Fachhochschule Gelsenkirchen

Spam – immer noch hoch im Kurs

Auswertung der 3.Umfrage zur E-Mail-Verlässlichkeit, 2006

Stand: Januar 2007

Autoren **Christian Dietrich**
dietrich@internet-sicherheit.de

Prof. Dr. Norbert Pohlmann
norbert.pohlmann@informatik.fh-gelsenkirchen.de

Institut für Internet-Sicherheit
Fachhochschule Gelsenkirchen
Fachbereich Informatik
Neidenburgerstr. 43
45877 Gelsenkirchen

Web www.internet-sicherheit.de

1. Einleitung

Der E-Mail Dienst ist einer der am weitesten verbreiteten und meist genutzten Dienste des Internets und wird heutzutage als Mittel zur einfachen, nachrichtenbasierten und zuverlässigen Kommunikation im Internet eingesetzt. Er ist inzwischen für unsere vernetzte Wissens- und Informationsgesellschaft eine nicht mehr wegzudenkende Anwendung.

Seit einigen Jahren jedoch beeinträchtigt insbesondere Spam das Medium E-Mail derart stark, dass die Frage zu stellen ist, ob E-Mail auch in Zukunft noch genauso einfach, unkonventionell, produktiv und vielfältig eingesetzt werden kann.

Um letztlich das Gefahrenpotential für die E-Mail Nutzung detaillierter einschätzen zu können, hat das Institut für Internet-Sicherheit der Fachhochschule Gelsenkirchen unterstützt durch das Bundesamt für Sicherheit in der Informationstechnik (BSI) eine Erhebung bei diversen Firmen, Organisationen sowie bei großen europäischen Internet Service Providern durchgeführt, die sowohl die aktuelle Bedrohungslage durch Spam und Viren als auch Maßnahmen zu deren Gefahrenabwehr analysiert (vgl. [DiPo05]).

2. Die Erhebungen

Die erste Untersuchung Ende des Jahres 2004 ist repräsentativ für über 40 Mio. E-Mail-Accounts und ein durchschnittliches monatliches E-Mail-Volumen von mehr als 2,3 Mrd. E-Mails. Die Ergebnisse der ersten Umfrage wurden in [DiPo05] veröffentlicht.

Um Trends und Entwicklungen erkennen zu können, wurde die Erhebung zwischen Juni und August 2005 sowie zwischen Oktober 2006 und Januar 2007 mit einigen Teilnehmern unter Verwendung von aktuellen Zahlen wiederholt. Im Rahmen der zweiten Erhebung wurde ein E-Mail-Volumen von 1 Mrd. E-Mails pro Monat und 18,5 Mio. E-Mail-Accounts berücksichtigt.

Die dritte Erhebung im Zeitraum Oktober 2006 bis Januar 2007 erfasste ein monatliches E-Mail-Volumen von 1 Mrd. E-Mails sowie 12,8 Mio. E-Mail-Accounts.

Im Rahmen der dritten Umfrage wurden 142 Personen angeschrieben. Darüber hinaus wurde die Umfrage auf diversen Webseiten beworben, u.a. die Webseiten des Instituts für Internet-Sicherheit www.internet-sicherheit.de, die Webseiten des Bundesamts für Sicherheit in der Informationstechnik www.bsi.de sowie das Anti-Spam-Portal www.antispam.de. Ferner wurde auf die Umfrage in der Zeitschrift iX des heise-Verlags hingewiesen.

Im Zeitraum Oktober 2006 bis Januar 2007 riefen 904 Personen die Webseite des Fragebogens auf. Es registrierten sich 106 Personen für die Online-Umfrage. 88 Teilnehmer haben sich mindestens einmal eingeloggt. Lediglich 23 Teilnehmer beantworteten mehr als 25% aller Fragen. Die effektive Rücklaufquote, d.h. der Anteil an Teilnehmern, die qualitativ hohe Daten bereitstellten beträgt somit 16,2%. Die geringe effektive Rücklaufquote liegt darin begründet, dass der Fragebogen sehr komplex und umfassend ist und für einige Unternehmen einen nicht unerheblichen Aufwand darstellt, um die gewünschten Zahlen zu eruieren. Darüber hinaus erheben einige Unternehmen derartige Messwerte nicht und konnten – trotz großen Interesses – den Fragebogen nicht ausfüllen. Ferner wurden keinerlei Anreize beispielsweise in Form von Verlosungsprodukten unter den Teilnehmern bereitgestellt.

