Hur hanterar KTH spam?

Mattias Amnefelt
mattiasa@kth.se

Varför får KTH mycket spam?

- KTH har haft epost sedan 1986.
- KTH har en öppen informationspolicy. Det är lätt att hitta t.ex. telefonboken på webben.
- Många användare publicerar sina adresser i samband artiklar, rapporter och liknande.

KTH:s organisation

- Central drift av domänen kth.se med tillhörande epostsystem
- Delegerad drift av de flesta institutioner, varje institution har en eller flera servrar.
- Totalt ca 100 epostservrar
- Ca 5000 brev/timme under normal last
- Vi ville ha ett spamfilter som kunde serva alla dessa underdomäner.

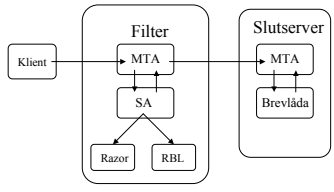
Krav på systemet

- Systemet skall fånga så mycket spam som möjligt
- Så lite icke-spam som möjligt skall fastna.
- Lättmodifierat
- Klara lasten
- Klara tillgänglighetskraven

Tillgänglighet och last

- Redundans – om filtersystemet går sönder skall inte alla underdomäner bli drabbade.
- Last – vi måste inte bara klara dagens last. Systemet måste också gå att skala upp.
- Spamhantering av epost är lätt att parallellisera. Det behöver inte göras på slutdestinationen.
- Epostprogramvaror är duktiga på att hantera parallella servrar.

Systemöversikt



Spamassassin

Spamassassin är ett programpaket som ger varje brev en poängsumma.

- SA matchar brevet mot ett antal regler.
- Ett brev får poäng för varje regel det matchar.
- Om poängsumman blir tillräckligt stor anses brevet vara spam.
- Poängen är genererade med genetisk algoritm.
- Frågar externa databaser. T.ex. razor, rbl.
- Stoppar in headrar i brevet om det är spam.

Spamassassin forts.

Exempel på regler

- Brevet har avsändare listad i osirusoft.
- Brevet innehåller mellan 30% och 40% HTML.
- Brevet innehåller mellan 80% och 90% HTML.
- Brevet innehåller bara bilder.
- Brevets checksumma finns med i razor.

Utvärdering av lösningen

- Positivt
  - Fångar en stor del av våra spam
  - Skalar väl
  - Stabil
- Negativt
  - Fångar brev som inte är spam
  - Kräver en del underhåll (whitelistor o. dyl.)

Vad gör man när SA tycker ett brev är SPAM?

Tre alternativ:

1. Låt användaren ta hand om det
   Kan upplevas som positivt
2. Lägg allting i en stor låda och gå igenom centralt.
   Tar mycket tid. Tråkigt för de som arbetar med det.
   Kostnadseffektivt. Bra möjligheter till utvärdering.
3. Kasta
   Kan vi som myndighet inte göra.

Konfiguration i praktiken

- Parallella maskiner, alla med namnet mx.kth.se

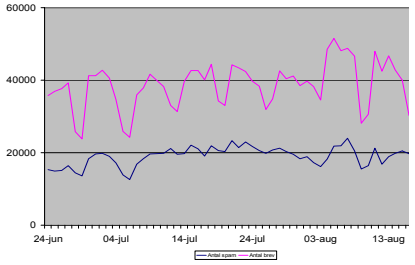
```
mx.kth.se.      IN A 130.237.48.98
mx.kth.se.      IN A 130.237.32.140
```

- MX-rr i DNS för de domäner som skall ha filtrering ex:

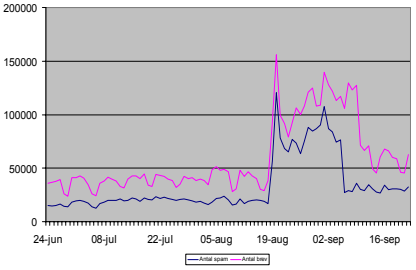
```
e.kth.se.       IN MX 10 mx.kth.se.
```

- Detta ger oss frihet att lägga till fler parallella maskiner utan att behöva kontakta alla som använder tjänsten.

Hur mycket spam får vi?



Vad hände när Sobig.F kom?



Frågor?

Mattias Amnefelt
IT-Enheten
KTH
100 44 Stockholm

mattiasa@kth.se

+46-8-790 8937