How dynamic are IP Addresses?

Benjamin Vahl
(damocles@cs.tu-berlin.de)

Seminar *Internet Routing* ,
Technische Universität Berlin

WS 2007/08 (Version vom 18. Januar 2008)

Inhaltsverzeichnis

Zusammenfassung

In diesem Paper wird ein neuartiger Algorithmus namens *UDmap* vorgestellt, der dem Identifizieren dynamischer IP-Adressen durch Anwendung verschiedener Analyseverfahren dient. Es wird darauf eingegangen, inwiefern dies zum Beispiel beim Filtern von Spam-mails verwendet wird und mit konventionellen Verfahren verglichen, wobei die Vorteile gegenüber bisherigen Techniken wie zum Beispiel DNS Blacklisting (DNSBL) herausgehoben werden.

1 *UDmap* Motivation

Die Motivation zum Untersuchen der charakteristischen Eigenschaften dynamischer IP-Adressen entstand für die Entwickler des *UDmap* Algorithmus aus statistischen Analysen, die belegten, dass derzeit etwa 95% aller Versender von Spam-Mails Mailserver benutzen, denen dynamische IP-Adressen zugewiesen sind. Desweiteren machen diese Server, die zu diesem spezifischen Zweck angelegt wurden, einen Groteil des gesamten Junkmail-Aufkommens (in diesem Fall beziehen sich die Autoren auf von Hotmail-Benutzern empfangene Spam-Mails) ausmachen. Da kontextbasierte Spam-Filter-Systeme relativ unzuverlässig sind und immer wieder durch diverse Methoden auszutricksen sind, gewinnt das Blockieren als für Spam-Versand bekannt gewordener IP-Adressen immer mehr Interesse. Hierbei ist anzumerken, dass diese Thematik neben Spam auch Möglichkeiten zum Vorgehen gegen Bot-Netzwerke, Virenverteiler oder Phishing-Seiten[1] abdeckt (welche auch mit hoher Wahrscheinlichkeit auf - möglicherweise kompromittierten - Hosts hinter dynamischen IP-Adressen laufen), auch weniger defensive Aufgaben wie das Erfassen von Webseiten durch Webcrawler, die möglichst Inhalte auf Servern ignorieren sollten, von denen aufgrund der dynamischen IP zu erwarten ist, dass sie nicht langfristig bestehen werden. Dagegen sollte nicht pauschal jede individuelle dynamische IP ge-blacklisted werden, da die Identitäten hinter diesen oft wechseln, wie zum Beispiel bei den Kunden von DSL-Anbietern, die bei jeder Neueinwahl eine beliebige freie IP-Adresse aus einem Pool verfgbarer Adressen zugewiesen bekommen.
Problematisch dabei ist, dass herkömmliche Umsetzungen davon (→ Kapitel 2: *Alternative Verfahren*) auf statischen Blacklisting-Diensten bestehen, deren Tabellen manuell gewartet werden müssen, was nicht nur einen enormen Aufwand verursacht, sondern zudem sehr lückenhaft und unzuverlässig ist. Hier kommt *UDmap* ins Spiel, welcher den ersten Ansatz zum automatischen Identifizieren von dynamisch zugewiesenen IP-Adressen liefert. Weitere immense Vorteile hierbei sind, dass als Hauptinformationsquelle von *UDmap* Inhalte von Logfiles dienen und kein Sammeln von Daten ausserhalb des Systems notwendig ist (→ Kapitel 3: *Vorteile von UDmap*). Darüberhinaus liefert der Algorithmus eine feingranularere Kategorisierung der Adressen als bisherige, statische Verfahren und verlangt vergleichsweise kaum Wartungsarbeit. Das Ziel der Autoren ist es, die Wichtigkeit und Möglichkeiten, die das automatisierte Identifizieren dynamischer IP-Adressen bietet, zu propagieren und einen Ansatz zum Verständnis der zugrundeliegenden Dynamik zu bieten.

[1] gefälschte Websites, die das Aussehen des Originals nachahmen, mit dem Ziel, Benutzern bestimmte Daten zu entlocken

2 Alternative Verfahren

Es gibt derzeit keine Verfahren, die auf demselben Prinzip wie *UDmap* arbeiten und überhaupt von der Annahme ausgehen, über hinter dynamischen IP-Adressen liegende Mail-Server versandte Mails wären gröβtenteils Spam. Alternative netzwerkbasierte Filtermethoden verwenden Blacklists, in denen bereits auffällig gewordene und als solche gemeldete (!) IP-Adressen oder -bereiche erfasst sind.

