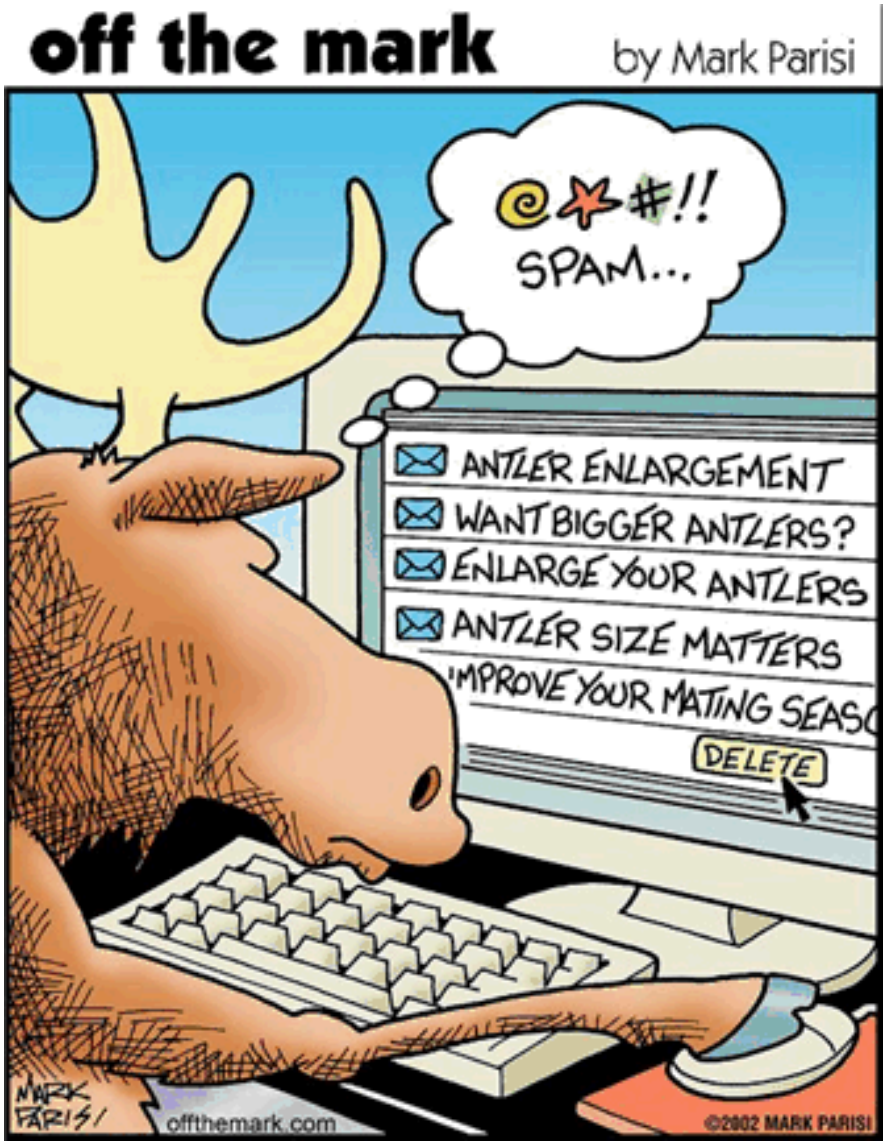
IT-Security 1

Chapter 12: SPAM

Prof. Dr.-Ing. Ulrike Meyer

WS 15/16

$


off the mark
by Mark Parisi
@*#!!
SPAM...
ANTLER ENLARGEMENT
WANT BIGGER ANTLERS?
ENLARGE YOUR ANTLERS
ANTLER SIZE MATTERS
'MPROVE YOUR MATING SEASC
DELETE
MARK PARISI
offthemark.com
©2002 MARK PARISI

