SPAMASSASSIN: MISE EN PLACE ET RÉGIONALISATION (FR)

Philippe Marion

Journées Mathrice 31 mars - 2 avril 2015, Marseille

Philippe.Marion@lmpa.univ-littoral.fr

mathrice

ulco L.M.P.A Joseph Liouville
UNIVERSITÉ DU LITTORAL CÔTE D'OPALE

AVANT-PROPOS: CONTEXTE

CONTEXTE / HISTORIQUE

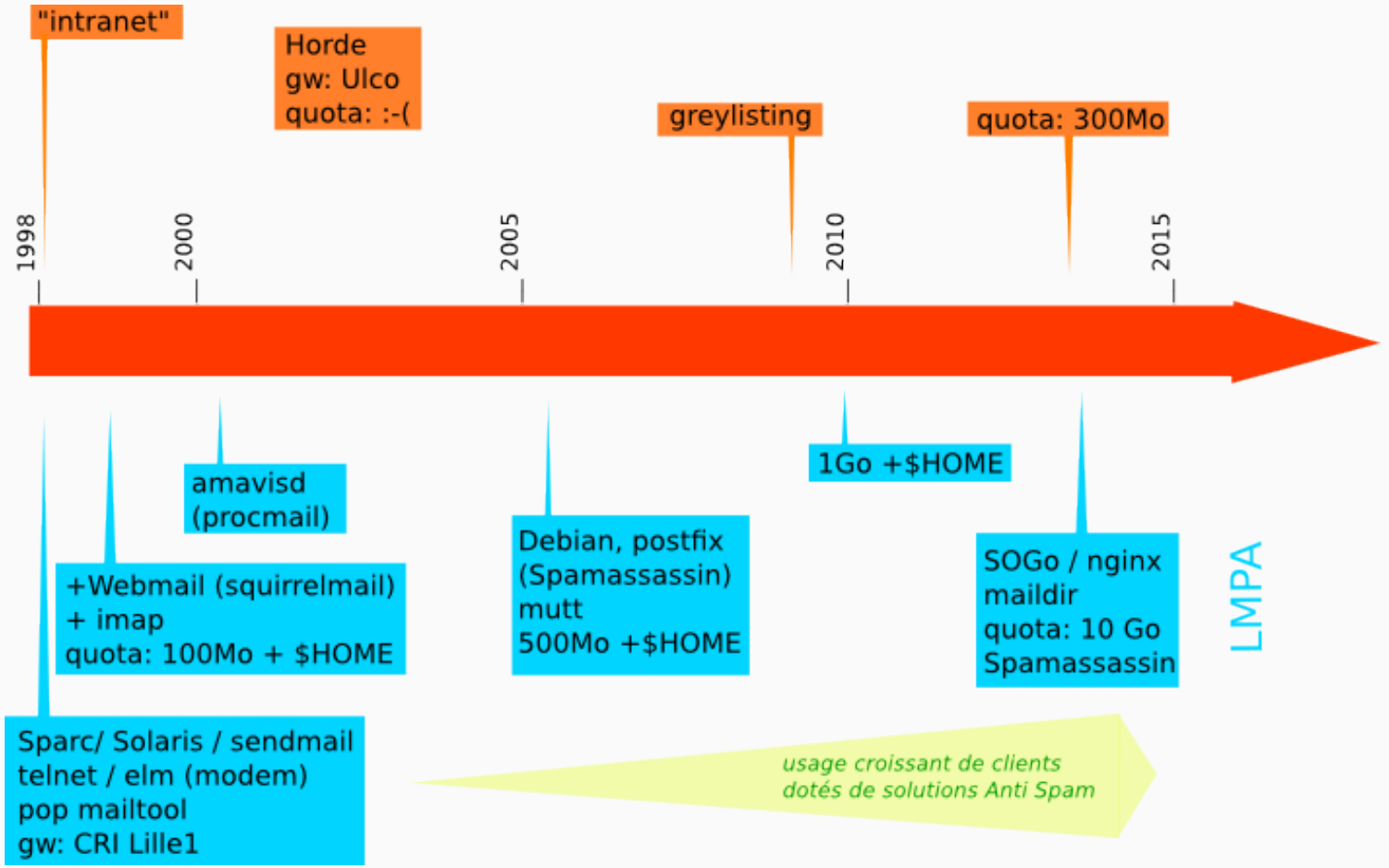


Figure: mail: historique LMPA

INSTALLATION DE SPAMASSASSIN

CONTRAINTES ET SOLUTIONS

Faire simple:

- existence d'une solution en amont,
- pas de BD supplémentaires,
- => exit la quadrilogie Postfix + postfixadmin + Dovecot + Mysql (couplée à amavisd-new Spamassassin, Clamav ... ou dspam)

solution:

