

Allgemeines und Hinweise

Verwendet für die Auswertung werden jeweils die Werte des Monatsletzten

In den Januarzahlen 2004 wurden noch keine unkategorisierten Server aufgeführt. Die dort angegebene Zahl ist willkürlich gewählt und nur eine grobe Schätzung

November 2005: Werte sind vom 25.11.; danach wurde die Datenerzeugung modifiziert und zusätzlich zum Newsscanning (MX-Records) noch eine Ripe-Hostliste zu den Servern hinzugefügt. Dadurch sind viele Server hinzugekommen, die nicht als MX für Domains eingetragen sind. Auffällig ist hierbei die extrem hohe Zunahme von Servern von Exchange und Postfix. Das deutet darauf hin, dass Exchange zwar sehr häufig eingesetzt wird, dabei nicht jedoch als Primary- oder Secondary MX :-)
Bei Postfix liegt die Vermutung nahe, dass dieser auf vielen Distributionen von Hause aus mitgeliefert wird und daher auf vielen Heimservern ebenfalls ausschließlich zum Versand eingesetzt wird.

Im April 2006 kam es aufgrund eines Hardwaredefekts zu Fehlern bei der Aktualisierung, weshalb viele Server nicht erreicht werden konnten und darum aus der Statistik gefallen sind :-)

Seit Mitte September 2006 werden täglich ca. 400000 zufällige IP-Adressen ausgewählt und geprüft. Dies führt zu einer sukzessiven Erhöhung der Datenmenge, die Liste enthält dadurch wiederum Server, die nicht als MX-Records eingetragen sind.

Im Juli 2007 waren zum ersten Mal mehr als 1 Million Server in der Liste verzeichnet!

Im Februar 2007 wurden alle Mailserver mit .mil am Ende in eine eigene Kategorie verfrachtet. Dadurch sanken die Anteile der Mailserver innerhalb der kategorisierten Mailserver. Hintergrund ist, dass bei .MIL anscheinend jeder Rechner auch als MX für sich eingetragen ist, es aber pro Stützpunkt nur einen zentralen Mailhost gibt, welcher dann SMTP abhandelt.

Im Januar 2008 wurden diese Rechner aus der Liste genommen und nur noch intern verwaltet. Dadurch erhöhten sich die Anteile der Server untereinander, und die Anzahl der kategorisierten Hosts wurde reduziert.

Im Februar 2008 wurden die Email-Adressen aus der Datenbank entfernt, und nur noch die Domains gespeichert. Ausserdem wurde eine Domainliste (ISC-Domain-Survey von 1997) importiert. Anscheinend gibt es aber keine aktuelleren Listen mehr - Spam sei Dank? :-)

Um ein paar Domains zur Erzeugung, wurden im April 2008 eine wordlist von crack verwendet, und die darin enthaltenen Worte mit .com, .net, .org, .info und .eu als in die Datenbank gepackt. Insgesamt wurden dadurch ca. 9 Millionen Domains auf MX-records geprüft.

Zusätzlich wurden anschließend noch einige wordlists mit allen Toplevel-Domains durchgegangen. Dadurch wurden ca. 100 Millionen Domainnamen geprüft, natürlich hat davon nur ein Bruchteil einen MX-Record. Die Domains .tk und .la haben zudem einen Wildcard-MX :-)

Im Juni 2008 wurden für die zufälligen IP-Adressen zusätzlich noch AS-Listen verwendet, um unnötige Scans abzufangen. Die Liste wird von <http://www.team-cymru.org> bereitgestellt. Der Scanner für IP-Adressen wurde auf ein kleines C-Programm umgestellt, dadurch können pro Tag ca. 2-3 Millionen IP-Adressen geprüft werden! IP-Adressen werden jetzt permanent in der Datenbank gespeichert, dadurch befanden sich Ende des Monats etwas über 20 Mio. Adressen darin, die Prüfung von fehlgeschlagenen Adressen wird gestaffelt durchgeführt.

Mitte Juli 2008 hat die Datenbank 30 Millionen IP-Adressen. Der Scanner wurde optimiert und kann dadurch bis zu 4 Millionen IP-Adressen täglich prüfen.

Durch Hardware-Update werden inzwischen bis zu 10 Millionen Adressen täglich gescannt.

Zwischen dem 24. Mai und 19. Juli wurden keine Statistiken in der Datenbank gespeichert :-)
Ab dem 19. Juli wurde die Speicherung der Datenbank modifiziert, so dass nicht erreichbare Hosts in der Datenbank verbleiben und nur gelegentlich neu gescannt werden. Dadurch müssen mehr Daten gespeichert werden, dafür werden weniger Adressen doppelt gescannt.

Oktober 2008: Die Tabelle mit den Server-IP-Adressen wurde partitioniert und in mehrere Tabellen mit getrennten Adressenbereichen gesplittet. Dadurch werden Auto-Analyse sowie und Auto-Vacuum nicht auf der kompletten Tabelle erforderlich. Einziges Manko: Die Funktion zum Einfügen von Datensätzen muss mittels eines komplexen Triggers realisiert werden.