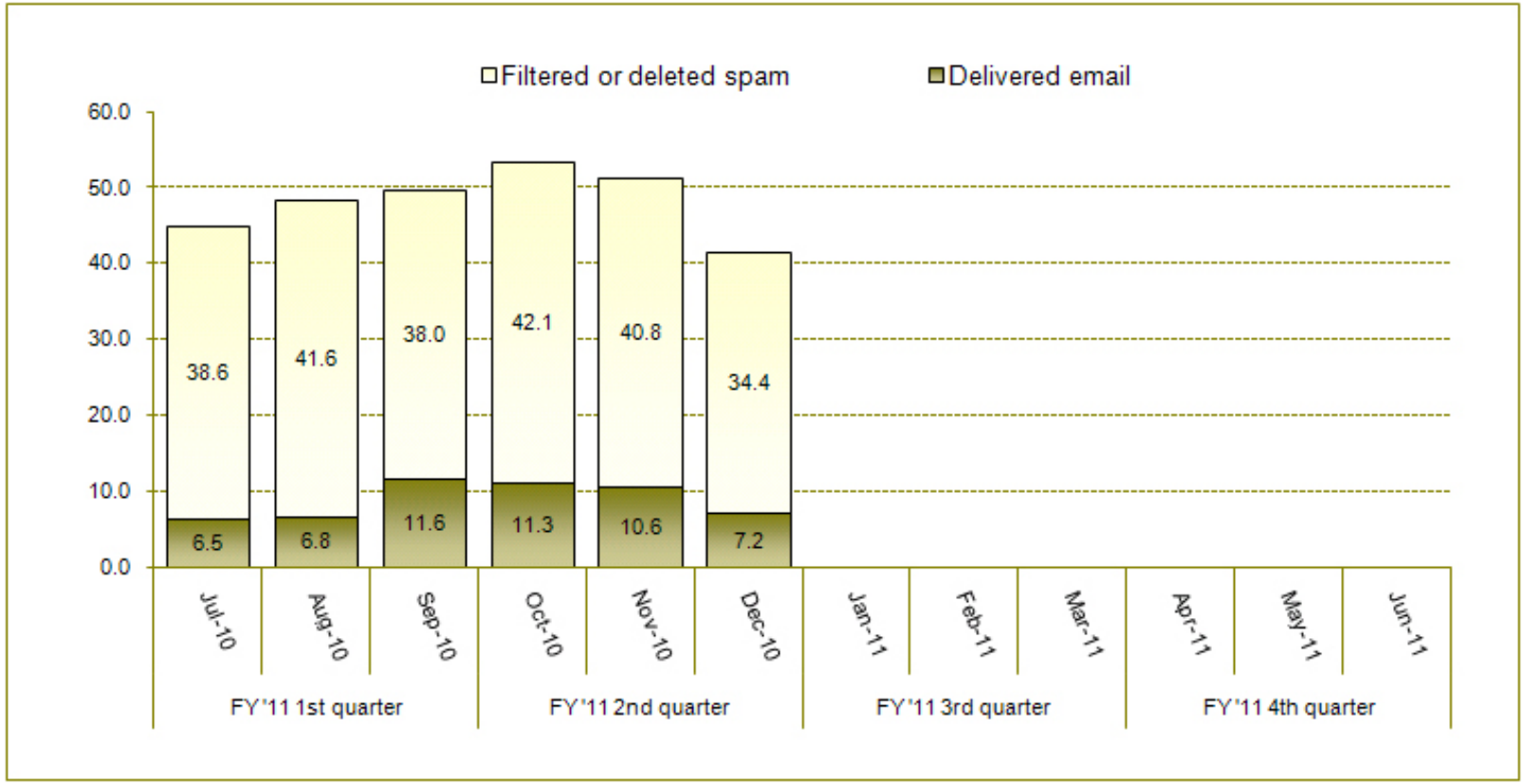
Chapter Overview

- What is Spam and how serious is it?
- Sources and Content of Spam
- Why Spam works in the first place
- Anti-Spam Mechanisms (non-exclusive)
  - Spam filters
  - Grey-listing
  - New protocols
  - Including authentication mechanisms in existing protocols
  - Increasing costs for spammers
  - Legal measures

What's Spam?

