Mail-Service im IMT

Technische Dokumentation

[Sebastian Porombka,

Sabine Mennen,

06.05.2010]

Inhalt

1 Hardware

Die folgende Hardware wird genutzt:

- Lastverteilung
  - 2x Cisco Content Service Switch CSS 11503
- IMAP- / POP3-Server
  - Sun X4100 2x 2,6 GHz Dual Core AMD Opteron, mit Sun StorEdge 3510 Raid-Array
  - Sun X4100 2x 2,6 GHz Dual Core AMD Opteron, mit HP StorageWorks 1000 Modular Smart Array
  - SAN
- E-Mail-Frontends (SMTP)
  - 3x Sun X4100 2x 2,4 GHz Dual Core AMD Opteron
  - sowie 3 virtuelle Server [1]
- Scan-Cluster
  - Sun X4100 M2 2x 2,8 GHz Dual Core AMD Opteron (als Datenbank Server)
  - 2x Sun X4100 2x 2,8 GHz Dual Core AMD Opteron
  - sowie 2 virtuelle Server
- Mailinglisten-Server
  - ein virtueller Server

[1] Virtuelle Server: Serverinstanzen auf dem ESX-Cluster. 6x HP ProLiant DL580 G5, ESX 3.5 (Stand: Januar 2010). Server mit zugesicherten Ressourcen.

Fehler! Verweisquelle konnte nicht gefunden werden. zeigt die Struktur des Mail-Servers. Dazu gehören die eingesetzten Komponenten, deren Einordnung in vier Klassen sowie die Abhängigkeiten von anderen Diensten. Dunkel eingefärbte Komponenten deuten dabei auf virtuelle Server auf dem ESX-Cluster hin. Helle Komponente stellen physikalische Maschinen dar.

Der Zugriff durch den Benutzer findet bis auf eine Ausnahme (Web-Zugriff auf den Mailinglisten-Server) ausschließlich über einen redundant ausgelegten Load Balancer (in der Aufstellung unter Lastverteilung zu finden) statt. Dies ermöglicht das Austauschen und Hinzufügen von Servern, ohne dass der Benutzer etwas merkt. Ebenso ermöglicht es dieses Setup transparent für den Benutzer unterschiedliche Dienste, verteilt auf mehrere Rechner, unter dem zentralen Hostnamen `mail.uni-paderborn.de` zu präsentieren.

