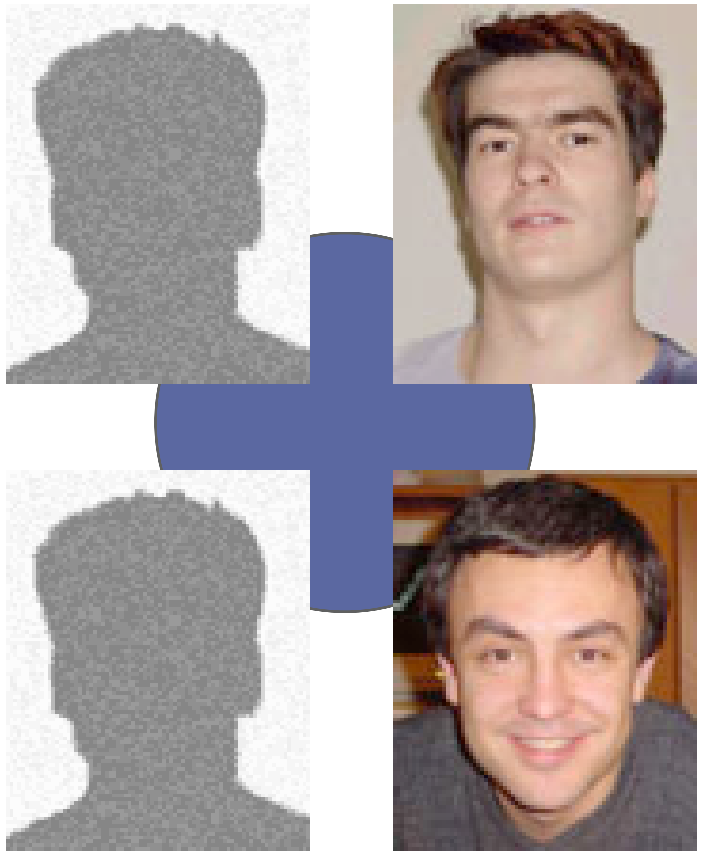
Genotype Spam Detection

Dmitry Samosseiko, SophosLabs

Virus Bulletin, Dublin, October 2005

Collaboration

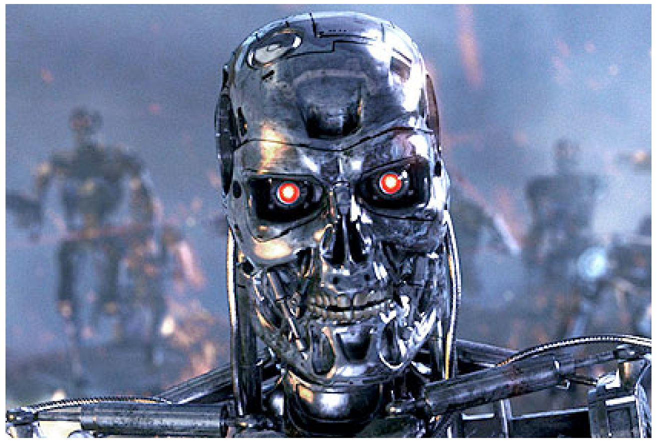
Malware Authors

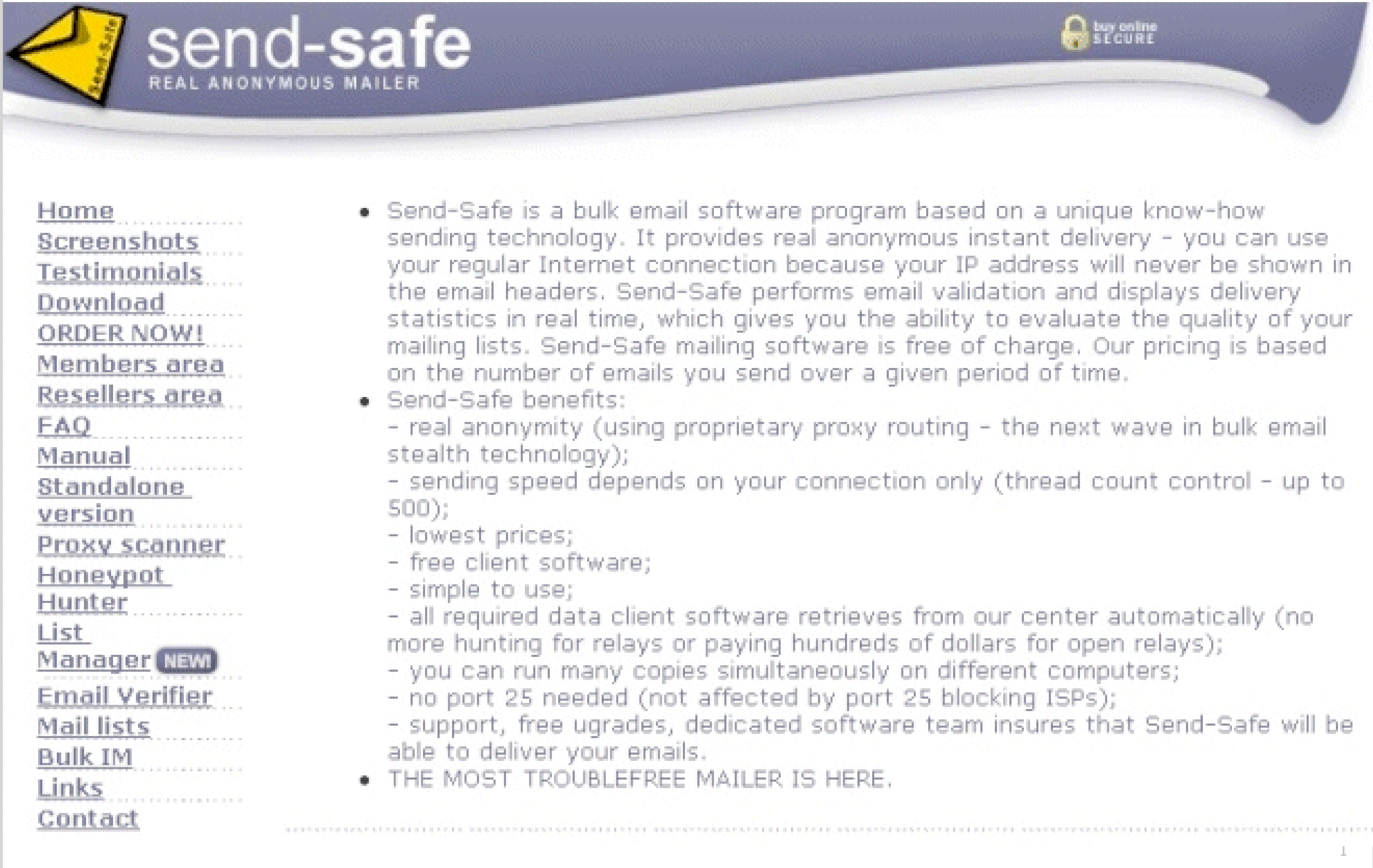


Spammers

pictures: courtesy of
www.spamhaus.org

Zombie / Bot
Trojans
Botnets





- Send-Safe is a bulk email software program based on a unique know-how sending technology. It provides real anonymous instant delivery - you can use your regular Internet connection because your IP address will never be shown in the email headers. Send-Safe performs email validation and displays delivery statistics in real time, which gives you the ability to evaluate the quality of your mailing lists. Send-Safe mailing software is free of charge. Our pricing is based on the number of emails you send over a given period of time.
- Send-Safe benefits:
  - real anonymity (using proprietary proxy routing - the next wave in bulk email stealth technology);
  - sending speed depends on your connection only (thread count control - up to 500);
  - lowest prices;
  - free client software;
  - simple to use;
  - all required data client software retrieves from our center automatically (no more hunting for relays or paying hundreds of dollars for open relays);
  - you can run many copies simultaneously on different computers;
  - no port 25 needed (not affected by port 25 blocking ISPs);
  - support, free ugrades, dedicated software team insures that Send-Safe will be able to deliver your emails.
- THE MOST TROUBLEFREE MAILER IS HERE.

One malware - One spam campaign

Sober-Q:

Ich bin immer noch kein Spammer! Aber sollte vielleicht einer werden...

a postcard?

You have just received a virtual greeting from a family member!

