CL-005

ダークネット観測における 特徴的な通信を持った送信元の挙動調査に関する研究

Investigating the behavior of sources with characteristic communications in darknet observation

ファン・アン・ソン † 中村 康弘 †
Son Pham Anh Yasuhiro Nakamura

1. 研究背景と目的

インタネット上では不特定多数の IP アドレスへ多数の不審なパケットが送信されている。その傾向とその中の脅威を早期に発見するために、未使用 IP アドレス空間であるダークネットで着信パケットを観測する。ダークネットへ着信するパケットはいくつかの理由で送信されると考えられる。アドレスの誤入力などの人為的誤り，マルウェアによる感染対象の探索活動，ボットネットや手動操作によるアドレススキャンやポートスキャンなどである。ダークネット観測システムではパケットのヘッダ情報から、送信元，送信先，ポート番号等のヘッダ情報しか得ることができない。このため、それらのパケットの特徴からその影響を推定することはできるが、送信者の真の意図不明である。また、NICTER の年間観測レポートによると、近年ではパケット量が急増する一方であり、送信者の意図を推移することが一段と困難になっている [1]。送信者の意図を推定するためには、送信側の挙動を観測しつつ送受信をエミュレートするハニーポットが有効であるが、多数のアドレス空間での同時観測はコストが高くなる。そこで、この研究では、最初の接続要求に対して擬似応答を行い、初期ペイロードのみを取得する観測システムを用いる。初期ペイロードを分析することで、パケットヘッダ情報よりも多くの情報を得ることができ、送信者の意図の推定に利用できる。

多数のアドレスを利用可能な送信者は、その走査効率を上げるために、複数の送信元アドレスを同時並行的に利用して、同一の走査プログラムを実行する場合がある。このような走査パケットの送信元アドレス間の相関を時系列的に観測・分析した研究は少ない。これは多数の送信元アドレスごとに長期にわたる多量の通信履歴を時系列で分析するには極めて多量の計算資源を必要とするためであると推測できる。そこでこの研究では、何らかの明確な意図を持った送信者が複数のアドレスを同時並行的に利用して走査活動を行う場合、そのアドレスの範囲は特定の AS 内に限られる場合が多いものと推定し、特定の AS 内の複数のアドレスから一定時間内に同一のペイロードを送信するような走査活動に着目する。このようなアドレスグループの挙動をパターン化して類別する手法を提案し、実データを用いた結果を示す。この結果を用いることにより、不正アクセスの兆候や走査意図の分析に利用できるものと期待できる。

†防衛大学校 理工学研究科

2. 関連研究

これまでに行われているダークネット観測に関する主な研究は、走査活動，可視化，協調関係の送信元検出，分散型スキャンの検出などがある。鈴木ら [2] はダークネットへ到達するパケットの初期ペイロードを Fuzzy Hash 値を算出した。その値とパケットのヘッダ情報を入力データとして自己組織化マップ (SOM) を作成し、初期ペイロードを種類毎に分類した。しかし、一部のペイロードは正しく分類できなかった。ハッシュ化した値以外の入力ベクトルを追加する必要があると述べている。さらに、鈴木ら [3] は簡易ハニーポットによるダークネット観測結果のペイロード分類手法も提案した。

一方、初期ペイロードを得ることでダークネットへの走査活動を分析する研究がある。中村ら [4] は観測中のセンサーアドレス空間を全走査するような一連の活動を抽出し、その走査順序を類型化するとともに、送付されたペイロードを取得することで、走査活動や意図の推定を行った。これにより全アドレスを走査する送信元は一部であり、走査パケットの密度を上げないように走査パケットの時間間隔を長めにした上で、隣り合う宛先アドレスの差分が一定にならないよに順序を変更しているものが多いことを明らかにした。

笹生ら [5] はダークネット観測結果から 24 時間毎の宛先アドレス数、パケット数による通信パターンの種別と OS フィンガープリントによる送信元を分類した。ダークネット全空間に対するスキャンを行う送信元の数は非常に少ないが、半分のパケットはこれらのホストから発生していることがわかった。また、1 回しか送信してこなかった送信元の数は非常に多かったが、着信したパケット全体の 4%にすぎないことを明らかにした。