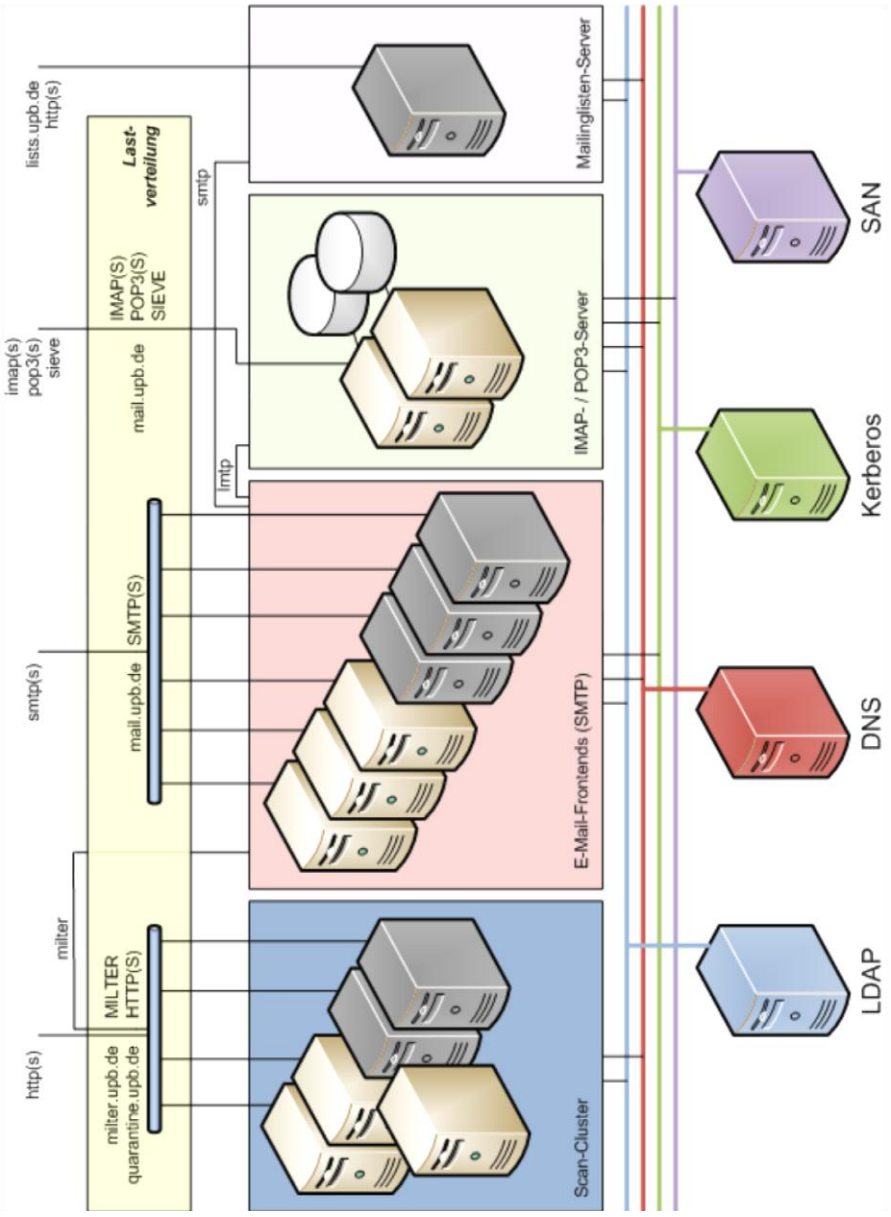

Abbildung 1: Struktur des Mailservers

2 Software

Auf den vier Rechnerklassen wird folgende Software eingesetzt:

- **Scan-Cluster**

  Zur Erkennung von Viren und Spam wird die von Sophos eingekaufte Softwarelösung PureMessage for Unix verwendet. Das System benötigt zwei Sorten von Rechnersystemen: die „Scanner"-Server und den Central-Server.
  Die Spam- und Virenerkennung erledigt dabei ein Cluster von Scannern, wohingegen der Central-Server die zentrale Koordinierung übernimmt. Die Scanner können 2 Stunden auf den Central-Server verzichten, danach werden aktualisierte Daten vom zentralen System benötigt.
  Der Zugriff auf die Rechner zur Analyse geschieht über den Load Balancer. Durch das Hinzufügen weiterer Scanner kann das System im Bedarfsfall skaliert werden.
  Updates von Viren und Spam Definitionen finden automatisch statt. Der aktuelle Softwarestand (Mai 2010) ist 5.5.9. Updates werden nach Prüfung zentral eingespielt.

- **E-Mail-Frontends (SMTP)**

  Die Rechner, die in diesem Dokument als „Frontends" bezeichnet werden, sind für die Annahme und Zustellung sowie die Zulieferung von lokalen E-Mails zum IMAP-Server zuständig. Dazu zählen die Verbindungsrichtungen
  • Universität extern → uni-interner Mailserver / lokales Postfach im IMT,
  • Universität intern → uni-interner Mailserver / lokales Postfach im IMT,
  • Universität intern → Universität extern.
  Darüber hinaus stellen sie für den Benutzer die Schnittstelle SMTP(S) bereit. Im Verlauf der Bearbeitung werden die E-Mails bei den Spam-/Virenfiltern eingereicht, und nach dortiger Bewertung passend weiter vermittelt. Als MTA wird Exim eingesetzt. Um den Scan-Cluster an die Frontends anzubinden, wurde die Milter-Schnittstelle[2] für den MTA exim selbst implementiert. Eine genauere Beschreibung des Lebenszyklus einer E-Mail, von der Annahme bis zum Einliefern bei einem uni-internen Mailserver oder dem IMT-Postfach, findet sich in Kapitel 3. Alle eingehenden E-Mails, passieren diese Systeme.