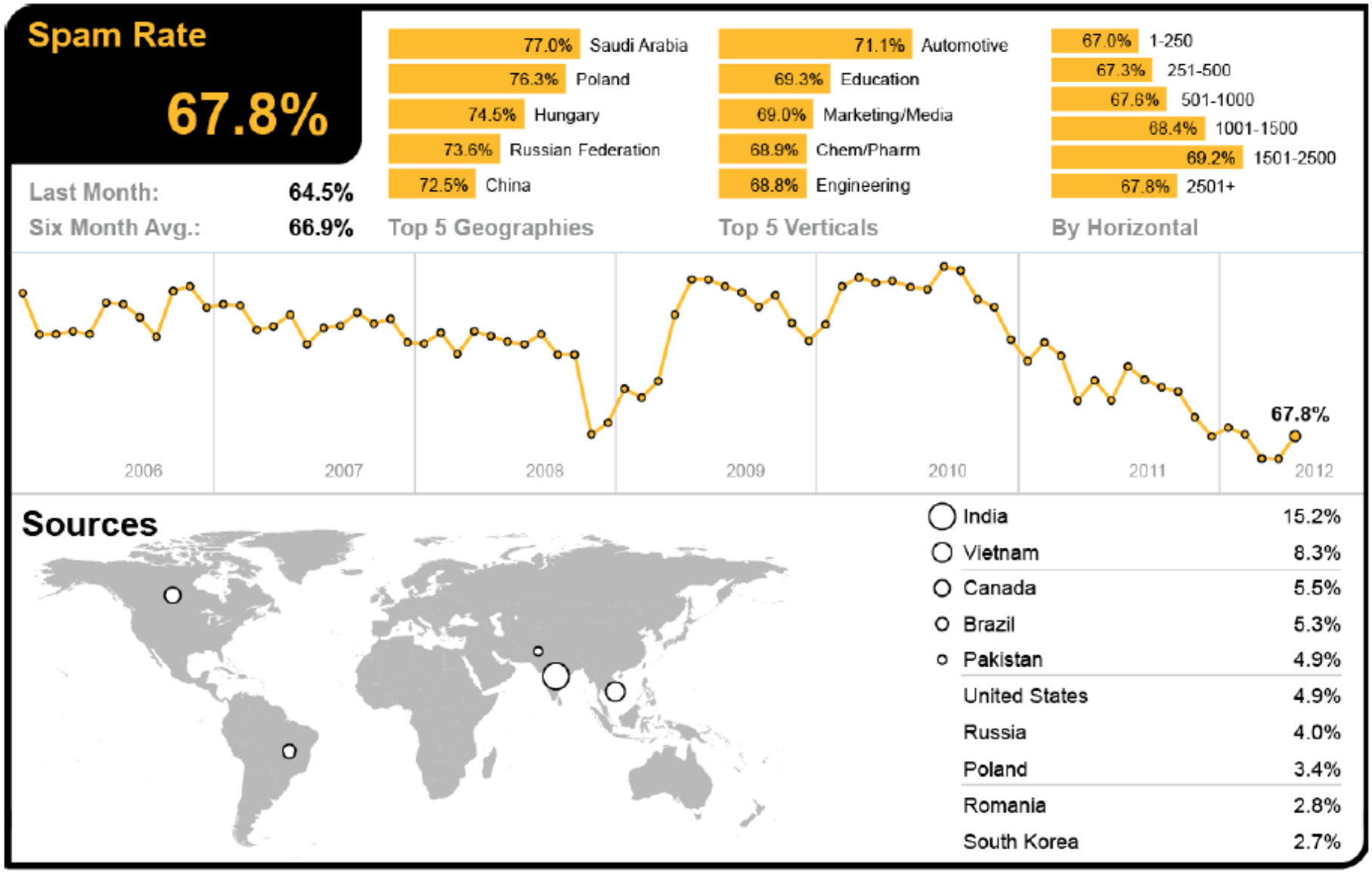
- Unsolicited commercial email or "junk email"
- SPAM stands for "SPiced hAM"
- The term originates from a Monty Python's Flying Circus scene
  - There the actors keep saying 'Spam, Spam, Spam and Spam' when reading options from a menu.
  - http://www.youtube.com/watch?v=cFrtpT1mKy8
- Today the spam-phenomenon has also spread to
  - Short messages
  - VoIP messages
  - ….