さらに、OS による分類結果は Windows と Linux2.4.x からのパケットが支配的であるとわかった。しかし、この分析法によると、24 時間以上の周期で通信するホストや時間をずらしながらスキャンを行う攻撃者の送信を捕捉することができない。

山門ら [6] は自組織内からダークネットアドレスへ送信するパケットは不適切な通信とみなし、通信ログと擬似応答により取得できた初期ペイロードを元にその通信を発生させたアプリケーションを自動的に特定する研究を行った。

3. 走査パケットの送信パターンの分類

攻撃者が継続的に走査活動を行うとき、対象のネットワーク管理者に検知・除外されにくいように様々な工夫が行なわれている。しかし、どうのような送信パターンを用いても、対象アドレス範囲全体をスキャンする目的を持つ限り、目的の宛先アドレス数と送信時間の関連性からペイロードごとの送信パターンを類別することが可能である。さらに、ダークネット内の複数アドレス範囲の同時観測が可能であれば、特定の攻撃者が対象アドレス範囲をどのような順序で走査したのかもわかる。

例えば、ある特定の宛先を標的とした場合、連続的に送信するものもあれば時間間隔を開けて送信を行うものもある。複数の宛先を標的とした場合、アドレス範囲を連続的に送信するものもあれば時間と宛先をずらしながら送信を行うものもある。特定の組織の全アドレス空間を標的としたとき、全空間を一気に走査するものもあれば、時間間隔を開けて宛先とその数を変更しつつ送信するものもある。これらの送信パターンの差異は、送信プログラムの宛先アドレスの選択アルゴリズムに依存して決定されるので、同一のプログラムを用いていれば、そのパターンやペイロードが一致あるいは類似するものと推定できる。また、個々のペイロードの送信パターンは、走査の意図や効率を考慮して決定されるものと考えられる。このため、これらのペイロードごとの送信パターンを分類することで、送信意図の推定に役立つ情報を提供できる。

この研究では、送信パターンを以下の尺度で分類できるものと仮定する。それぞれについては表 1 に定義をする。

4. 提案手法

ペイロードごとの送信パターンを分類する手順を次に示す。

4.1. 前処理

- ステップ 1： D 日分の観測データを 1 日ごとに分割。
- ステップ 2：1 日ごとのデータを 10 分間を単位時間としてパケットを分割し、次の情報を抽出する。ペイロードのハッシュ値 (Hash)、AS 番号、宛先 IP アドレス、送信元 IP アドレス、送信時間。
- ステップ 3： 1 日の前後比較で特定 AS からしか送ってこないペイロードを抽出する。
- ステップ 4： それぞれのペイロードについて以下の情報を求める。ペイロードのハッシュ値 (Hash)、AS 番号、宛先のアドレス数 (Dst)、最初の単位時間における宛先アドレス数 (F_Dst)、時間単位の送信回数 (Time)、最大送信可能時間単位数 (M_Time)。

4.2. 送信パターンの分類

パターン分類の段階では前処理で作成したデータを用いて、ペイロードごとに 3 つのパラメータを持つベクトルを作成する。各ベクトルを表 2, 3 の条件で判定する。

以上のような判定条件により特徴ベクトルを作成し、これにより分類を行い、3 つのパラメータの組み合わせで送信パターンの種類は表 4 となる。

5. 実験と結果

5.1. 実験対象データ

実データを用いて送信パターンの分類を行うために、既存の AS に割り当てられたアドレス範囲で 2017 年 1 月 01 日から 2017 年 1 月 10 日の 10 日間に観測されたデータを用いた。データの容量は約 10GB である。観測対象の未使用の IP アドレス数は約 1400 である。

5.2. 実験結果

ペイロードごとの送信元 AS 番号と送信パターンを類別した結果を表 5 に示す。

また、表 6 はそれぞれのパターンの AS 数と割合である。Type 11-S1 の数が極めて多数でほとんどを占め、Type 12-SM が次いで多いことがわかった。Type 313-A1A, Type 323-AMA, Type 331-AA1 と Type 333-AAA は出現しなかった。ダークネット全空間へ送信されたペイロードとその送信元 AS が存在したことが確認できた。この結果から特定のひとつのアドレスを宛先とする通信がほとんどであることがわかった。