Die Auswertung und damit einhergehend die Interpretation der Erhebung erfolgt aus mehreren Perspektiven. Insbesondere bei der Betrachtung der Anteilsverteilung von E-Mail ist die Sichtweise entscheidend für das Verständnis der Zahlen. Aus der Perspektive eines E-Mail-System-Betreibers spielt beispielsweise der Anteil, der bereits im SMTP-Dialog abgewiesen wird eine große Rolle (siehe Abbildung 1). Für den E-Mail-Nutzer hingegen sind die Verhältnisse von erwünschter E-Mail zu Spam und Viren von Bedeutung (siehe Abbildung 4). Aus diesem Grund werden im Folgenden diese beiden Sichtweisen unterschieden.

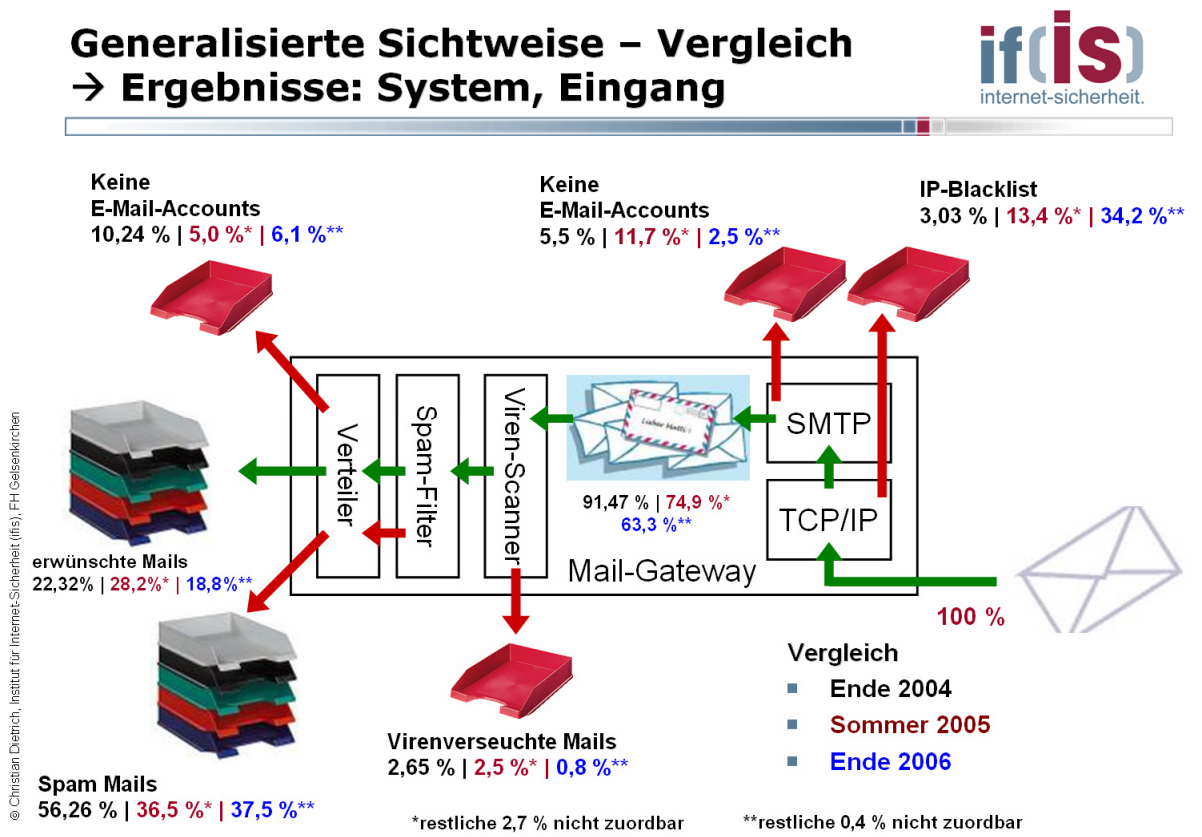

Abbildung 1: Anteilsverteilung des E-Mail-Volumens aus Perspektive des E-Mail-Systems, sog. Systemsicht, Christian Dietrich, Institut für Internet-Sicherheit

