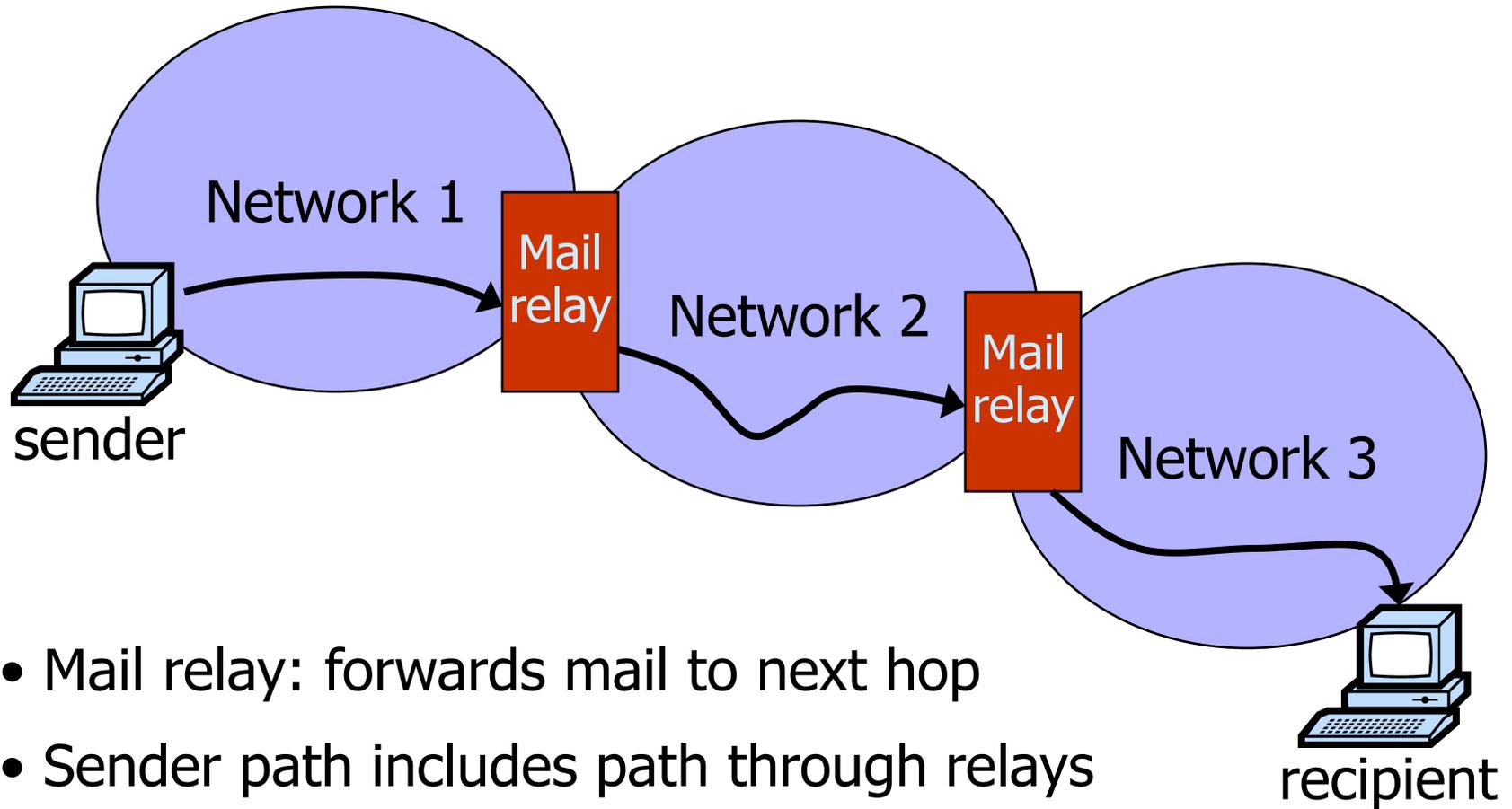




Vitaly Shmatikov

# Email in the Early 1980s



# Email Spoofing

---

- ◆ Mail is sent via SMTP protocol
  - No built-in authentication
- ◆ MAIL FROM field is set by the sender
- ◆ Recipient's mail server only sees the IP address of the direct peer from whom it received the message

# Mail Relays

---

- ◆ An SMTP **relay** forwards mail to destination
  1. Bulk email tool connects via SMTP (port 25)
  2. Sends list of recipients via RCPT TO command
  3. Sends email body (once for all recipients!)
  4. Relay delivers message
- ◆ Honest relay adds correct Received: header revealing source IP
- ◆ Hacked relay does not

# A Closer Look at Spam

Received: by 10.78.68.6 **Inserted by relays** hua;

Mon, 12 Feb 2007 06:43:30 -0800 (PST)

Received: by 10.101.10.10 MTP id l18mr17307116agc.1171291410432;  
Mon, 12 Feb 2007 06:43:30 -0800 (PST)

**Bogus!**

Return-Path: <\

Received: from onelinkpr.net ([203.169.49.172])

by mx.google.com with ESMTP id 30si1171291410432c.2007.02.12.06.43.18;

Received: from onelinkpr.net ([203.169.49.172]) by mx.google.com with ESMTP id 30si1171291410432c.2007.02.12.06.43.18; **Puerto Rico** **Mongolia** r permitted nor

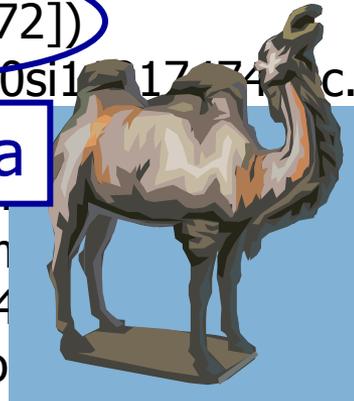
by best guess record for domain onelinkpr.net

Message-ID: <050057765.stank.203.169.49.172@onelinkpr.net>

From: "Barclay Morales" <wvnlwee@aviva.ro>

To: <raykwatts@gmail.com>

Subject: You can order both Viagra and Cialis.



