Spam and Phishing

Ramin Sadre

Spam

- https://www.youtube.com/watch?v=_bW4vEo1F4E

- Spam = Unsolicited messages, sent in bulk to many recipients
- Typically, the sender address is forged to prevent counter measures
- Works with all messaging services where it is difficult for the recipient to filter such messages (mail, chat, SMS,...)

Why spamming?

1. Advertising products and good
   - Illegal products (weapons, drugs,…)
   - Fake products, stolen products,…

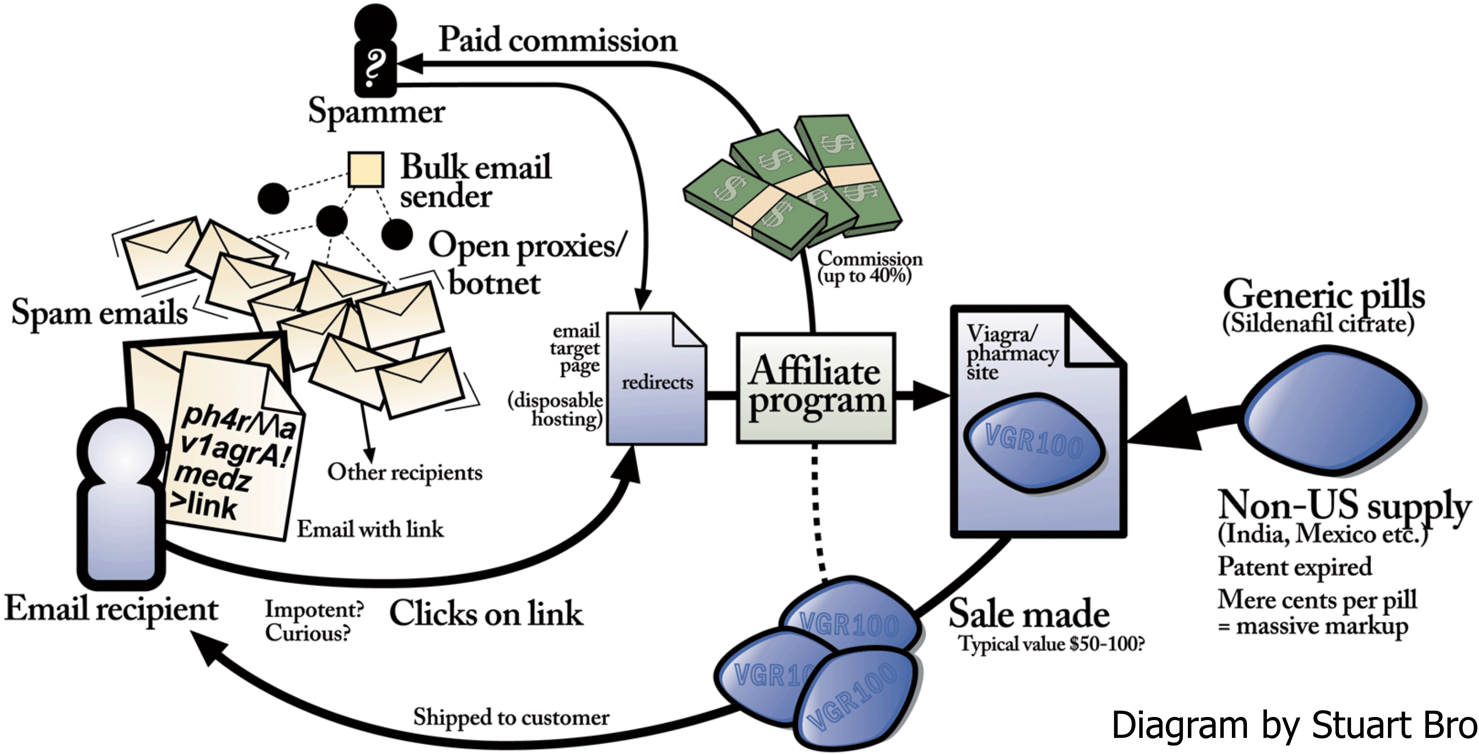


Diagram by Stuart Brown
http://modernlifeisrubbish.co.uk

Why spamming? (2)