Die sog. Systemsicht stellt die Anteilsverteilung des E-Mail-Volumens nach Anzahl an E-Mails im Gesamtüberblick dar. Die Systemsicht ist in der Regel für die Dienstanbieter interessant, da hier erkennbar ist, welcher Anteil an E-Mails die Kunden erreicht und welche Anteile im Rahmen der Verarbeitung durch die E-Mail-Systeme eliminiert werden. Darüber hinaus ist ersichtlich, welche Anteile in den jeweiligen Stufen der E-Mail-Sicherheitsmechanismen erfasst werden.

Demgegenüber hilft die sog. Nutzersicht (siehe Abbildung 4), die Perspektive der E-Mail-Anwender zu verdeutlichen. Hierbei tritt die Anteilsverteilung der E-Mails wie sie E-Mail-Anwender in ihren E-Mail-Postfächern vorfinden in den Vordergrund.

3. Trends und Entwicklungen

Im Vergleich zur ersten Erhebung fällt negativ auf, dass aktuell ein geringerer Anteil an erwünschten E-Mails versandt und empfangen wird (siehe Abbildung 1 und Abbildung 2). Statt 22,3% (im Jahr 2004) beträgt der Anteil an erwünschten E-Mails im Jahr 2006 18,8%. Hierbei ist eine Trendwende im Vergleich zur Untersuchung in 2005 zu erkennen, bei der der Anteil an erwünschten E-Mails bei 28,1% lag.

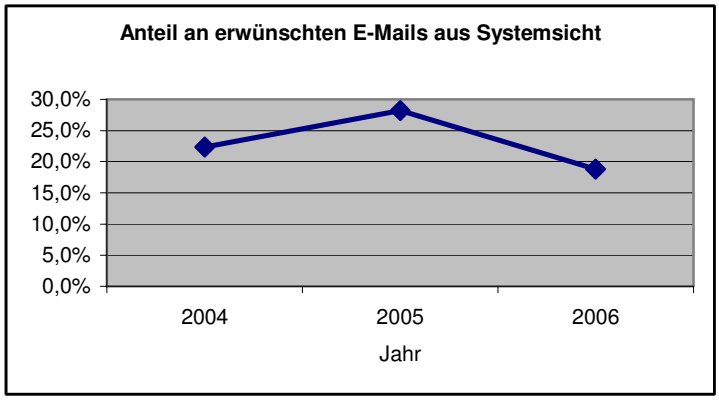

Abbildung 2: Anteil an erwünschten E-Mails aus Systemsicht
Christian Dietrich, Institut für Internet-Sicherheit

Dies ist insbesondere vor dem Hintergrund, dass der Anteil an angenommenen E-Mails über alle 3 Erfassungszeitpunkte hinweg deutlich gesunken ist ein unerwartetes Ergebnis. Im Jahr 2004 wurden 91,47% aller E-Mails angenommen, im Jahr 2005 betrug dieser Wert 74,9% und im Jahr 2006 lag er bei 63,3%.

Erhielten E-Mail-Nutzer Ende des Jahres 2004 durchschnittlich 69,3% Spam, so verringerte sich dieser Anteil bereits bis Mitte 2005 auf lediglich 52,3% (vgl. Abbildung 4). Dieser Trend setzte sich jedoch nicht weiter fort, sondern der Anteil an angenommenen Spam-E-Mails stieg geringfügig auf 37,5% in der Systemsicht.