How Much SPAM is there?

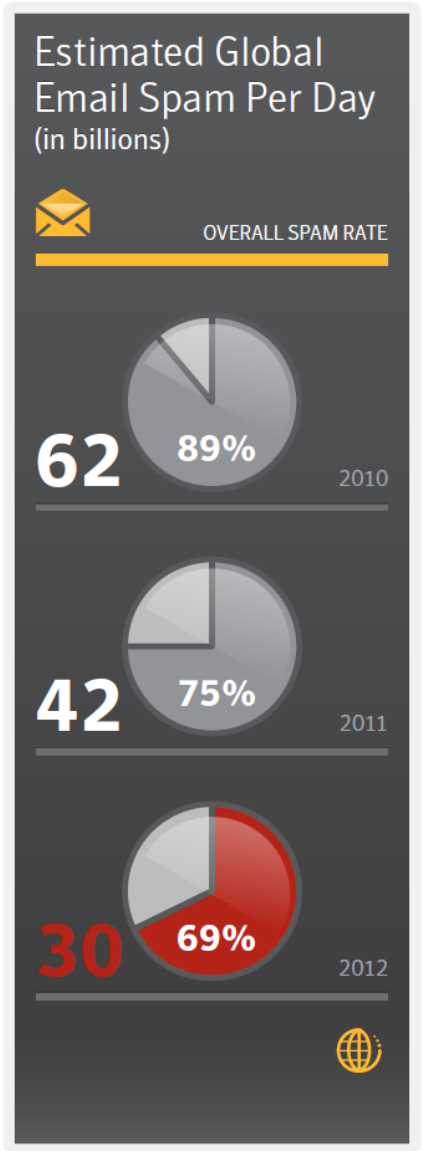


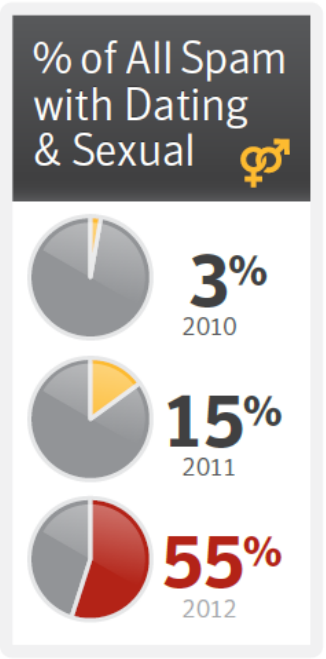
- Statistic shows spam/email gathered by Yale's IT-Services
- Numbers are in million emails (corresponds to 70-85% spam)