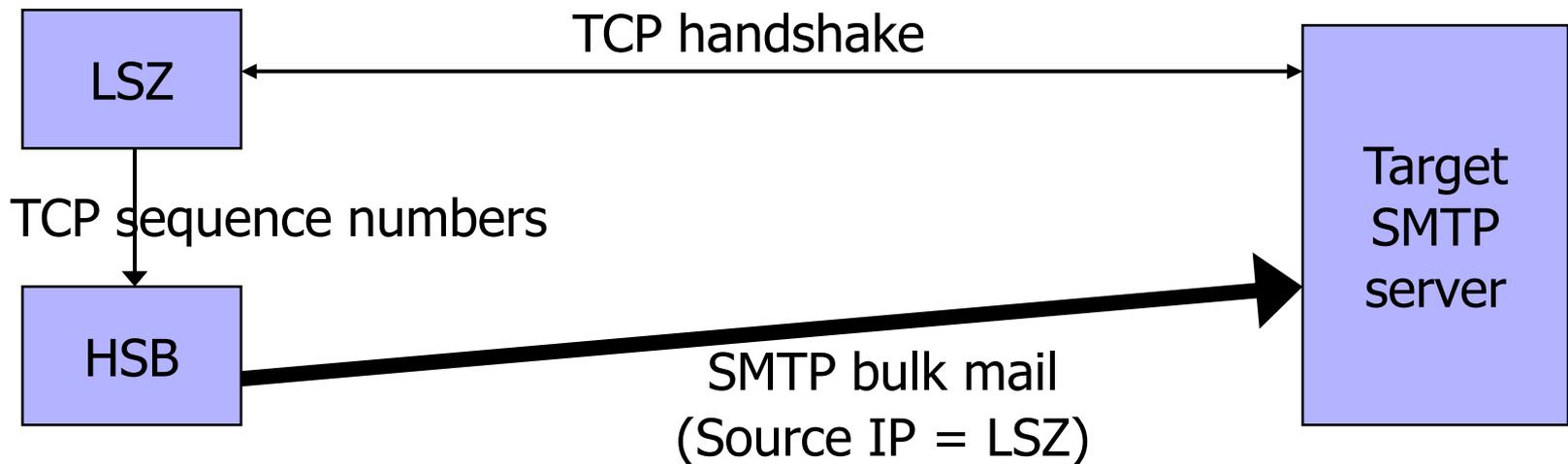
# Why Hide Sources of Spam?

---

- ◆ Many email providers blacklist servers and ISPs that generate a lot of spam
  - Use info from [spamhaus.org](http://spamhaus.org), [spamcop.net](http://spamcop.net)
- ◆ Real-time blackhole lists stop 15-25% of spam at SMTP connection time
  - Over 90% after message body checks
- ◆ Spammers' objective: evade blacklists
  - Botnets come very handy!

# Thin Pipe / Thick Pipe

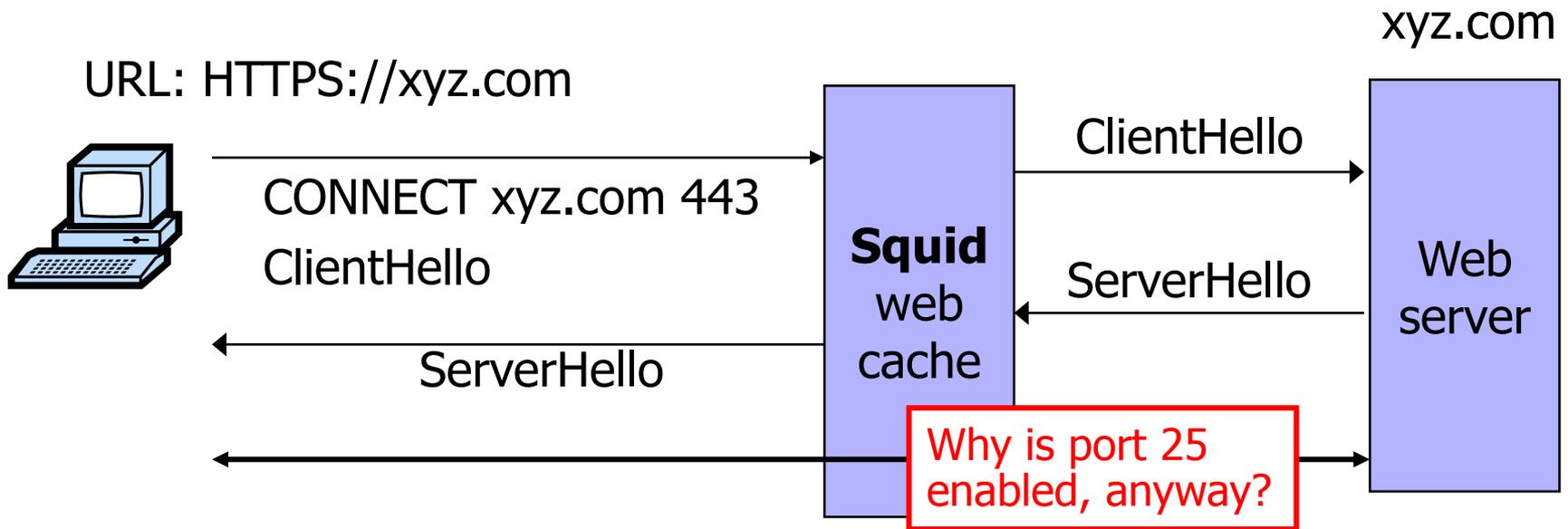
- ◆ Spam source has high-speed broadband machine (HSB) and controls a low-speed zombie (LSZ)



- ◆ Hides IP address of HSB; LSZ is blacklisted

# Open HTTP Proxies

- ◆ Web cache (HTTP/HTTPS proxy), e.g., squid



- ◆ To spam: `CONNECT <Victim's IP> 25`, then issue SMTP Commands
  - Squid becomes a mail relay