2. Fraud schemes
   - Pen-pal relationships
   - Nigerian prince
   - “Earn money at home”
3. Recruiting for illegal activities
   - Money mules: Allow others to use your bank account for fraudulent financial transfers
   - Re-shippers: Reship stolen/illegal goods
4. Infection (recruiting machines for botnet)
   - Lure user to malicious web sites and infect user’s computer by exploiting vulnerabilities in the browser or in plugins
   - Convince users to download malicious software

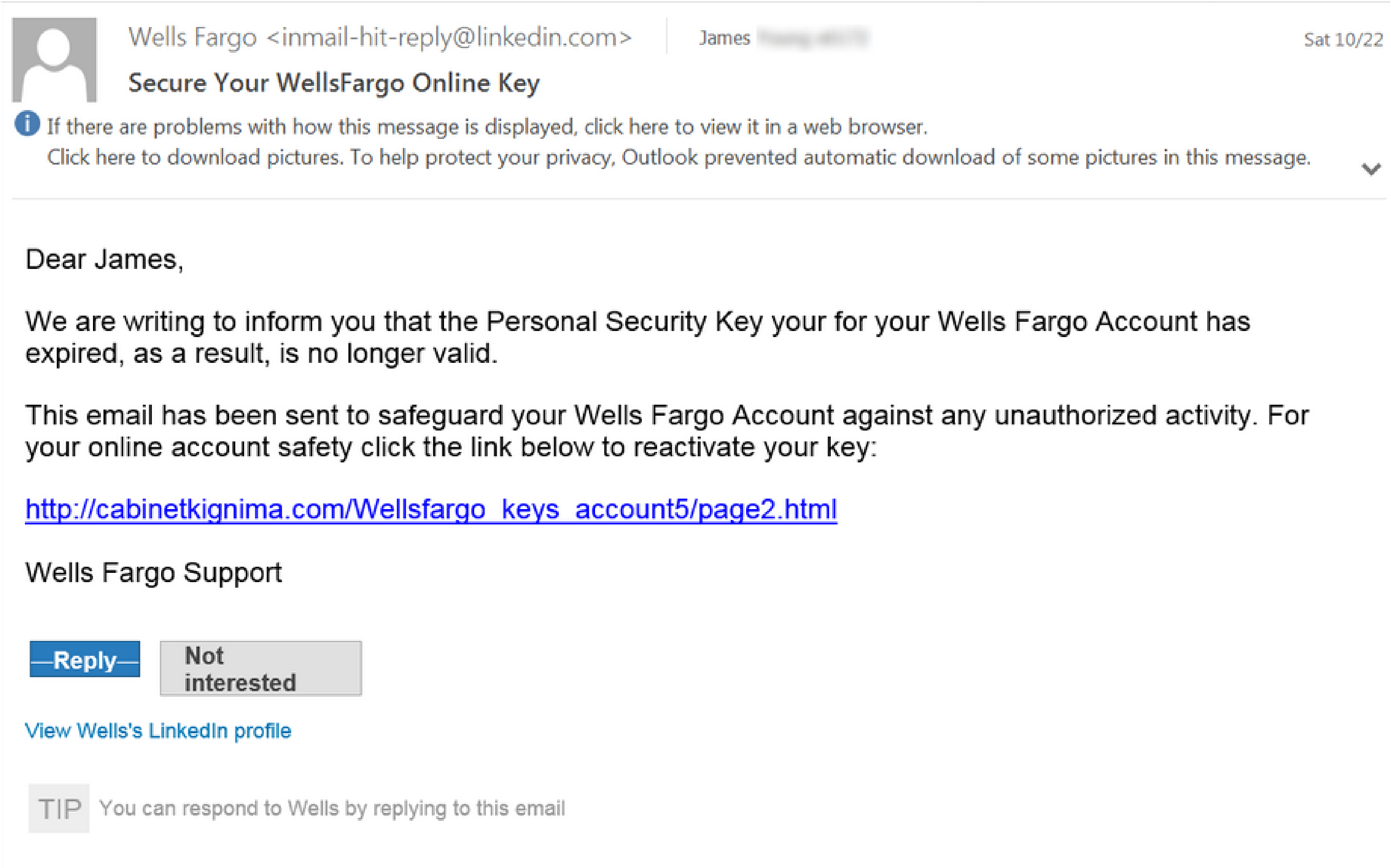
     “Microsoft has an important patch for you. Click here.”

