CSE227 – Graduate Computer Security

Spam and e-Crime

UC San Diego

Housekeeping

General course things to know

- **Final presentations happen in week 10!**
  - Keep making consistent progress, and feel free to meet with me if you have things to discuss and work on
- Most midpoint check-ins are done, just a few more today!
  - Instructive, really impressed with the quality of the work — keep it going
- **No class next Thursday!**

Housekeeping

General course things to know

- Final presentations will be final week of class
- Presentation Details
  - 10 minutes for presentation, 2 mins for questions (I will cut you off!)
  - Talk should include introduction to the problem, your research questions, your methodology (including data collection), and your results
  - All team members must speak approximately an even amount
  - **Number 1 Rule: Not boring!**

Today's lecture

Learning Objectives

- Learn about the spam and e-crime ecosystem and the spam value chain, understand the levers and knobs that go into creating a successful (or unsuccessful) spam campaign
- Discuss the click-trajectories paper
- Discuss the Re: CAPTCHAs paper

Preliminaries

What is email spam?

What is email spam?

Email spam: Unwanted, unsolicited email that are sent in bulk for, typically, fraudulent purposes

Why does spam exist?

Why does spam exist?

In short: Because it makes money! If it didn't, why would anyone do it? *How does spam typically make money*?

Why does spam exist?

In short: Because it makes money! If it didn't, why would anyone do it? *How does spam typically make money*?

Another chain – malware, phishing, or *affiliate marketing*

Modern Spam value Chain

Advertising, Click Support, Realization

- What is the advertising phase?

Modern Spam value Chain

Advertising, Click Support, Realization

- What is the advertising phase?
  - How do we get eyes and clicks to the spam? Is email the only way spam is sent?

Modern Spam value Chain

Advertising, Click Support, Realization

- What is the advertising phase?
  - How do we get eyes and clicks to the spam? Is email the only way spam is sent?
- What is the click support phase?

Modern Spam value Chain

Advertising, Click Support, Realization

- What is the advertising phase?
  - How do we get eyes and clicks to the spam? Is email the only way spam is sent?
- What is the click support phase?
  - All the infrastructure involved in helping spammers actually *run* websites (domains, redirection, name servers, web servers, affiliates)

Modern Spam value Chain

Advertising, Click Support, Realization

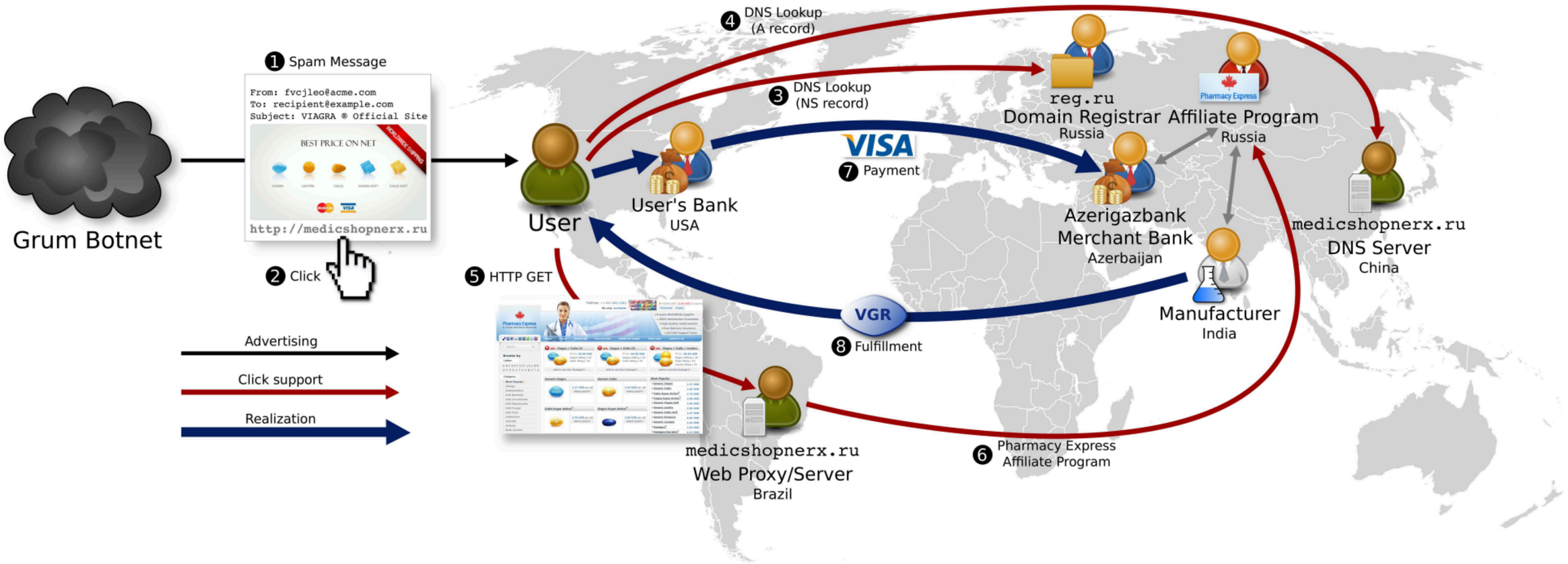
- What is the advertising phase?
  - How do we get eyes and clicks to the spam? Is email the only way spam is sent?
- What is the click support phase?
  - All the infrastructure involved in helping spammers actually *run* websites (domains, redirection, name servers, web servers, affiliates)
- What is the realization phase?

