則は，文献[1]を拡張したものを用いた(表1).

表1: 問い合わせ名のクエリパターン分類規則

分類番号	分類パターン
分類1	ウイルス対策ソフトのドメイン名である
分類2	未登録のTLDのドメイン名である
分類3	Spam対策関連のドメイン名である
分類4	dlv.isc.orgで終わるドメイン名である
分類5	local,wpadなど設定用の単語を含む
分類6	RFC1018等の規則に違反する文字を含む
分類7	ランダム性が高いと判断された文字列を含む
分類8	ローカルネットワークで使用する単語を含む
分類9	(IPアドレス)+(TLD)とTLDの重複があると判断されたドメイン名である
分類10	分類1-9以外の特徴的な文字列を含む

3　結果と考察

10月16日14時-10月23日14時のDNSトラフィック変動を図1に示す．SINET内キャッシュリゾルバの問い合わせクエリタイプ上位4項目はA(61.4 %), AAAA(24.8 %), PTR(10.2 %), MX(1.1 %)，SINET外キ

An analysis and classification of DNS queries.
†Yuta KAZATO　†Waseda university.
‡Kensuke FUKUDA　‡NII.
†Toshiharu SUGAWARA

ャッシュリゾルバは A(61.9 %), AAAA(15.4 %), PTR(8.3 %), MX(8.3 %) であった．図 1 のラベル A(10/22/17-18) では，18 分間で 1 つのキャッシュリゾルバから 15,716,086 のクエリが 1 つの権威 DNS サーバに送信される異常が確認された．

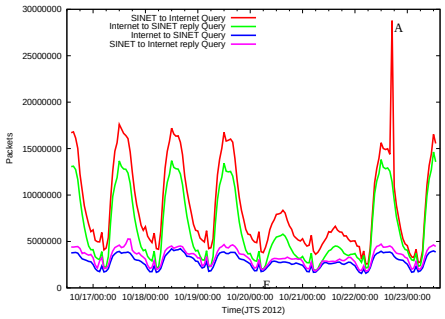

図 1: パケット数推移

一方，応答クエリの回答エラーは，SINET 内キャッシュリゾルバは回答成功率 31.1 %，回答エラー率 13.7 %，SINET 外キャッシュリゾルバは成功率 68.2 %，エラー率 12.9 %であった．SINET 内キャッシュリゾルバのエラー上位 3 項目であった回答エラー数の変化を図 2 に，SINET 外のキャッシュリゾルバの回答エラー数の変化を図 3 に示す．SINET 内キャッシュリゾルバでは，図 2 の B(10/17:13-14) における Server Error のクエリの内，98.8 %が 1 つの権威 DNS サーバから 2 つのキャッシュリゾルバに対するエラークエリであった．このような DNS クエリの異常現象は図 2 の C(10/18:20-21) と図 3 の D(10/19:7-8) での Refused Error でも確認された．また Name Error では，SINET 内キャッシュリゾルバへ I と K のルート DNS サーバからのエラー，SINET 外キャッシュリゾルバへ SINET 内に存在する JP DNS サーバからのエラーが多数存在した．これらの結果は，ごく少数のキャッシュリゾルバの振る舞いが DNS トラフィックに大きな影響を与えることを示している．

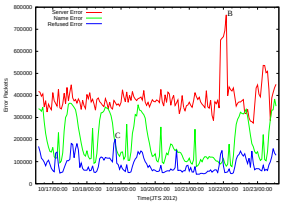

図 2: SINET 内キャッシュリゾルバエラー統計

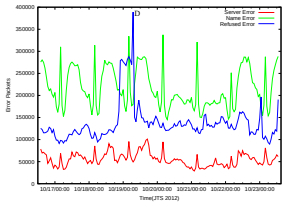

図 3: SINET 外キャッシュリゾルバエラー統計

次に，表 1 のクエリパターン分類により，6 日分の各データの SINET 内キャッシュリゾルバのクエリの問い合わせ名を分類した平均値を表 2 に示す．表 2 の分類結果より，回答エラーの 90.7 %を分類し，ウイルスやスパム対策に関連するクエリパターンが半数近く (52.2 %) を占めていることがわかった．また，問い合わせ名の TLD が正しくないクエリパターン (20.2 %) や，設定ミスにより回答エラーとなったクエリパターン (17 %) が存在し，これらが回答エラーの主な原因と言える．

表 2: クエリパターン分類結果

分類規則	回答エラーデータ		回答 OK データ	
	個数	(%)	個数	(%)
初期数	845,985		1,645,552	
分類 1	270,296	32.0	63,759	3.87
分類 2	269,824	31.9	208	0.01
分類 3	171,268	20.2	17,390	1.06
分類 4	57,927	6.8	68	＞ 0.01
分類 5	31,998	3.8	456	0.03
分類 6	23,112	2.7	482	0.03
分類 7	55,122	6.5	1,017	0.06
分類 8	34,169	4.0	69	＞ 0.01
分類 9	45,271	5.4	1,116	0.07
分類 10	2,500	0.3	13	＞ 0.01
分類数	766,956	90.7	84,221	5.12

4 まとめと今後の課題

本研究では，パッシブ計測による DNS クエリの収集・解析とクエリパターンの分類によって，DNS クエリの異常送信を行うキャッシュリゾルバの検出が可能であり，SINET のインターネットバックボーンでのクエリ送信方向による DNS トラフィックの傾向の違いと，ウイルス・スパム対策やサーバの設定ミスなど，特定の原因によって回答エラーのクエリが生成されることがわかった．この手法を応用していくことで，ネットワーク内の DNS の異常検出や，回答エラーとなるクエリの送信を防ぐことが可能である．今後の課題として，クラス分類の規則細分化による精度向上と，より現実的な異常検出システムを目指すために，DNS クエリ解析から得られた特徴量データを利用したクラスタリング手法によって，悪質な攻撃や異常な動作をする DNS サーバやキャッシュリゾルバの異常検出を進める予定である．

参考文献

[1] D. Wessels and M. Fomenkov. Wow, that ' s a lot of packets. In *Proceedings of Passive and Active Measurement Workshop (PAM)*, 2003.

[2] S. Hao, N. Feamster, and R. Pandrangi. Monitoring the initial DNS behavior of malicious domains. In *Proceedings of the 2011 ACM SIGCOMM Internet Measurement Conference (IMC'11)*, pp. 269–278, 2011.

[3] B. Zdrnja, N. Brownlee, and D. Wessels. Passive Monitoring of DNS Anomalies. In *Conference on Detection of Intrusions Malware, and Vulnerability Assessment (DIMVA) 2007*, pp. 129–139, 2007.

[4] F. Weimer. Passive DNS replication. In *Conference on Computer Security Incident Handling (FIRST'05)*, 2005.