You can pick up your postcard at the following web address:

http://www.postcards1001.com/?a91-valets-cloud-187

If you can't click on the web address above, you can also
visit E-Greetings at http://www.postcards1001.com/
and enter your pickup code, which is: a91-valets-cloud-187

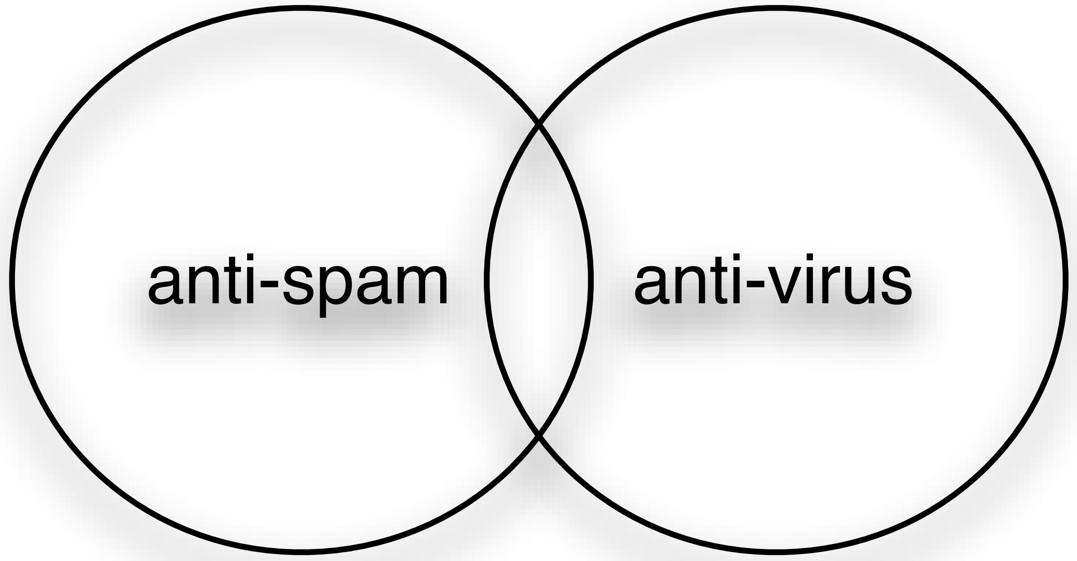
(Your postcard will be available for 60 days.)

Oh -- and if you'd like to reply with a postcard,
you can do so by visiting this web address:
http://www.postcards1001.com/
(Or you can simply click the "reply to this postcard"
button beneath your postcard!)

We hope you enjoy your postcard, and if you do,
please take a moment to send a few yourself!

Regards,
1001 Greetings and Postcards
http://www.postcards1001.com/

http://loveyoudear.home.ro/postcard.gif.exe



anti-spam

anti-virus

Computer Security Labs

- Anti-virus and anti-spam experts working together

THIS CABINET MUST BE DECONTAMINATED
SERVICING FILTER AREAS, THE FAN/MOTOR
DO NOT USE TOXIC, EXPLOSIVE, OR FLAMMABL

the techniques commonly used to detect spam and malware are ... **different**

Anti-spam vs. anti-virus

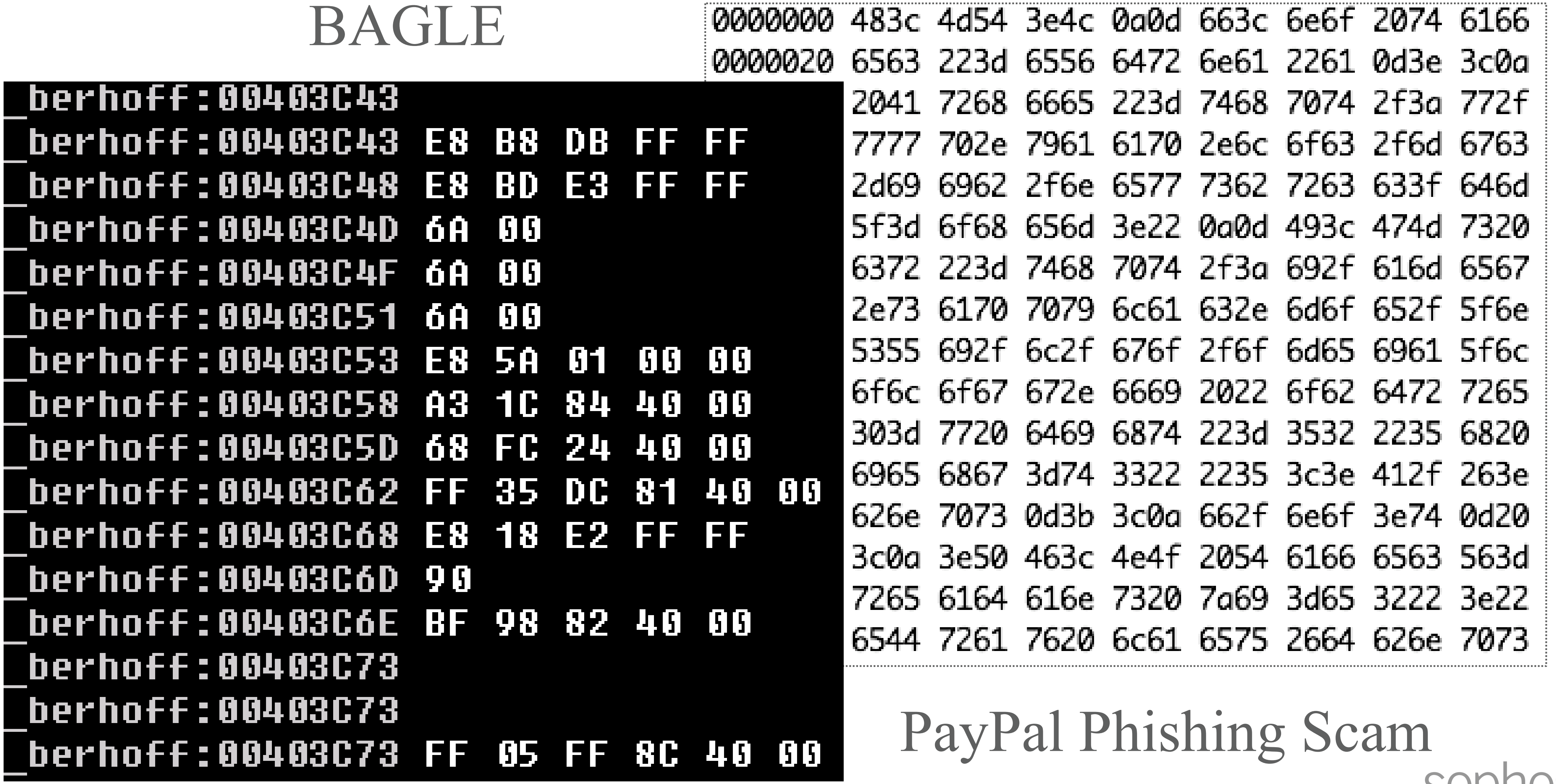
- Anti-Virus
  - Virus signatures / Definitions
  - Generic detection / Heuristics
  - Behavior-based detection
- Anti-Spam
  - IP blacklists, sender reputation
  - URI / phone / IM blacklists
  - Heuristic rules
  - Checksums
  - “Bayesian”
  - ...

WHY?