Symantec SPAM and Phishing Report 2012

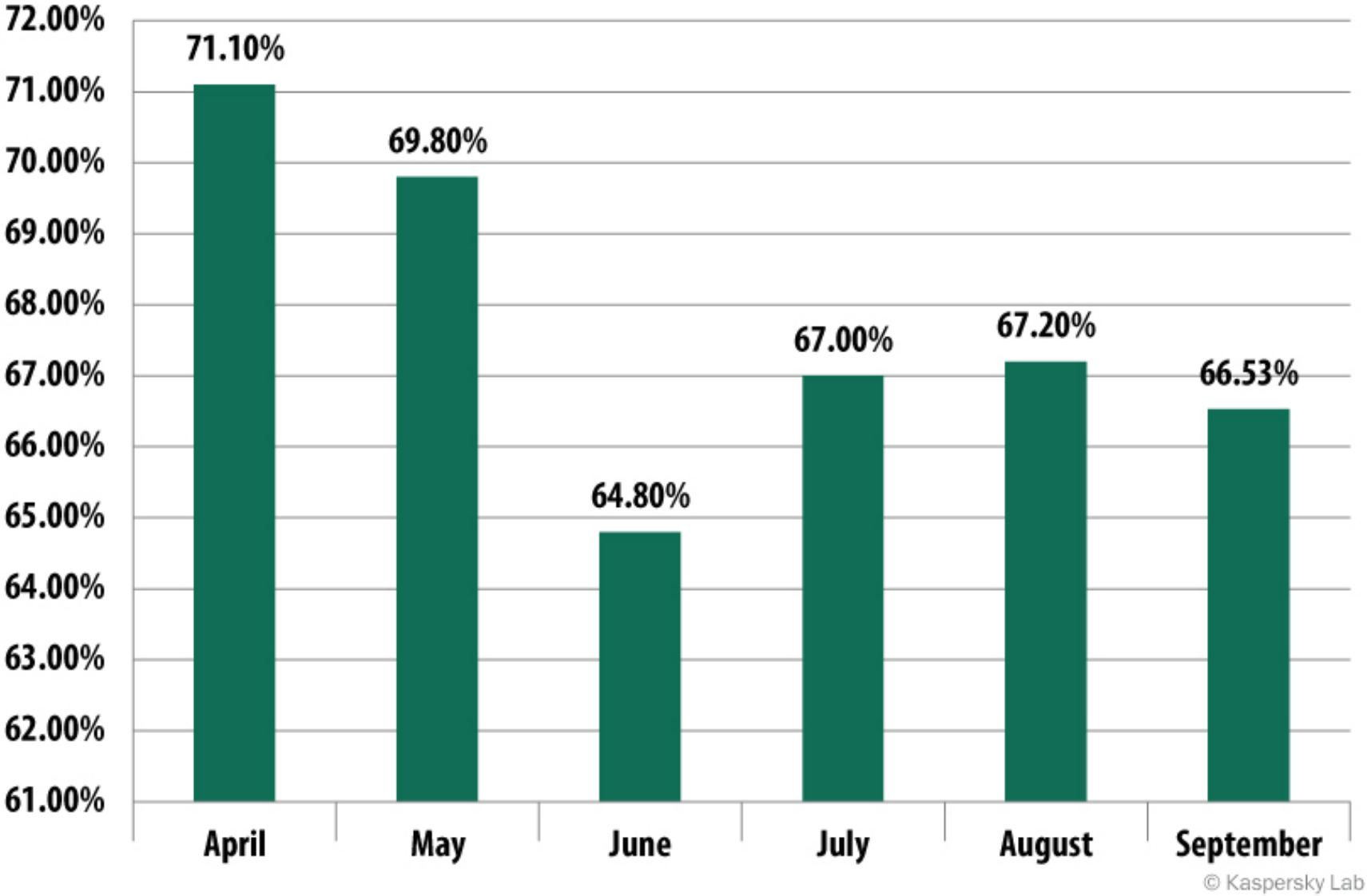


SPAM Statistics from Symantec Threat Report 1H 2013





Kaspersky Labs, SPAM Report 2014



- Proportion of spam in email traffic in 2014, 2015: 54% in Nov.

How Much Does SPAM Cost?

- Ferris Research estimates the total cost of spam for 2009 to US $130 billion
  - User productivity cost (deleting spam, looking for false positives, etc.): **85%**
  - Help desk cost (IT helping end users deal with spam): **10%**
  - Spam control software/hardware/service (licensing fees, amortized capital costs, etc.): **5%**

How Do Spammers Make Money?

