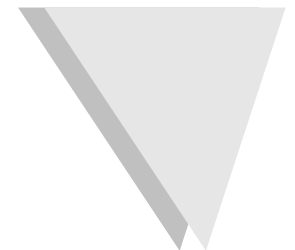
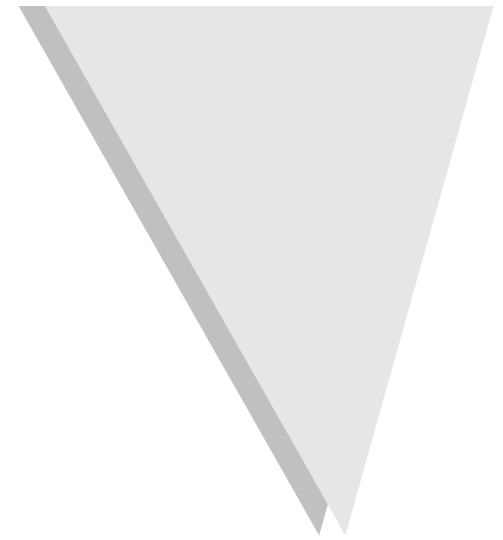


# Spam – und wie man ihn los wird

Tipps zur komfortablen Spam-Entsorgung  
[andreas@rittershofer.de](mailto:andreas@rittershofer.de)  
Mai 2003



# Was ist Spam?

- Unter Spam versteht man i.a. alle unerwünschten Mails.
- Eine genauere Unterscheidung ist meist nicht notwendig – Schrott ist und bleibt Schrott.

# Warum gibt es soviel Spam?

- Es ist die billigste Möglichkeit, Werbung oder Weltanschauungen zu verteilen.
- Gefördert wird die Verbreitung durch offene Mail-Relays, das sind falsch konfigurierte Mailserver, die beliebige Mails akzeptieren und weiterschicken.

# Adressangaben in Spam

- Die Adressangaben in Spam-Mails sind in aller Regel gefälscht.
- Sie können sich auf **nichts** verlassen!
- Auch im Empfängerfeld zeigt manchmal der Mail-Client eine fremde Adresse an, dennoch haben Sie die Mail erhalten – alles gefälscht!

# Was tun mit Spam?

- **Löschen!**
- In aller Regel ist es ganz falsch, auf Mails zu antworten oder sich vermeintlich aus einem Verteiler auszutragen – Sie bestätigen damit nur, dass diese Mail gelesen wurde und erhalten noch mehr Spam.

# Mailfilterung

- Viele moderne Mailclients sind in der Lage, eingehende Mails nach vom Anwender selbst festgelegten Regeln zu filtern.
- Nicht jeder Mailclient ist allerdings flexibel genug – evtl. ist es Zeit für einen Umstieg.
- Der Mailclient muss auch nach beliebigem Text im Header und seinen Feldern filtern können. Der Header wird im Regelfall nur verkürzt angezeigt, ist aber sehr wichtig.

# Mails über BelWü

- Recht einfach ist der Fall bei Mails, die über das BelWü hereinkommen, z.B. für den Landesbildungsserver oder von dort versorgte Schulen.
- BelWü filtert nicht nur Viren und Würmer aus.
- Jede Mail wird nach komplexen Regeln untersucht, ob es sich um Spam handeln könnte – falls die Wahrscheinlichkeit groß genug ist, erfolgt eine entsprechende Markierung.

# X-Spam-Level

- Bei mutmaßlichem Spam wird im Header der Mail ein Feld ergänzt: X-Spam-Level
- Dahinter folgen fünf oder mehr x – darunter erfolgt ohnehin keine Markierung.
- Es ist Sache des Empfängers, dieses Headerfeld auszuwerten.