2.1 Blacklisting-Dienste

Bisherige Ansätze zum Klassifizieren von IP-Adressen basierten (und basieren noch) auf der Verwaltung immens groβer Tabellen, welche von den Betreibern des Dienstes gewartet und von einer Vielzahl von Personen (insbesondere Administratoren gröβerer Netzwerke) äktuellgehalten werden, die IP-Adressen oder -Blöcke selbst dort eintragen oder entfernt[2] werden können. Im Paper beziehen sich die Autoren auf den Dienst *Dynablock*, der zum gegenwärtigen Zeitpunkt nicht mehr unter dem ursprünglichen Namen läuft, sondern als *Spamhaus Project*[2][3] betrieben wird, dass auf dem Dynablock-Prinzip aufbaut und in mehrere Kategorien (Spamhaus Blacklist, Policy Blacklist) unterteilte Listen dynamischer IP-Blöcke/Adressen verwaltet. Ist eine Adresse fälschlicherweise gemeldet worden, besteht die Möglichkeit sie per Hand über eine entsprechende Funktion auf der Seite aus den Listen zu entfernen, wonach diese Änderung nach spätestens 15 Minuten wirksam wird.
`Beispiel:` Suchen einer IP-Adresse aus einer Spam-Mail in den Blacklists:
`Received: from host189-188-dynamic.180-80-r.retail.telecomitalia.it (80.180.188.189)`

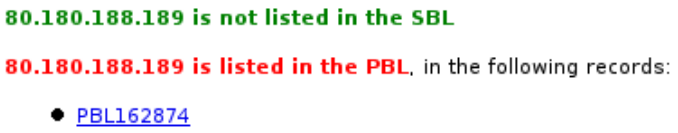


Abbildung 1: Prüfen einer IP auf Vorhandensein in einer Blacklist

2.2 Nachteile

Im Januar 2007 umfassten die Dynablock-Listen 192 Millionen dynamische IP-Adressen. Diese manuell zu warten verursacht nicht nur einen unglaublichen Aufwand. Die Effektivität steht und fällt mit dem Eintragen und Melden von IP-Blöcken/-Adressen durch Systemadministratoren, was bedeutet, dass solch ein System nicht autonom funktionieren kann.
Durch die ständigen Änderungen der Internet-Topologie und IP-Zuweisung werden viele der in den von Dynablock-Tabellen gehaltenen Daten schnell redundant, wodurch das zu verwaltende Datenaufkommen zusätzlich steigt. Desweiteren sind selbstverständlich viele neu entstandenen IP-Bereiche noch nicht auf

[2] für den Fall, dass eine IP fälschlicherweise gemeldet wurde
[3] The Spamhaus Project: `http://www.spamhaus.org/`

den Blacklists erfasst. Hinzu kommt, dass die meisten Netzwerkbetreiber ihre Struktur der IP-Vergabe nicht offenlegen wollen, was das Abgrenzen der betroffenen IP-Adressen in Bereiche zusätzlich erschwert.
Um gezielt an von Spam-Sendern benutzte IP-Adressen zu kommen und diese in sogenannten *DNS Blacklists* (DNSBLs) zu benutzen, wurden Mailserver-Logs und Honeypot-Projekte[4] eingesetzt, was wiederum Spammer dazu bewegte, sich gegen solche Verfahren zu schützen, indem sie eine groβe Anzahl an Zombie-Hosts zum Spam-Versand verwenden, was einerseits für gesteigerten Durchsatz sorgt und zum anderen das Blacklisting erschwert. In einem Paper *Can DNS-based Blacklists Keep Up with Bots?*[2] berichtete der Autor Ramachandran, dass nur 6% aller von seinem Team gefundenen zu Botnet-Clients gehörenden IP-Adressen tatsächlich in bekannten Blacklists auftauchten.