表 1: ペイロードごとの送信パターンの類別

送信パターン	挙動	記号
一つの宛先に一回のみ送信する	Single Destination one time	S1
一つの宛先に時間を開けて繰り返し送信する	Single Destination Mutiple time	SMA
一つの宛先に連続的に送信する	Single Destination All time	SA
複数宛先へ一回のみ送信する	Multiple Destination one time	M1
複数宛先へ時間を開けて繰り返し送信する	Multiple Destination Multiple time	MM
複数宛先へ連続的に送信する	Multiple Destination All time	MA
最初は一つの宛先 IP アドレスのみへ送信し、時間を開けて宛先を変えながら全空間へ送信する	All Destination from One with Multiple time	A1M
最初は一つの宛先 IP アドレスのみへ送信し、連続的に宛先を変えながら全空間へ送信する	All Destination from One with All time	A1A
最初は複数の宛先 IP アドレスへ送信し、時間を開けて宛先を変えながら全空間へ送信する	All Destination from Multiple with Multiple time	AMM
最初は複数の宛先 IP アドレスへ送信し、連続的に宛先を変えながら全空間へ送信する	All Destination from Multiple with All time	AMA
全宛先 IP アドレスへ一回のみ送信する	All Destination with All One time	AA1
全宛先 IP アドレスへ送信後、時間を開けて繰り返す	All Destination with Multiple Time	AMA
全宛先 IP アドレスへの送信を連続的に繰り返す	All Destination with All Time	AAA

表 2: 宛先アドレス数の判定条件

宛先 IP アドレス数	第 1 と 2 のパラメータ
宛先数 = 1	-1, -1
1 宛先数 All	0, 0
All 宛先数/1	1, -1
All 宛先数/Ran	1, 0
All 宛先数/All	1, 1

表 3: 時間単位による送信頻度の判定条件

送信時刻数	第 3 パラメータ
Time または M_Time = 1	-1
Time M_Time	0
Time = M_Time	1

表 4: 時間による第 3 パラメータの判定条件

ベクトル	タイプ	説明
(-1, -1, -1)	Type 11	S1
(-1, -1, 0)	Type 12	SM
(-1, -1, 1)	Type 13	SA
(0, 0, -1)	Type 21	M1
(0, 0, 0)	Type 22	MM
(0, 0, 1)	Type 23	MA
(1, -1, 0)	Type 312	A1M
(1, -1, 1)	Type 313	A1A
(1, 0, 0)	Type 322	AMM
(1, 0, 1)	Type 323	AMA
(1, 1, -1)	Type 331	AA1
(1, 1, 0)	Type 332	AAM
(1, 1, 1)	Type 333	AAA

5.3. 送信パターンの検証

ここでは、上で分類された送信パターンのそれぞれについて、実パケットの送信時刻を参照し、想定した送信パターンでパケットが着信していることを検証する。

- Type 13 - SA の検証
  図 1 は、観測対象全時間を横軸、宛先のダークネットのアドレス空間を縦軸とし、Type 13 に分類された AS 番号・着信パケットをプロットしている。この結果、AS 1722 から送信されたペイロード（ハッシュ値は 7fbd68b1d6996c9e2c8875c57eac4961）がひとつの宛先のみに連続して送信しているため、分類されたグループ特徴に一致することが確認できた。
- Type 22 - MM の検証
  図 2 は、AS 63199 から同一のペイロードが複数の宛先へ時間を開けて複数回送信してきている様子が確認でき、分類されたパターンの特徴と一致した。
- Type 23 - MA の検証

表 5: 実験結果

ハッシュ値	AS	分類結果
7321d0e452f186b7636ff8d1a8e7539a	4837	Type 312
0357387366776bedfe9d173570a83b38	3462	Type 22
2e3e48b6b8732a8d70b6e6f5709f3162	3462	Type 13
f4368b29058ab813bd86c99afe572761	42003	Type 22
305165b7cf5ccab83d3a8daf71d06e63	45090	Type 331
…	…	
668d4392e5a28cc1f6f0f2f77a9e7e35	10439	Type 22
66cd587d27975a1834725ffd646a1d09	3462	Type 22
e73913fee76f8ccec9b19ceaff95e977	4134	Type 22
d67600889d9219b53b55d35b654babe3	4134	Type 23
7afdd3b033e920ed81f69330872576b8	38019	Type 22
…	…	