Modern Spam value Chain

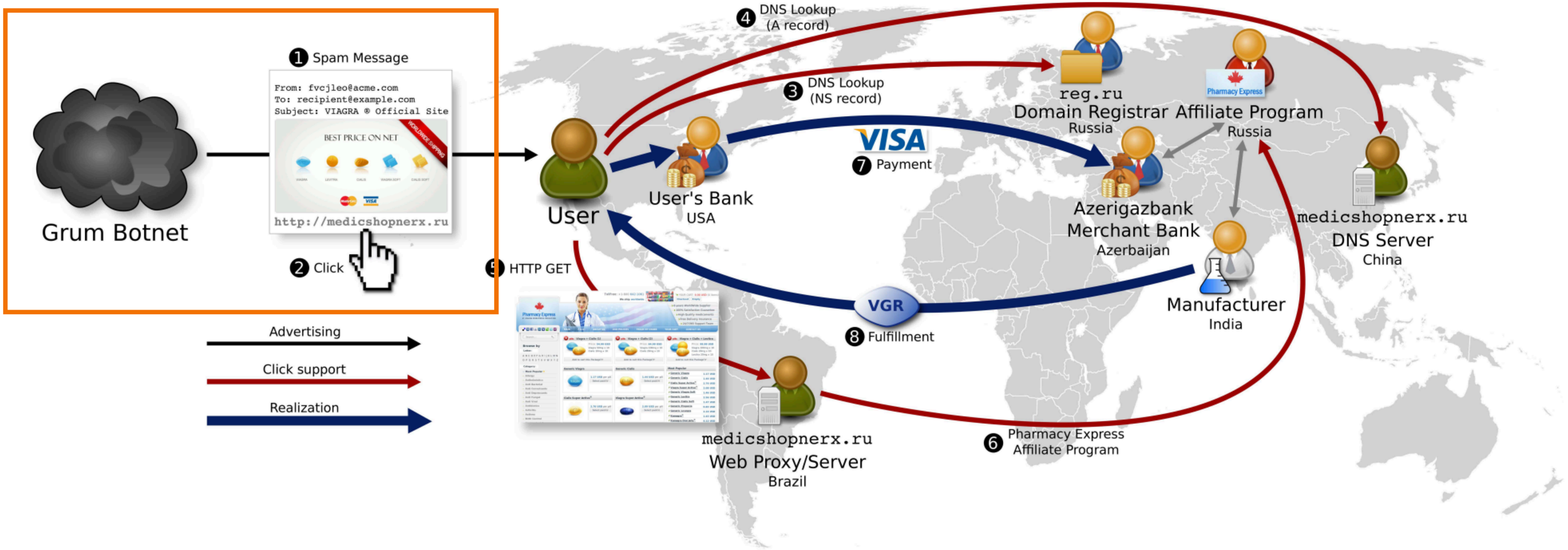
Advertising, Click Support, Realization

- What is the advertising phase?
  - How do we get eyes and clicks to the spam? Is email the only way spam is sent?
- What is the click support phase?
  - All the infrastructure involved in helping spammers actually *run* websites (domains, redirection, name servers, web servers, affiliates)
- What is the realization phase?
  - Everything needed to fulfill the “order,” i.e., whatever the spammer is selling or putting on your computer

