Towards Proactive SPAM Filtering

DIMVA 2009

Laboratory for Dependable Distributed Systems

UNIVERSITÄT MANNHEIM

Survey

- Motivation
- Sandnet Setup
- Template Creation
- Preliminary Results
- Summary & Future Work

Motivation

- SPAM is unwanted
- Why templates for filtering:
  - Templates more precise than current methods? (Bayes Filter, Reputation based, ...)
  - Templates send to Bots are encrypted
    - Retrieve template from memory of running bot - too complex?

Example Template I

```
Received: by 192.168.54.34 with SMTP id nacZcMBB;
        for <{%MAIL_TO}>; Wed, 30 Aug 2006 01:40:03 -0700
Message-ID: <000001c6cc0f$e36f8470$2236a8c0@amjit>
Reply-To: "{%NAME_FROM}" <{%MAIL_FROM}>
From: "{%NAME_FROM}" <{%MAIL_FROM}>
To: {%MAIL_TO}
Subject: Re: tiRXda
Date: Wed, 30 Aug 2006 01:40:03 -0700
MIME-Version: 1.0
Content-Type: multipart/alternative;
        boundary="----=_NextPart_000_0001_01C6CBD5.3710AC70"
X-Priority: 3
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 6.00.2800.1106
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2800.1106
```

In this example the body is fixed

Example Template 2

Example: Command {file "body.html", quoted printable} tells the bot to substitute the body.html file

Xarvester Botnet

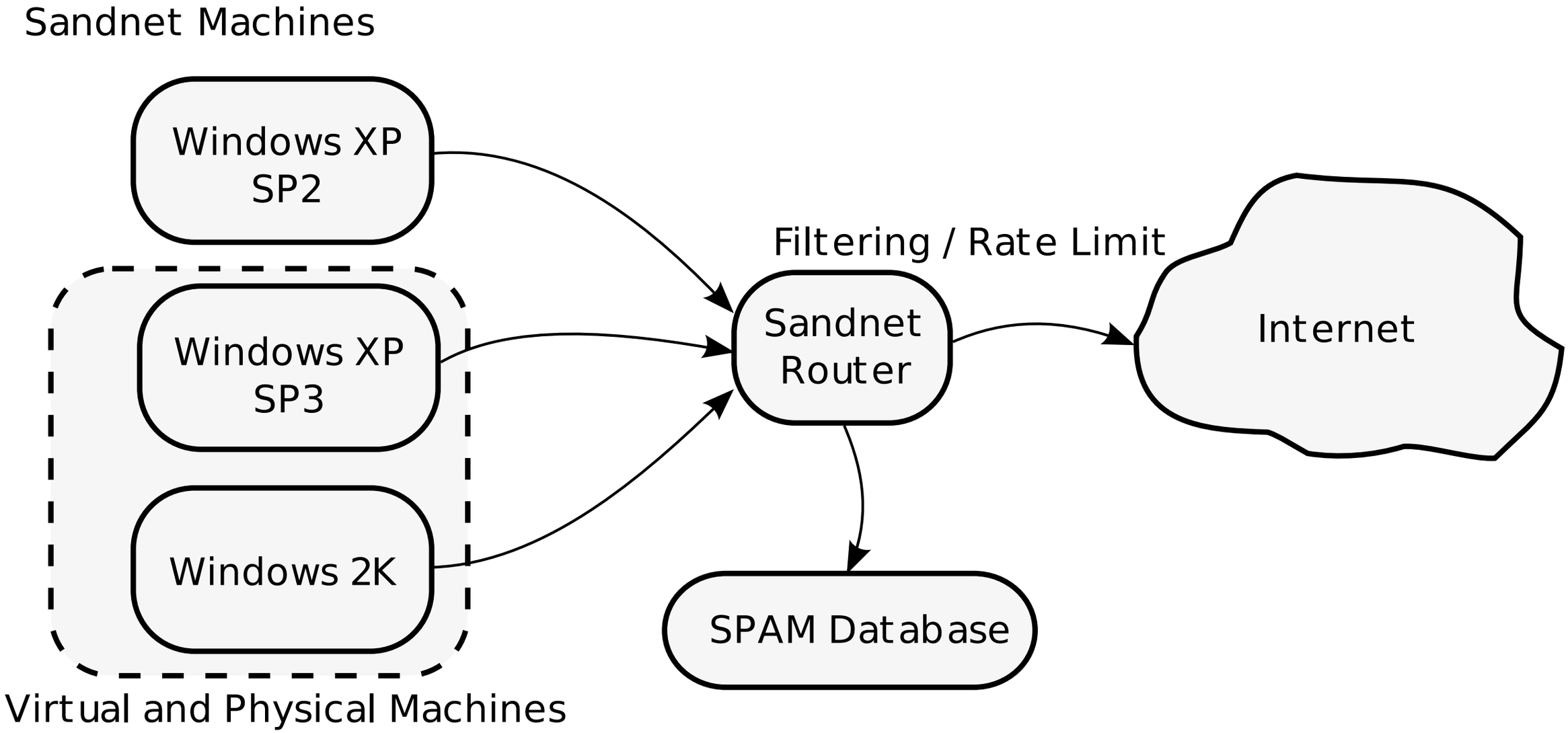
```
This is a multi-part message in MIME format.

--{teml:var boundary}
Content-Type: text/plain;
    charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

{file "body.html", html-plain-quoted-printable}
--{teml:var boundary}
Content-Type: text/html;
    charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

{file "body.html", quoted-printable}

--{teml:var boundary}--
```