図 3 では、AS 9808 から同一のペイロードが複数の宛先へ連続的に送信している様子が確認でき、分類されたグループ特徴を示していることがわかる。

- Type 322 - AMM の検証
  図 4 では AS 4134 からの特定のペイロードの送信状況を表す。複数の宛先へ複数回送信しており、全空間を走査したことから、分類結果の Type 322 に当てはまる。
- Type 332 - AAM の検証
  図 5 では AS 4816 から同一のペイロードの着信

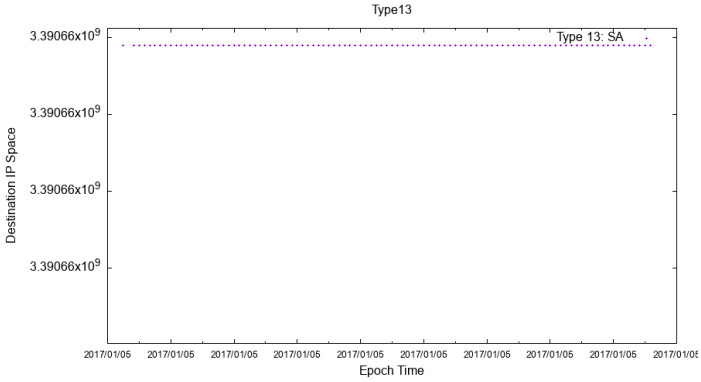

図 1: Type 13 hash=7fbd68b1d6996c9e2c8875c57eac4961

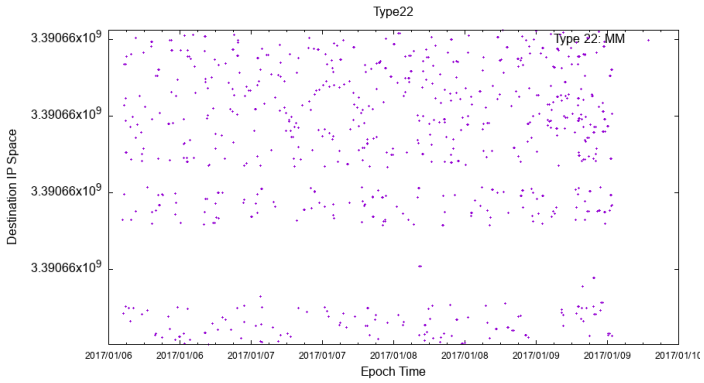

図 2: Type 22 hash=46fb59e6f55a684ac47f48e446ff2550

表 6: 各種類の割合

タイプ	数
Type 11 - S1	5552984
Type 12 - SM	19502
Type 13 - SA	486
Type 21 - M1	817
Type 22 - MM	1505
Type 23 - MA	25
Type 312 - A1M	2
Type 313 - A1A	0
Type 322 - AMM	17
Type 323 - AMA	0
Type 331 - AA1	0
Type 332 - AAM	2
Type 333 - AAA	0
総数	5575340

の様子を表しており、Type 332 - AAM の送信パターンに当てはまることがわかる。

6.AS ごとの送信パターンとその特徴

6.1.AS ごとの送信パターン

ペイロードの送信パターンから、各 AS の送信パターンを類別することができる。ここでは、Type11, 12,21,22,31 の送信パターンを除いて分類した。

表 7: AS の送信パターンの類別結果

AS 番号	送信パターン
2516	322
3462	13,23
4134	322,13,23
27699	13
...	...
56044	13
4808	322,13
23650	322

表 8: AS の送信パターンの分類

送信パターン	AS 数
Type 13	48
Type 23	34
Type 322	12
Type 332	1

6.2. 全空間をスキャンする AS の特徴

AS 42570 は 5 つのペイロードを使って Type 322 の送信パターンに類別された。それぞれの送信元アドレス数は 94, 90, 105, 95, 94 であった。これら 5 つのペイロードの着信時刻と着信アドレスの関係を図 6 に示す。