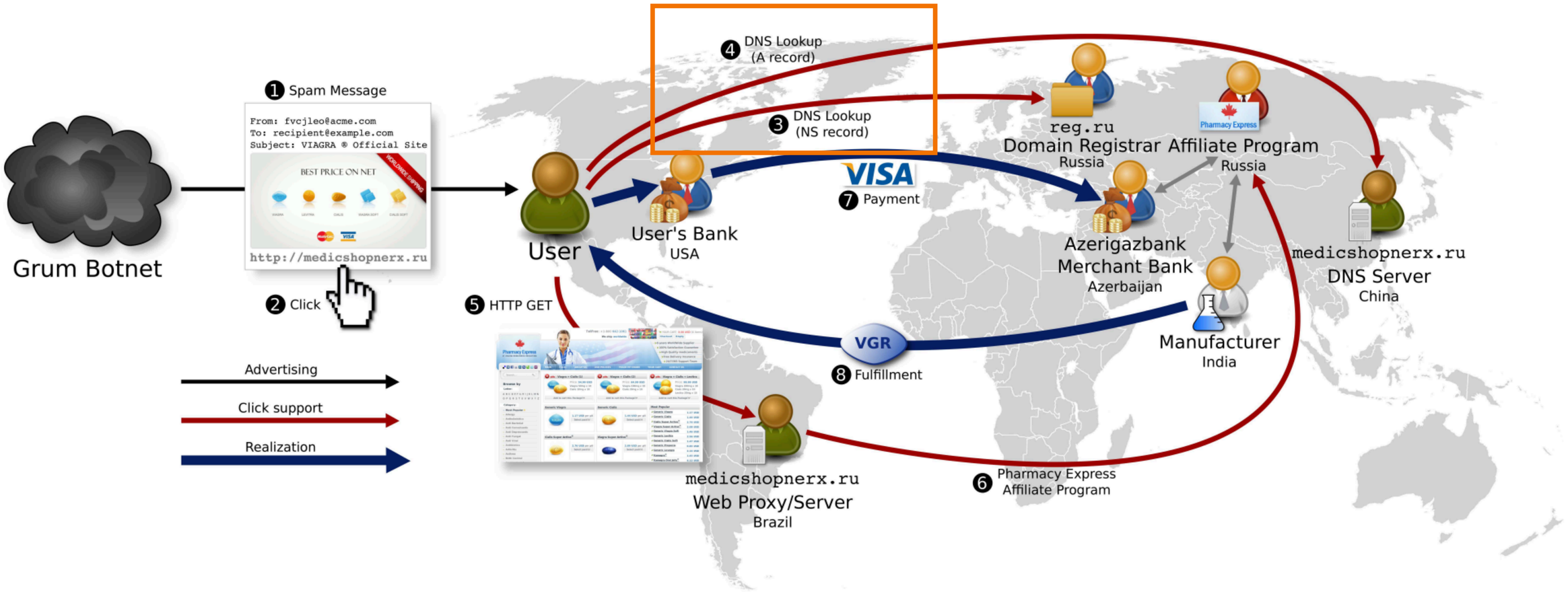
Pharmacy Express: An Example



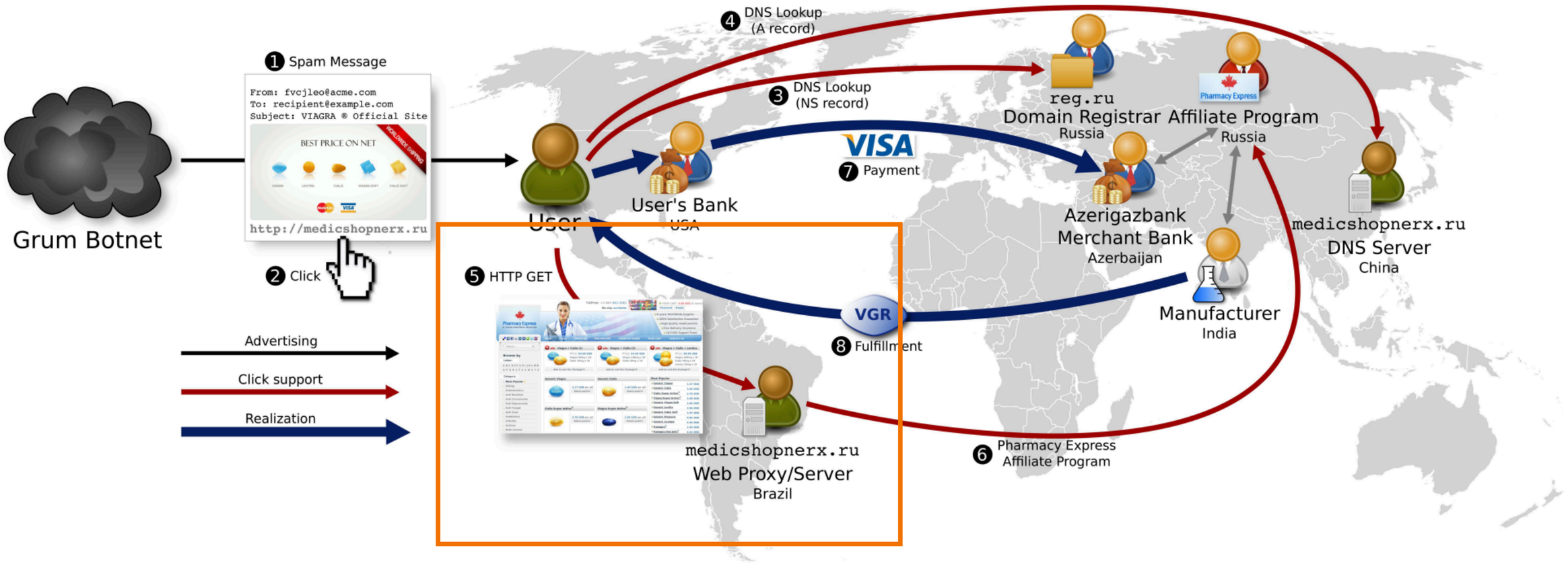
Pharmacy Express: An Example