Quelle: www.marshal8e6.com

Sandnet Setup

Running Spam Bots

Sandnet I


Sandnet Machines
Windows XP SP2
Windows XP SP3
Windows 2K
Virtual and Physical Machines
Filtering / Rate Limit
Sandnet Router
Internet
SPAM Database

Sandnet 2

- Spam Email are collected at the gateway (mbox)
- Filtering of malicious traffic + rate limit
- How to handle test emails send by bots?
  - Currently blocked
- Our current setup runs the bots only for a limited time

Generating Templates

The Algorithm

Template Creation I

- The Template Creation Algorithm:
  - Take first email as starting template
  - Sort emails according to their length
  - Take next email as comparing template
    - Common Substring Extraction
  - Add emails to the template as long as threshold is not exceeded

Template Creation 2

```
function genTempl(emails, treshold):
emails_in_template = List()
email = emails.pop()

alpha.learn(email)
emails.sort(alpha)      # sort emails by editDist to alpha
updated = True

while updated do
  updated = False
  while "emails not empty" do
    email = emails.pop()
    beta = alpha
    beta.learn(email)
    if (#beta/#alpha) > threshold        # non-placeholder
    then
      emails.append(email)
    else
      alpha = beta
      emails_in_template.append(email)
      updated = True

return alpha
```

Example Template I

```
Subject\:\ your\ new\ job\
X\-Mailer\:\ Microsoft\ Outlook\ Express\ 6\.00\.([\d]){4,4}\.([\d]){3,4}\
Body\:\
Hello\,\ I\ am\ Jennifer\.\
I\ am\ manager\ of\ Russian\ reselling\ company\ Startmaster\.\
http\:\/\/startmaster\.ru\
Our\ company\ need\ US\ partners\ for\ dropshipping\.\
We\ buy\ staff\ in\ the\ USA\ and\ resell\ it\ to\ our\ clients\ in\ Eastern\ Europe\ \(including\ Russi
We\ have\ contracts\ with\ US\ resellers\ but\ unfortunately\ some\ of\ these\ resselers\ don\'t\ ship\
If\ you\ are\ interested\ in\ cooperation\ we\ offer\ the\ following\ conditions\:\
You\ recieve\ a\ package\
Then\ we\ send\ you\ pre\-paid\ shipping\ label\ \(we\ have\ our\ own\ USPS\ account\)\,\ you\ should\ p
Then\ you\ go\ to\ the\ nearest\ USPS\ office\ and\ ship\ this\ package\ as\ soon\ as\ possible\.\
We\ will\ pay\ for\ your\ work\ via\ Paypal\.\
The\ first\ month\ of\ work\ you\ will\ get\ \$20\ per\ package\ \(it\ is\ some\ kind\ of\ verification\
\
If\ you\ are\ interested\ please\ provide\ the\ following\ details\ at\:\ resume\@startmaster\.ws\ to\ g
Name\
Age\
Martial\ status\
Telephone\ number\
Address\ for\ receiving\ package\ \(will\ be\ delevered\ 9\:30\ am\ \-\ 17\:30\ pm\)\
\
\
If\ you\ are\ interested\ in\ this\ offer\ please\ write\ at\:\ resume\@startmaster\.ws\ for\ more\ info
\
Thanks\ for\ your\ time\.\
\
Jennifer\.\
```