# Open Relays vs. Open Proxies

---

## ◆ Open proxy

- Spammer must send message to each recipient through the proxy

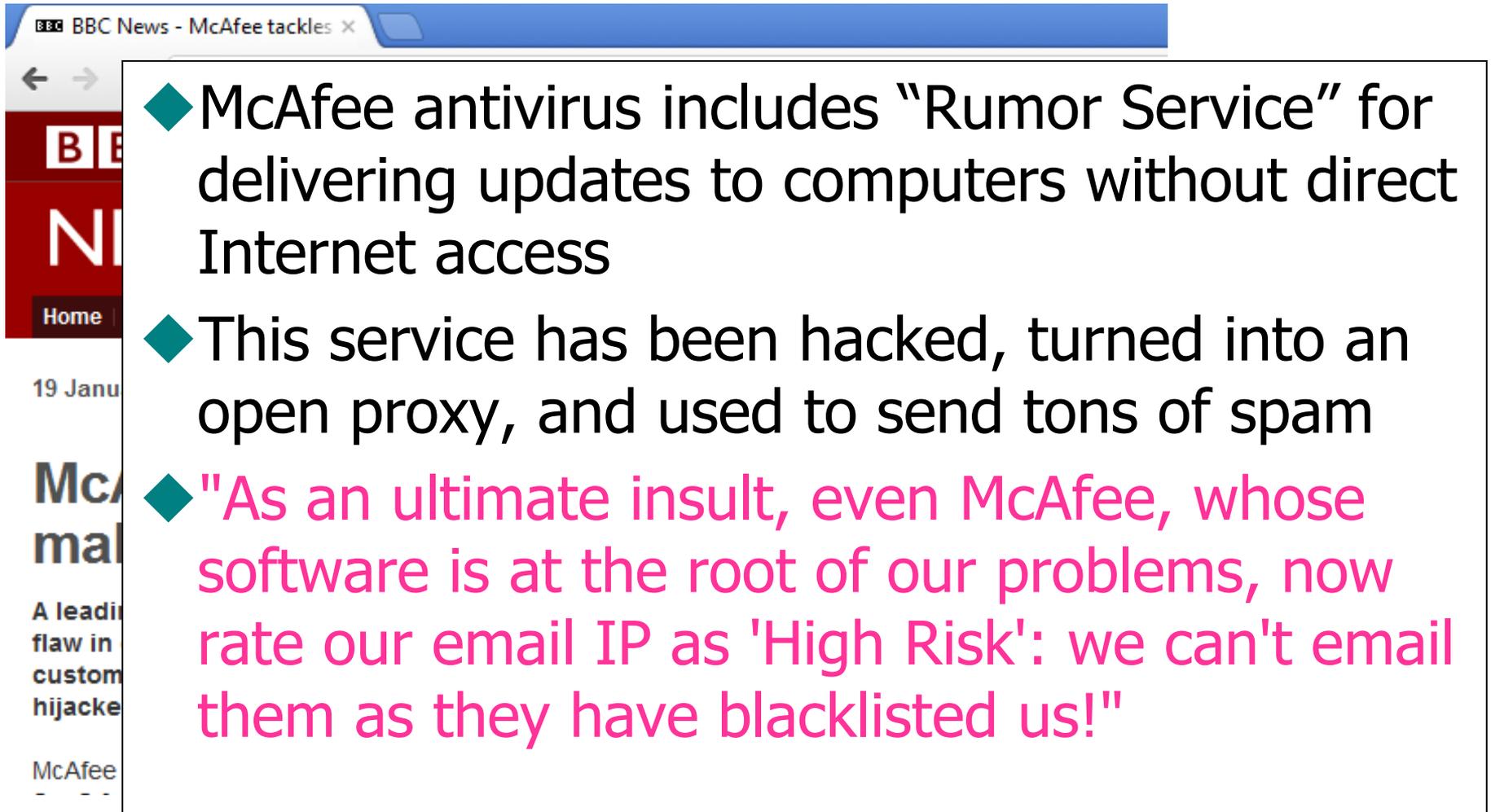
## ◆ Open relay

- Takes a list of addresses and sends to all
- Can host an open relay on a zombie

## ◆ Listing services for open proxies and relays (many appear to be defunct as of 2010)

- <http://www.multiproxy.org/>  
<http://www.stayinvisible.com/>  
<http://www.openproxies.com/> (\$20/month)

# McAfee Spam Hijack (Jan 2012)

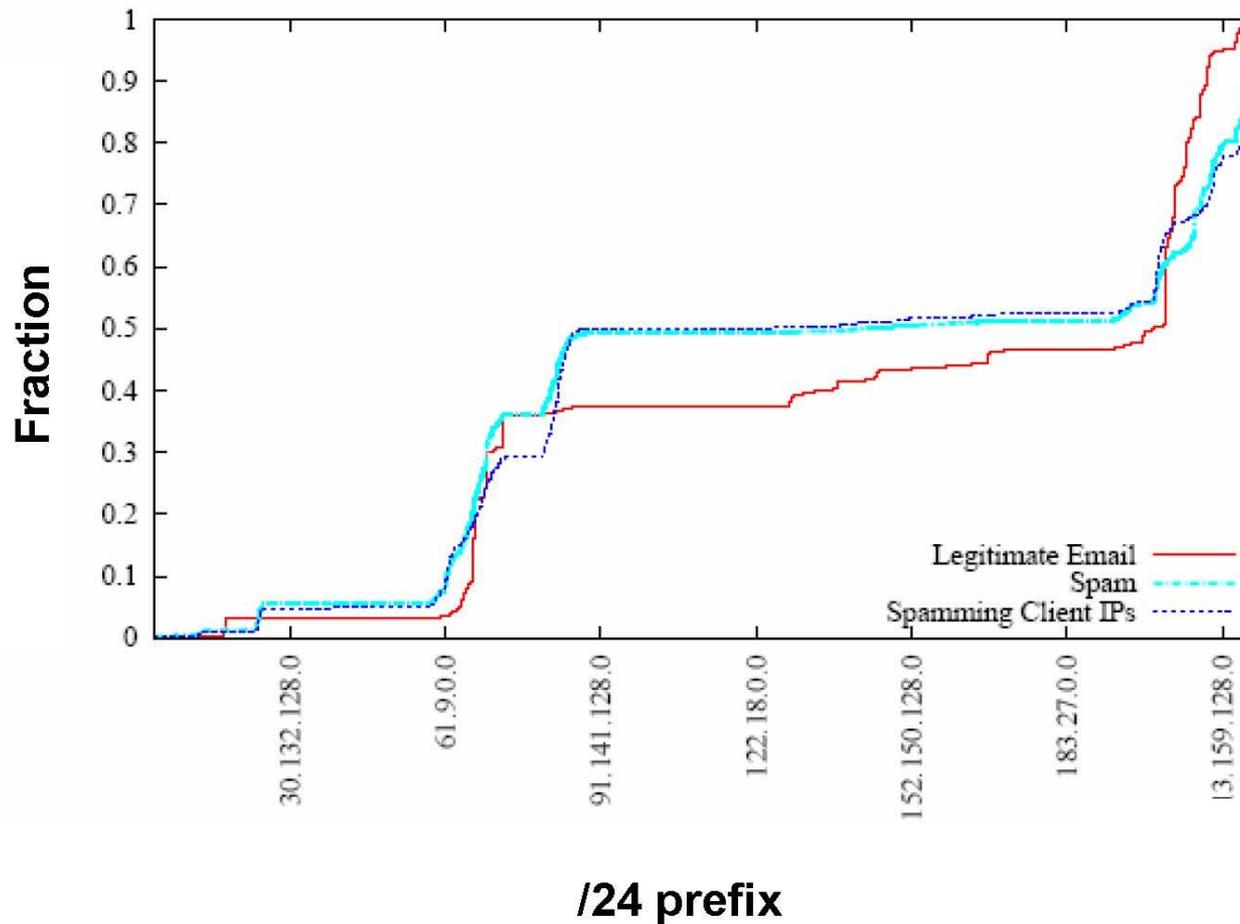


The image shows a screenshot of a web browser displaying a BBC News article. The browser's address bar shows "BBC News - McAfee tackles". The article's main heading is "McAfee mal". Below the heading, the text reads: "A leading flaw in custom hijacked". The article content is partially obscured by a large text box containing three bullet points. The first two bullet points are in black text, and the third is in pink text.

- ◆ McAfee antivirus includes "Rumor Service" for delivering updates to computers without direct Internet access
- ◆ This service has been hacked, turned into an open proxy, and used to send tons of spam
- ◆ "As an ultimate insult, even McAfee, whose software is at the root of our problems, now rate our email IP as 'High Risk': we can't email them as they have blacklisted us!"