Betrachtet man diesen Sachverhalt aus der Perspektive eines E-Mail-Nutzers, indem alle Anteile eliminiert werden, die der E-Mail-Anwender nicht zu Gesicht bekommt, so zeigt sich im Jahr 2006 ein deutlich schlechteres Verhältnis von Spam- zu erwünschten Nachrichten als noch im Jahr zuvor. Während in 2005 aus

Nutzerperspektive 52,3% der E-Mails im Postfach als Spam klassifiziert wurden, so sind dies im Jahr 2006 65,6%. Gleichzeitig sinkt der Anteil an erwünschten Nachrichten aus Nutzerperspektive ebenfalls von 40,3% im Jahr 2005 auf 33% im Jahr 2006 (siehe Abbildung 4). Vergleicht man diese aktuellen Werte mit der Untersuchung aus dem Jahr 2004, so erreichen sie fast wieder das Niveau von vor zwei Jahren.

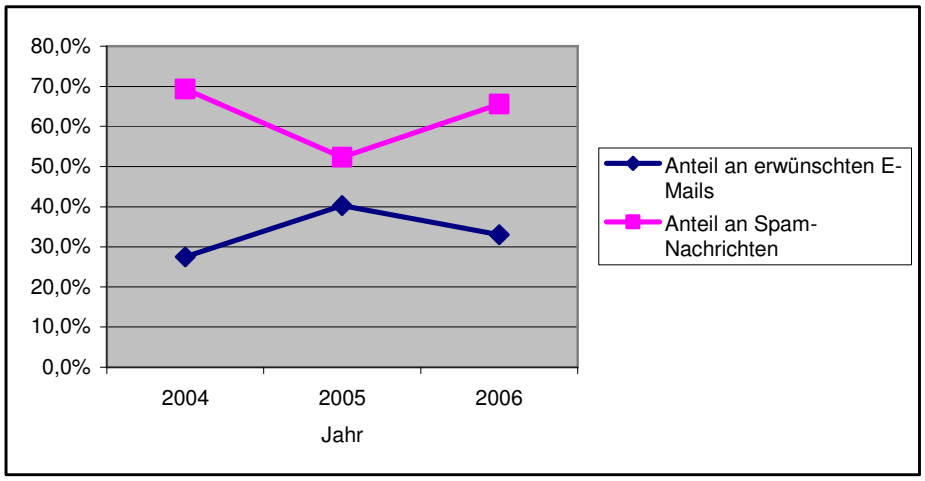

Abbildung 3: Anteil an erwünschten und Spam-Nachrichten aus Nutzerperspektive, Christian Dietrich, Institut für Internet-Sicherheit

Es lässt sich ein deutlicher Rückgang des Virenvolumens verzeichnen. Im Rahmen der ersten Befragung konnte aus der Systemsicht ein Virenanteil von 2,65% ermittelt werden, im Jahr 2005 betrug er im Durchschnitt 2,5%, zum aktuellen Zeitpunkt liegt er bei 0,8% aller E-Mails, die den E-Mail-Server erreichen. Eine mögliche Ursache stellt die Tatsache dar, dass im Vergleich zur Erhebung 2004 durchschnittlich deutlich weniger E-Mails durch den E-Mail-Server angenommen werden (91,47% (2004) bzw. 74,9% (2005) bzw. 63,3% (2006); vgl. Abbildung 1). Diesbezüglich können jedoch nur Vermutungen angestellt werden.

Dieser Rückgang wirkt sich auch auf die Nutzerperspektive aus. Aus der Sicht eines E-Mail-Nutzers liegt das Virenvolumen dadurch derzeit mit 1,3% deutlich unter den Werten der ersten beiden Erhebungen (3,3% bzw. 3,5%) (siehe Abbildung 4).