- Seulement taguer les courriers,
- Spamassassin (Consorsium Apache),
- spampd (et non spamd/spamc)
  - efficacité proxy,
  - intégration à Postfix,
  - filtrage en amont (content_filter)

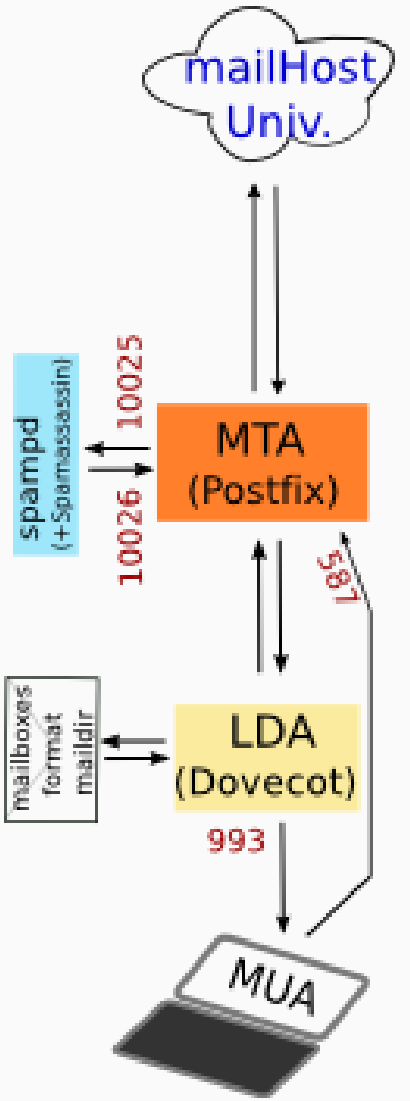
POSTFIX ET SPAMPD

spampd: `/etc/default/spampd`

- démarrage du service,
- listenport, destport
- droits: user, group

postfix:

- `main.cf`: content_filter=scan[127.0.0.1]:10025
- `master.cf`:
  scan unix - - y - 10 smtp
  localhost:10026 inet n - y - 10 smtpd
  -o content_filter=
  -o ...la suite!



SPAMASSASSIN: MISE EN PLACE

installation: package Debian (binaires + 40^{aine} de fichiers règles.cf)

initialisation:

- `/usr/bin/sa-learn –ham $MAIL/.Ham/cur`
- `/usr/bin/sa-learn –spam $MAIL/.Spam/cur`

crontab:

```
/usr/bin/sa-update && /etc/init.d/spampd restart >
/dev/null 2&>1
 apprentissage.sh > /dev/null 2&>1
```

performances ?

les premiers résultats sont très médiocres, un paramétrage local s'impose! ...

PARAMÉTRAGES, RÉGIONALISATION (FR)

PARAMÉTRAGES /etc/spamassassin/local.cf

les premières pistes:

- baisser la limite `required score` : 5 -> 3.5
- UTF8 ? (mauvaise idée)
- identification des spammers
  - les adresses récurrentes (blacklist?),
  - les faux-positifs (whitelist?),
  - les classiques: jeux, médicaments, voyances, rencontres,
  - les commerciaux: bons plans (réductions etc.), mutuelles, optimisation fiscale

sources:

- WRITING RULES - SA APACHE
- JOHN GALLET - SAPHIRTECH

/etc/spamassassin/local.cf ...

Outils: expressions régulières -perl-

règle simple

```
body LOCAL_BODY_reduction /\b(r.+duction|remise)\b/i
score LOCAL_BODY_reduction 0.3
describe LOCAL_BODY_reduction  offre réduction
```

règle multiple

```
meta LOCAL_MULTIPLE_Exemple (((0.8 * LOCAL_ex1)
                      + (0.5 * LOCAL_ex2)) > 1)
score LOCAL_MULTIPLE_Exemple 2
describe LOCAL_MULTIPLE_Exemple  exemple1 + exemple2
```