# Distribution of Spam Sources

[Ramachandran, Feamster]



# Distribution Across Domains

[Ramachandran, Feamster]

<i>AS Number</i>	<i># Spam</i>	<i>AS Name</i>	<i>Primary Country</i>
766	580559	Korean Internet Exchange	Korea
4134	560765	China Telecom	China
1239	437660	Sprint	United States
4837	236434	China Network Communications	China
9318	225830	Hanaro Telecom	Japan
32311	198185	JKS Media, LLC	United States
5617	181270	Polish Telecom	Poland
6478	152671	AT&T WorldNet Services	United States
19262	142237	Verizon Global Networks	United States
8075	107056	Microsoft	United States
7132	99585	SBC Internet Services	United States
6517	94600	Yipes Communications, Inc.	United States
31797	89698	GalaxyVisions	United States
12322	87340	PROXAD AS for Proxad ISP	France
3356	87042	Level 3 Communications, LLC	United States
22909	86150	Comcast Cable Corporation	United States
8151	81721	UniNet S.A. de C.V.	Mexico
3320	79987	Deutsche Telekom AG	Germany
7018	74320	AT&T WorldNet Services	United States
4814	74266	China Telecom	China

# Where Does Spam Come From?

 [Ramachandran, Feamster]

- ◆ IP addresses of spam sources are widely distributed across the Internet
  - In tracking experiments, most IP addresses appear once or twice; 60-80% not reachable by traceroute
- ◆ Vast majority of spam originates from a small fraction of IP address space
  - Same fraction that most legitimate email comes from
- ◆ Spammers exploit routing infrastructure
  - Create short-lived connection to mail relay, then disappear
  - Hijack a large chunk of unallocated “dark” space

# CAN-SPAM Act (passed in 2003)

<http://www.ftc.gov/spam>

## ◆ Legal solution to the problem

- Bans email harvesting, misleading header information, deceptive subject lines, use of proxies
- Requires opt-out and identification of advertising
- Imposes penalties (up to \$11K per violation)

## ◆ FTC report on effectiveness (Dec 2005)

- 50 cases pursued in the US
- No impact on spam originating outside the US (60%)
- Open relays hosted on botnets make it difficult to collect evidence

# Bobax Worm

---

- ◆ Infects machines with high bandwidth
  - Exploits Windows LSASS buffer overflow vulnerability
- ◆ Slow spreading (and thus hard to detect)
  - On manual command from operator, randomly scans for vulnerable machines
- ◆ Installs hacked open relay on infected zombies
  - Once the spam zombie added to blacklist, spread to another machine
  - Interesting detection technique: look for botmaster's DNS queries (trying to determine who is blacklisted)

# Major Spambots in 2008

<http://www.marshal.com/trace/traceitem.asp?article=615>

SRIZBI	43.7%
RUSTOCK	17.5%
MEGA-D	16.5%
HACKTOOL	6.8%
PUSHDO	5.1%
STORM	1.4%
OTHER SOURCES	9.0%



# McColo



- ◆ McColo was a San Jose-based hosting provider
- ◆ Hosted command-and-control servers of the biggest spam botnets
  - Rustock, Srizbi, Mega-D, Pushdo/Cutwail, others
- ◆ Disconnected by upstream providers on Nov 11, 2008  $\Rightarrow$  75% reduction of spam worldwide
- ◆ Resurrected for 12 hours on Nov 20, 2008
  - Through a backup connection (soon terminated)
  - During this time, 15MB/sec of traffic to Russia – botmasters getting data to regain control of botnets

# Srizbi



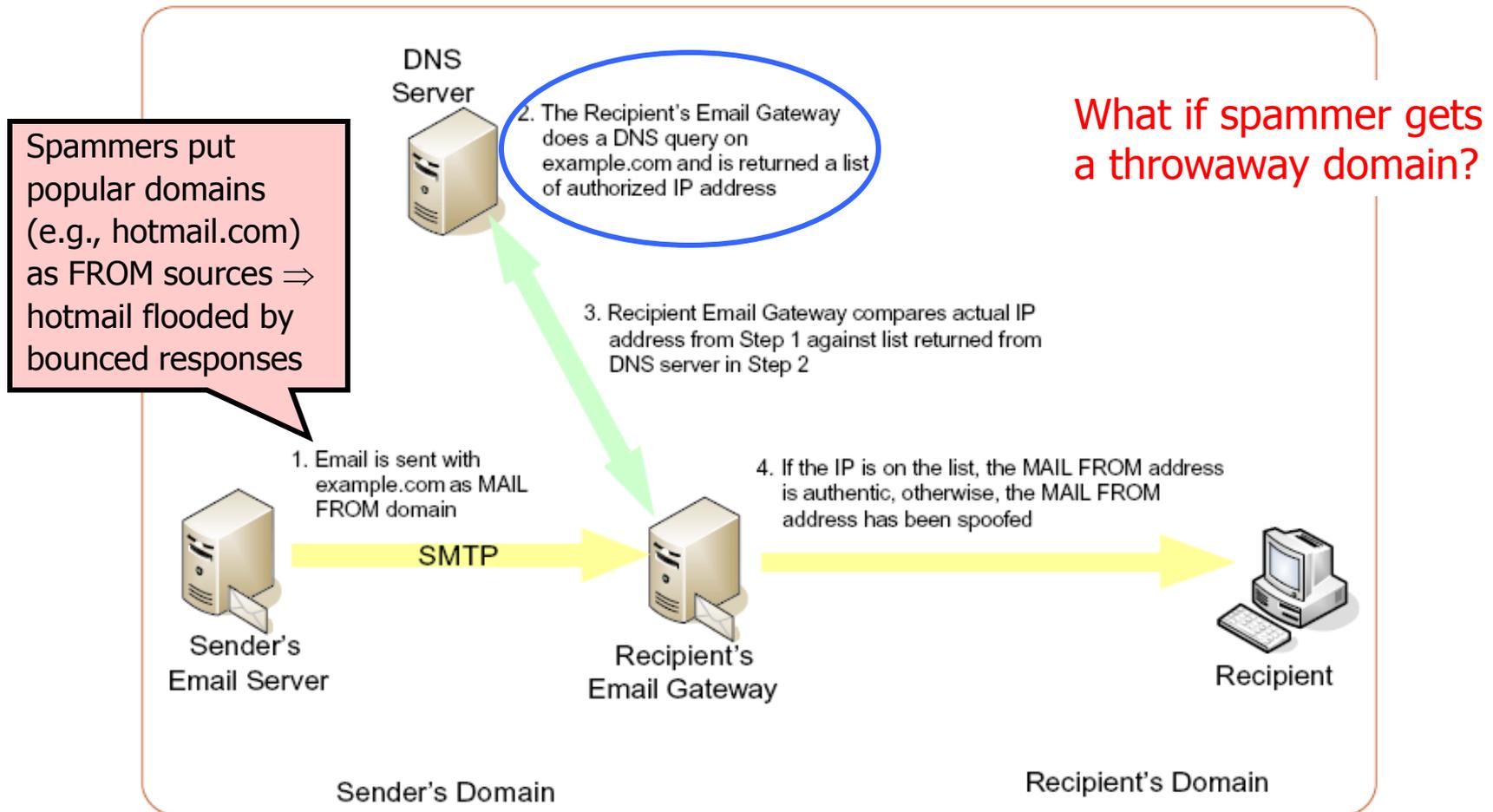
- ◆ Rootkit + sophisticated spam mailer
- ◆ 500K zombies, 60 billion spam messages daily
  - More than half of all spam worldwide
- ◆ After McColo takedown, fail-safe code inside bots started generating names of backup domains
  - [ypouaypu.com](http://ypouaypu.com), [oryitugf.com](http://oryitugf.com), [prpoqpsy.com](http://prpoqpsy.com) ...
  - Botmasters regained control by registering these domains (through a Russian registrar) and hosting new C&C servers in Estonia – shut down later