3 Vorteile von *UDmap*

Um der Problematik des bisherigen statischen Blacklistings entgegenzuwirken, beschäftigten sich die Autoren des Papers umfassend mit bestimmten Charakteristika dynamischer IPs, wodurch der *UDmap* -Algorithmus nicht nur Funktionalität zum dynamischen Identifizieren von IPs bietet sondern auch in der Lage ist, präzisere Informationen über einen Host herauszufinden, etwa ob es sich beispielsweise um einen Heimcomputer, Proxy oder Internet-Cafe-Rechner handelt. *UDmap* schätzt ausserdem eine Zeit für die Benutzungsdauer einer IP-Adresse ab, die sogenannte *ip volatility* [5].
Folgende Voraussetzungen sind nötig, um den Algorithmus anwenden zu können:

- eine beliebige Log-Datei, welche Host-Identiäten (zum Beispiel über login-Namen manifestiert) bestimmten IP-Adressen zuordnet (die *UDmap* - Autoren benutzten zu diesem Zweck ein Logfile mit Hotmail-Session Informationen, welches den Zeitraum von einem Monat umfasste). Es ist entscheidend, dass der betreffende Host über eine möglichst hohe Anzahl an Nutzern verfügt, die einen Dienst regelmäβig verwenden (wie zum Beispiel ein gröβerer Mail-Provider), da *UDmap* sonst aufgrund fehlender Informationen keine vollständigen Ergebnisse liefern kann.

- spezifische Daten zur Verteilung von IP-Adressräumen, zum Beispiel BGP-Routing-Tabellen oder CIDR-IP Prefixe [6]. Hierdurch erhält der Algorithmus nähere Informationen zur Netzwerktopologie, was zum Beispiel wichtig ist, um Sub-Netze als zusammengehörigen Adressbereich (Block) ansehen zu können, dem man wiederum Charakteristika dynamischer IP-Vergabe zuordnen kann (es ist wahrscheinlich, dass wenn man zwei auseinanderliegende IP-Adressen aus einem Block als dynamisch identifiziert, alle zwischen diesen beiden auch dynamisch vergeben werden).

[4] Honeypots: zu Analysezwecken als Angriffsziel aufgestellte Server

[5] eine von den Autoren eingeführte Maβeinheit, welche die Dauer bis zur Neuvergabe einer IP-Adresse angibt

[6] CIDR: **C**lassless **I**nter-**D**omain-**R**outing, Verfahren zur besseren Einteilung des 32-Bit IP-Adressraums, gibt im Suffix die Länge der Netzklasse an, z.B.: 192.168.10.100/24 ← 24-Bit-Netzmaske, äquivalent zu 255.255.255.0

Daraus lassen sich die entscheidenden Vorteile, die *UDmap* gegenüber Blacklisting bietet, erklären:

- **generell anwendbar**, funktioniert für beliebige Logdateien, welche über die notwendigen Informationen, sprich Benutzeridentität zu IP-Adresse, verfügen, es ist also auch keine Änderung der Client-Software der betroffenen Benutzer notwendig

- läuft **autonom**, d.h. ist nicht auf Informationen anderer Domains und Administratoren angewiesen

- liefert **hohen Detailgrad**, da es nicht unbedingt ganze IP-Prefixe kategorisiert/generalisiert, sondern versucht, die Dimensionen solcher Blöcke möglichst exakt zu bestimmen. Desweiteren wird bei den gemessenen Werten versucht, false positives [7] wie zum Beispiel IP-Adressen von Proxy-Servern oder Rechnern, die vielen Personen zugänglich sind, als solche zu identifizieren und herausfallen zu lassen.

- Ergebnisse sind **aktuell** und hängen nicht davon ab, wann bestimmte IP-Bereiche als dynamisch gemeldet werden, *UDmap* bestimmt diese instantan durch Berechnung aus den bereits gesammelten Daten.

UDmap identifizierte zum Beispiel aus dem Hotmail Session-Logfile 102 Mio. dynamische IPs, von denen die meisten zu DSL-Anschlüssen gehörten. Interessanterweise tauchten davon über 50 Mio. nicht in existierenden Blacklists auf. Im Fall von Spam-Versand durch Rechner eines zu einer Universität gehörenden Subnetzes tauchte keine der eindeutig als Spam-Mail-Server erkannten Rechner (und zugehörigen IP-Adressen in den Dynablock-Tabellen auf.