Why spamming? (3)

4. Phishing
   - Lure users to a spoofed websites and convince them to enter passwords, credit card numbers,…

- Spear-phishing = Variant of phishing that does not rely on (mass) phishing
  - Mails tailored to recipient
  - Works well with the information you got from hacking databases and PCs (names, client numbers,...)

Phishing example (from www.phishing.org)



Wells Fargo <inmail-hit-reply@linkedin.com> James [redacted] Sat 10/22

Secure Your WellsFargo Online Key

If there are problems with how this message is displayed, click here to view it in a web browser.
Click here to download pictures. To help protect your privacy, Outlook prevented automatic download of some pictures in this message.

Dear James,

We are writing to inform you that the Personal Security Key your for your Wells Fargo Account has expired, as a result, is no longer valid.

This email has been sent to safeguard your Wells Fargo Account against any unauthorized activity. For your online account safety click the link below to reactivate your key:

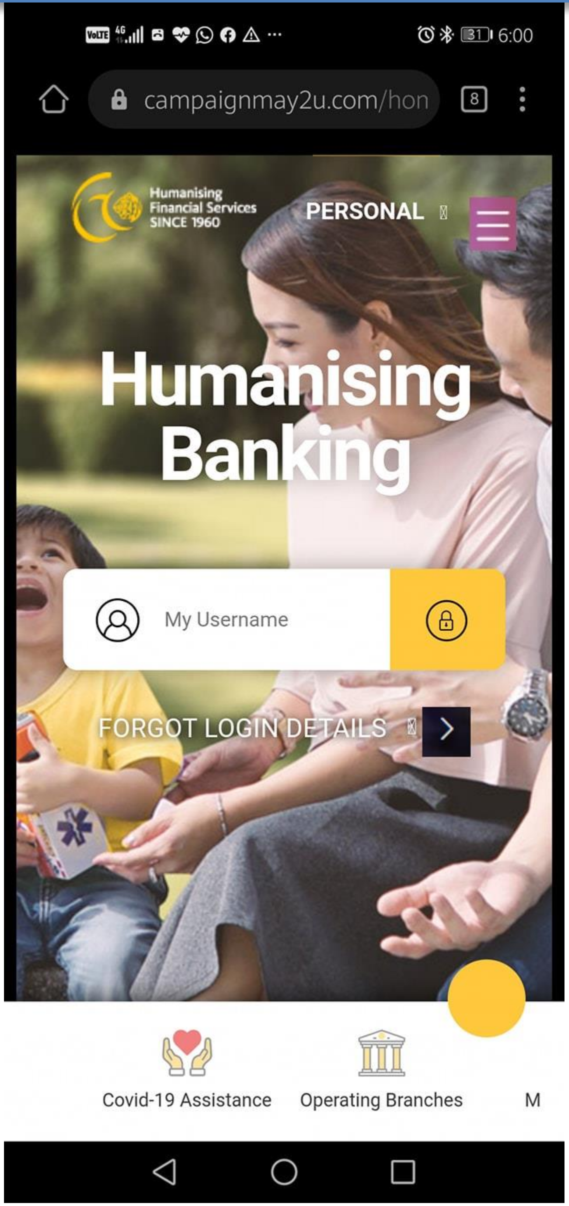
http://cabinetkignima.com/Wellsfargo_keys_account5/page2.html