# Rustock

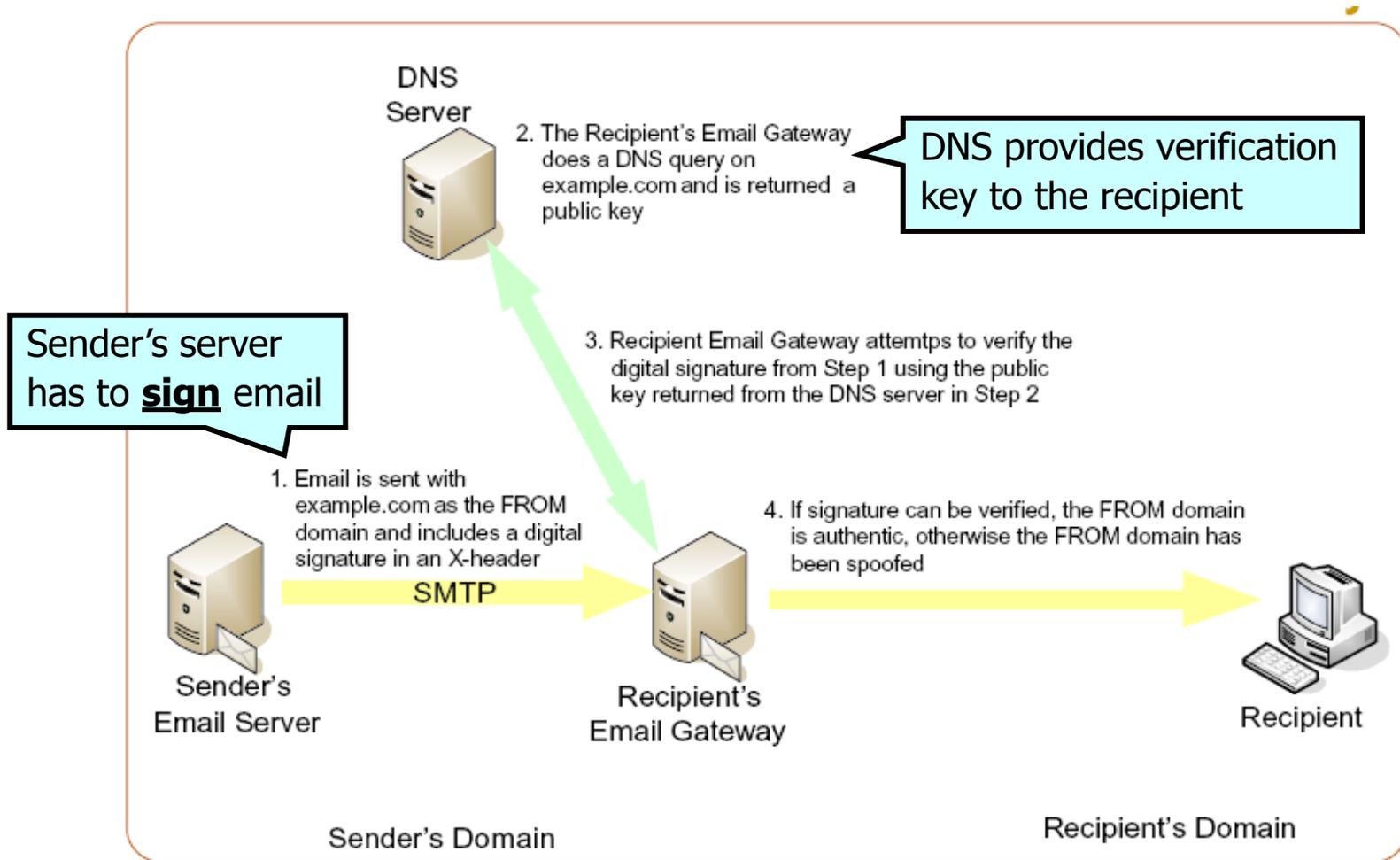
---

- ◆ Responsible for 40% of all spam in 2010
- ◆ Between 1 and 2.5 million infected computers
  - Up to 240,000 messages daily from each host
- ◆ Based on a fairly elaborate rookit
- ◆ C&C servers taken down on March 16, 2011
  - Investigation by Microsoft, Pfizer, FireEye, and security researchers from the University of Washington
  - “John Doe” lawsuit against botnet operators
  - Coordinated seizure of C&C servers in the US
  - 33% decline in spam afterwards

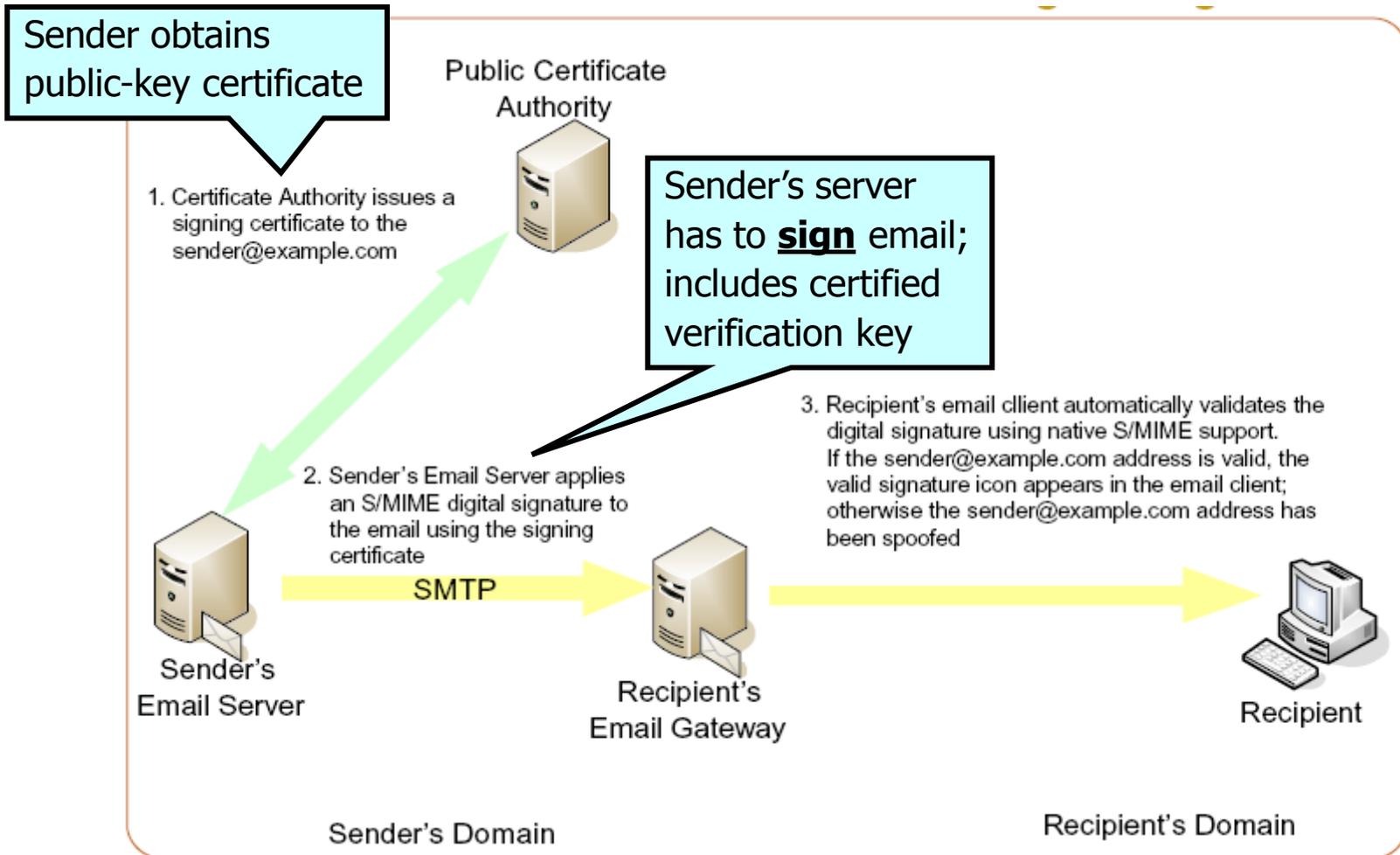
# SPF (Sender Policy Framework)