[7] fälschlicherweise als zutreffend erkannte Werte, hier dyn. IP-Adressen

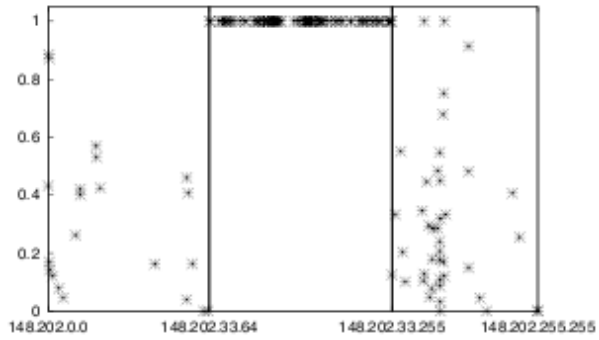


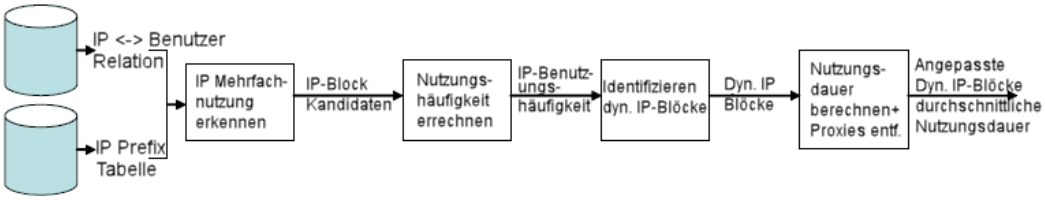
Abbildung 2: Bereich der Spam-versendenden Hosts. Hier wird der zusammenhängende Bereich dynamischer IP-Adressen innerhalb des Pools deutlich

4 Identifizieren dynamischer IP-Adressen

Das simpelste Vorgehen beim Klassifizieren von IP-Adressen ist das Analysieren des rDNS-Lookups[8]. Dabei erhält man in vielen Fällen genauere Informationen über den Host, wie von den Autoren gezeigt:
`Beispiel:` Die rDNS-Anfrage für eine IP-Adresse ergab: `adsl-dc-305f5.adsl.wanadoo.nl`, was darauf schliessen lässt, dass es sich hierbei um einen (sehr wahrscheinlich privat genutzten) ADSL-Anschluss bei einem niederländischen ISP handelt. Leider verfügen nur etwa 50-60% aller IP-Adressen über rDNS-Einträge. Dies fiel auch bei der Studie des IP-Bereiches auf, welcher einer Universität in Mexiko gehört und 65,535 Adressen umfasst. Das Hauptinteresse richtete sich dabei auf eine Vielzahl an Servern aus diesem Bereich, die regelmäßig Spam an Hotmail-Benutzer versandten. Insgesamt wurden 136 SMTP-Server gezählt, von denen 75 aussschliesslich zum Absetzen von Junk waren, was bei den anderen nur zum Teil der Fall war. Keine der betreffenden IPs kam in den Blacklists von Dynablock vor, über das rDNS-Verfahren liessen sich nur bei 33 Servern Aussagen treffen, für den Rest fehlte ein entsprechender rDNS-Record. Nur 3 der 33 Server konnten mit zweifelhafter Sicherheit als legitime Mailserver identifiziert werden (durch das Schlüsselwort m̈ailïn der DNS-Adresse). Bei der Suche auf ausreichend frequentierten Blacklisten konnten nur 8 der betreffenden SMTP-Hosts als dynamisch kategorisiert werden, was sich entweder auf einen absichtlich niedrig gehaltenen Durchsatz (um unauffällig zu bleiben) oder das zufällige Wechseln der dynamischen IP zurückführen lässt.
Erst die Analyse von *UDmap* auf dem gesamten IP-Range der Universität identifizierte insgesamt über 7000 dynamische IPs, darunter auch 73 der Spam-Mail-Server. In den meisten Fällen trifft die Aussage, dass hinter dynamischen IPs laufende SMTPs nur für Junk-Mail aufgesetzt wurden, zu, da der Sinn bei legitimen Mail-Servern normalerweise auch der Empfang ist, was für Administratoren heisst, dass diese praktischerweise mit statisch zugewiesenen Adressen versehen werden.