AS 4134 は 1 つのペイロードを 144,511 回送信し、送信パターンは Type 322 であった。送信元アドレス

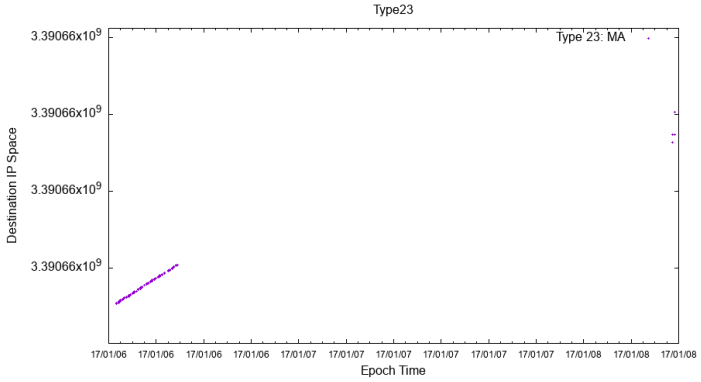

図 3: Type 23 hash=7e3025a5880f464d69f7b668e67dd0d1

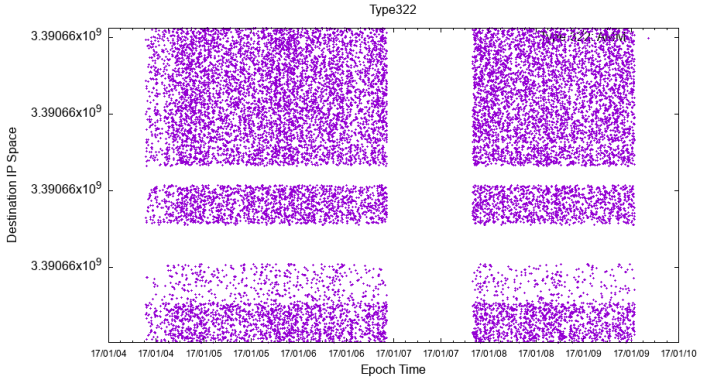

図 4: Type 322 hash=06fca64621482c286d415519070985e9

表 9: AS の送信パターンの分類

AS	ペイロード	送信元数
42570	751ff3d9ba93b16ce7a3101df5978e79	105
4808	5cb31ff7a56f8fab6f44cdcac2b572a7	54
6939	af1ffdbde3533831eefadab495762971	11
4134	06fca64621482c286d415519070985e9	2
4816	5649e9b4f150c743d89c697aa10350c0	1
36375	76ce75b86913aa438293e28e36766885	1
36375	32dceb058d6d439d82f1acb13a83a265	1
36375	27b9cb26823baabb70adebb34e95a7b5	1
45758	aa36aab0a265203de2bc8557a1283ec4	1
2516	dc3a4054a03d445cb26ff184d1fe0819	1
23650	ee939be90b8e904dd26d57da04b7af8b	1

数は 2 であり、時間・宛先アドレスの関係を図 4 に示す。AS 4134 は 1 つのペイロードを 299,482 回送信し、そのパターンは Type 322 であった。送信元アドレス数は 54、時間・宛先アドレスの関係を図 7 に示す。

該当するすべての AS の送信パターンは、表 9 のように分類することができた。全空間をスキャンする AS の送信元のアドレス数は少数であることが確認ができる。

7. まとめ

本研究により、ある特定の AS のみから送信されたペイロードの送信パターンを分類することができた。さらに各 AS が行う走査活動の送信パターンを分類することができた。全観測アドレス空間を走査した AS についてはその送信元アドレス数を調査した。しかし、これらのペイロードの危険性や走査の意図を推定するためには、さらにペイロードの内容を分析する必要がある。今後、ペイロードの内容に基づいた尺度を用いるとともに、分類処理の自動化を目指す。

参考文献

[1] 情報通信研究機構 (NICT)、NICTER 観測レポート、https://www.nict.go.jp/press/2021/02/16-1.html

[2] ダークネットへ到達するパケットの初期ペイロードの自己組織化分類手法鈴木悠太、中村康弘、Computer Security Symposium 2015, 2015/10/21-23.

[3] 簡易なハニーポットによるダークねっt お観測とペイロード分類手法の提案鈴木悠太、後藤　洋一、中村康弘、情報処理学会第 77 回全国大会、4X-04、3-483 483 ページ、2015/03.

[4] 宛先アドレス順序とペイロードに着目したネットワーク走査活動簿分析中村康弘、梶川慶太、芦野佑樹、鮫島礼佳 Computer Security Symposium 2018, 2018/10/22-25.

[5] 通信源ホストの分類を利用したダークネット通信解析笹生憲、森達哉、後藤滋樹、Computer Security Symposium 2013, 2013/10/21-23.