[2] Siehe http:\\www.milter.org

Der aktuelle exim-Softwarestand (Mai 2010) ist 4.63.

- **IMAP- / POP3-Server**

  Der IMAP-Server übernimmt die Verwaltung der Postfächer. Auf diesem Rechner werden E-Mails gespeichert, um sie per IMAP(S) und POP3(S) für die Benutzer bereitzustellen. Dabei werden neben den eigentlichen E-Mails zusätzliche Datenstrukturen (Meta-Daten) vorgehalten, um Zugriffzeiten und Antwortzeiten zu optimieren. Daneben existiert ein weiterer Speicherbereich, der für Archive vorgesehen ist. Die eigentlichen E-Mails sind auf per Fibre Channel angeschlossenen Raid-Arrays gespeichert, während die Meta-Daten und das Archiv im SAN gespeichert sind. Die Anbindung des SANs erfolgt redundant über das ISCSI Protokoll.

  Neben dem IMAP-Server, auf dem die Benutzer arbeiten, existiert ein Replikat, welches eine Kopie der Benutzer- und Meta-Daten sowie des Archivs vorhält. Im Desaster-Fall kann auf diese Kopie umgeschaltet werden. Die Übertragung der Daten vom produktiven IMAP-Server zum Replikat erfolgt asynchron. Die Verzögerung beträgt dabei ca. 5 Minuten. Als IMAP-Server wird Cyrus eingesetzt. Der aktuelle Softwarestand (Januar 2010) ist 2.3.13-6.

- **Mailinglisten-Server**

  Der Mailinglisten-Server ist für die Koordination der Mailinglisten sowie den Versand an die Abonnenten zuständig. Als Mailing-Listen-Software wird die Open-Source-Software Mailman 2 eingesetzt. Die Benutzerauthentifizierung für Benutzer des IMT findet per LDAP statt. Die Listenadministration erfolgt über eigene Passwörter, die in einem eigenen Format auf dem Server selbst gespeichert werden. Für die Bereitstellung des Web-Interfaces ist ein Apache-Webserver installiert, auf den per HTTPS-Protokoll zugegriffen werden kann. Für Benutzer ist eine Administration ihrer Listen nur per Webinterface oder E-Mail möglich. Sowohl die Annahme als auch die Zustellung von E-Mails geschieht über die Mail-Frontend-Rechner. Der Mailinglisten-Server ist für die Subdomain lists.uni-paderborn.de zuständig. Zusätzliche Adressen für Mailinglisten werden im LDAP eingetragen und von den E-Mail-Frontends ausgewertet. Um Mitgliederlisten von Mailinglisten per LDAP zu erzeugen, wurde eine Erweiterung entwickelt, um diesen Typ von Verteilern in Mailman abzubilden.

Um den Administrationsaufwand zu minimieren, werden alle Benutzerdaten mehrfach gespiegelt in einem LDAP-Verzeichnis gehalten. Auf diesem Verzeichnis basieren automatisierte Verfahren, mit denen Studierende und Mitarbeiter ein Postfach erhalten oder Postfächer gesperrt werden können. Um den Mailbetrieb in die Single-Sign-On-Umgebung des IMT einzubetten, ist die Anmeldung an den Frontends sowie dem IMAP- / POP3-Server per Kerberos möglich.

Für die Wiederherstellung des E-Mail-Systems im Katastrophen-Fall werden alle Daten täglich im Zeitraum von 20.00 bis 08.00 Uhr gesichert. Als Sicherungssoftware wird der Tivoli Storage Manager (TSM) eingesetzt. Die Sicherungspolicy für die Mailbox-Sicherung legt fest, eine E-Mail noch 90 Tage nach dem Löschen in der Sicherung zu belassen. Danach wird sie aus der Sicherung entfernt. Die Sicherung dient ausschließlich der Restaurierung des gesamten Systems, etwa nach einem Systemcrash oder -ausfall.

3 Umgang mit eingehenden E-Mails

3.1 Behandlung auf dem Frontend-System

Das Frontend-System umfasst die E-Mail-Frontends für die SMTP-Behandlung und den Scan-Cluster.