[8] reverse DNS: findet den zur IP-Adresse zugehörigen domain-Namen

5 Funktionsweise von *UDmap*



Da dynamische IP-Adressen immer in Blöcken auftreten, wird *UDmap* versuchen, sich bei den Resultaten auf solche festzulegen und diese sinnvoll einzugrenzen. Ziel ist es, nur aktiv benutzte Adressen zu berücksichtigen, um keine redundanten Daten anfallen zu lassen. Es wird dabei davon ausgegangen, dass zwischen Benutzer und vergebenen IP-Adressen eine m:n-Relation besteht, es kann also sein, dass ein Benutzer sich von mehreren Rechnern mit unterschiedlichen Adressen (Büro, zuhause) aus verbindet, genauso wie mehrere Benutzer dieselbe IP besitzen können, wenn sie aus demselben lokalen Netz heraus mit dem Internet verbunden sind.
UDmap untersucht zuerst bestimmte Grundkriterien:

- Wenn der Benutzer sich von mehreren Systemen aus anmeldet, haben diese in der Regel IP-Adressen aus grundlegend verschiedenen Subnetzen, d.h. befinden sich nicht in gemeinsamen Blöcken. Es ist daher unwahrscheinlich, dass jemand mehrere Adressen aus ein und demselben Block benutzt.
- Dagegen ist anzunehmen, dass jeder normale Benutzer im Laufe der Zeit eine Vielzahl von dynamischen Adressen innerhalb eines Routing Table Prefixes erhält, wenn er zum Beispiel die Verbindung zu seinem Internet Service Provider neu aufbaut.

Unter Anwendung dieser Voraussetzungen fängt *UDmap* an, IP-Blöcke abzuschätzen.

5.1 Mehrfachnutzung von IP-Bloecken erkennen

Wenn eine IP innerhalb eines längeren Zeitraumes mehreren verschiedenen Benutzern gehört, ist davon auszugehen, dass sie zu einem dynamischen Block gehört, in dem dieselbe Adresse mehrfach vergeben wurde. Dieses Kriterium kann allerdings nicht als robust gelten, da sehr wahrscheinlich nie *alle* Adressen innerhalb des betreffenden Blocks in der Logfile auftreten und verschiedene Adressen innerhalb des dynamischen IP Bereiches trotzdem statisch zu sein scheinen, obwohl sie es nicht sind (weil die Verbindung lange aufrechterhalten wird, eventuell durch einen Router).
UDmap geht in diesem Fall trotzdem von einem durchgehend dynamischen IP-Block aus, für den allerdings die folgenden Voraussetzungen erfüllt sein müssen:

1. alle Adressen im Block gehören zum selben AS[9] (und haben dasselbe Prefix innerhalb der Routing-tabelle)
2. in jedem Block von IP_1 bis IP_m müssen mindestens k Adressen im Log aufgetreten sein, sodass gilt: $m \geq k$

[9] Autonomous System

3. die Anfangs- und Endadresse des Bereiches sind ebenfalls im Log enthalten, es dürfen ausserdem keine zu groβen Lücken im Block mit g oder mehr aufeinanderfolgenden IP-Adressen vorhanden sein.

Durch die erste Eigenschaft wird dafür gesorgt, dass die Adressen von der Topologie her nah beieinander liegen, während 2 und 3 zusichern, dass eine ausreichender Anteil der Adressen im Block wirklich dynamisch ist. Abhängig davon, wie die Parameter k und g gewählt werden, liefert der Algorithmus unter Umständen stark voneinander abweichende Ausgaben: durch kleine k erreicht man insgesamt eine breitere Abdeckung, da die Wahrscheinlichkeit bei kleinen Blöcken, alle Eigenschaften zu erfüllen, größer ist. Dagegen werden für groβe k eher mit Sicherheit aktiv benutzte, umfangreiche Blöcke erfasst. Kleine, maximal erlaubte Anzahlen an aufeinanderfolgenden Lücken in einem Block g, sorgen dafür, dass groβe Blöcke stark fragmentiert (zerlegt) werden wohingegen aus groβen g eine höhere Wahrscheinlichkeit an false positives resultiert.