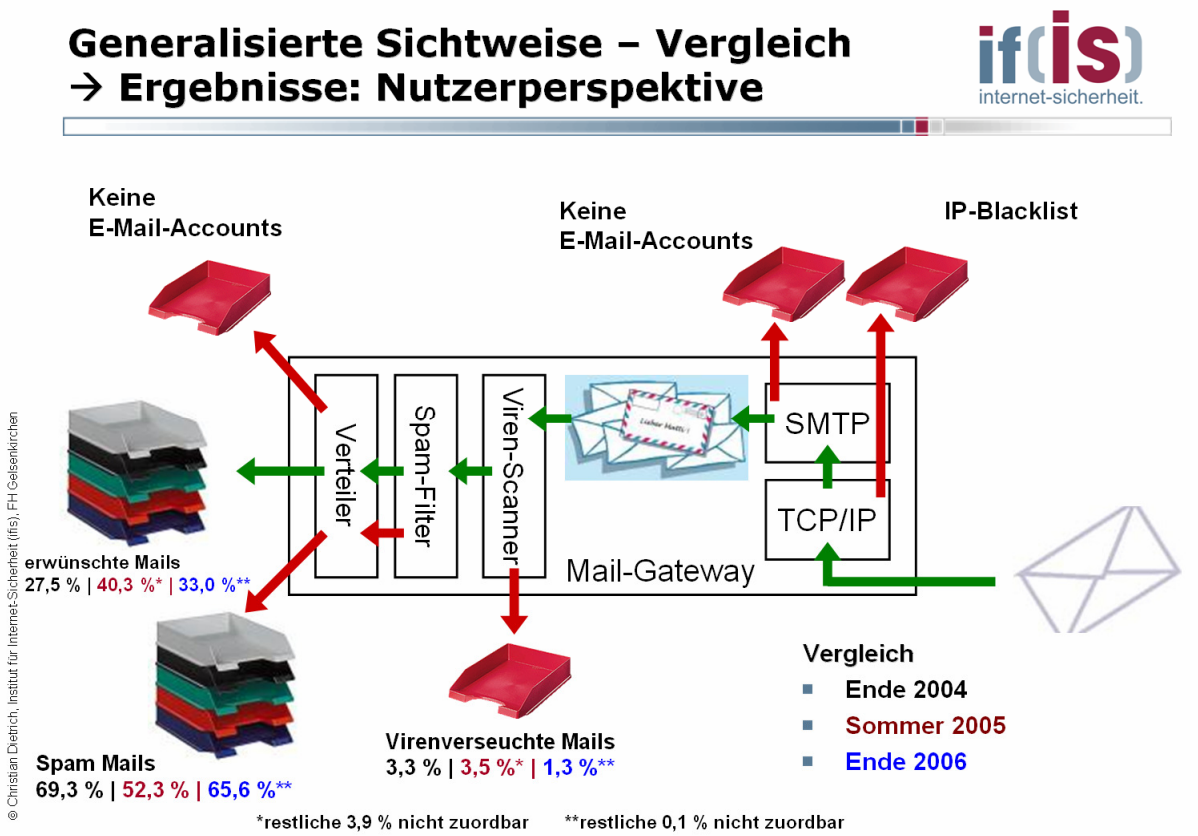

Abbildung 4: Anteilsverteilung des E-Mail-Volumens aus Perspektive des E-Mail-Benutzers, Christian Dietrich, Institut für Internet-Sicherheit

4. Spam-Abwehr

Der Anteil an E-Mails mit nicht-existierender Empfänger-E-Mail-Adresse hat sich insgesamt lediglich im Vergleich der ersten beiden Erhebungen zur letzten Untersuchung auffällig verändert. Während er 2004 und 2005 bei 15,7% (10,2% + 5,5%) respektive 16,7% (5% + 11,7%) lag, beträgt er in der aktuellen Untersuchung 8,6% (6,1% + 2,5%). Es handelt sich also um eine deutliche Abnahme an E-Mails mit nicht-existierender Empfänger-E-Mail-Adresse.