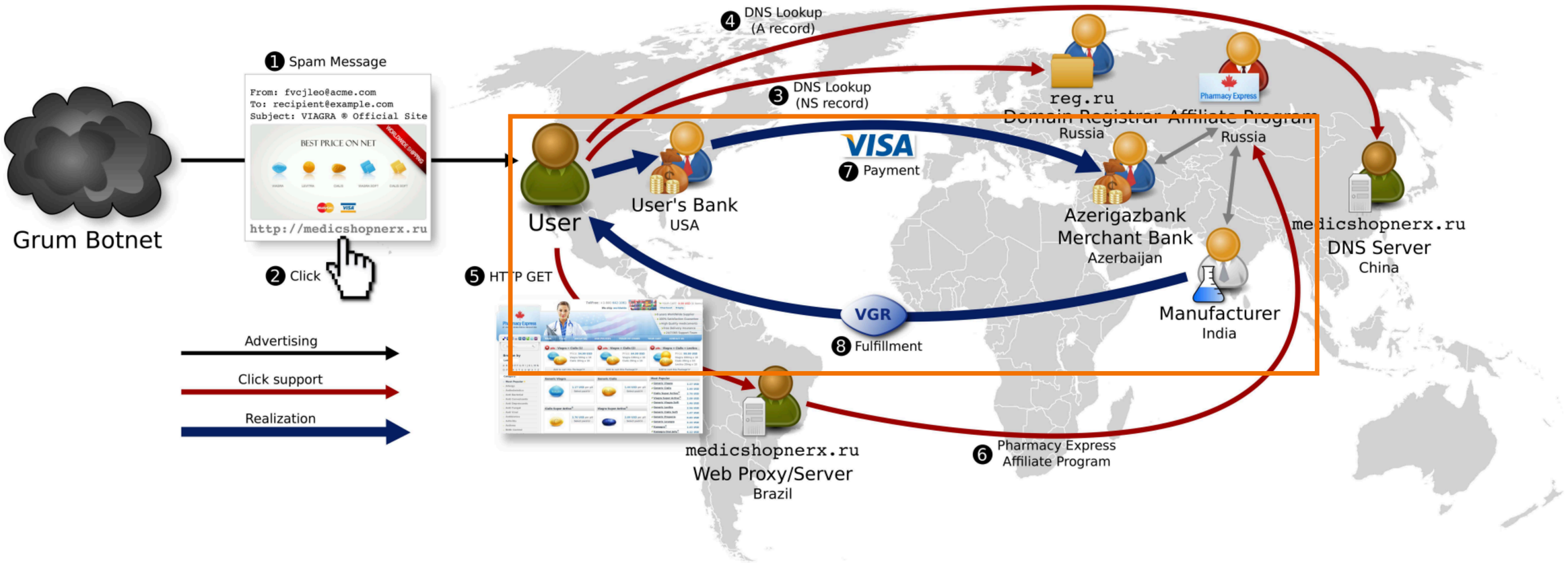
Pharmacy Express: An Example



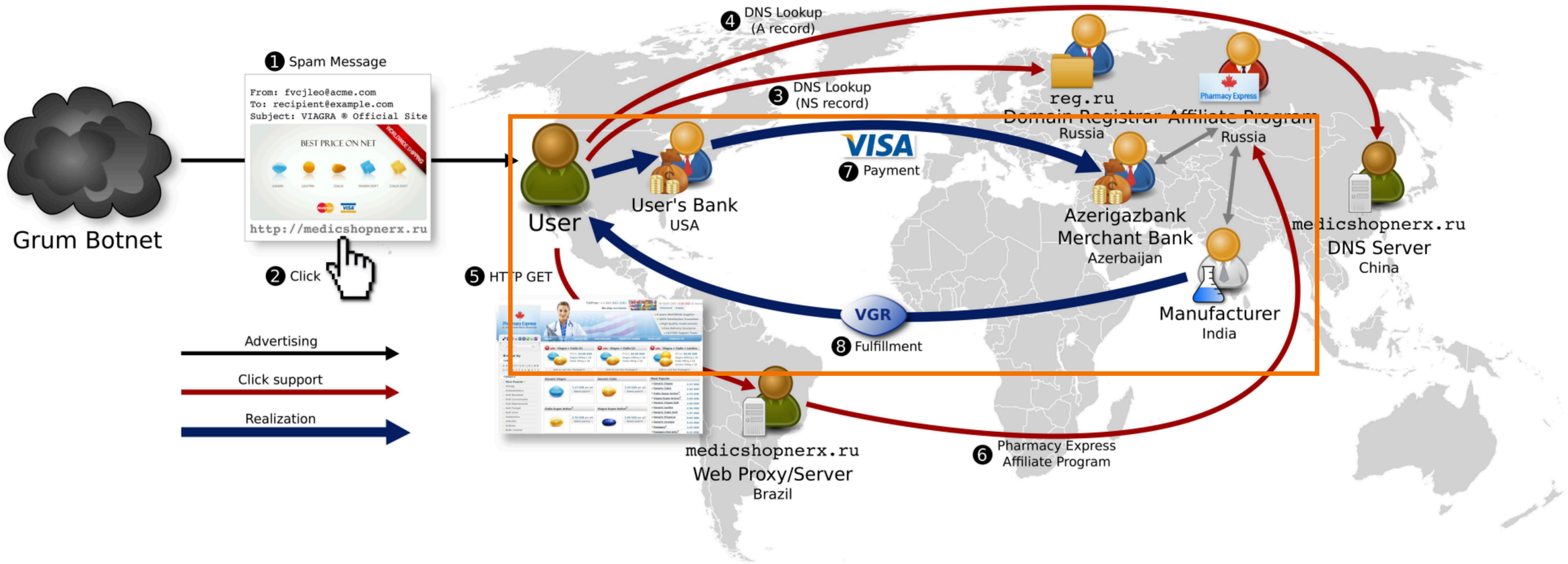
Pharmacy Express: An Example



Pharmacy Express: An Example



Pharmacy Express: An Example



Money changes hands at every part of the chain!

Discussion point: Why should we study spam?