Eingehende E-Mails passieren verschiedene Stationen, um die uni-internen Mailserver und die Benutzer vor Spam und Viren zu schützen. **Fehler! Verweisquelle konnte nicht gefunden werden.** stellt ein Ablaufdiagramm dar, an dem die einzelnen Stationen einer eingehenden E-Mail abgelesen werden können.

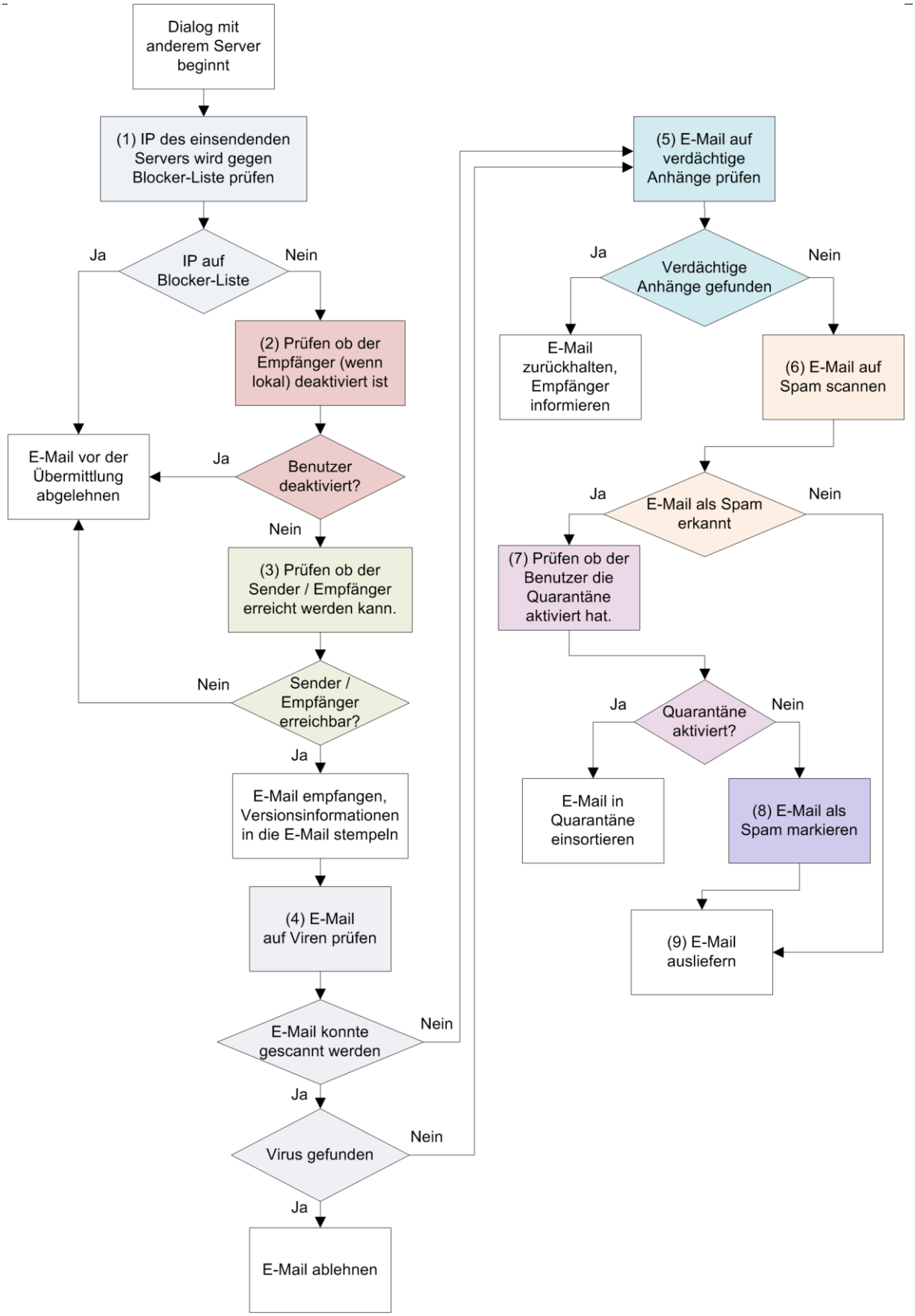

Abbildung 2: Behandlung einer E-Mail

(1) Nach dem Verbindungsaufbau wird die IP des einsendenden Servers gegen eine Blocker-Liste geprüft. In dieser Liste sind IPs von bekannten Spam-Versendern verzeichnet. Diese Liste wird von der Firma Sophos erzeugt und auf aktuellem Stand gehalten. Wird die IP in der Liste geführt, wird die Verbindung vor Annahme der E-Mail mit einem permanenten Fehler abgelehnt. Ist der Zugriff auf die Liste kurzzeitig gestört, wird ein temporärer Fehler signalisiert.

(2) Postfächer können für den E-Mail-Empfang deaktiviert werden. Dazu gehören Postfächer, die zu abgelaufenen Accounts gehören, sowie Postfächer die seit langer Zeit übergelaufen sind. Ist der E-Mail-Empfang für ein Postfach deaktiviert, werden eingehende E-Mails abgewiesen.

(3) Es wird geprüft, ob die E-Mail-Adresse des Empfängers und die des Absenders erreichbar sind. Ist dies nicht der Fall, kann die E-Mail nicht ausgeliefert werden und wird abgelehnt.

(4) Durch den in PureMessage für UNIX integrierten Virenscanner wird die E-Mail auf Viren geprüft. Die Signaturen werden von der Firma Sophos erzeugt und auf aktuellem Stand gehalten. Enthält die E-Mail einen Virus, wird diese mit einer endgültigen Fehlermeldung abgelehnt. Ist der PureMessage für Unix-Cluster vorübergehend nicht erreichbar oder antwortet er fehlerhaft, wird dem einliefernden System ein temporärer Fehler signalisiert[3].