- Direct Income
  - Act as marketing companies and obtain money for campaigns
- Web banner revenues
  - Get paid for every visit on a web page advertised in a spam mail
- Validation of contact information
  - Find valid email addresses and sell them
- Sell spam business models
  - Collect and sell information collected from responses to spam
- Scams
  - Phishing, gambling, dubious job offers, investment swindles
- Product selling
  - Spammers that sell the product they advertise

Why Spam Works

- SMTP does not require sender authentication or any authenticated routing information
  - Only the address of the receiver has to be correct
- In particular: from field can easily be manipulated to contain any email address
  - Spammers can thus easily hide their identity
- Spammer typically do not use their own ISP to send out spam
  - They try to connect to the destination email server directly or
  - Use open email relays
  - Use compromised client machines

Anti-Spam Methods

- Can be divided into
  - Pre-sent mechanisms
    - Increase spammers costs
      - Micropayments or digital stamps
    - Increase spammers risks
      - Legislation such as CAN-SPAM Act
  - Post-sent mechanisms
    - Content-based filtering
    - Source address-based
      - Filtering (blacklisting, whitelisting)
      - Authentication based on certificates and signatures
      - Challenge-response based mechanisms
  - New protocols

Increasing Spamming Costs

- Increasing Monetary Costs
  - Digital stamps and other micropayments
    - Require sender of an email to pay a small monetary amount for each email sent
    - Does not hurt regular users but mass-mailers
    - Problem: mailing-lists and other wanted bulk-mail
  - Using tokens
    - Same idea but non-monetary tokens
  - Using puzzles that require human interaction
    - Before an email is accepted the sender is required to solve a puzzle that requires human interaction
    - Problem: what about wanted automatically generated mail such as order confirmations?

Increase Spammers Risk

- Take legal measures and prosecute spammers
- May help to deter spammers
- Example of spam legislation
  - The US CAN-SPAM-Act
  - The European regulations
  - Both were created in 2003 and are effective since 2004

The CAN-SPAM Act of 2003 (1)

- **C**ontrolling the **A**ssault of **N**on-**S**olicited **P**ornography **A**nd **M**arketing Act of 2003
- Defines commercial e-mail as
  - Email, the primary purpose of which is
    - commercial advertisement or promotion of a commercial product or service
- Allows for commercial e-mail to be sent to recipients only if
  - Information in header is not false or misleading
  - Subject headings are not deceptive
  - Return e-mail address is functional
  - Opt-out mechanism included
    - Enables recipients to indicate they do not wish to receive future commercial e-mail messages from that sender
  - Email must be clearly and conspicuously identified as an advertisement

The CAN-SPAM Act of 2003 (2)

- Sexually oriented commercial e-mail must include, in the subject heading, a “warning label”
- Violators may be sued by Federal Trade Commission, state attorneys, and ISPs (but not by individuals)
- Violators are subject to statutory damages of up to $250 per e-mail, to a maximum of up to $6 million
- Violators may be fined, or sentenced to up to 3, or five years in prison, or both, for
  - accessing someone else’s computer without authorization and using it to send multiple commercial e-mail messages,
  - or materially falsifying header information in multiple commercial email messages
  - …

The CAN-SPAM ACT Does Not…

- … Create a “Do Not Email registry” where consumers can place their e-mail addresses in a centralized database to indicate they do not want commercial e-mail
- … Require that consumers “opt-in” before receiving commercial email
- … Require commercial e-mail to include an identifier such as “ADV” in the subject line to indicate it is an advertisement

EU Directive 2003-13

- Article 13(1) of the Privacy and Electronic Communications Directive requires Member States to
  - Prohibit the sending of unsolicited commercial communications by
    - Email or other electronic messaging systems such as SMS and MMS
  - Unless prior consent of the receiver has been obtained (opt-in system)
  - Unless there is an existing customer relationship, in which case, the sender must provide an opt-out option
- Article 13(4) prohibits direct marketing messages by e-mail or SMS which
  - Conceal or disguise the identity of the sender
  - Do not include a valid address to which recipients can send a request to cease such messages
- The EU directive sets the broad policy, but each member nation must pass its own law as to how to implement it