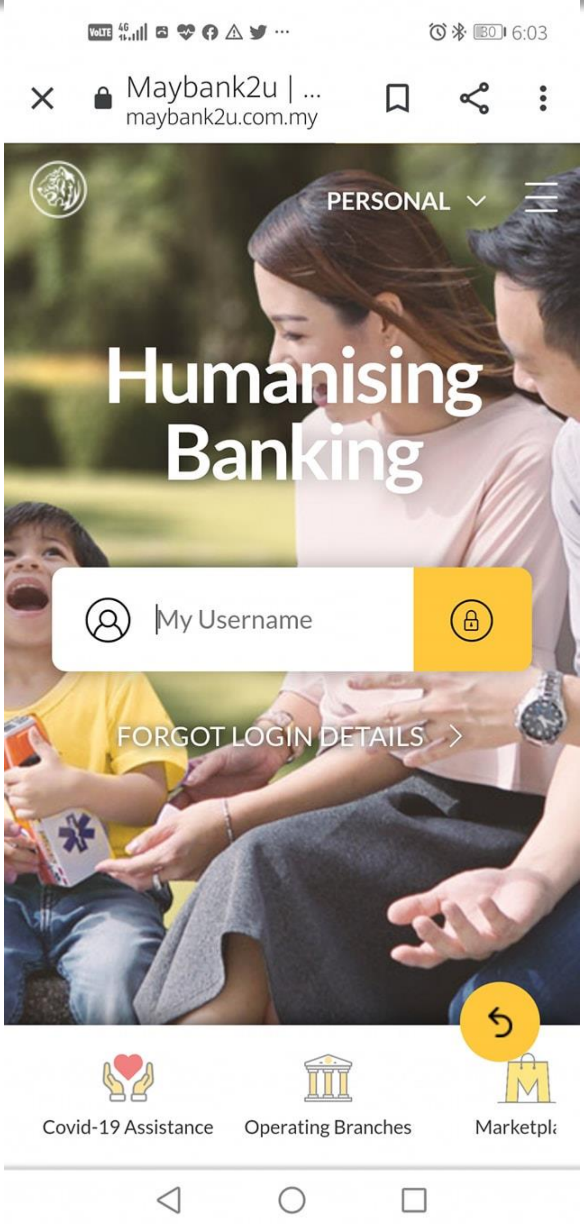
Wells Fargo Support

View Wells's LinkedIn profile

TIP You can respond to Wells by replying to this email

Spoofed website: Example found by Azizul Osman





URL obfuscation

- A spoofed website tries to have a genuine looking URL
- Examples:
  - http://www.cnn.com@evilwebsite.com
  - http://%65%76%69%6C%77%65%62%73%69%74%65%2E%63%6F%6D
  - Homograph attacks:
    - https://www.paypal.com
    - https://wikipedia.org
    - See https://en.wikipedia.org/wiki/IDN_homograph_attack

How to get e-mail addresses for spamming?

- Crawl web pages
- Join mailing lists, forums, … to get e-mail addresses of other members
- Hack servers of discussion forums, online communities,…
- Infect a PC/smartphone and get the entries from user’s address bock
- Guess (dictionary)

  john.doe@hotmail.com

  johndoe@hotmail.com

  jdoe@hotmail.com

  …

How to get e-mail addresses for spamming? (2)

- Or just buy them …
- Price depends on “quality”
- E-mails from hacks of company or community databases are particularly valuable for phishing attacks because they look real
  - Contain other information, e.g. client number