# Filtern, verschieben, löschen

- Im Mailclient kann nun dafür gesorgt werden, dass Mails mit der Zeichenkette „X-Spam-Level: xxxxx“ (damit sind auch mehr als fünf x erfasst) gefiltert werden.
- Möglich sind: Verschieben in einen gesonderten Ordner zur manuellen Nachkontrolle oder
- sofortiges Löschen.
- Nicht jeder Mailclient ist dazu in der Lage – „richtige“ Mailclients können das ;-)

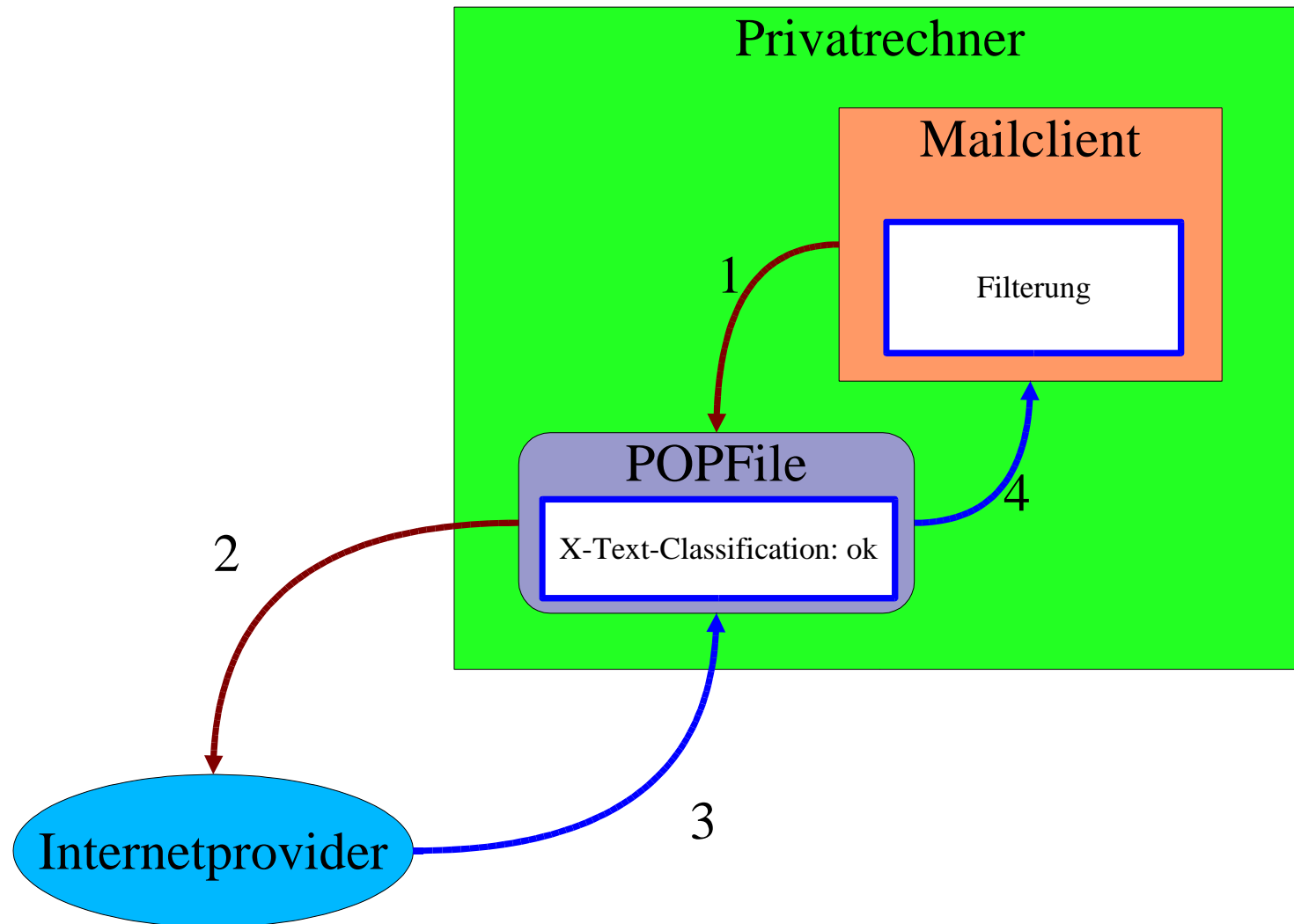
# POPFile

- Zur Erfassung *aller* eingehenden Mails kann z.B. POPFile <http://popfile.sourceforge.net/> installiert werden.
- Es ist ein POP3-Proxy, der alle eingehenden Mails durchsieht und nach selbstgelernten Regeln markiert.
- Die Bedienung von POPFile erfolgt über ein Browser-Interface auf <http://127.0.0.1>
- Anschließend kann im Mailclient wieder gefiltert werden.

# Funktionsprinzip von POPFile

- Der Mailclient holt seine Mails durch POPFile hindurch.
- In POPFile müssen „Buckets“ angelegt werden.
- Neu geholte Mails müssen zunächst manuell klassifiziert werden, damit POPFile lernen kann, wie die Mails den Buckets zuzuordnen sind.
- Künftig muss nur noch bei falscher Zuordnung korrigiert werden, um die Treffergenauigkeit der automatischen Klassifizierung zu erhöhen.

# Funktionsdiagramm



# X-Text-Classification

- Bei mutmaßlichem Spam wird in der Mail ein Headerfeld ergänzt: X-Text-Classification
- Dahinter folgt der Name des Buckets, dem diese Mail zugeordnet wurde.
- Es ist nun Sache des Mailclients, entsprechend zu reagieren.

# Buckets in POPFile

History Buckets Magnets Configuration

## Summary

Bucket Name	Word Count	Unique Words	Subject Modification	Quarantine
<a href="#">antw</a>	41,154	2,017	Disabled globally	Off <input type="button" value="Turn On"/>