# Domain Keys (DKIM)



# S/MIME



# Graylists

---

- ◆ Recipient's mail server records (sender email, recipient email, peer IP) triple in its database
  - Each triple kept for 3 days (configuration parameter)
- ◆ First time (triple not in DB): 421 reply "Busy"
  - Records triple in the database
- ◆ Second time (after 5 minutes): let email pass
- ◆ What is this defense based on?
- ◆ Easily spoofable, but works against current spammers

# Puzzles and CAPTCHAs

---

- ◆ Generic defenses against spam and DoS
- ◆ Basic idea: sender must solve a “puzzle” before his email or connection request is accepted
  - Takes effort to solve, but solution easy to check
  - Sender has to “pay” in computation time
    - Example (Hashcash): find collision in a short hash
- ◆ CAPTCHA: prove that the sender is human
  - Solve a “reverse Turing test”
  - Only in application layer (e.g., Web)
- ◆ Both are difficult to deploy (why?)



SMALL

# Worst CAPTCHA Ever?

<http://depressedprogrammer.wordpress.com/2008/04/20/worst-captcha-ever/>

No premium user. Please enter all letters having a  below.



Four letters with a  :

[Download via Cogent](#)

# Gone in Six Seconds

---

- ◆ Spammers like to create a large number of Gmail and Hotmail accounts, use them to send spam
  - DKIM and SPF don't help (why?)
  - But CAPTCHAs do (how?)
- ◆ Botnet = massive distributed computing platform, so spammers can use them to solve CAPTCHAs
  - 2008: 6 seconds per CAPTCHA, 10-15% success
  - After Microsoft upgraded their CAPTCHA system: 20 seconds per CAPTCHA, 12-20% success

# CAPTCHA Cracking in Progress

<http://securitylabs.websense.com/content/Blogs/3063.aspx>

92.168.197.213:1131	208.66.119.42	TCP	16	20:41:26	20:38:19.921	54	0	
92.168.197.213:1133	119.42	TCP:http	9	20:38:55	20:38:19.921	166	112	GET /gen_name.cgi HTTP/1.1 User-Agent: M
92.168.197.213:1135	65.54.119.42	TCP	76	20:38:40	20:38:19.921	54	0	
92.168.197.213:1136	119.42	TCP:http	9	20:39:00	20:38:38.109	1514	1460	HTTP/1.1 200 OK Date: Mon, 07 Apr 2008 10
92.168.197.213:1138	119.42	TCP:http	9	20:39:06	20:38:38.109	801	747	88793Rosie
92.168.197.213:1139	65.54.119.42	TCP:http	23	20:38:51				
92.168.197.213:1140	65.54.119.42	TCP	30	20:38:52				

```
Server: Apache/2.2.3 (Debian) PHP/5.2.0-8+etch7
Transfer-Encoding: chunked
Content-Type: text/plain; charset=WINDOWS-1251
```

```
7e4
Susan:Wha
Kate:Stro
Jeannine:
Brandy:Lea
William:Ma
Denver:Ca
Norma:Moy
Lane:Cran
Sasha:Van
Efren:Stro
Dino:Voge
Tonya:Cal
Ophelia:E
```

# Using Humans to Solve CAPTCHAs

<http://old.post-gazette.com/pg/03278/228349.stm>

“But at least one potential spammer managed to crack the CAPTCHA test. Someone designed a software robot that would fill out a registration form and, **when confronted with a CAPTCHA test, would post it on a free porn site.** Visitors to the porn site would be asked to complete the test before they could view more pornography, and the software robot would use their answer to complete the e-mail registration.”

# Solve CAPTCHAs for Fun and Profit

- ◆ Third-world “data entry specialists” will solve CAPTCHAs for 60 cents an hour

The screenshot shows a project listing on the GetAFreelancer.com website. The project is titled "Data Entry - Solve CAPTCHA's" and is currently closed. The project details include a budget of \$30-100, a creation date of 08/30/2006, and a bidding end date of 10/02/2006. The project creator is identified as "afmatt" with a 5-star rating and 25 reviews. The description states that the project involves solving CAPTCHA's for a 50-hour week. The job type is listed as "Data Entry" and "Data Processing". The database and operating system are both listed as "None". The bid count is 58, and the average bid is \$57. The page also features a "Related project" section for "Virtual Assistant - Long Term" and a "FREE trial project" offer. At the bottom, there is a table of service providers, with "Gudoc" being the only one listed, offering a bid of \$100 for a 4-day delivery within a bid time of 08-31-2006 13:50. The provider has a 5-star rating and 132 reviews.

**GET a FREELANCER.com** Sign Up | Log in | Top Rated Users | Browse projects | Post Project | RSS feeds | Articles

Find projects

### Data Entry - Solve CAPTCHA's

Data Entry - Solve CAPTCHA's is project number 85170 posted at [GetAFreelancer.com](#). [Click here](#) to post your own project.

**Closed**  
(Cancelled by the Site Administrator)

Status:	Closed
Budget:	\$30-100
Created:	08/30/2006 at 13:34 EDT
Bidding Ends:	10/02/2006 at 13:34 EDT
Project Creator:	<a href="#">afmatt</a> <input type="button" value="View PM"/> <input type="button" value="Post PM"/> Buyer Rating: ★★★★★★ (25 reviews)
Description:	I will provide a piece of software that will display CAPTCHA's - you will provide the service of solving them for one 50 hour week. Post your price and internet connection type. <input type="button" value="Report Violation"/>
Job Type:	<ul style="list-style-type: none"><li>• Data Entry</li><li>• Data Processing</li></ul>
Database:	(None)
Operating system:	(None)
Bid count:	58
Average bid:	\$ 57

Messages Posted: 12

If you are the project creator or one of the bidders  for more options

Service Providers	PMB	Bid	Delivery Within	Time of Bid	Provider Rating
<a href="#">Gudoc</a> <input type="button" value="View PM"/> <input type="button" value="Post PM"/>		\$ 100	4 days	08-31-2006 13:50	★★★★★★ (132 reviews)