SPAM != VIRUS

Spam vs. Virus

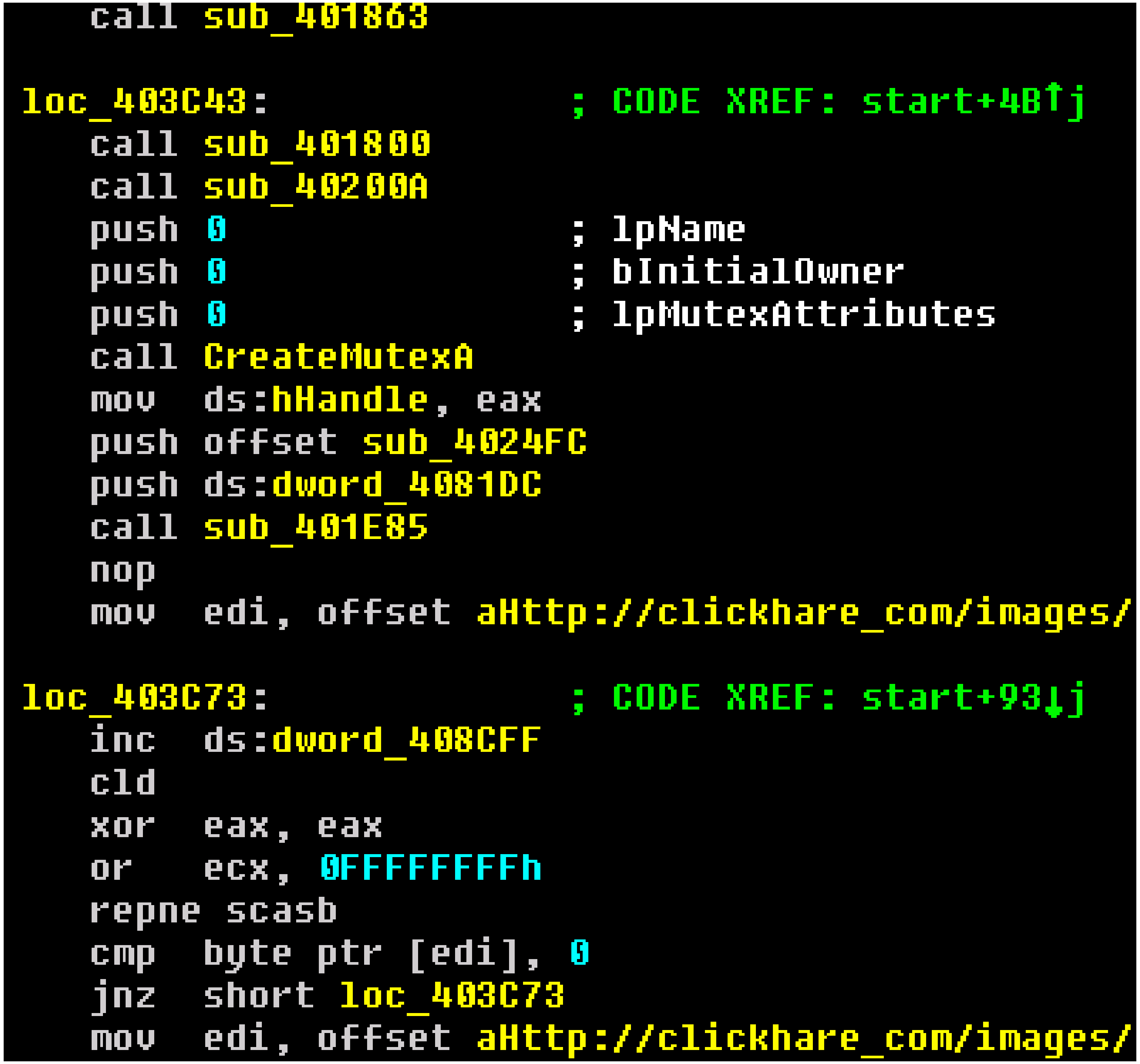
BAGLE



```
_berhoff:00403C43
_berhoff:00403C43 E8 B8 DB FF FF
_berhoff:00403C48 E8 BD E3 FF FF
_berhoff:00403C4D 6A 00
_berhoff:00403C4F 6A 00
_berhoff:00403C51 6A 00
_berhoff:00403C53 E8 5A 01 00 00
_berhoff:00403C58 A3 1C 84 40 00
_berhoff:00403C5D 68 FC 24 40 00
_berhoff:00403C62 FF 35 DC 81 40 00
_berhoff:00403C68 E8 18 E2 FF FF
_berhoff:00403C6D 90
_berhoff:00403C6E BF 98 82 40 00
_berhoff:00403C73
_berhoff:00403C73
_berhoff:00403C73 FF 05 FF 8C 40 00
```

```
0000000 483c 4d54 3e4c 0a0d 663c 6e6f 2074 6166
0000020 6563 223d 6556 6472 6e61 2261 0d3e 3c0a
        2041 7268 6665 223d 7468 7074 2f3a 772f
        7777 702e 7961 6170 2e6c 6f63 2f6d 6763
        2d69 6962 2f6e 6577 7362 7263 633f 646d
        5f3d 6f68 656d 3e22 0a0d 493c 474d 7320
        6372 223d 7468 7074 2f3a 692f 616d 6567
        2e73 6170 7079 6c61 632e 6d6f 652f 5f6e
        5355 692f 6c2f 676f 2f6f 6d65 6961 5f6c
        6f6c 6f67 672e 6669 2022 6f62 6472 7265
        303d 7720 6469 6874 223d 3532 2235 6820
        6965 6867 3d74 3322 2235 3c3e 412f 263e
        626e 7073 0d3b 3c0a 662f 6e6f 3e74 0d20
        3c0a 3e50 463c 4e4f 2054 6166 6563 563d
        7265 6164 616e 7320 7a69 3d65 3222 3e22
        6544 7261 7620 6c61 6575 2664 626e 7073
```

PayPal Phishing Scam



```
   call sub_401863

loc_403C43:             ; CODE XREF: start+4B↑j
   call sub_401800
   call sub_40200A
   push 0               ; lpName
   push 0               ; bInitialOwner
   push 0               ; lpMutexAttributes
   call CreateMutexA
   mov  ds:hHandle, eax
   push offset sub_4024FC
   push ds:dword_4081DC
   call sub_401E85
   nop
   mov  edi, offset aHttp://clickhare_com/images/

loc_403C73:             ; CODE XREF: start+93↓j
   inc  ds:dword_408CFF
   cld
   xor  eax, eax
   or   ecx, 0FFFFFFFFh
   repne scasb
   cmp  byte ptr [edi], 0
   jnz  short loc_403C73
   mov  edi, offset aHttp://clickhare_com/images/
```

```
<HTML>
<font face="Verdana">
<A href="http://www.paypal.com/cgi-bin/webscr?cmd=_home">
<IMG src="http://images.paypal.com/en_US/i/logo/email_logo.gif" border=0
width="255" height="35"></A>&nbsp;
</font>
<P><FONT face=Verdana size="2">Dear valued&nbsp;<STRONG><SPAN
style="FONT-SIZE: 10pt; COLOR: black; FONT-FAMILY:
Verdana">PayPal<SUP></SUP></SPAN>&nbsp;</STRONG>member:&nbsp;<BR></FONT>
<font face="Verdana"><BR></font></P>
<P><FONT face=Verdana size=2><SPAN style="FONT-SIZE: 10pt; COLOR: black;
FONT-FAMILY: Verdana"><STRONG>PayPal<SUP></SUP></STRONG></SPAN></
FONT><tt><font face="Verdana">
is committed to maintaining a safe environment for its community of <br>
buyers and sellers. To protect the security of your account, PayPal
employs <br>some of the most advanced security systems in the world and
our anti-fraud <br>
teams regularly screen the PayPal system for unusual activity. <br>
<br>
```

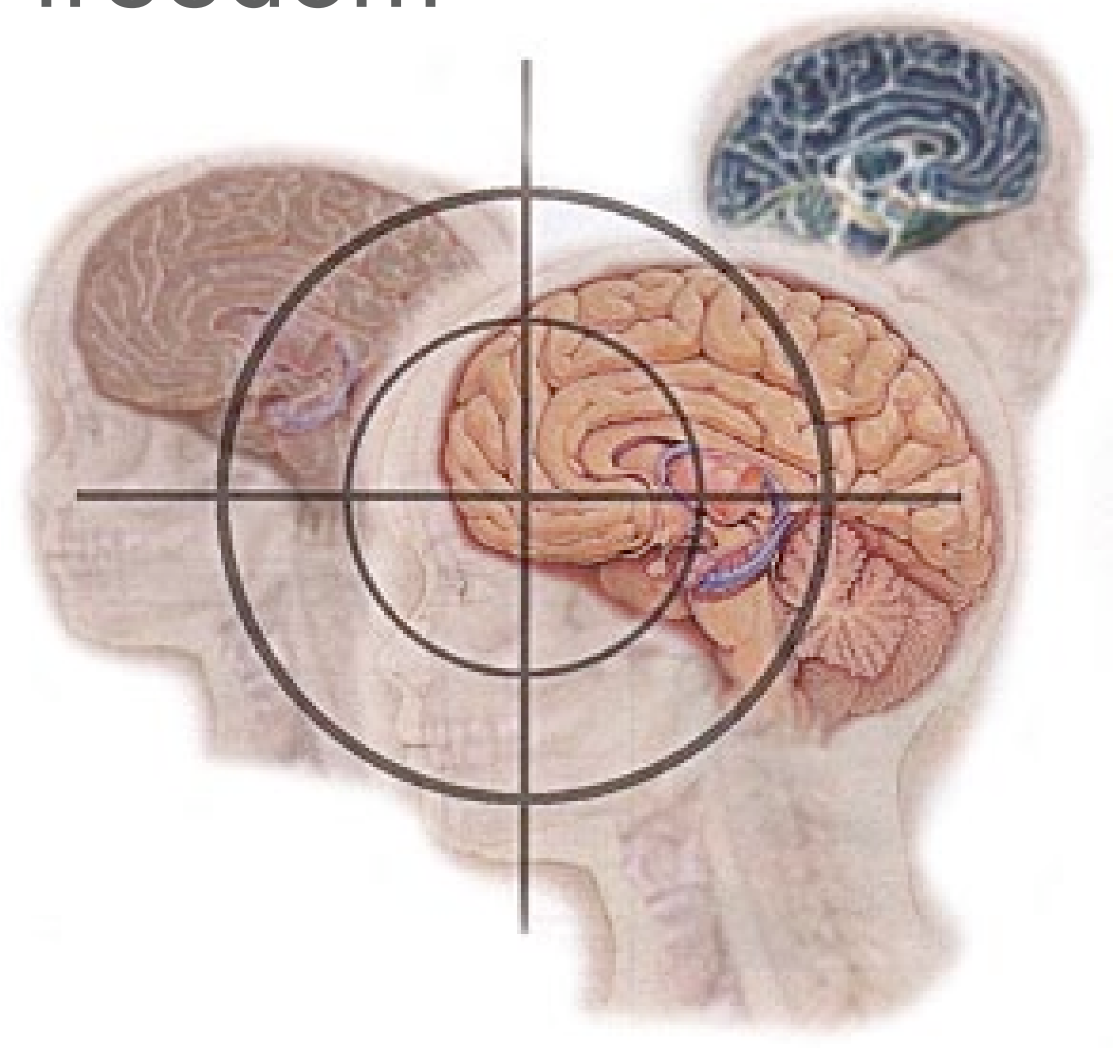
Malware

- Targets a computer platform (h/w, OS, application)
- Must be EXECUTABLE on the targeted platform