Blacklisting

- Can be centralized or local
- Centralized blacklisting typically uses DNS
  - IP addresses of suspected spammers are entered in a centrally maintained database
  - Standard DNS lookups are used to query the database at the time of the SMTP connection
    - SMTP client connects to an SMTP server
    - Server checks SMTP client's IP address using a DNS lookup mechanism (blacklist check) to check if the client's address is on the blacklist
- Advantages
  - DNS lookups have low CPU overhead
  - Spam can be blocked before it arrives
- Disadvantages
  - Number of overall DNS lookups increases
  - Blacklist maintainer needs to be trusted

Whitelisting

- Is typically local
- Indicates a list of IP addresses of SMTP clients which are to be trusted
- Advantages
  - Emails coming from IP addresses on the whitelist can bypass further spam filters
- Disadvantages
  - Whitelists tend to be overly restrictive when used alone
  - If a spammer is able to send email from an IP address on the whitelist his spam may bypass further spam filters
- Whitelists are typically combined with other spam blocking techniques

Challenge / Response Mechanisms

- Advanced versions of whitelists
- Incoming messages from IP addresses not on the whitelist trigger an automated reply (challenge) to the sender
- Sender needs to prove that he is a real user and not an automated mailer
  - E.g. sender is required to click on a link in the challenge message
- Only after the proof is obtained the original email is delivered to the intended recipient
- Advantages
  - Help to protect against the large amount of false positives generated by whitelists
- Disadvantages
  - Potential deadlock if sender and receiver applies such a mechanism
  - Unclear how to deal with legitimate automated emails like order confirmations

Requiring Authentication

- Sender is required to authenticate to the receiver, otherwise his email is not accepted
- Mechanisms suggested are either based on simple IP address legitimacy checking, or based on certificates and digital signatures
- Digital signatures and certificates require a working public key infrastructure with trusted authorities
  - Signatures are generated and checked by mail servers rather than senders and receivers, public keys are distributed via DNS
  - Example for such a mechanism: Domain Key Identified Mail developed by IETF

Greylisting

- Very aggressive method of blocking Spam
- Based on the observation that Spammers typically do not try to resent messages as they do not know if the recipient addresses really exist
- When a client connects to an SMTP server that uses greylisting
  - Server records IP address of SMTP client, sender address, recipient address
  - Checks if entry for the record already exists
    - If yes, it accepts the message
    - If no it stores the record and returns a temporary reject message
  - Typically the mail agents of the sender reacts by trying to resent the message

Rule-based Filters

- Are content-based filters
- Were popular until 2002 when Bayesian filters became largely available
- Search each email message for patterns that indicate spam
  - E.g. specific words or phrases, large amounts of capital letters or exclamation marks
- Detection of specific patterns attributes an amount of points to an email message
- If the points gathered for an email message exceeds a certain threshold it is classified as spam

- Problems
  - Static rule sets can easily be defeated by spammers e.g. by misspelling words

Bayesian Spam Filters

- Originally introduced in 1998
- Work by analyzing the words inside an email to calculate the probability that it is spam
- Probability is based on
  - Those words that provide evidence that a message is spam
  - Those words that provide evidence that a message not spam
- Bayesian filters need to be trained with a set of email messages that are correctly classified into spam and “ham” (= legitimate email)
- Provided they receive ongoing training, Bayesian filters are constantly self-adapting
  - Can potentially also stop new spam
- Disadvantage
  - Bayesian filters required the entire message to be received before analysis can begin

How Bayesian Spam Filters Work (1)

- Require "priors", i.e. a priory probabilities, for a message being spam or ham
  - E.g. P(spam) = : 80%, P(ham) = : 20%
- Require "feature vectors" that describe certain properties of an email message
  - E.g. feature vector x may be a Boolean variable that is
    - 1 if message contains both the words "Viagra" and "buy"
    - 0 if message does not contain both "Viagra" and "buy"
- Requires estimates for the conditional probabilities
  - P(x|spam) and P(x|ham)
  - These can be computed, with the help of a training set of correctly classified email messages