(5) Nach dem Check durch den Virenscanner werden die Attachments auf potentiell gefährliche Inhalte geprüft. Dazu werden die Dateiendungen[4] und die Anzahl der Leerzeichen in den Dateinamen überprüft. Wird ein Anhang als verdächtig eingestuft, wird die Datei zurückgehalten und der Empfänger informiert.

(6) Um unerwünschte digitale Werbepost zu minimieren, wird jede E-Mail von einem umfangreichen Regelwerk auf ihre Spam-Wahrscheinlichkeiten untersucht. Dazu wird im Inhalt der E-Mail nach verschiedenen Mustern gesucht. Zu jedem Muster ist ein

[3] Diese Einstellung schützt vor einem Denial of Service.
[4] Abgefangen werden Dateinamen mit folgenden Endungen: ade, adpx, app, bas, bat, chm, cmd, com, cpl, crt, exe, fxp, hlp, hta, inf, ins, isp, js, jse, lnk, mda, mdb, mde, mdt, mdw, mdz, msc, msi, msp, mst, ops, pcd, pif, prf, prg, reg, scf, scr, sct, shb, shs, url, vb, vbe, vbs, wsc, wsf, wsh, xsl (Stand: Januar 2010)

numerischer Wert gespeichert, der angibt, mit welcher Wahrscheinlichkeit das gefundene Muster in einer Spam-Mail vorkommt. Aus den gefundenen Merkmalen wird ein Wert errechnet, der angibt, mit welcher Wahrscheinlichkeit es sich bei dieser E-Mail um unerwünschte Werbung handelt. Dieser Wert sowie Hinweise auf die gefundenen Muster, werden im Header der E-Mail gespeichert. (X-IMT-Spam-Score, X-PerlMx-Spam)

(7) Hat der Benutzer die Quarantäne aktiviert, werden E-Mails mit einer Wahrscheinlichkeit von ≥ 50% als Spam einsortiert. In einem täglich an den Benutzer verschickten Digest werden alle einsortierten E-Mails mit ihrer Betreffzeile vermerkt. Der Digest enthält auch Informationen über die nachträgliche Zustellung ausgewählter Nachrichten.