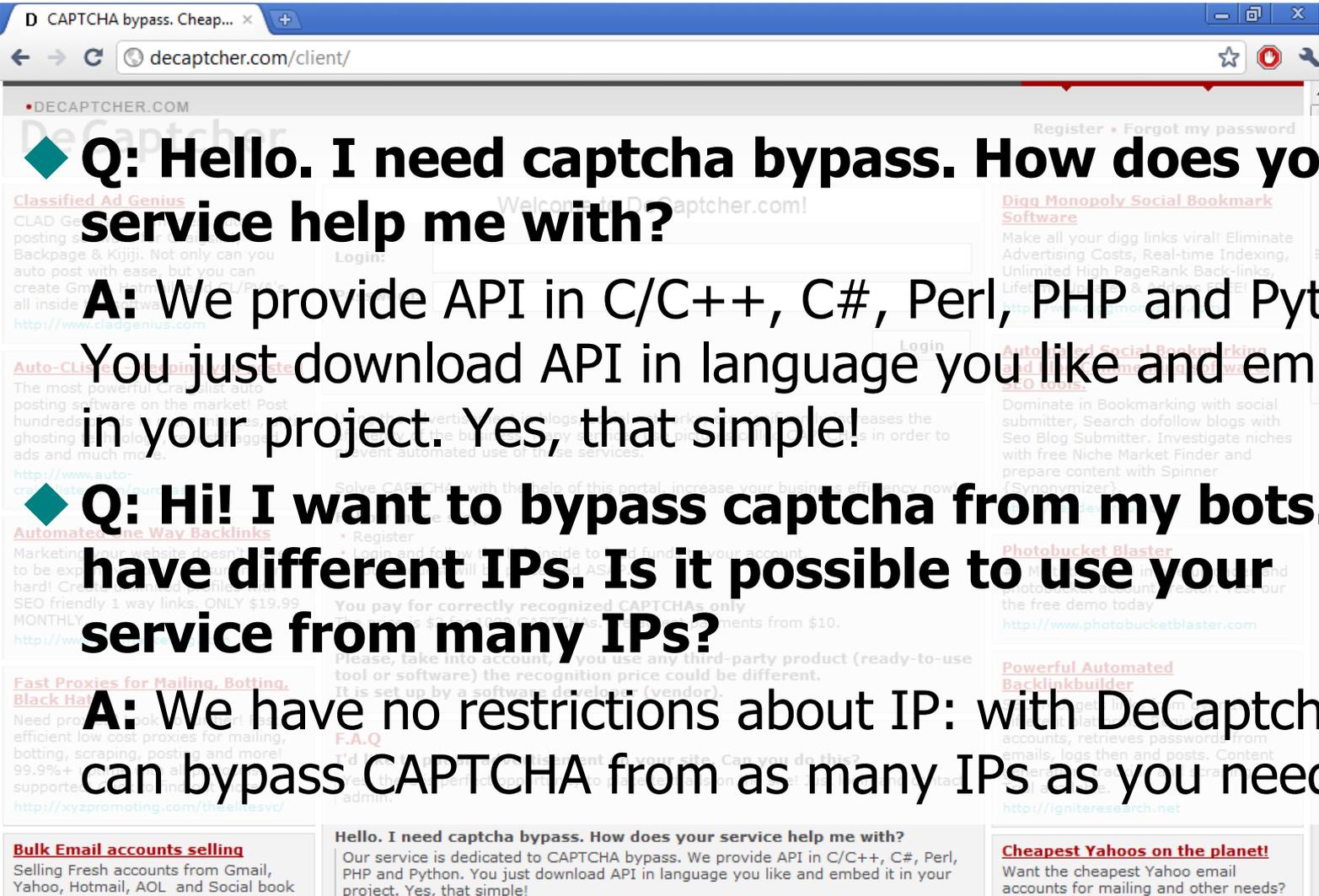
Ready to start immediately on your project, as we are the experts in data entry. Waiting for your selection. Thanks!

# CAPTCHA-Solving Services

[Motoyama et al. "Understanding CAPTCHA-Solving Services in an Economic Context" ]

<b>Service</b>	<b>\$/1K Bulk</b>	<b>Dates (2009–2010)</b>	<b>Requests</b>	<b>Responses</b>
Antigate (AG)	\$1.00	Oct 06 – Feb 01 (118 days)	28,210	27,726 (98.28%)
BeatCaptchas (BC)	\$6.00	Sep 21 – Feb 01 (133 days)	28,303	25,708 (90.83%)
BypassCaptcha (BY)	\$6.50	Sep 23 – Feb 01 (131 days)	28,117	27,729 (98.62%)
CaptchaBot (CB)	\$1.00	Oct 06 – Feb 01 (118 days)	28,187	22,677 (80.45%)
CaptchaBypass (CP)	\$5.00	Sep 23 – Dec 23 (91 days)	17,739	15,869 (89.46%)
CaptchaGateway (CG)	\$6.60	Oct 21 – Nov 03 (13 days)	1,803	1,715 (95.12%)
DeCaptcher (DC)	\$2.00	Sep 21 – Feb 01 (133 days)	28,284	24,411 (86.31%)
ImageToText (IT)	\$20.00	Oct 06 – Feb 01 (118 days)	14,321	13,246 (92.49%)

# DeCaptcher.com (Now Defunct)



The screenshot shows a web browser window with the address bar displaying "decaptcher.com/client/". The website content includes a navigation bar with "Register" and "Forgot my password" links. The main area features several advertisements and service listings, such as "Classified Ad Genius", "Auto-Clicker", "Automated One Way Backlinks", "Fast Proxies for Mailing, Botting, Black Hat", "Bulk Email accounts selling", "Digg Monopoly Social Bookmark Software", "Automated Social Bookmarking", "Photobucket Blaster", "Powerful Automated Backlinkbuilder", and "Cheapest Yahoos on the planet!".

**◆ Q: Hello. I need captcha bypass. How does your service help me with?**

**A: We provide API in C/C++, C#, Perl, PHP and Python. You just download API in language you like and embed it in your project. Yes, that simple!**

**◆ Q: Hi! I want to bypass captcha from my bots. Bots have different IPs. Is it possible to use your service from many IPs?**

**A: We have no restrictions about IP: with DeCaptcher you can bypass CAPTCHA from as many IPs as you need.**

**Hello. I need captcha bypass. How does your service help me with?**  
Our service is dedicated to CAPTCHA bypass. We provide API in C/C++, C#, Perl, PHP and Python. You just download API in language you like and embed it in your project. Yes, that simple!

**Cheapest Yahoos on the planet!**  
Want the cheapest Yahoo email accounts for mailing and other needs? As low as \$1.50/10! Unlimited stock.