5.2 Berechnung der IP-Benutzungs-Entropie

[10] Hat *UDmap* erst einmal mögliche Kandidaten für IP-Blöcke gesammelt, muss unterschieden werden zwischen dynamischen IP-Adressen, die oft neu vergeben wurden oder statisch zugewiesene, die aber von mehreren Benutzern verwendet werden. Also ist es über die Zeit verteilt gesehen auch zu erwarten, dass im Fall einer dynamischen IP der Benutzer mehrere über den Pool von IPs verteilte Adressen gehabt haben muss.
Die Benutzungs-Entropie $H(j)$ einer bestimmten einzelnen IP wird durch einen Wert zwischen 0 und 1 repräsentiert, je näher dieser an der 1 liegt, desto höher ist die Wahrscheinlichkeit, dass eine dynamische Adresse vorliegt. Zur Berechnung wird eine Matrix $A^{U(j)\times m}$, also mit den Dimensionen Anzahl der Nutzer $U(j)$ [11] aus der Menge aller User U und der Anzahl an Adressen im Block, m. Die Elemente sind jeweils $\in \{1, 0\}$, wobei die 1 überall dort steht, wo eine Aktivität des Benutzers auf dieser IP-Adresse geloggt wurde. $H(j)$ berechnet sich wie folgt:
$H(j) = \sum_{k=1}^{m}(\frac{a_k}{z_j} log_2(\frac{a_k}{z_j}))$,
wobei a_k die k-te Spalte von A_j ist und z_j die Summe aller Elemente in A_j (es geht also um einen Vergleichswert, wievielen Benutzern diese IP zuzuordnen war). Da die Blockgröβe m stark variiert, wird die darstellung der Benutzungs-Entropie mit den folgenden Formeln normiert:
$H_B(j) = \frac{H(j)}{log_2 m}$
$H_U(j) = \frac{H(j)}{log_2 |C(j)|}$
$H_B(j)$ (normalisierte Benutzungs-Entropie) gibt an, ob die Wahrscheinlichkeit die Nutzer im Block, diese IP $IP(j)$ zu erhalten, gleich verteilt ist, $H_U(j)$ (normalisierte, ausgewählte Benutzungs-Entropie) dagegen gibt diesen Wert nur für die Menge von IP-Adressen $C(j)$ an, die der Benutzer in der überwachten Dauer tatsächlich mindestens einmal besessen hat. Im Idealfall läge die normali-

[10] Beschreibt die Mächtigkeit an Informationen, die über bestimmte IPs vorliegen, also wie präzise die Datenmenge insgesamt ist

[11] umfasst alle, die von der IP-Adresse IP_j aus verbunden waren

sierte Benutzungs-Entropie fast bei 1, d.h. die Auswahl der IP-Adressen wäre vollständig zufällig.

5.3 Zusammenfassen von dynamischen IPs als Block

Liegen die Werte für die Benutzungs-Entropie vor, könnte man davon ausgehen, dass IP-Adressen mit $H(j) \approx 1$ unbedingt dynamisch sind. Dies muss jedoch unter der Bedingung betrachtet werden, dass dynamische IP-Adressen immer nur in Blocks auftreten. Daher werdem im nchsten Bearbeitungsschritt von *UDmap* die Mehrbenutzer-IP-Blöcke nochmals in Substrukturen (Teilblöcke) unterteilt, wobei die Teilblöcke so ausgewählt werden, dass die Mehrheit der Adressen über einem vorher festgelegten Schwellenwert H_e für die Entropie liegt. Um dies zu erreichen, wendet *UDmap* eine Signalglättungsfunktion auf die einzelnen Werte an, um den Einfluss von Ausreiβern zu reduzieren.