<a href="#">md</a>	43,193	15,610	Disabled globally	Off <input type="button" value="Turn On"/>
<a href="#">ok</a>	62,545	7,274	Disabled globally	Off <input type="button" value="Turn On"/>
<a href="#">spam</a>	84,736	24,740	Disabled globally	Off <input type="button" value="Turn On"/>
<b>Total</b>	231,628			


---

### Classification Accuracy

Emails classified: 6,198  
Classification errors: 417

---

Accuracy: 93.27%




0% 100%

(Last Reset: Mon Apr 14 22:48:09 2003)

### Emails Classified

Bucket	Classification Count
<a href="#">antw</a>	562 (14.58%)
<a href="#">md</a>	-1 (-0.02%)
<a href="#">ok</a>	1,407 (36.5%)
<a href="#">spam</a>	1,886 (48.93%)



100%

# Klassifizierung in POPFile

History Buckets Magnets Configuration Security Adva

## Recent Messages

Search From/Subject:  Find Filter By:

ID	From	Subject	Classification	Should be
1	Andreas Rittershofer <andreas@rittersho...	Testmail Spam Sex Penis Enlargement Mo...	ok	<input type="text"/>
2	Andreas Rittershofer <andreas@rittersho...	testmail ok	ok	<input type="text"/>
3	MP3-Downloads <mp3download79xbbn@lycos...	LEGALE MP3 Downloads	unknown class	<input type="text"/>
4	tt_info@transtec.de	Email-News/190503 (ID: 533550)	ok	<input type="text"/>
5	H.Rattinger@t-online.de	Arbeitsbereich MMB	ok	<input type="text"/>

Reclassify

To remove entries in the history click:

# Qualität der Klassifizierung

- POPFile lernt durch den Nutzer, wie dessen spezifische Mailzusammensetzung zu klassifizieren ist.
- Die Qualität der Klassifizierung ist mit der Zeit – nach ein paar Wochen – sehr gut.
- Die Entlastung der INBOX ist deutlich spürbar.
- Spams werden nach einer kurzen Durchsicht gelegentlich entsorgt.

**Ende**