Die Überprüfung der Zustellbarkeit einer E-Mail (im Sinne der Existenz eines Ziel-Postfachs zu einer E-Mail-Adresse) kann zu zwei verschiedenen Zeitpunkten während der Verarbeitung durch einen empfangenden E-Mail-Server erfolgen (siehe Abbildung 1). Zum einen kann ein MTA eine E-Mail vollständig akzeptieren und daraufhin überprüfen, ob ein entsprechendes Postfach zu der angegebenen Empfänger-Adresse existiert. Ist dies nicht der Fall, so generiert der E-Mail-Server eine Informationsnachricht über die fehlgeschlagene Zustellung, eine sog. Non Delivery Notification. Dieses Verfahren war historisch bedingt durch den sog. Store-and-forward-Mechanismus bei E-Mail sehr weit verbreitet. Zum anderen kann bereits während des SMTP-Dialogs die Zustellung abgebrochen werden, falls kein Ziel-Postfach zugeordnet werden kann. Die letztere Methode ist ressourcenschonender als die erstere, da in diesem Fall die E-Mail gar nicht erst übertragen wird und somit Kosten für Bandbreite, Speicherplatz etc. eingespart werden.

Es zeigt sich, dass im Jahr 2005 den E-Mails mit nicht-existierender Empfänger-Adresse in höherem Maße bereits im SMTP-Dialog die Annahme verweigert wird als noch ein Jahr zuvor. Ca. 35% aller E-Mails mit nicht-existierender Empfänger-E-Mail-Adresse wurden zum ersten Erhebungszeitpunkt bereits im SMTP-Dialog abgewiesen. Im Jahr 2005 hat sich der Anteil auf 70% erhöht – bei insgesamt in etwa gleich gebliebenem Anteil an E-Mails mit nicht-existierender Empfänger-Adresse. Das Ergebnis zum aktuellen Zeitpunkt zeigt, dass weniger E-Mails mit nicht-existierender Empfänger-Adresse im SMTP-Dialog abgewiesen werden (siehe Abbildung 1). Allerdings hat sich auch hier insgesamt der Anteil an E-Mails mit nicht-existierender Empfänger-Adresse von rund 16% auf 8,6% deutlich verringert. Ein möglicher Grund hierfür könnte die mangelnde Nachrüstung von bestehenden E-Mail-Systemen sein, sodass solche Systeme noch nicht die Fähigkeit haben, derartige E-Mails überhaupt im SMTP-Dialog abzuweisen. Diesbezüglich lässt sich allerdings keine empirische Bestätigung liefern.

Nicht nur basierend auf SMTP-Adressinformationen, sondern auch aufgrund von IP-Adressmerkmalen kann die Spam-Abwehr effizient ermöglicht werden. Unter einer Blacklist wird eine Liste negativ aufgefallener (IP-)Adressen in Bezug auf die E-Mail-Nutzung verstanden. Ebenso können in einer Blacklist Adressen von Rechnern, die nicht für die direkte E-Mail-Einlieferung vorgesehen sind – wie beispielsweise dynamische IP-Adressen – aufgeführt werden. Eine Blacklist entscheidet damit nicht zwangsläufig über Zulassung oder Ablehnung einer Verbindung. Die Entscheidung, ob eine Kommunikation aufgrund eines Eintrags einer entfernten Partei in einer Blacklist abgebrochen wird, liegt weiterhin beim Empfänger und muss durch eine eigene Policy auf der Empfängerseite bestimmt werden. Der Eintrag eines Adressdatums in einer Blacklist kann – wenn ein Missbrauch des eigenen E-Mail-Dienstes durch Spam verhindert werden soll – als Indiz gegen das Zustandekommen einer Verbindung gewertet werden.

Diejenigen Befragten, die anhand von Merkmalen der IP-Schicht blockieren, lehnen auf diese Art bis zu über 90% der gesamten Anzahl an E-Mails, deren Zustellung versucht wurde ab. Jedoch wenden lediglich ca. 43,5% der Befragten IP-Adress-basierte Blacklists an. Im Vergleich zur Erhebung im Vorjahr hat sich dieser Anteil nicht auffällig verändert. In 2005 haben rund 45% der Befragten IP-Adress-basiertes Blacklisting angewendet. Dies ist immerhin deutlich mehr als der Verbreitungsgrad von 30% bei der Befragung im Jahr 2004. Damit hat sich der Anteil an blockierten E-Mails auf IP-Ebene aus der Systemsicht sehr positiv auf 34,2 % erhöht.