Ebenfalls im Oktober wurden alle AS-Bereiche mit /22 bis /30 **komplett** in der Datenbank abgelegt. Die zufälligen Adressen werden nur noch für die Masken /8 bis /21 erzeugt. Dadurch verdoppelte sich die Anzahl der gespeicherten IP-Adressen.

April 2009: Durch einen Bug in der Datenerzeugung wurden nicht (mehr) erreichbare Server nicht entfernt, dies wurde korrigiert. Dadurch sind die Zahlen generell niedriger als im Vormonat. Betroffen sind vermutlich alle Werte ab ca. Februar 2009

Mai 2009: .mil wurde komplett aus der Datenbank der erreichbaren Server entfernt. Ranges /21 und /20 werden sukzessiv komplett in die Datenbank übernommen.

Juli/August 2009: /19-/17 werden **komplett** in die Datenbank übernommen. Dadurch wächst die Datenbank auf knapp 650 Millionen IP-Adressen...

Juli 2009: Die Zahlen sind zu hoch, da aufgrund interner Umstellungen der Bug von April 2009 wieder aufgetreten ist. Fehler wurde behoben und Zahlen für August sollten wieder korrekt sein.

September/Dezember 2009: Aufgrund von internen Umstellungen/Optimierungen wurden die nicht mehr erreichbaren Server nicht gelöscht (siehe April 2009). Der Fehler wurde im Dezember behoben, die Zahlen für Ende 2009 - Anfang 2010 sind daher rückläufig.

Dezember 2009: Am 27.12.2009 erreicht die Datenbank **Milliarde** Einträge. Der Bereich /16 ist noch nicht komplett enthalten, dies wird Anfang Januar der Fall sein.

1.1.2010: Die Bereiche /16 und /15 sind **komplett** in der Datenbank enthalten.
5.1.2010: Der Bereich /14 ist **komplett** in der Datenbank enthalten.

Januar 2010: Die Datenbank wurde neu durchgezählt, demnach sind **5 Milliarden** Hosts vorhanden.

April 2010: Alle AS-Ranges in Datenbank, zufällige Adresserzeugung abgeschaltet.
20.04.10: Alle AS-Ranges sind **komplett** in der Datenbank enthalten!

März 2011: Es werden auf Basis der frei verfügbaren RIPE/ARIN/APNIC-Datenbank, sowie der Delegation-Statistik der ARIN alle Netzwerke und Hosts hinzugefügt, die als ASSIGNED deklariert sind. Dadurch steigt die Anzahl der Adressen in der Datenbank **außer 3,5 Milliarden**

April 2011: Die Haupt-Scanning-Engine wurde erweitert. Es wird jetzt zunächst eine SYN-Probe abgesetzt, und erst wenn diese erfolgreich ist, ein kompletter CONNECT-Durchlauf mit Prüfung der SMTP-Antworten durchgeführt. Dadurch können deutlich mehr Hosts abgefragt werden, zur Zeit ist das aber auf 400/Sekunde limitiert, also ca. 32 Millionen Hosts/Tag. Anscheinend werden aber dadurch mehr Hosts gefunden, die Zahlen von erreichbaren Servern ist deutlich gestiegen.

Dezember 2011: Die Failcount-History wurde erweitert, und ist jetzt nicht mehr auf Level 0..15, 100 beschränkt. Level 15-99 werden ca. alle 80 Tage gescannt, dadurch werden die Statistiken jetzt detaillierter für die nächsten 20 Jahre, danach muss ich mir ein neues Schema überlegen :-)
Zur Zeit werden im Schnitt ca. 50 Millionen Einträge pro Tag gescannt, also alle 3 Monate einmal das komplette IPv4-Segment...

Januar 2013: Historischer Augenblick: sendmail ist das nicht mehr unter den Top 5. Der fünfte Platz wird jetzt von PowerMTA belegt. Die Top 5 sind jetzt also eine Top 6 :-)

Bekannte Ungenauigkeiten:

Viele Server haben mehrere IP-Adressen, und tauchen in solchen Fällen natürlich auch mehrfach in der Liste auf. Dies abzufangen ist fast unmöglich, da es genügend identische Server-Responses gibt, die aber auf völlig unterschiedlichen Servern landen...

In manchen Netzen wird Port 25 anscheinend zentral auf einen Rechner weitergeleitet, so dass dadurch der eingesetzte Server deutlich häufiger auftritt als er tatsächlich vorhanden ist. Beispiel hierfür ist unter anderem alles, was auf .mil endet...

Rechner, die in dynamischen Bereichen hängen, tauchen ebenfalls mehrfach auf und erhöhen dadurch den Anteil der abzufangenden Server. Auch dies könnte man nur über den Namen im HELO abfangen; da dieser jedoch aber nicht immer auftritt, kann man auch dies nicht abfangen.

(c)2004-2013 by Creative Chaos Computing, Jochen Erwied, mack at smtpscan.erwied.eu