Click Trajectories: End-to-End Analysis of the Spam Value Chain

A few words on these papers…

- These are both UCSD papers (seeing a theme in this class?)
- Geoff Voelker / Stefan Savage are the lead faculty authors
  - In a long line of work on spam, scams, e-crime, etc.
- Some of my favorite papers of all time
- Trajectories paper won Test of Time award at IEEE S&P 2022
  - Think about why.

What’s the goal of this paper?

What's the goal of this paper?

…Indeed, each distinct path through this chain—registrar, name server, hosting, affiliate program, payment processing, fulfillment—directly reflects an "entrepreneurial activity" by which the perpetrators muster capital investments and business relationships to create value. ***Today we lack insight into even the most basic characteristics of this activity. How many organizations are complicit in the spam ecosystem?*** Which points in their value chains do they share and which operate independently? How "wide" is the bottleneck at each stage of the value chain—do miscreants find alternatives plentiful and cheap, or scarce, requiring careful husbanding?

What’s the goal of this paper? (In my words)

Fall for all spam email lures and see what happens (mess around and find out)

Measuring Spam Value Chain

- What was the author's strategy in measuring the spam value chain?

Measuring Spam Value Chain

- What was the author's strategy in measuring the spam value chain?
  - Direct engagement with the chain through clicking, purchasing, and recording the chain

Measuring Spam Value Chain

- What was the author's strategy in measuring the spam value chain?
  - Direct engagement with the chain through clicking, purchasing, and recording the chain
- How else might you try and measure the spam value chain *without* direct engagement?

Measuring Spam Value Chain

- What is the “feed collection” step?
- What is the URL extraction step?
- What is the DNS & Web Crawling step?
  - What is fast flux DNS?
- What is the Content Clustering step?
- What is the Content Tagging step?
- What is the Selective Purchasing step?

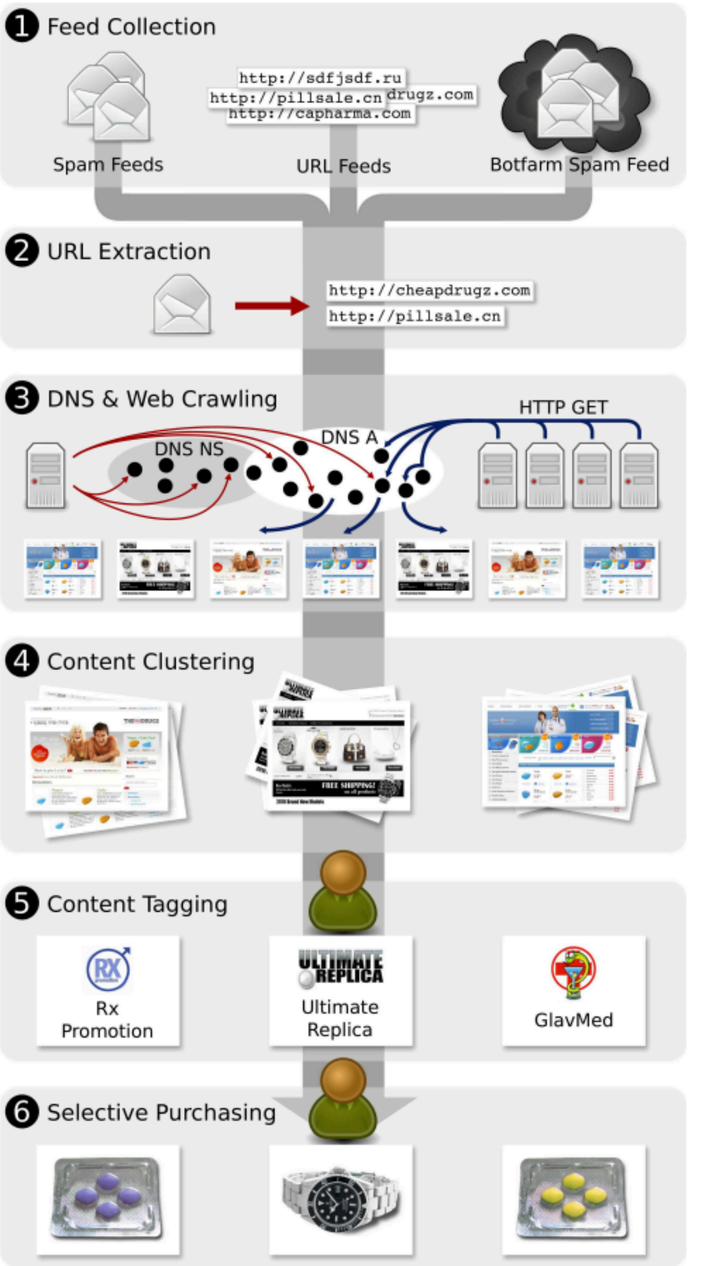


Figure 2: Our data collection and processing workflow.