Spam & Phishing

- target humans
- infinite degree-of-freedom

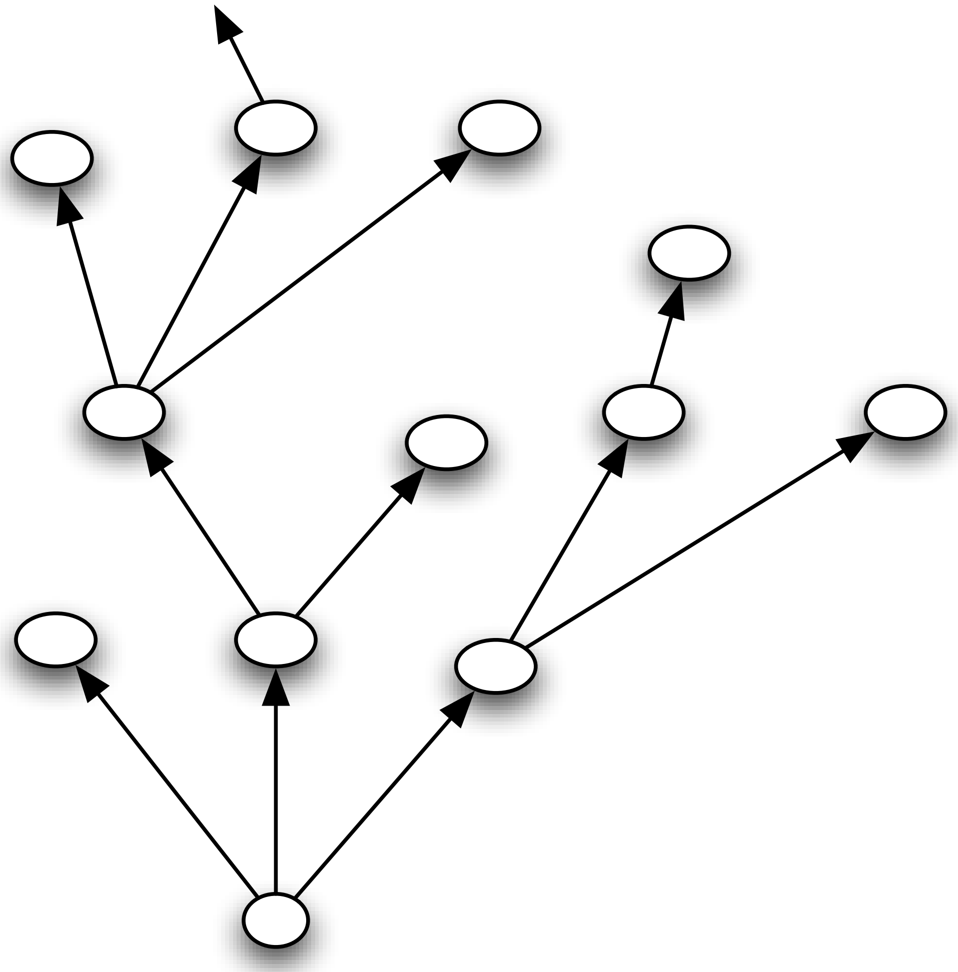
v1@grA
VIIAGARA
V<td>I<div>AGRA...



Volume

- Hundreds of new spam campaigns each day
- Dozens of new viruses or modifications

Spread: malware



Spread: spam

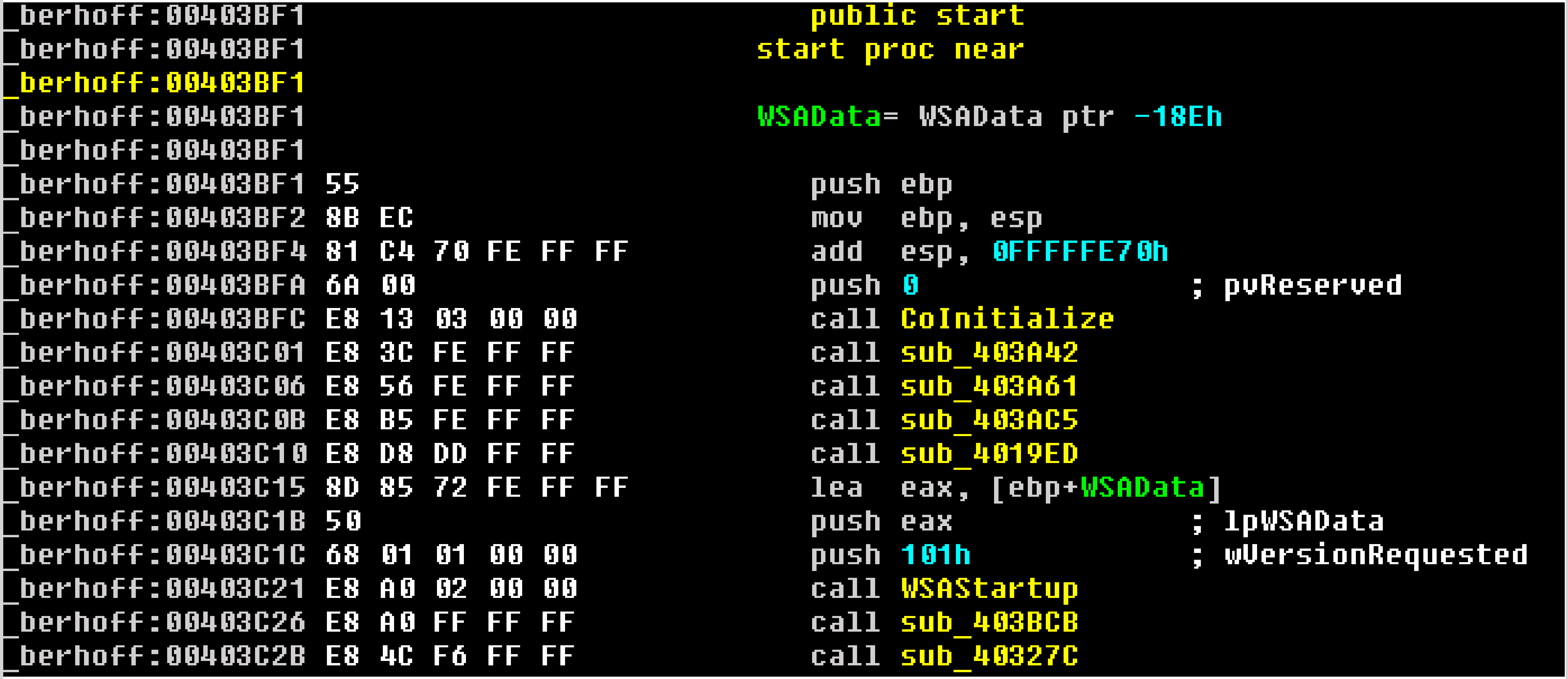
Malware =

Standard requirements

- AV
  - 100% catch rate for known threats
  - detect on gateways and desktops
- AS
  - 98-99+%
  - stop at the gateway

Virus Analysts

THIS CABINET MUST BE DECONTAMINATED
SERVICING FILTER AREAS, THE FAN/MOTOR
DO NOT USE TOXIC, EXPLOSIVE, OR FLAMMABL
SterilGARD II