    “Dear Ramin Sadre (client-number #1234),

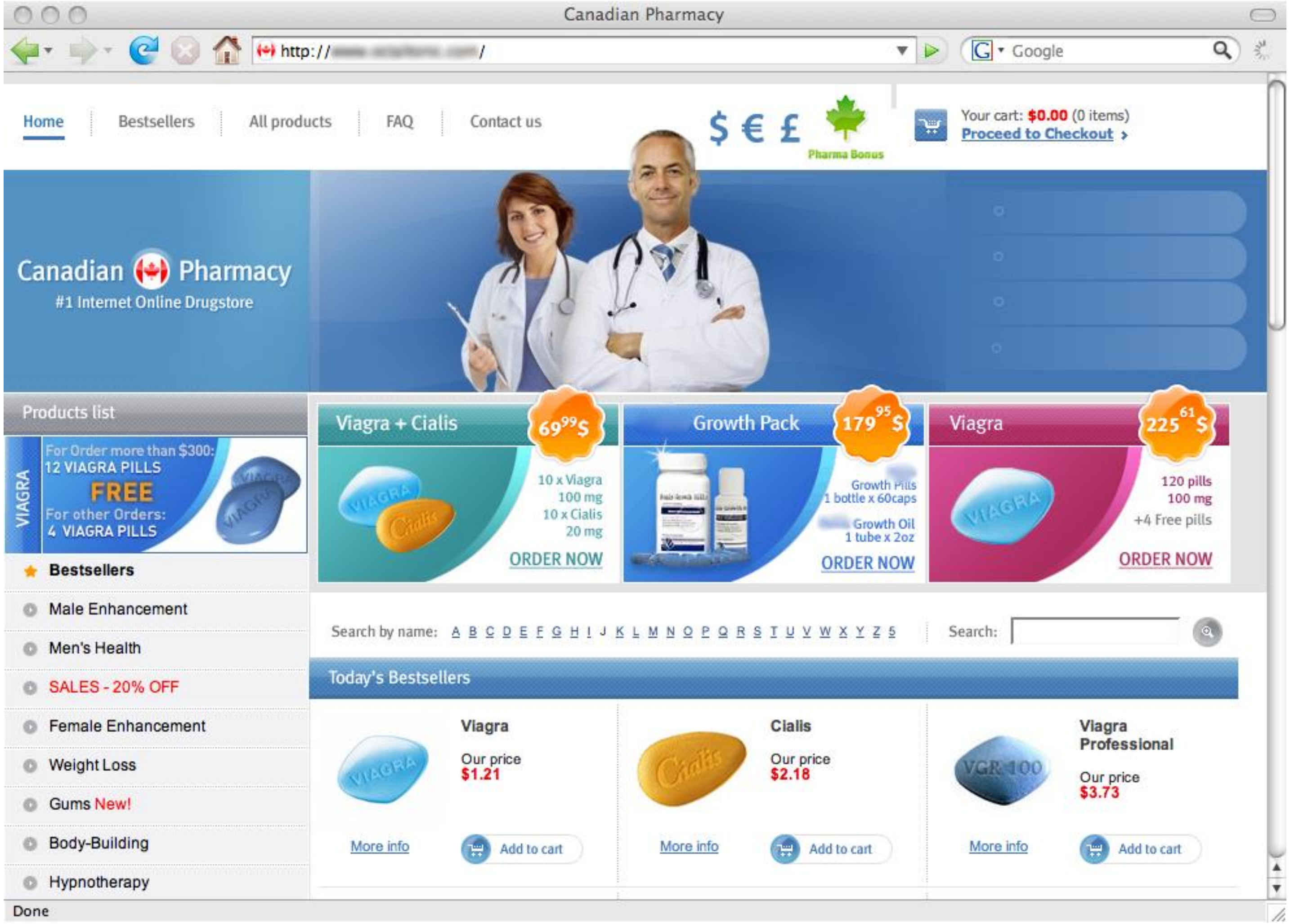
    …

    “
  - Spoofed sender address from somebody you trust

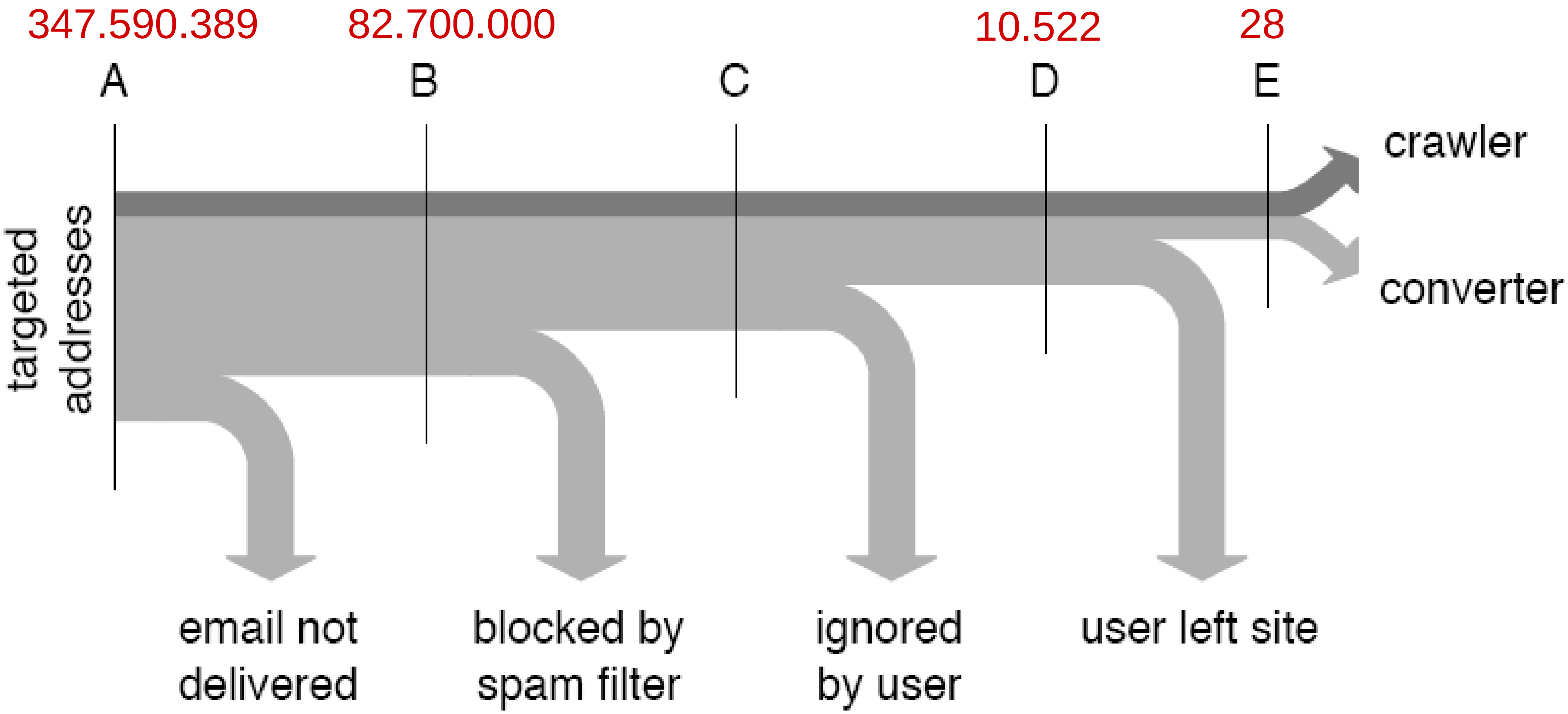
How successful is it?

