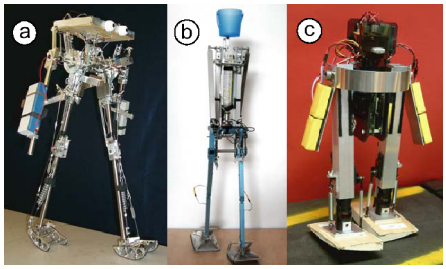
1

## Robots run amok



3

## How to protect a website?

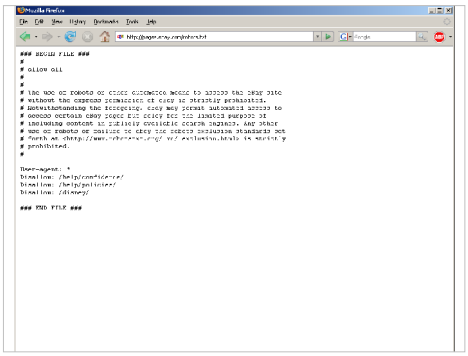
- law
- code
- markets
- norms

4

## I, Robot

- http://www.robotstxt.org/robotstxt.html
- Robot exclusion standard: A robot wants to vists a Web site URL, say http://www.example.com/welcome.html. Before it does so, it first checks for http://www.example.com/robots.txt, and finds:
  - User-agent: *
  - Disallow:
- There are two important considerations when using /robots.txt:
  - robots can ignore your /robots.txt. Especially malware robots that scan the web for security vulnerabilities, and email address harvesters used by spammers will pay no attention.
  - the /robots.txt file is a publicly available file. Anyone can see what sections of your server you don't want robots to use.
- E.g.
  - http://www.whitehouse.gov/robots.txt
  - http://bulk.resource.org/robots.txt

5



## CheepTix

- CheepTix.com wants to set up an airfare search engine to help travelers find discount flights to vacation spots, offering alternative destinations if it can't find the one requested. CheepTix gets its listings by "crawling" airline websites and "screen-scraping" the fare information they list.
- Just before launch, the CheepTix GC comes to you for advice. What questions do you have for her in assessing the risk of a lawsuit?



- You (a different you) have been hired by Gamma Airlines to develop the company response to screen-scraping robots. Help your client to evaluate the business and legal considerations here.



## eBay v. Bidder's Edge

## Intel v. Hamidi

Former And Current Employees of Intel



## Register.com v. Verio

The data in Register.com's WHOIS database is provided to you by Register.com for information purposes only, that is, to assist you in obtaining information about or related to a domain name registration record. Register.com makes this information available "as is," and does not guarantee its accuracy. By submitting a WHOIS query, you agree that you will use this data only for lawful purposes and that, under no circumstances will you use this data to: (1) allow, enable, or otherwise support the transmission of mass unsolicited, commercial advertising or solicitations via direct mail, electronic mail, or by telephone; or (2) enable high volume, automated, electronic processes that apply to Register.com (or its systems). The compilation, repackaging, dissemination or other use of this data is expressly prohibited without the prior written consent of Register.com.

- Organization:
  Register.Com, Inc.
  Domain Registrar
  575 Eighth Avenue, 11th Floor
  New York, NY 10018
  US
  Phone...212-798-9100
  Fax...212-798-5464
  Email: domainregistrar@register.com

  Registrar Name....: Register.com
  Registrar Whois...: whois.register.com
  Registrar Homepage: http://www.register.com

  Domain Name: REGISTER.COM

  Created on.....: Wed, Aug 04, 1999
  Expires on.....: Tue, Aug 04, 2015
  Record last updated on...: Fri, Jun 09, 2006

- Administrative Contact, Technical Contact, Zo
  Register.Com, Inc.
  Domain Registrar
  575 Eighth Avenue, 11th Floor
  New York, NY 10018
  US
  Phone...212-798-9100
  Fax...212-798-5464
  Email: domainregistrar@register.com

  Domain servers in listed order:

  NS1.REGISTER.COM 216.21.234.96
  NS2.REGISTER.COM 216.21.226.96
  NS3.REGISTER.COM 216.21.234.97
  NS4.REGISTER.COM 216.21.226.97

  Visit AboutUs.org for more information about
  AboutUs: register.com

## EF Cultural Travel v. Zefer

Zefer (contractor)



## How far is too far?

- Sending a worm that exploits holes, guesses passwords, and bogs down computer networks.
- Scraping screens from a publicly accessible website
- Even if tipped off to the organization of those screens by a misappropriated trade secret?



## How might we stop spam?

- CAN it?
- Block it?
- Ignore it?