Only X-Mailer Changes

Generated from 1175 emails

Example Template 2

```
Subject\:\ ([\!\-\?\,\%\s\w]){16,49}\
X\-Mailer\:\ Microsoft\ Outlook\ Express\ 6\.00\.([\d]){4,4}\.([\d]){3,4}\
Body\:\
Es\ lauft\ im\ Bett\ nicht\ mehr\ wie\ frueher\?\ Haben\ Sie\ das\ Gefuehl\,\ dass\ i
d\ intensiveren\ Sex\?\
\
Das\ Leben\ ist\ zu\ kurz\ \-\ geniessen\ Sie\ das\ in\ vollen\ Zuegen\.\
Mit\ Geld\ kann\ man\ nicht\ alles\ kaufen\!\ Die\ Potenz\ und\ ueber\ 20\ Minuten\ S
\
Mit\ unserem\ Produkt\ vergessen\ die\ Potenzprobleme\ und\ haben\ wieder\ Spass\ am\
genau\ das\ Richtige\ fuer\ Sie\!\
Das\ Geld\ kommt\ und\ geht\ \-\ unvergessliches\ Sex\-Erlebnis\ bleibt\!\
\
Bestellen\ Sie\ jetzt\ und\ vergessen\ Sie\ Ihre\ Enttaeuschungen\,\ anhaltende\ Vers
\
Jetzt\ bestellen\ und\ naechste\ Woche\ erhalten\ \-\ 12\ Tb\.\ umsonst\ zum\ Weihnac
\
http\:\/\/ageclothe\.com\
\
Frohe\ Weihnachten\
```

Only Subject and X-Mail change

Generated from 4741 emails

Example Template 3

```
Subject\:\ ([\!\@\$\.\,\-\:\#\%\|\?\s\w]){19,59}\
X\-Mailer\:\ Mediacomm\ Communicator\ 1\.([\d]){1,1}1\
Body\:\
Ever([A-Za-z]){0,1}y\ single\ med\ is\ ([\<E0>A-Za-z]){1,1}vailab([\|A-Za-z]){1,1}e\ here\ i
\
Pr([\<EC>A-Za-z]){1,1}c([\<E9>A-Za-z]){1,1}s\ are\ direct([A-Za-z]){0,1}\ wholesal([A-Za-z])
-Za-z]){0,1}oorstep\.\
\
\-\ Pa([\<EC>A-Za-z]){1,1}nkillers([A-Za-z]){0,1}\
\-\ A([\<F1>A-Za-z]){1,1}t([A-Za-z]){0,1}i\-depressants\
\-\ ED\ meds\ \(for\ massive([A-Za-z]){0,1}\ bedr([\<F5>A-Za-z]){2,2}m\ acti([\<F5>\<F1>A-Za
\-\ Antibx\
\-\ ([\[\]\<E0>A-Za-z]){2,4}gic([A-Za-z]){0,1}\ Blue\ p([\<EE>A-Za-z]){2,2}l\ \(from\ j([\<F
\
and\ many\ oth([\<E9>]){0,1}e([A-Za-z]){0,1}r\ class([\<E9>A-Za-z]){1,1}s\ of\ med([A-Za-z])
All\ cred([\<EE>A-Za-z]){1,1}t([A-Za-z]){0,1}\ c([A-Za-z]){0,1}ards\ accepted\ via\ sec([\<B
\ e([\<D7>A-Za-z]){1,1}press([A-Za-z]){0,1}\ ([\<E7>A-Za-z]){1,1}our([\<ED>A-Za-z]){1,1}er\
\
F([\<F6>A-Za-z]){1,1}rget\ de([A-Za-z]){1,2}ays\ and\ costly\ doct([\<F3>A-Za-z]){1,1}r\ vis
<F4>A-Za-z]){1,1}da([A-Za-z]){0,1}y\.\
\
http\:\/\/([A-Za-z]){2,5}s([A-Za-z]){2,5}\.cn\
```