- Actually, it's hard…
- You have to
  - Pass the spam filters of the mail server and/or client
  - Convince user to not directly delete the mail
  - Convince user to read the mail
  - Convince user to click on a link in the mail
  - Convince user to stay on the website
  - Convince user to buy something (shopping) or enter sensitive information, like the credit card number or a password (phishing)
- In 2007, researchers made a fake pharmacy website and hijacked the Storm botnet to send 347.590.389 spam mails. Only 28 users entered their credit card number on the website.

Their fake website



Conversion Rate



(Source: Spamalytics: An Empirical Analysis of Spam Marketing Conversion, Kanich *et al.*, 2008)

How to send spam?

Different ways possible:

1. Use your own/ISP’s SMTP server
2. Create fake mail accounts at free web mail providers
3. Use an open mail-relay server
   - Mail servers that forward to any e-mail to (thousands of) other mail servers
   - Nowadays, intentional open relay servers are less common and are often quickly blocked (blacklisted) by other mail servers
4. Use infected computers (botnets) (the most popular way)

Mail Filtering

Spam filters

- Spam filters run in the mail server or the mail client
- Modern filters analyze several aspects of the mail
  - Suspicious source IP (by blacklists)
  - Keywords (“money”) in the subject and body
  - Suspicious mail addresses (special characters, only numbers,…)
  - Suspicious URLs
  - …
- The filter calculates a “score” for each mail. If the score is above a configurable threshold, the mail is marked as spam

Example: SpamAssassin report

- https://spamassassin.apache.org/

```
X-Spam-Status: No, score=1.3 required=6.0 version=3.3.2
X-Spam-Report:
  *  0.4 URIBL_GREY Contains an URL listed in the URIBL greylist  [URIs: list-manage1.com]
  * -0.1 RCVD_IN_DNSWL_NONE RBL: Sender listed at http://www.dnswl.org/, low trust
  *                                    [205.201.128.128 listed in list.dnswl.org]
  * -1.5 SPF_HELO_PASS SPF: HELO matches SPF record
  *  0.0 HTML_IMAGE_RATIO_08 BODY: HTML has a low ratio of text to image area
  *  0.3 HTML_MESSAGE BODY: HTML included in message
  *  1.0 BAYES_50 BODY: Bayes spam probability is 40 to 60%  [score: 0.5000]
  *  0.9 MIME_QP_LONG_LINE RAW: Quoted-printable line longer than 76 chars
  *  0.1 DKIM_SIGNED Message has a DKIM or DK signature, not necessarily valid
  *  0.2 T_DKIM_INVALID DKIM-Signature header exists but is not valid
```

“HELO matches SPF record”

- SPF = Sender Policy Framework
  - allows the owner of a domain (e.g. uclouvain.be) to specify which computers are authorized to send emails with that sender address
  - can be queried over DNS

- When a message sender (SMTP client) sends a message to a message receiver (SMTP server), the client first sends a HELO message with its identity, e.g.