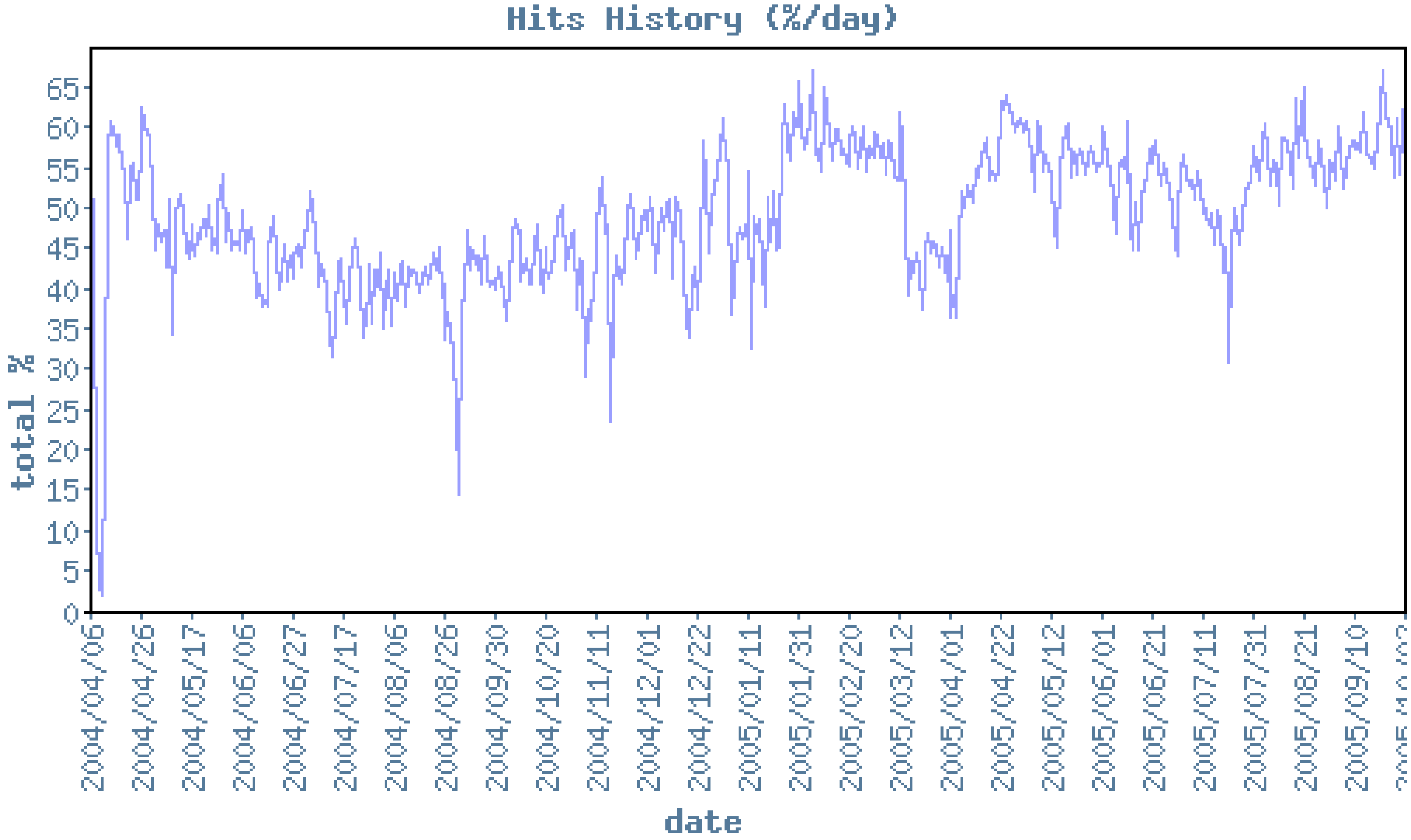


```
_berhoff:00403BF1                            public start
_berhoff:00403BF1                         start proc near
_berhoff:00403BF1
_berhoff:00403BF1                         WSAData= WSAData ptr -18Eh
_berhoff:00403BF1
_berhoff:00403BF1 55                         push ebp
_berhoff:00403BF2 8B EC                      mov  ebp, esp
_berhoff:00403BF4 81 C4 70 FE FF FF          add  esp, 0FFFFFE70h
_berhoff:00403BFA 6A 00                      push 0               ; pvReserved
_berhoff:00403BFC E8 13 03 00 00             call CoInitialize
_berhoff:00403C01 E8 3C FE FF FF             call sub_403A42
_berhoff:00403C06 E8 56 FE FF FF             call sub_403A61
_berhoff:00403C0B E8 B5 FE FF FF             call sub_403AC5
_berhoff:00403C10 E8 D8 DD FF FF             call sub_4019ED
_berhoff:00403C15 8D 85 72 FE FF FF          lea  eax, [ebp+WSAData]
_berhoff:00403C1B 50                         push eax             ; lpWSAData
_berhoff:00403C1C 68 01 01 00 00             push 101h            ; wVersionRequested
_berhoff:00403C21 E8 A0 02 00 00             call WSAStartup
_berhoff:00403C26 E8 A0 FF FF FF             call sub_403BCB
_berhoff:00403C2B E8 4C F6 FF FF             call sub_40327C
```

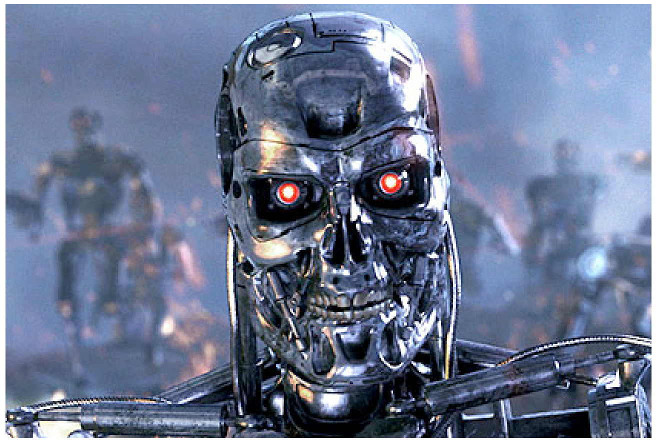
Volume, speed, “degree-of-freedom”