[6] ダークネットあて通信分析によるネットワーク管理者支援、山門彩、佐藤聡、 新城靖 、情報処理学会研究報告 Vol.2018-IOT-40 No.31, 2018/3/6.

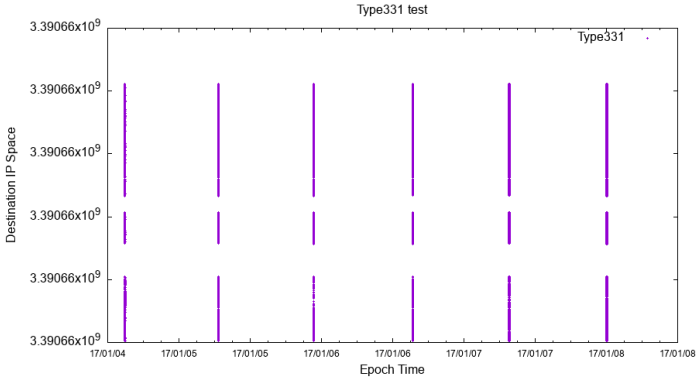

図 5: Type 332 hash=3a86697e08de45f9483e8ed81c4f5735

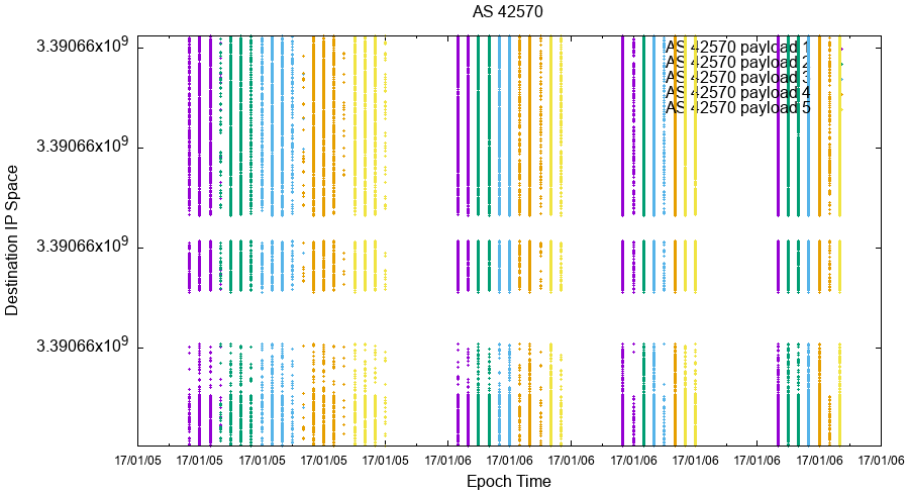

図 6: AS 42570

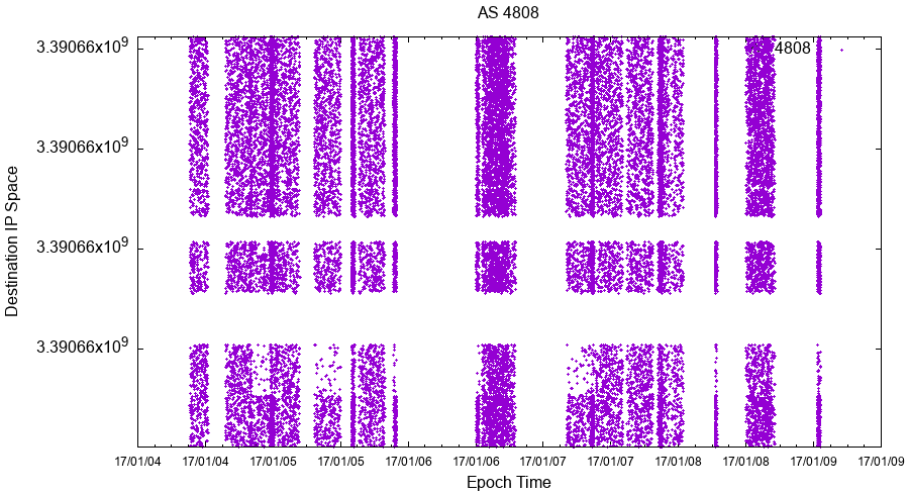

図 7: AS 4808