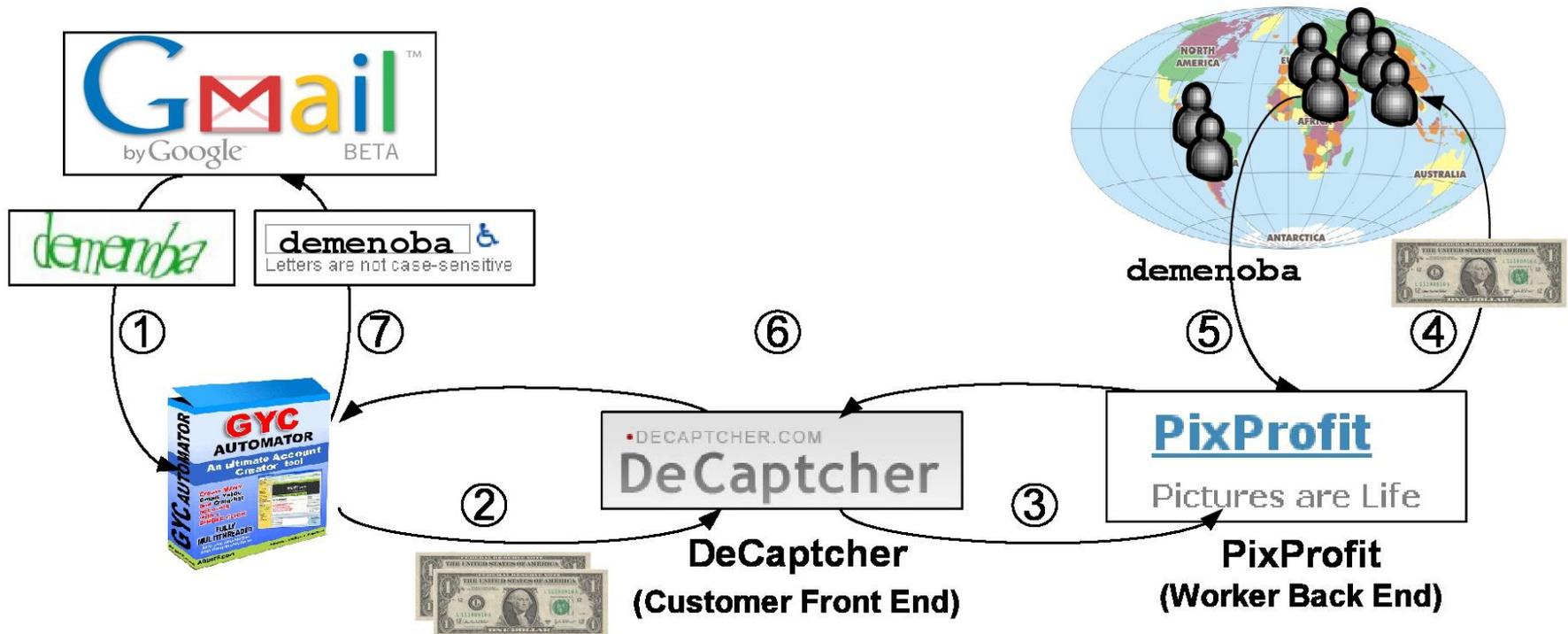
# India's CAPTCHA Solving Economy

<http://www.zdnet.com/blog/security/inside-indias-captcha-solving-economy/1835>

- ◆ 24/7 support still like. We have 30 pc 90 worker & we have 300 captcha team. Your any captcha project we done quickly. We have high experience captcha worker
- ◆ Hello Sir, I will kindly introduce myself.. This is shivakumar.. we have a team to type capcthas 24/7 and we can type more than 200k captchas per day
- ◆ WE ARE PROFESSIONAL CAPCHA ENTRY OPEATORS AND WE CAN DO EVEN 25000 ENTRIES PER DAY AS MY COMPANY IS A 25 SEATER FIRM SPEALISED IN DATA ENTRY
- ◆ We having more then 10 teams, we are oping 24/7 data entry works and delivering 700k/day captchas d !!
- ◆ My rate \$4.00 per 1k My team can work 24/7. They are jobless now

# CAPTCHA-Solving Economy

[Motoyama et al. "Understanding CAPTCHA-Solving Services in an Economic Context" ]



# Support Tools

<http://www.zdnet.com/blog/security/inside-indias-captcha-solving-economy/1835>

Main menu

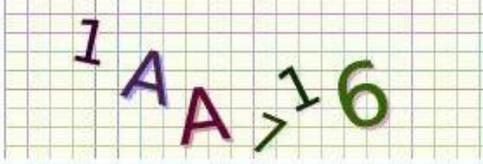
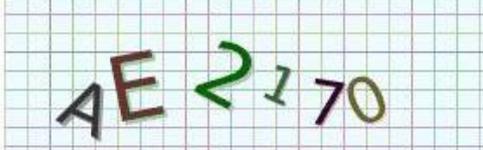
- Home
- Contact Us

---

- Help
- Work
- Practice
- Qualify to Work
- Tests made**
- Statistics
- Profile
- Logout

Start time	Items completed / total	Success Rate (%)	Items OK	Items Failed	Duration	Items per hour	
2008-08-29 12:26:30	4 / 5	%	3	1	00:00:00		<b>Failed</b>
2008-08-29 12:25:48	0 / 5	%	0	0	00:00:00		Failed

**You have failed to qualify.**  
 Minimum required average rating: 75%

CAPTCHA	Text	Your solution	Result
	BKZRLZ		
	DPHYXQ		
	AX5EW	ax5ewa	Length mismatch: 6 (should be 5)
	AJVBA	ajvba	OK
	1aa716	1aa716	OK
	ae2170	ae2170	OK

# Where Do CAPTCHA Solvers Live?

[Motoyama et al. "Understanding CAPTCHA-Solving Services in an Economic Context" ]

Language	Example	AG	BC	BY	CB	DC	IT	All
English	one two three	51.1	37.6	4.76	40.6	39.0	62.0	39.2
Chinese (Simp.)	一 二 三	48.4	31.0	0.00	68.9	26.9	35.8	35.2
Chinese (Trad.)	一 二 三	52.9	24.4	0.00	63.8	30.2	33.0	34.1
Spanish	uno dos tres	1.81	13.8	0.00	2.90	7.78	56.8	13.9
Italian	uno due tre	3.65	8.45	0.00	4.65	5.44	57.1	13.2
Tagalog	isá dalawá tatlo	0.00	5.79	0.00	0.00	7.84	57.2	11.8
Portuguese	um dois três	3.15	10.1	0.00	1.48	3.98	48.9	11.3
Russian	один два три	24.1	0.00	0.00	11.4	0.55	16.5	8.76
Tamil	ஒன்று இரண்டு மூன்று	2.26	21.1	3.26	0.74	12.1	5.36	7.47
Dutch	een twee drie	4.09	1.36	0.00	0.00	1.22	31.1	6.30
Hindi	एक दो तीन	10.5	5.38	2.47	1.52	6.30	9.49	5.94
German	eins zwei drei	3.62	0.72	0.00	1.46	0.58	29.1	5.91
Malay	satu dua tiga	0.00	1.42	0.00	0.00	0.55	29.4	5.23
Vietnamese	một hai ba	0.46	2.07	0.00	0.00	1.74	18.1	3.72
Korean	일 이 삼	0.00	0.00	0.00	0.00	0.00	20.2	3.37
Greek	ένα δύο τρία	0.45	0.00	0.00	0.00	0.00	15.5	2.65
Arabic	واحد اثنين ثلاثة	0.00	0.00	0.00	0.00	0.00	15.3	2.56
Bengali	এক দুই তিন	0.45	0.00	9.89	0.00	0.00	0.00	1.72
Kannada	ಒಂದು ಎರಡು ಮೂರು	0.91	0.00	0.00	0.00	0.55	6.14	1.26
Klingon	ᑭ ᑭᑭ ᑭᑭᑭ	0.00	0.00	0.00	0.00	0.00	1.12	0.19
Farsi	یک دو سه	0.45	0.00	0.00	0.00	0.00	0.00	0.08

Table 2: Percentage of responses from the services with correct answers for the language CAPTCHAs.