(8) Bei E-Mails mit einer Spamwahrscheinlichkeit >= 70% wird der Text „{SPAM!}“, bei E-Mails mit Spamwahrscheinlichkeit zwischen 50% und 69% wird der Text „{SPAM?}“ in die Betreffzeile eingefügt.

(9) E-Mails die nicht als Viren abgelehnt oder als Spam in der Quarantäne einsortiert wurden, werden ausgeliefert. E-Mails, die in ein lokales Postfach im IMT einsortiert werden, übergibt das Frontend per LMTP direkt dem IMAP- / POP3 Server. Alle anderen E-Mails werden per SMTP an externe E-Mail-Server, den Mailinglisten-Server oder uni-interne Mailserver übergeben.

3.2 E-Mail-Behandlung auf dem IMAP- /POP3 Server

Wenn eine E-Mail an eine lokale E-Mail-Adresse des IMT adressiert ist, wird sie nach erfolgter Prüfung per LMTP an den IMAP-/POP3 Server übergeben. Während der Übergabe prüft der IMAP-/POP3 Server, ob das Postfach vorhanden oder bereits überfüllt ist. Wenn das Postfach überfüllt ist, erhält das einliefernde E-Mail-Frontend eine temporäre Fehlermeldung, so dass die E-Mail in der Warteschlange auf dem Frontend verbleibt und in länger werdenden Abständen Zustellversuche unternommen werden[5]. Schlagen alle Zustellversuche fehl, erhält der Absender eine Fehlermeldung.

[5] Bis zu 7 Tagen

Sollte das Postfach nicht vorhanden sein, erhält das einliefernde Frondend eine endgültige Fehlermeldung, was eine Fehlermeldung an den Absender der E-Mail auslöst.
Nachdem die E-Mail vom IMAP- /POP3 Server angenommen wurde, erfolgt eine Prüfung der serverseitig hinterlegten Sieve-Regeln. Die E-Mail wird entsprechend den vorhandenen Regeln einsortiert, gelöscht oder weitergeleitet. Liegen keine Regeln vor oder enthalten sie syntaktische Fehler, wird die E-Mail in den Posteingang des Postfachs einsortiert. Semantische Fehler kann der Server nicht prüfen, sodass bei semantisch fehlerhaften Sieve-Regeln E-Mails verloren gehen können.

4 Verwaltung von Unterdomänen

Zusätzliche Unterdomänen werden in die Softwarekonfigurationen eingetragen. Darüber hinaus werden die Adressen der Unterdomänen in die zentrale Datenbasis eingetragen, so dass sie per LDAP verfügbar sind.

5 Relayfunktion

Um unkontrolliertes Verschicken von E-Mails, z.B. von mit Viren und Bots infizierten Rechnern, in der Universität einzudämmen, können E-Mails nur über das zentrale Mail-System von beliebigen Computern verschickt werden. Dazu stellt das IMT ein Mail-Relay zur Verfügung. E-Mails von nachgelagerten Systemen werden, anstatt sie direkt an externe Server auszuliefern, beim Mail-Relay aufgegeben. Bei E-Mails von externen Mail-Servern läuft dieser Prozess analog ab. Möchte ein nachgelagertes System für seine verwaltete Subdomäne E-Mails empfangen, wird das zentrale Mail-System im DNS als MX eingetragen. In dem zentralen Mailsystem wird eine Route eingerichtet, die für die passende Weiterleitung sorgt. Die durchgeleiteten E-Mails werden, wie vorher beschrieben, auf die Spamwahrscheinlichkeit untersucht und markiert sowie auf Viren gescannt.

6 Webmail

Der Zugriff per https://webmail.uni-paderborn.de erfolgt mit Hilfe der Horde-Suite[6]. Die Software ist auf dem Web-Cluster des IMT installiert und von den Webmastern betreut. Der

[6] http://horde.org

Zugriff auf die Postfächer erfolgt per IMAPS. Zum Senden per SMTP hat der Webmailer das Recht, die Absenderadresse beim Versenden zu setzen.

Der installierte Webmailer unterstützt das Sieve-Protokoll. Die Verbindung zum IMAP-/POP3 Server ist per TLS gesichert.