Sender IP blacklists

- SBL, XBL, CBL, NJABL
- > 60% effective with a very low FP rate

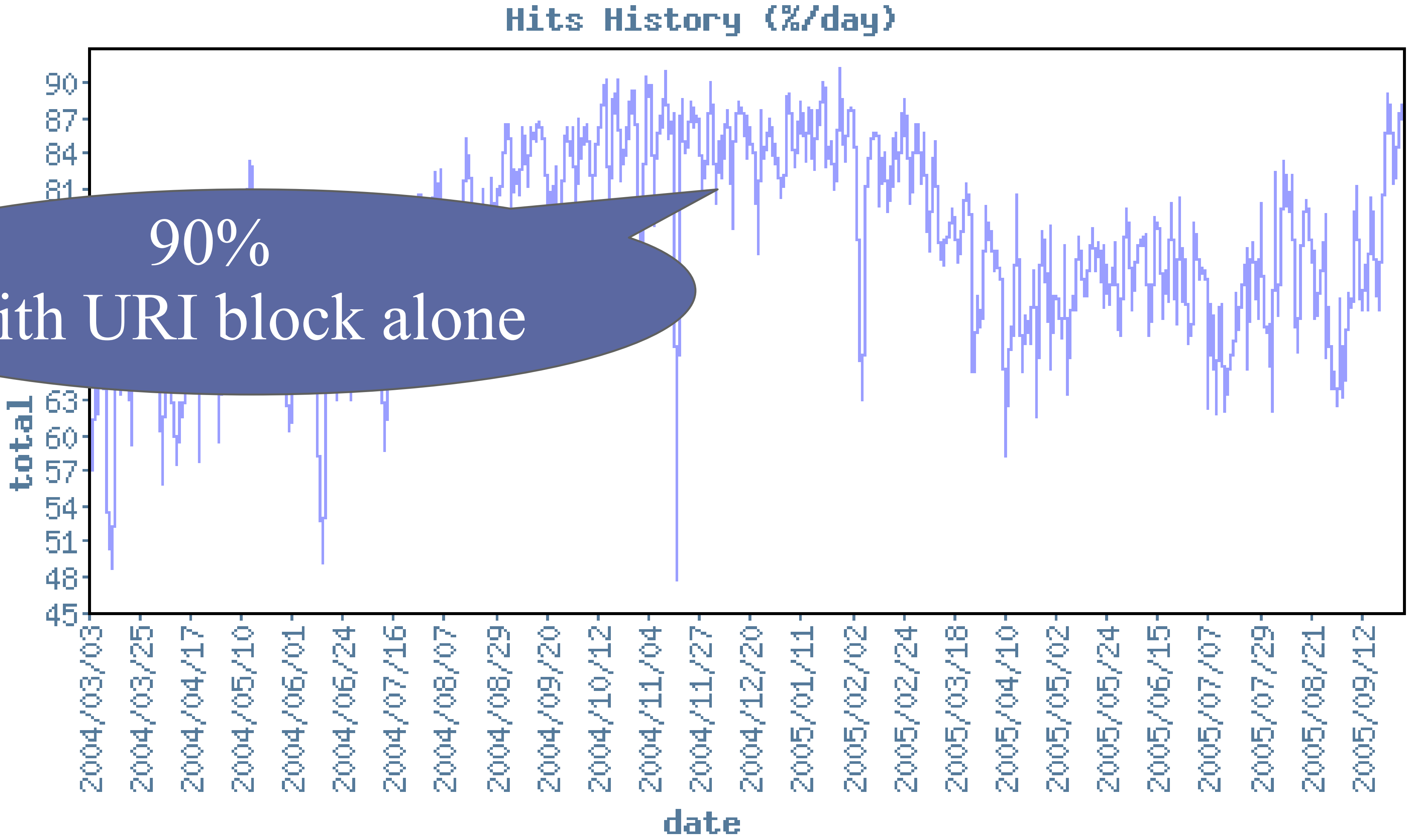


Zombie/Bot
Trojans
Botnets



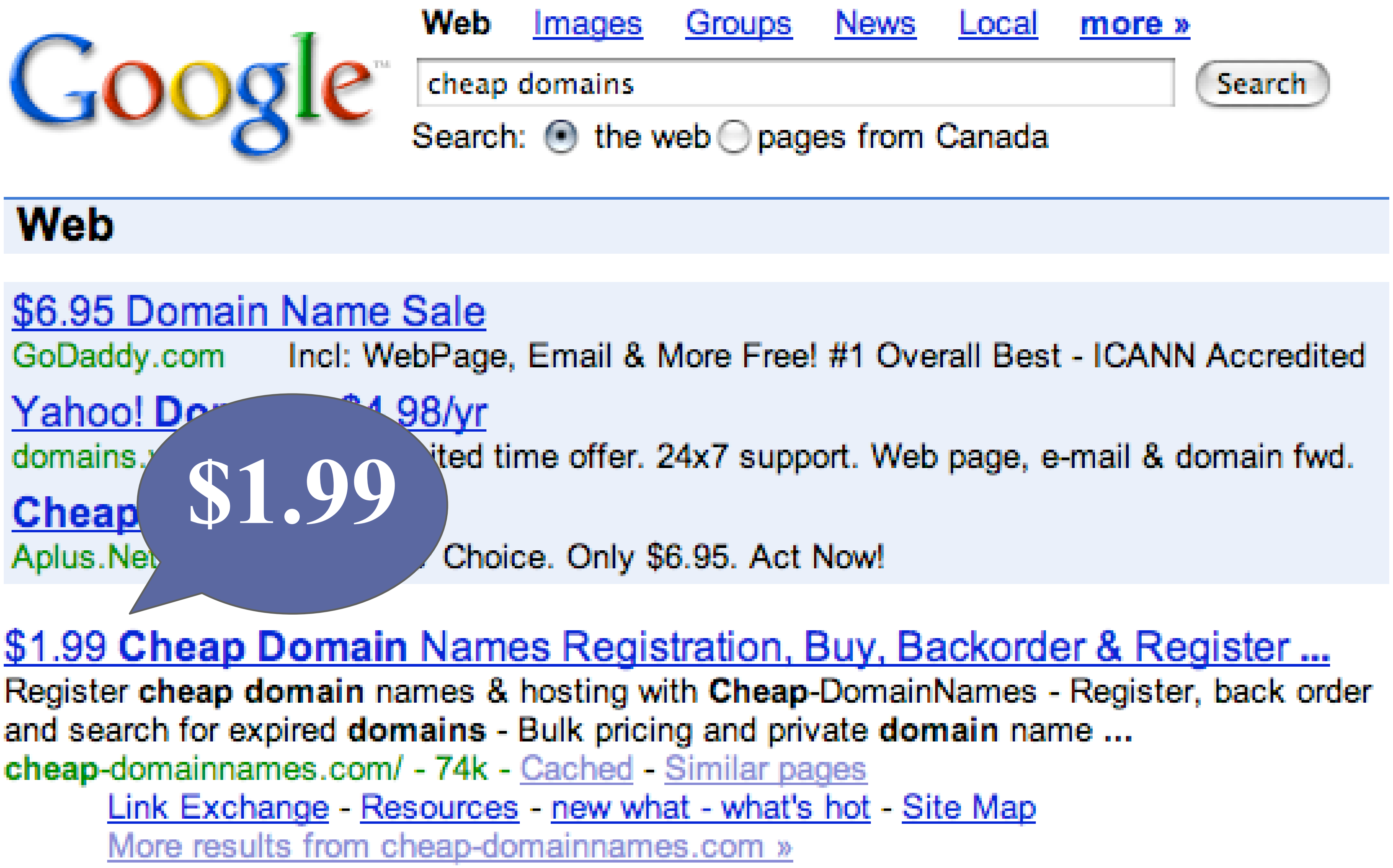
Call-to-action blocks

- domains, URIs, mailto:, phone, IM



90%
with URI block alone

Domains are dirt cheap



Web Images Groups News Local more »

cheap domains

Search

Search: ● the web ○ pages from Canada

Web

$6.95 Domain Name Sale

GoDaddy.com Incl: WebPage, Email & More Free! #1 Overall Best - ICANN Accredited

Yahoo! Do $1.98/yr

domains. ited time offer. 24x7 support. Web page, e-mail & domain fwd.

Cheap

Aplus.Ne Choice. Only $6.95. Act Now!

$1.99

$1.99 Cheap Domain Names Registration, Buy, Backorder & Register ...

Register cheap domain names & hosting with Cheap-DomainNames - Register, back order and search for expired domains - Bulk pricing and private domain name ...

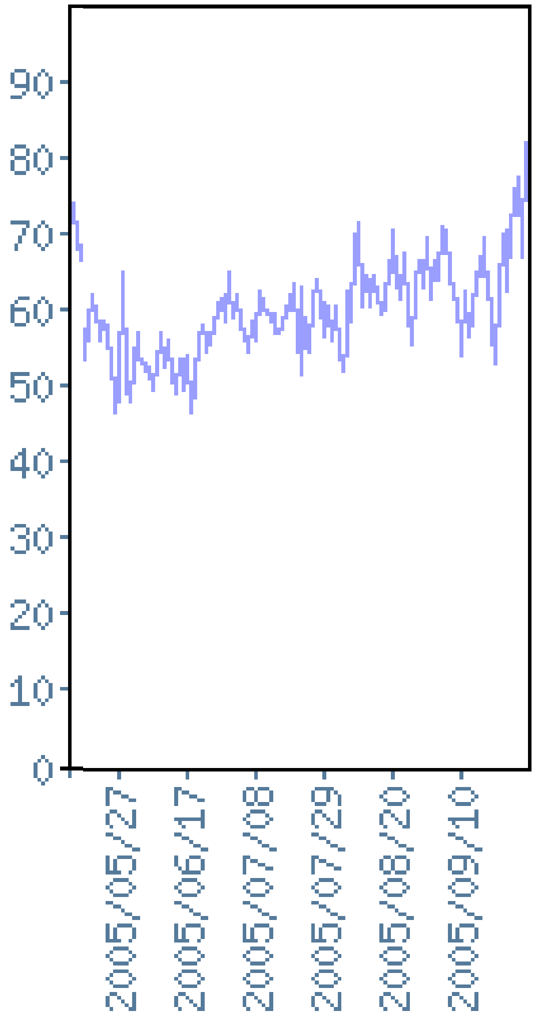
cheap-domainnames.com/ - 74k - Cached - Similar pages

Link Exchange - Resources - new what - what's hot - Site Map

More results from cheap-domainnames.com »

Content Checksums

- DCC, Razor, ..., commercial implementations



e.g. DCC catches up to 70-80%

Content randomization

- Infinite degree of freedom
- “Hashbusters” / “word salad”
- Letter transpositions
- Image unifications
- etc.

Heuristic

- Header forgeries, common phrases, styles
- Address spam problem in general
- Hard to maintain and use as a reactive tool

Spam Genotype concept

- Apply virus signature concept to spam campaigns
- Analyze similarities and unique features
- Write a template that combines this features and uniquely identifies all messages within this campaign with no false positives

Genotype match samples

- ? Body size between 514 and 546 bytes.
- ? Has a text/plain part followed by an HTML part encoded with base64.
- ? The sender’s name contains only lowercased characters
- ? The URL found ends with an .aspx string followed by a question mark and 5 to 7 digits.
- ? There is an image of a certain size embedded in the HTML table.
- ? There is a particular string sequence found in the second paragraph.

Objective

- Separate changeable spam campaign attributes from the static ones
- Create a “blueprint” or a “genetic constitution” of a campaign
- Use it in the cases, where traditional techniques do not provide a desired level of protection

Rationale

- Spam campaigns are repetitive. Some run for months
- Spammers use tools and templates. Tools have bugs
- Spammers make mistakes
- Spammers repeat successful tactics

Deterministic

if messages matches a spam genotype,
it’s classified as spam

Scope

- all message headers
- decoded/encoded content
- raw html
- “invisible” content
- text
- a paragraph
- a line
- a MIME part
- a URL

/^Regular\s{1,3}Expressions$/

the special regex library is used to minimize the risk of a bad expression

Life Cycle

from hours to days to months and maybe years

24/7 research and analysis

FUSSP?