5.4 Durchschnittliche Nutzungsdauer und Entfernen von Proxies

Sobald die Blöcke dynamischer IP-Adressen in dieser Form vorliegen, müssen nun noch jeweils die Wiederverwendungshäufigkeiten einzelner IPs (*ip volatility*) berechnet werden, wobei zwei Werte von Bedeutung sind: die Anzahl der Nutzer, welche einer einzelnen IP über die Zeit des Logs hinweg zugeordnet waren und die mittlere Dauer, bis der Benutzer dieser IP wechselte.
Welchen Sinn die Berechnung der durchschnittlichen Benutzungsdauer hat, wird klar, wenn man sich überlegt, unter welchen Voraussetzungen *UDmap* fälschlciherweise Adressbereiche als dynamisch erkennen würde.
Zum einen würden darunter Proxy-Server fallen, die dem Lastausgleich beim Senden aus einem lokalen Netzwerk dienen, was zur Folge hätte, dass hier Benutzer ebenfalls oft mit einer anderen IP aus demselben Adress-Pool auffallen, eigentlich aber an ein und demselben Rechner sitzen. Dasselbe trifft auch für Internet-Cafés und Netzwerke in Bildungsinstitutionen zu. Der einzige Unterschied liegt darin, dass bei den Proxy-Server mehrere Benutzer gleichzeitig dieselbe IP haben können, im Fall der Internet-Café Rechner wäre dies nur sequenziell möglich.
Unabhängig davon haben beide Fälle die charakteristischen Eigenschaften dynamischer IP-Blöcke. Um nun zu verhindern, dass IPs fälschlicherweise als solche erfasst werden, überprüft *UDmap* für das erste Beispiel (Proxy-Server) den *ip volatility* Zeitwert und kann solche bei einer sehr hohen Rate (was den entscheidenden Unterschied zu normalen Multi-User IP Blöcken macht) von gleichzeitigen Zugriffen ausschliessen.

6 Validierung der identifizierten IP-Adressen

Es ist nicht in jedem Fall eindeutig feststellbar, ob die von *UDmap* gefundenen IPs tatsächlich dynamisch sind, meistens lässt sich dies auch nicht nachvollziehen, da Netzbetreiber ihre Prinzipien der IP-Vergabe nicht offenlegen.

Um eine ungefähre Aussage über die Korrektheit der gefundenen Adressen treffen zu können, werden die Adressblöcke mit den bestehenden Daten von Dynablock abgeglichen und dabei 6 Fälle für jeweils einen IP-Block betrachtet:

1. **identisch**
   (trifft auf 0.11% der von *UDmap* gefundenen Adressen zu)
2. **Teilmenge** eines Blocks in den Blacklists
   (47.93%, wobei die IP-Adressen, die für eine identische Menge fehlen, durch die nicht vorhandene Nutzung im vorliegenden Logfile nicht identifiziert werden konnten)
3. **Übergeordnete Menge**, d.h. der in Dynablock erfasste IP-Block ist eine Teilmenge der identifizierten Adressen (1.6%)
4. **neu**, der Block ist also noch nicht erfasst (48.06%)
5. **fehlt**, der Block ist in Dynablock-Tabellen eingetragen, wurde von *UDmap* aber nicht erkannt
   (5.87%, wobei dies wieder mit fehlendem Vorkommen in den Eingabedaten zu erklären ist)
6. **teilweise überlappt**, die Dynablock Adressen und *UDmap* IPs überschneiden sich teilweise (2.3%)

Die restlichen 50.19% werden, sofern möglich, über rDNS verifiziert. Da dies wie oben erwähnt längst nicht für alle IP-Adressen angegeben ist, kann über den verbleibenden Teil nicht mit letzter Sicherheit eine konkrete Aussage getroffen werden.

7 Revision

Durch das zunehmende Interesse an netzwerk-basiertem Filtern von Spam, Botnet-IPs, Phishing-Sites oder ähnlichen Anwendungsgebieten wird spätestens beim Vergleich von *UDmap* Algorithmus gegenüber der bisherigen Verwaltung von Blacklists wird klar, von welcher Wichtigkeit die in diesem Paper dargestellten Erkenntnisse über das Identifizieren von dynamisch vergebenen IPs sind. Gerade weil es sich in 90% der insgesamt versendeten Mails um Spam handelt und es bisher keinen wirklich wirksamen Ansatz zum effektiven Vorgehen gegen dieses Problem gegeben hat (da aus den oben genannten Gründen natürlich auch netzwerk-basiertes Filtern über das klassische Blacklisting sehr ineffektiv ist), liesse sich hier durch einen verbreiteten Einsatz von auf *UDmap* aufbauenden Filterlisten bei vielen namhaften Mail-Providern dieser Prozentsatz möglicherweise drastisch reduzieren.

Literatur

[1] http://www.spamhaus.org

[2] A. Ramachandran, D. Dagon, and N. Feamster In Conference on Email and Anti-Spam, 2006.

[3] Anirudh Ramachandran, Nick Feamster, and Santosh Vempala CCS '07: Proceedings of the 14th ACM conference on Computer and communications security Pages: 342 - 351