/etc/spamassassin/local.cf …

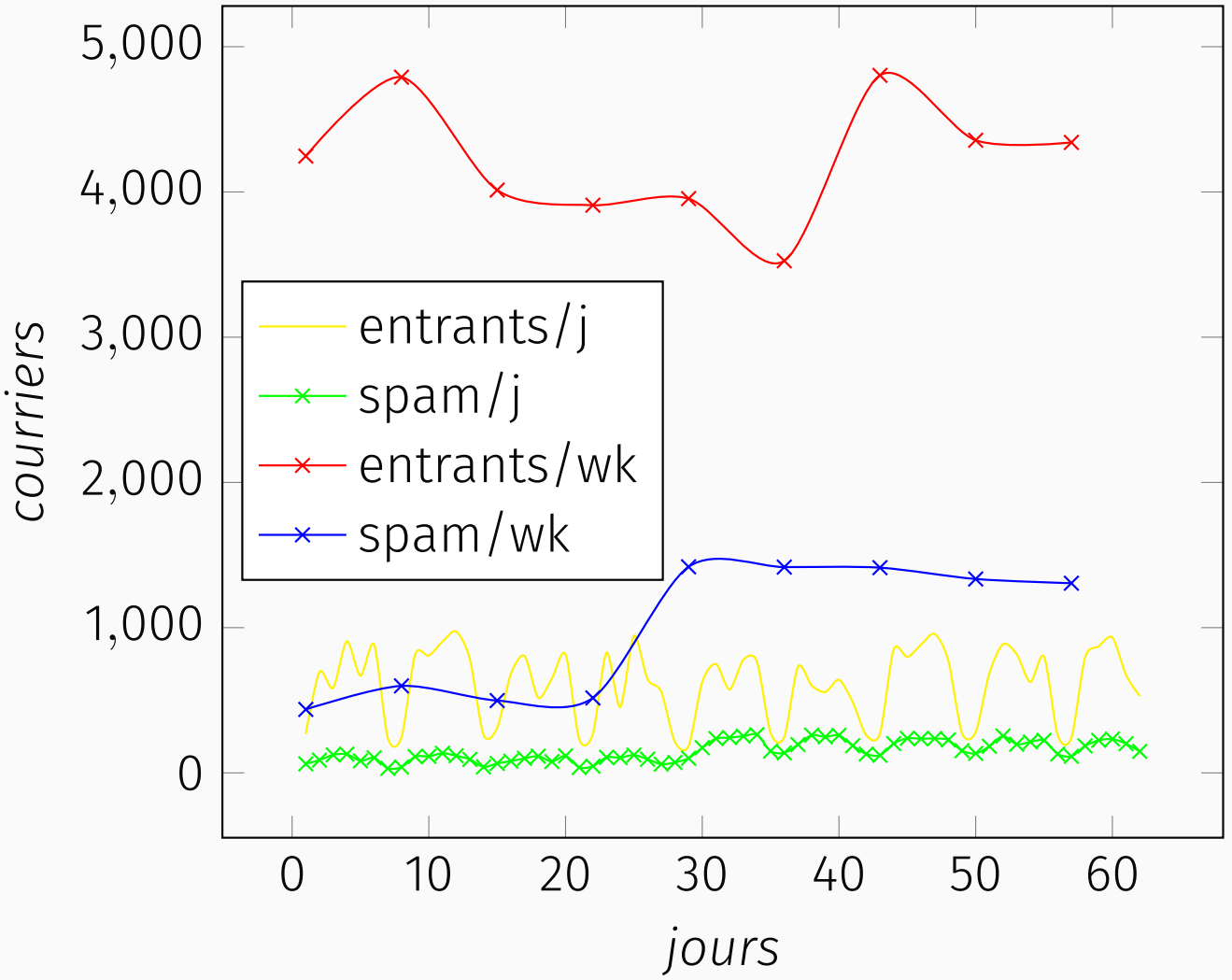
Posologie :

- ne pas trop charger le `local.cf` (CPU)
- soulager avec un minimum de *whitelist* (univ., fournisseurs) & *blacklist*
- spécificités françaises: CNIL, lois de Robien, Girardin, mailer,
- récupérer les faux-positif: Math, Académie, LateX
- réévaluer la taille max. d'analyse de message

EXTRAITS local.cf

```
body LOCAL_math /\b(comit.+ scientifique|Math.+mat.+
  |Universi(d|t).{1,3}|s.+minair(e|es)? de math.+)\b/i
score LOCAL_math -3
body LOCAL_ac /\b(coll.+ges|lyc.+(e|es)|acad.+miqu(e|es)
   |ED\s{1,2}SPI|doctorant)\b/i
score LOCAL_ac -3
body FR_SPAMISLEGAL /\b(Conform.+ment|En vertu).{0,5}
      (article.{0,4}34.{0,4})?la loi\b/i
body FR_PAYLESSTAXES /\b(paye|calcul|simul|r.+dui|investi
     .{1,7}(moins|vo|ses).{0,5}imp.+t(s)?\b/i
body FR_REALESTATE_INVEST /\b(loi)? (de.robien|girardin)
  .{1,15}(neuf|recentr.+|ancien|IR|IS|imp.+t(s)?|indust
   riel(le)?)\b/i
```

ANALYSE LOGS



CONCLUSION

QUELQUES RÉFÉRENCES

http://spamassassin.apache.org/
http://wiki.apache.org/spamassassin/WritingRules
http://www.saphirtech.com/spamassassin_fr.txt
http://wiki.apache.org/spamassassin/
IntegratePostfixViaSpampd
http://www.postfix.org/SMTPD_PROXY_README.html

Questions?