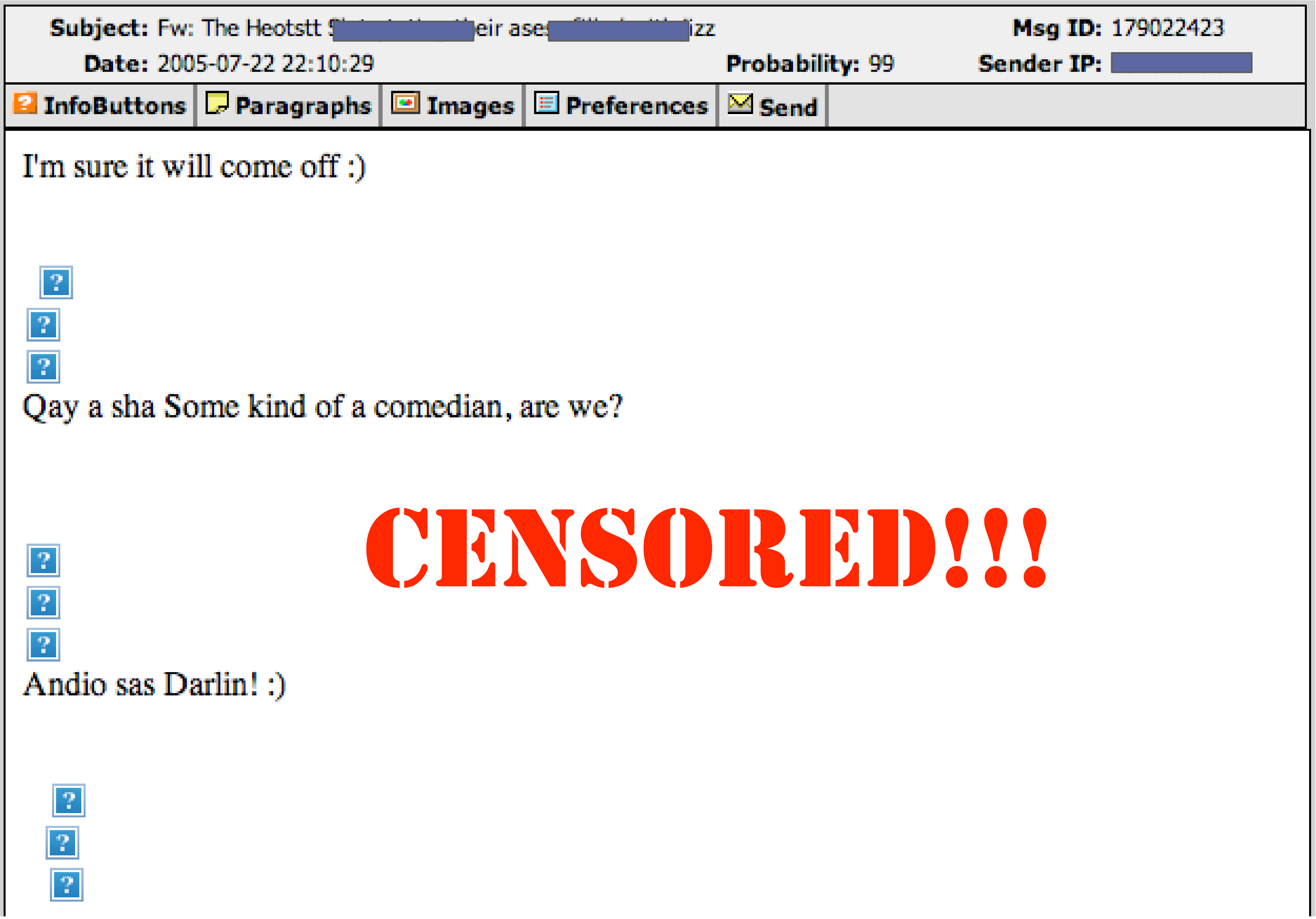
No:

- takes relatively long **time** to build
- is **expensive** to build

98-99% is not good enough

Sexually explicit content **is not tolerable** in a corporate environment...

Employee protection issues.



Subject: Fw: The Heotstt [redacted]eir ase[redacted]izz **Msg ID:** 179022423

Date: 2005-07-22 22:10:29 **Probability:** 99 **Sender IP:** [redacted]

I'm sure it will come off :)

?
?
?

Qay a sha Some kind of a comedian, are we?

CENSORED!!!

?
?
?

Andio sas Darlin! :)

?
?
?

99.999[9]% is not good enough either

Anything you can do to block this without introducing a FP risk or delays in message delivery **is worth doing**...

August 2004

```
 From: "Booker O. Belaboring" <poems@e-....net>
Subject: FW: Fetish Slaves Hard Inrestiosn
MIME-Version: 1.0
Content-Type: multipart/alternative;
    boundary="----=_NextPart_000_0031_7CD5220F.669EB952"
X-Priority: 3
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 6.00.2800.1437
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2479.0006

------=_NextPart_000_0031_7CD5220F.669EB952
Content-Type: text/plain
Content-Transfer-Encoding: 7bit

hi every one salams

Expansion means complexity and complexity decay.
[ ... ]

http://seiwara.net/gWKNoD7PmjDhlGWLJOfldics1/Ag8HJxweFCIHARo6GAIUFgcWC3ONCAo=.htm

Philosophers, for the most part, are constitutionally timid, and dislike the unexpected. Few of them
would be genuinely happy as pirates or burglars. Accordingly, they invent systems which make the
[ ... ]

http://seiwara.net/gWKNoD7PmjDhlGWLJOfldics1/Ag8HJxweFCIHARo6GAIUFgcWC3ONCAo=.html

------=_NextPart_000_0031_7CD5220F.669EB952
Content-Type: text/html
Content-Transfer-Encoding: quoted-printable

<META HTTP-EQUIV=3d"Content-Type" CONTENT=3d"text/html;charset=3diso-8859=
-1">
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4=2e0 Transitional//EN">
<HTML><HEAD>
<META HTTP-EQUIV=3d"Content-Type" CONTENT=3d"text/html; charset=3dus-asci=
i">
<META content=3d"MSHTML 6=2e00=2e2800=2e1437" name=3dGENERATOR>
<STYLE></STYLE>
</HEAD>
<BODY bgColor=3d#ffffff>
<DIV>
heeey!
The absent are never without fault=2e Nor the present without excuse=2e <
```

Anti-anti-spam techniques:

- sent though “zombies”
- “URI/domain rotation” -- new domains every 2 hours
- random textual content
- random Subject lines

From: John M. Doe <...@...>

```
From: "Percentiles U. Rewards" <iniquity@bru....de>
From: "Potboiler V. Bust" <Angelou@ca...com>
From: "Replete S. Watchband" <redraws@ads...net>
...
From: "Railleries G. Preserve" <Wade@cir....net>
```

URL

http://<host>/<hash>/<hash>.htm

http://tinchentm.net/WTyz4XGzXE/Ag8HJxweFCIHARo6GAIUFgcWC30NCAo=.htm

http://seiwara.net/LNoD7PmjEhlGYLJ1flfics1/Aw8POgInBgYCEAsgHUkEDQs=.htm

http://sumianayv.net/LNoD7PmjEhlGYLJ1flfics1/Aw8POgInBgYCEAsgHUkEDQs=.htm

email address

Forged headers

```
X-Mailer: Microsoft Outlook Express 6.00.2800.1437

X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2505.0000
```

“Bug”

Legitimate Outlook email:

```
        <META HTTP-EQUIV=3D" ...
```

Spam email:

```
        <META HTTP-EQUIV=3d" ...
```

charset = ?

```
Content-Type: multipart/alternative;
          boundary="----=_NextPart_000_0031_7CD..6692"
```

Is probably a work of a certain “spamware” and/or spam gang...

“Five Elements, Inc” ?

Writing a definition

Size ranges
min / max

Message part: X-MimeOLE header

Regular Expression:

```
/Produced By Microsoft MimeOLE V6\.00\.2\d{3}\.\d{4}/
```

Message part: X-Mailer

Regular Expression:

```
/Microsoft Outlook Express 6\.00\.2\d{3}\.\d{4}/
```

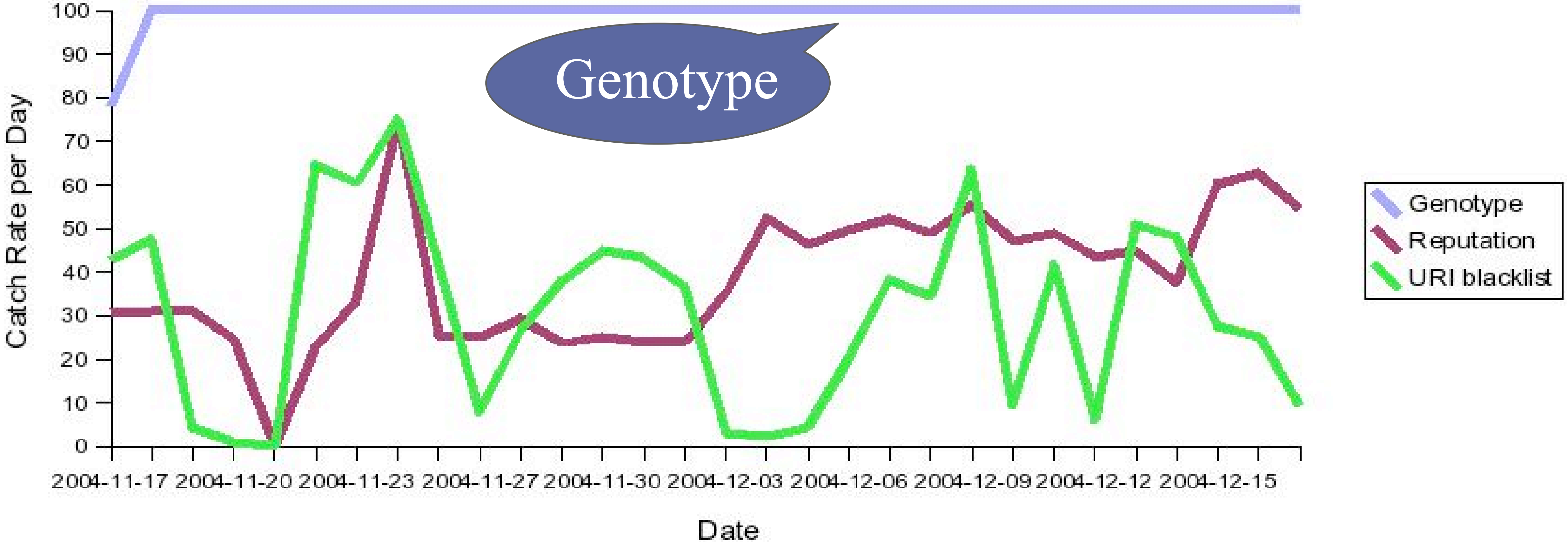
Message part: URIs
Regular Expression:

```
/[a-z]{5,20}\.[a-z]{2,4}\/[a-zA-Z0-9]{2,40}\/(?:.....
```

MIME structure

Review
Test
Ship

Genotype vs. URI and IP blacklists



Genotype's age: 11 months.