Generated from 172 emails

More complex due to word mutations in the emails

Preliminary Results

Euro Dice Casino Case Study

Euro Dice Casino 1

- We generated a Template from 71 emails all collected during a single day in October 2008

```
Subject\:\ Euro\ Dice\ Casino\.\ 2\.500\$\ Extra\.\ Worauf\ warten\ Sie\?\
X\-Mailer\:\ Microsoft\ Outlook\ Express\ 6\.00\.2720\.3000\
Body\:\
Er\&\#246\;ffnen\ Sie\ noch\ heute\ ein\ Konto\ bei\ Europas\ f\&\#252\;hrendem\ \=\
Onlinekasino\,\ und\ Sie\ erhalten\ einen\ unglaublichen\ Bonus\ im\ Wert\ von\ bis\ \=\
zu\ 2\.500\ \&\#8364\;\!\ Leisten\ Sie\ ganz\ einfach\ Einzahlungen\ auf\ Ihr\ Konto\,\ \=\
und\ der\ Kasinokassierer\ schreibt\ sie\ Ihrem\ Konto\ zusammen\ mit\ einem\ \=\
Spielbonus\ in\ H\&\#246\;he\ von\ 100\%\ im\ Wert\ von\ bis\ zu\ 500\ \&\#8364\;\ f\&\#252\;r\
jede\ der\ ersten\ f\&\#252\;nf\ Einzahlungen\ gut\.\
\
Euro\ Dice\ Casino\ ist\ genauso\ sicher\ wie\ ein\ landbasiertes\ Kasino\,\ \=\
verf\&\#252\;gt\ aber\ \&\#252\;ber\ eine\ gr\&\#246\;\&\#223\;ere\ Auswahl\ an\ \=\
realistischen\ und\ aufregenden\ Spielen\ und\ ist\ daher\ die\ erste\ Wahl\ unter\ \=\
den\ Online\-Spielern\.\ Sprechen\ Sie\ noch\ heute\ mit\ anderen\ Spielern\ und\ \=\
\&\#252\;berzeugen\ Sie\ sich\ selbst\!\
\
Verpassen\ Sie\ nicht\ Ihre\ Chance\;\ klicken\ Sie\ hier\,\ um\ zu\ Euro\ Dice\ \=\
Casino\ zu\ gelangen\!\
eurocasinokg\.comNext\ Body\ Part\:\
```

Euro Dice Casino 2

- We collected SPAM emails advertising the casino during June 2008 till April 2009
- A total of 493 emails advertising the Euro Dice Casino were collected at our spamtraps (some free email accounts)
- Checking against our previously generated template revealed a detection rate of only 5.3%
- All matches are emails received at the spamtraps during October 2008

Euro Dice Casino 3

- We added a randomly chosen email from the spamtrap emails to our template generation process

```
Subject\:\ Euro\ Dice\ Casino\.\ 2\.500\$\ Extra\.\ Worauf\ warten\ Sie\?\
X\-Mailer\:\ Microsoft\ Outlook\ Express\ 6\.00\.2720\.3000\
Body\:\
Er\&\#246\;ffnen\ Sie\ noch\ heute\ ein\ Konto\ bei\ Europas\ f\&\#252\;hrendem\ \=\
Onlinekasino\,\ und\ Sie\ erhalten\ einen\ unglaublichen\ Bonus\ im\ Wert\ von\ bis\ \=\
zu\ 2\.500\ \&\#8364\;\!\ Leisten\ Sie\ ganz\ einfach\ Einzahlungen\ auf\ Ihr\ Konto\,\ \=
und\ der\ Kasinokassierer\ schreibt\ sie\ Ihrem\ Konto\ zusammen\ mit\ einem\ \=\
Spielbonus\ in\ H\&\#246\;he\ von\ 100\%\ im\ Wert\ von\ bis\ zu\ 500\ \&\#8364\;\ f\&\#25
jede\ der\ ersten\ f\&\#252\;nf\ Einzahlungen\ gut\.\
\
Euro\ Dice\ Casino\ ist\ genauso\ sicher\ wie\ ein\ landbasiertes\ Kasino\,\ \=\
verf\&\#252\;gt\ aber\ \&\#252\;ber\ eine\ gr\&\#246\;\&\#223\;ere\ Auswahl\ an\ \=\
realistischen\ und\ aufregenden\ Spielen\ und\ ist\ daher\ die\ erste\ Wahl\ unter\ \=\
den\ Online\-Spielern\.\ Sprechen\ Sie\ noch\ heute\ mit\ anderen\ Spielern\ und\ \=\
\&\#252\;berzeugen\ Sie\ sich\ selbst\!\
\
Verpassen\ Sie\ nicht\ Ihre\ Chance\;\ klicken\ Sie\ hier\,\ um\ zu\ Euro\ Dice\ \=\
Casino\ zu\ gelangen\!\
eurocasino([A-Za-z]){2,2}\.comNext\ Body\ Part\:\
```