How Bayesian Spam Filters Work (2)

- If a new message is received with feature vector x, then a decision for ham or spam depends on
  - $P(\mathrm{ham}|\mathrm{x})$ and $P(\mathrm{spam}|\mathrm{x})$
- These probabilities can be computed with the help of Bayes' Theorem:

$$P(\mathrm{ham}|\mathrm{x}) = \frac{P(\mathrm{x}|\mathrm{ham})\ P(\mathrm{ham})}{P(\mathrm{x})}$$

$$P(\mathrm{spam}|\mathrm{x}) = \frac{P(\mathrm{x}|\mathrm{spam})\ P(\mathrm{spam})}{P(\mathrm{x})}$$

How Bayesian Filters Work (3)

- Recall

$$P(\text{ham}|x) = \frac{P(x|\text{ham})\,P(\text{ham})}{P(x)}$$

$$P(\text{spam}|x) = \frac{P(x|\text{spam})\,P(\text{spam})}{P(x)}$$

- Decide for "spam" if $P(\text{spam}|x) > P(\text{ham}|x)$

- This is equivalent to

$$P(x|\text{spam})\,P(\text{spam}) > P(x|\text{ham})\,P(\text{ham})$$

- or

$$\frac{P(x|\text{spam})}{P(x|\text{ham})} > \frac{P(\text{ham})}{P(\text{spam})}$$

Decision Threshold

How Bayesian Filters Work (4)

- Up till now we considered false classification as spam as bad as false classification as ham
- However, this is generally not true
  - False classification of ham message as spam is much worse than receiving a spam message once in a while
- This can be captured with the help of a loss function

Decision

Truth	ham	spam
ham	0	$L_{hs} = 1000$
spam	$L_{sh} = 1$	0

Loss on falsely classifying ham as spam

Loss on falsely classifying spam as ham

- Goal: minimize the expected loss during decision

How Bayesian Filters Work (5)

- Decision that minimizes the expected loss

$$L_{sh}P(\text{spam}\,|\,x) > L_{hs}P(\text{ham}\,|\,x)$$

- Using Bayes Theorem again, we get

$$L_{sh}P(x\,|\,\text{spam})\,P(\text{spam}) > L_{hs}P(x\,|\,\text{ham})\,P(\text{ham})$$

- Or equivalently

$$\frac{P(x\,|\,\text{spam})}{P(x\,|\,\text{ham})} > \frac{L_{hs}P(\text{ham})}{L_{sh}P(\text{spam})}$$

Decision Threshold taking loss into account

How Bayesian Filters Work (6)

- Naturally Bayesian spam filters do use many feature vectors x_1 $x_2 \ldots$
- The Naïve Bayes approach assumes that all of these feature vectors are independent such that

$$P(x_1,\ldots,x_n \mid \text{spam}) = \Pi_i P(x_i \mid \text{spam})$$

- A decision for spam is therefore taken if

$$\Pi_i \frac{P(x_i \mid \text{spam})}{P(x_i \mid \text{ham})} > \frac{L_{hs} P(\text{ham})}{L_{sh} P(\text{spam})}$$

Examples for Feature Vectors

- Standard Bayesian Spam Filters
  - Each word in the header, subject and body is a feature
- Token grab bag
  - A sliding window of five words is moved across the input text
  - All combinations of those five words are taken in an order-sensitive way as a feature
- Token Sequence Sensitive
  - Sliding window of five words moved across the text
  - All combinations of word deletions are applied
  - Each resulting combination of words is taken as feature
- ....

Further Reading and Resources

- Marcia S. Smith, “Spam”: An Overview of Issues Concerning Commercial Electronic Mail, 2006
- W. Gansterer et al., “Anti-Spam-Methods – State of the Art”, 2005
- D. Cook et al., “Catching Spam Before it Arrives”, 2006
- Domain key identified mail: http://www.dkim.org/ietf-dkim.htm provides a nice overview on the IETF work