Number of false positive reports: 0

Catch rate: 100% after the definition was released

Limited use

Not every spam campaign **deserves** a definition

Very **few** spam campaigns require a genotype

Needs special **skills** and **expertise**

Reaction Time

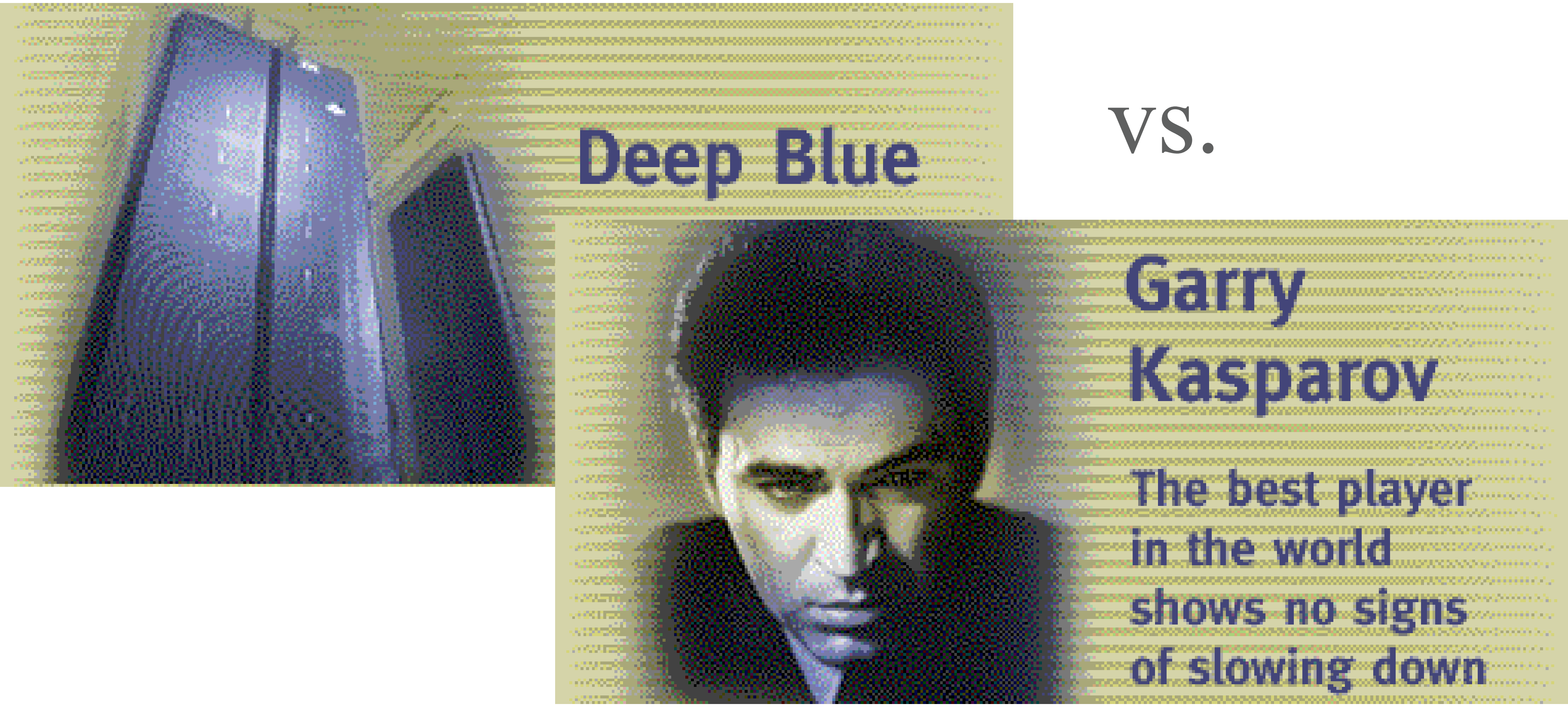
- Might take from 20 min to a few hours to build a genotype
- Takes minutes to deliver

Genotype vs. Heuristic rules

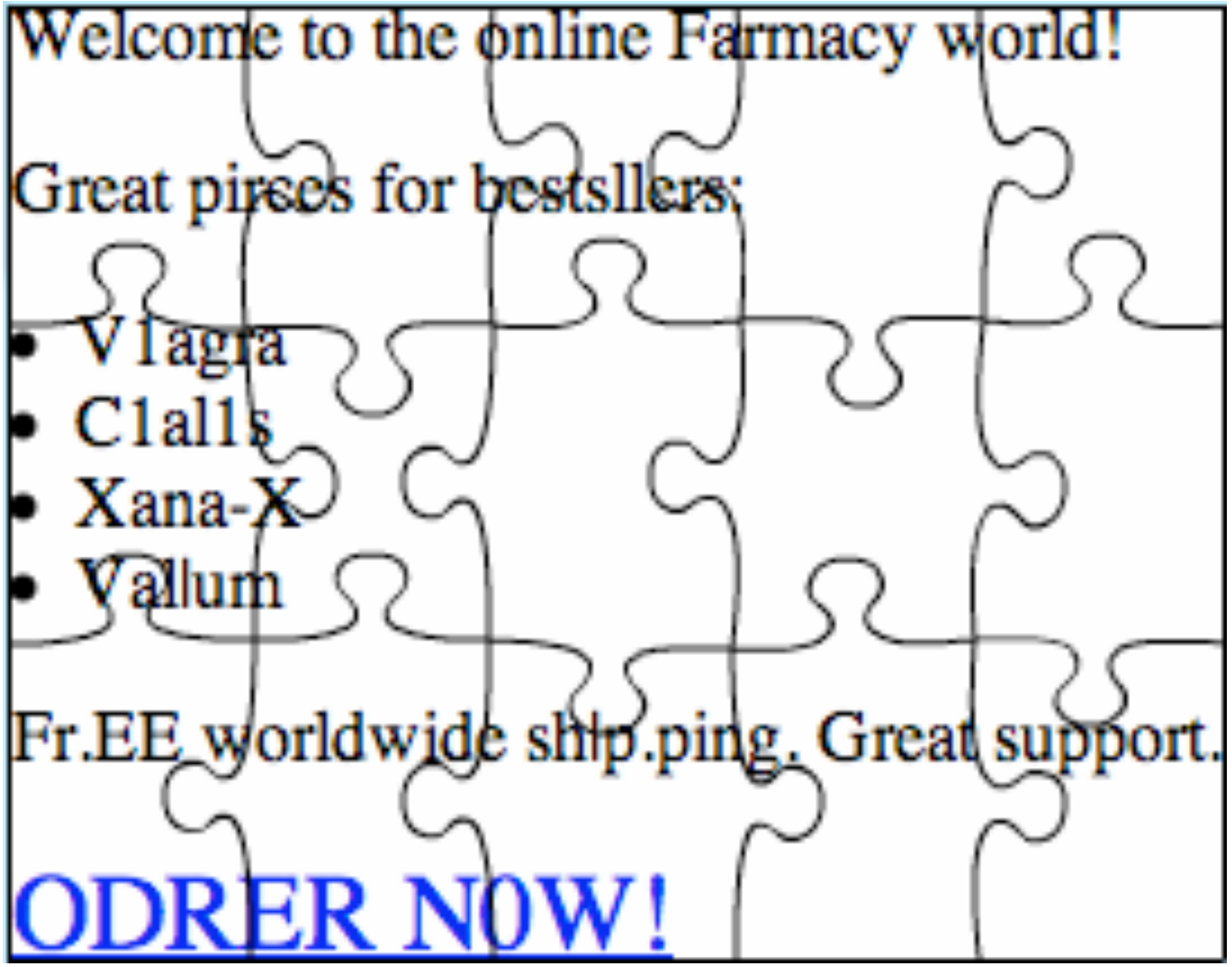
- Deterministic. No guess work
- Targets a specific campaign vs. spam problem in general

Genotype vs. Automated signatures

vs.



Spam Analysts enjoy it!



Welcome to the online Farmacy world!

Great pirces for bestsllers:

- V1agra
- C1al1s
- Xana-X
- Vallum

Fr.EE worldwide shlp.ping. Great support.

ODRER N0W!

The technology is simple. The key is in the **execution**.

Dmitry Samosseiko
dmitry.samosseiko@sophos.com