## How might we stop spam?

- CAN it? (law)
  - Place liability on the sender
  - Place liability on the advertiser
  - Place liability on the ISP
- Block it?
  - Where?
    - At the ISP level
    - At the user level
  - How?
    - Blacklists
    - Black hole lists
    - Whitelists
    - Trusted senders
    - Sender confirmation
    - Do-not-email registry
- Ignore it?



## How should we evaluate anti-spam?

- Who is regulating / regulated?
- Who feels the impact?
  - Who apart from the target?
- How effective is the regulation?
  - Where are its errors?
- How easy to administer?
- How transparent is the regulation?



## CAN-SPAM

- What's prohibited?
- Who can enforce?
- Against whom?

## What are the effects of spam-blocks?

- On spammers?
- On email recipients?
- On non-spamming email senders?
- On developers of email software?
- On …



## Special Medzz

- Bruno's Botanicals wants to introduce people to its new "healing compound." Can it harvest email addresses of potential customers from natural healing weblogs?
- Can it send emails to its prior customers?
  - If Bruno hasn't asked for their consent?
  - From bruno@amazon.com, because that's more likely to get opened?
  - If the list includes customers who have specifically opted out of email communications?



## Technical measures

- MAPS RBL
  - Blackhole lists of IP addresses "that have been shown to send spam and/or allow their resources to be used by those who send spam."
  - "Loss of connectivity hurts us all. Spam hurts us all even more."
- DomainKeys sender authentication
  - Signatures linking email messages to the domain from which they are sent
  - "Finally, you could choose to send unauthenticated mail…. If you choose this path, you should carefully monitor the amount of authenticated mail over time to ensure that this strategy does not impact the deliverability of your email."



## Technical measures…

- Local spam blockers
  - Spamassassin
  - DSPAM
  - Bayesian filtering
- Other sender verification
  - Bonded sender
  - Challenge-response
  - Whitelist



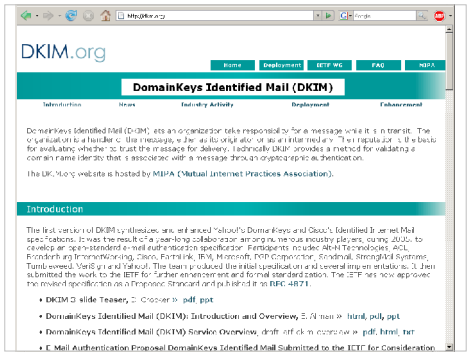
## OptInRealBig v. SpamCop

- SpamCop runs a service to which users can forward spam they have received. SpamCop identifies ISPs from header information and URLs in the message, and forwards complaints mechanically to ISP abuse departments.
- OptInRealBig is a "sender of bulk commercial emails" whose ISPs limit or terminate service after complaints received via SpamCop.
- Can OptInRealBig recover from SpamCop?



## How should we evaluate anti-spam?

- Who is regulating / regulated?
- Who feels the impact?
  - Who apart from the target?
- How effective is the regulation?
  - Where are its errors?
- How easy to administer?
- How transparent is the regulation?



DKIM.org

Home | Deployment | IETF WG | FAQ | MAAWG

DomainKeys Identified Mail (DKIM)

Introduction | News | Industry Activity | Deployment | Enhancement

DomainKeys Identified Mail (DKIM) lets an organization take responsibility for a message while it is in transit. The organization is a handler of the message, either as its originator or an intermediary. Technology is the basis for evaluating whether to trust the message for delivery. Technically, DKIM provides a method for validating a domain name identity that is associated with a message through cryptographic authentication.

The DKIM.org website is hosted by MIPA (Mutual Internet Practices Association).

## Introduction

The first version of DKIM synthesizes and enhanced Yahoo!'s DomainKeys and Cisco's Identified Internet Mail specifications. It was the result of a yearlong collaboration among several industry players, during 2005, to develop an open-standard e-mail authentication specification. Participants include AT&T Technologies, AOL, Brandenburg InternetWorking, Cisco, FarmLink, IBM, Microsoft, PGP Corporation, Sendmail, StrongMail Systems, TumbleweedMailSign and Yahoo!. The team produced the initial specification and several implementations. It then submitted the work to the IETF for further enhancement and formal standardization. The IETF has now approved the revised specification as a Proposed Standard and published it as RFC 4871.

- DKIM Slide Teaser, D. Crocker » pdf, ppt
- DomainKeys Identified Mail (DKIM): Introduction and Overview, E. Allman » html, pdf, ppt
- DomainKeys Identified Mail (DKIM) Service Overview, draft-ietf-dkim-overview » pdf, html, txt
- E-Mail Authentication Proposal DomainKeys Identified Mail Submitted to the IETF for Consideration