Euro Dice Casino 4

- Adding a single slightly different email resulted in a detection rate of 26% (previously 5.3%)
- We now match emails of this campaign ranging from September to November 2008
- All that changed is the URL
  - eurocasinokg.com
  - eurocasino([A-Za-z]){2,2}.com

Euro Dice Casino 5

- Adding another email:

```
Subject\:\ Euro\ Dice\ Casino\.\ 2\.500\$\ Extra\.\ Worauf\ warten\ Sie\?\
X\-Mailer\:\ Microsoft\ Outlook\ Express\ 6\.00\.2720\.3000\
Body\:\
Er\&\#246\;ffnen\ Sie\ noch\ heute\ ein\ Konto\ bei\ Europas\ f\&\#252\;hrend
Onlinekasino\,\ und\ Sie\ erhalten\ einen\ unglaublichen\ Bonus\ im\ Wert\ vo
zu\ 2\.500\ \&\#8364\;\!\ Leisten\ Sie\ ganz\ einfach\ Einzahlungen\ auf\ Ihr
und\ der\ Kasinokassierer\ schreibt\ sie\ Ihrem\ Konto\ zusammen\ mit\ einem\
Spielbonus\ in\ H\&\#246\;he\ von\ 100\%\ im\ Wert\ von\ bis\ zu\ 500\ \&\#83
jede\ der\ ersten\ f\&\#252\;nf\ Einzahlungen\ gut\.\
\
Euro\ Dice\ Casino\ ist\ genauso\ sicher\ wie\ ein\ landbasiertes\ Kasino\,\
verf\&\#252\;gt\ aber\ \&\#252\;ber\ eine\ gr\&\#246\;\&\#223\;ere\ Auswahl\
realistischen\ und\ aufregenden\ Spielen\ und\ ist\ daher\ die\ erste\ Wahl\
den\ Online\-Spielern\.\ Sprechen\ Sie\ noch\ heute\ mit\ anderen\ Spielern\
\&\#252\;berzeugen\ Sie\ sich\ selbst\!\
\
Verpassen\ Sie\ nicht\ Ihre\ Chance\;\ klicken\ Sie\ hier\,\ um\ zu\ Euro\ Di
Casino\ zu\ gelangen\!\
([\.A-Za-z]){0,16}Next\ Body\ Part\:\
```

Euro Dice Casino 6

- Adding another email raises the detection rate to 99%
- Again only the URL changes:
  - eurocasino([A-Za-z]){2,2}.com
  - ([\.A-Za-z]){0,16}
- The number of distinct emails of a campaign determines the quality of a template
- In this case a total of 3 emails suffices for a 99% detection rate of the email campaign

Summary

...and future work

Summary

- Sandnet (run bots periodically)
- Offline template generation
  - Common Substring Algorithm
- First results are promising

Future Work

- Rebuild the Sandnet to run bots endlessly
- Construct templates while collecting the SPAM from the running bots (realtime)
- Build a Mail-Client Plugin for template filtering
- Evaluate the approach

Jan Göbel
http://pi1.informatik.uni-mannheim.de/
goebel@informatik.uni-mannheim.de

Questions ?

Pi1 - Laboratory for Dependable Distributed Systems

UNIVERSITÄT MANNHEIM