  HELO mail.uclouvain.be

- The server can use SPF to verify whether the sending IP address is authorized to send e-mails for @uclouvain.be

Blacklists

- Recipient’s mail server (or mail client) compares sender’s source IP against a blacklist
  - (Sender mail address is useless, can be spoofed)
- That’s the reason why spammers like using hacked mail servers or botnets
  - IP address of innocent users’ PCs likely not (yet) in the blacklist
- Greylisting:
  - First attempt is rejected
  - Real mail servers will try a second time, spam servers will probably not retry (to save resources)
  - Several minutes between the two attempts. Gives time for blacklists to register the spam campaign.

Blacklists (2)

- Example: http://www.spamhaus.org/
  - SBL: list of IP addresses of known spammers
  - XBL: list of IP addresses of hijacked PCs (infected by botnet software)

- Lists can be downloaded or accessed via DNS
  - Mail server sends a DNS query to the blacklist with the source IP of the mail
  - Blacklist replies with a DNS response whether the IP is on the list

Blacklists (3)

- How do blacklists get their entries?
- Spamhaus operates *spam traps* (mail honeypots)

  = E-mail addresses that do not belong to real users

  Usually hidden on web pages, so they are only found by the crawlers of the spammers
- Blacklists sometimes list an entire subnetwork (e.g. /24) if many of its IP addresses send spam

Statistical filters (Bayesian filtering)

- Statistical filter proposed in 2002 by Paul Graham

  http://www.paulgraham.com/spam.html

- Principle
  1. Take a large corpus of (a) spam mails and (b) non-spam mails
  2. Split the mails into tokens (words) and calculate the token frequencies in (a) and (b)
  3. Based on the result from step 2, calculate for each token the probability that a mail containing the token is spam
  4. For a new mail, look at all its tokens and calculate the probability that the mail is spam

Step 3

- Probability that an e-mail is spam if it contains token t:

$$P(is\ Spam \mid t)$$

- Bayes theorem:

$$= \frac{P(t \mid is\ Spam) \cdot P(is\ Spam)}{P(t)}$$

$$= \frac{P(t \mid is\ Spam) \cdot P(is\ Spam)}{P(t \mid is\ Spam)\ P(is\ Spam) + P(t \mid is\ not\ Spam) P(is\ not\ Spam)}$$

- Paul Graham's original mail filter assumed that $P(is\ Spam) = P(is\ not\ Spam)$ and got the simple formula:

$$P(is\ Spam \mid t) = \frac{P(t \mid is\ Spam)}{P(t \mid is\ Spam) + P(t \mid is\ not\ Spam)}$$

Step 4

- If a mail m consists of n tokens $t_1, t_2, \ldots$, what is the probability that the mail is spam?
- If we assume that tokens appear independently in emails ("Naive Bayes classifier"):

$$P(\, m \; is \; Spam) \coloneqq \frac{\prod p_i / P(is \; Spam)}{\prod p_i / P(is \; Spam) + \prod (1 - p_i) P(is \; not \; Spam)}$$

- Again, if we assume that $P(is \; Spam) = P(is \; not \; Spam)$:

$$P(\, m \; is \; Spam) \coloneqq \frac{\prod p_i}{\prod p_i + \prod (1 - p_i)}$$

where $p_i = P(is \; Spam \mid t_i)$

Practical implementation of step 3

In practice, $P(is\ Spam) < P(is\ not\ Spam)$.
Correction factor 2 avoids too many false positives

g := $2 \cdot$ token freq. in good mails (or 0 if no occurrence)
b := token freq. in bad mails (or 0 if no occurence)

```
if $g + b >= 5$ then
    $goodratio := \min\left(1, \frac{g}{total\ \#good\ mails}\right)$

    $badratio := \min\left(1, \frac{b}{total\ \#bad\ mails}\right)$

    $result := \max\left(0.01, \min\left(0.99, \frac{badratio}{goodratio+badratio}\right)\right)$
```

Only consider tokens with enough data

Limits if token only appears in good/bad mails

Remarks on statistical filters

- Advantages:
  - We get a concrete probability instead of an abstract score
  - Learning the token spam probabilities happens *once* in training
- Ways to improve the algorithm:
  - Ignore frequent tokens like “the” with spam probability close to 0.5. They don’t carry much information.
  - Don’t assume independence of tokens. Instead of using single words as tokens, combine two neighbor words to a token (*bigrams*). The words “buy” and “cheap” don’t carry much information, but the combined token “buy cheap” is more interesting
  - Decide how to handle new words that didn’t appear in your training data. Should they get high or low spam probability?