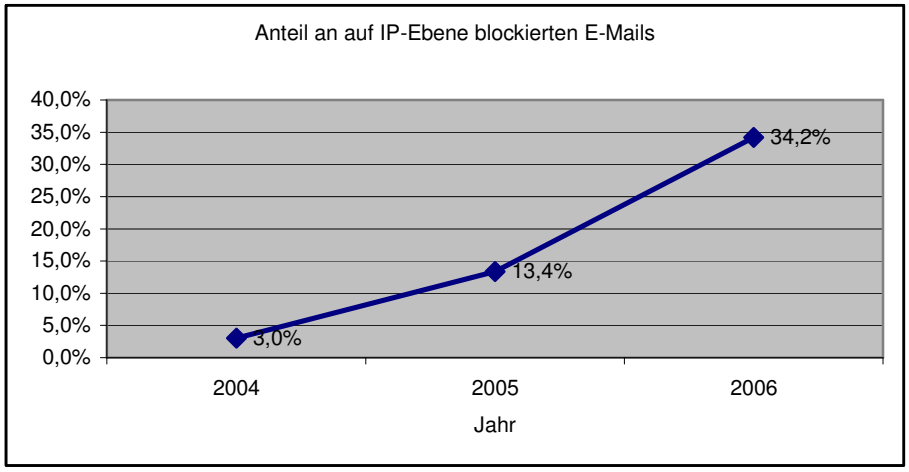

Abbildung 5: Anteil an E-Mails, die auf IP-Ebene blockiert werden, Christian Dietrich, Institut für Internet-Sicherheit

5. Fazit

Seit der ersten Erhebung Ende 2004 hat sich der Anteil an erwünschten E-Mails zwar zwischenzeitlich deutlich vergrößert, allerdings liegt er derzeit auf einem ähnlichen Niveau wie zum Ausgangszeitpunkt. Die wichtigste Interpretation hierfür ist ein sehr hoher Anteil an Spam-Nachrichten von derzeit rund 65% aus der Perspektive der E-Mail-Anwender. Betrachtet man diejenigen Anteile an E-Mails, die durch die Verarbeitung der E-Mail-Systeme aussortiert werden ebenfalls als Spam, so ergibt sich insgesamt eine Spam-Rate von 80,4%.

Positiv zu erkennen ist eine deutliche Verringerung des Anteils an Viren-behafteten E-Mails. Seit der ersten Untersuchung im Jahr 2004 reduzierte sich dieser Anteil konstant auf mittlerweile lediglich 1,3% aus Sicht der E-Mail-Anwender. Für Systembetreiber beträgt der Anteil an Viren-Nachrichten lediglich 0,8%.

Positiv zu erkennen ist ebenfalls die Tatsache, dass der Anteil an E-Mails, die auf IP-Ebene blockiert werden konstant zunimmt und in der aktuellen Untersuchung im Durchschnitt bereits mehr als ein

Drittel aller E-Mails beträgt. Einzelne Teilnehmer erreichen mit dieser Methode eine Filterung von mehr als 90% aller E-Mails und als Konsequenz deutlich geringere Spam-Raten. Dieser Trend ist zu begrüßen und sollte fortgeführt werden.

Der Anteil an aufgrund von nicht-existierenden Empfänger-E-Mail-Adressen nicht zustellbaren E-Mails ist deutlich gesunken. Während bei den ersten beiden Untersuchungen rund 16% nicht zustellbar waren, beträgt dieser Anteil in der aktuellen Erhebung lediglich 8,6%.

Die angestiegene Spam-Rate ist jedoch ein alarmierendes Zeichen dafür, dass der Kampf gegen Spam bei weitem nicht entschieden ist. Im Laufe der Untersuchungen ist klar geworden, dass Mechanismen existieren, mit Hilfe derer Spam reduziert werden kann. Hierunter fällt in erster Linie IP-Blacklisting unterstützt durch IP-Adress-basierte Reputations-Systeme.

6. Literatur

[DiPo05] C. Dietrich, N. Pohlmann: E-Mail-Verlässlichkeit: Auswertung der Umfrage Ende 2004, 2005, http://www.internet-sicherheit.de/center-berichte.html