Purchasing stuff

- Why did the authors decide to purchase anything at all?

Purchasing stuff

- Why did the authors decide to purchase anything at all?
  - Insight into the **realization** phase
    - Payment info (via relationship w/ card issues)
      - **Bank Identification Number (BIN) of acquiring bank**
    - Fulfillment
      - Did you actually get anything?
      - Where was it shipped from?
      - What's in it?

What does a credit card authorization record look like?

Bank Identification Number (BIN) 448314

Merchant descriptor “Smart rt online”

Card Acceptor ID “8875236”

Merchant Category Code (MCC) 5912


GLOBAL
EXPRESS
MAIL
UNITED STATES POSTAL SERVICE
Arrival
For Exchange Office use only
AMC of Arrival
Dispatch Number
SPO 5349
CN
EMS
Addressee Copy
For Inbound EMS Items Only
Item
Number
EE248975418CN
Delivery
Scan as appropriate. Obtain recipient signature on
Form 3849, Delivery Receipt
Delivery Attempt
Time
Employee Signature
Mo.
Day
AM
PM
PS Form 5626X, October 2002
Deliver By 3:00 PM Today
EMS
速递服务专家


DECLARATION EN DOUANE
MEDICINE


20
30
40
50
SUBMARINER
SUPERLATIVE CHRONOMETER
OFFICIALLY CERTIFIED

600+ orders later...

Infrastructure Sharing

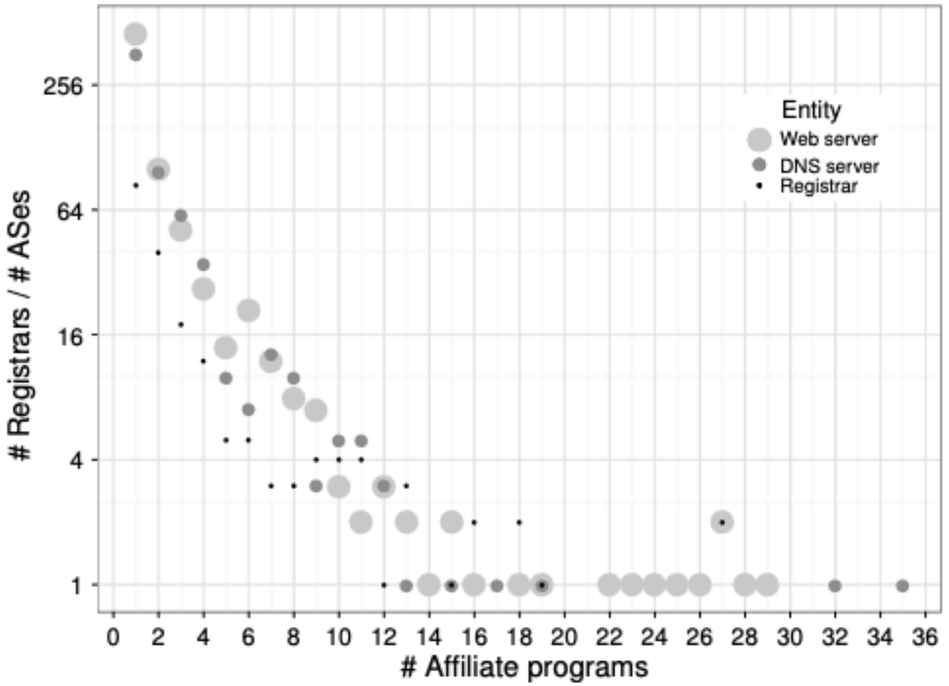


Figure 3: Sharing of network infrastructure among affiliate programs. Only a small number of registrars host domains for many affiliate programs, and similarly only a small number of ASes host name and Web servers for many programs. (Note y-axis is log scale.)

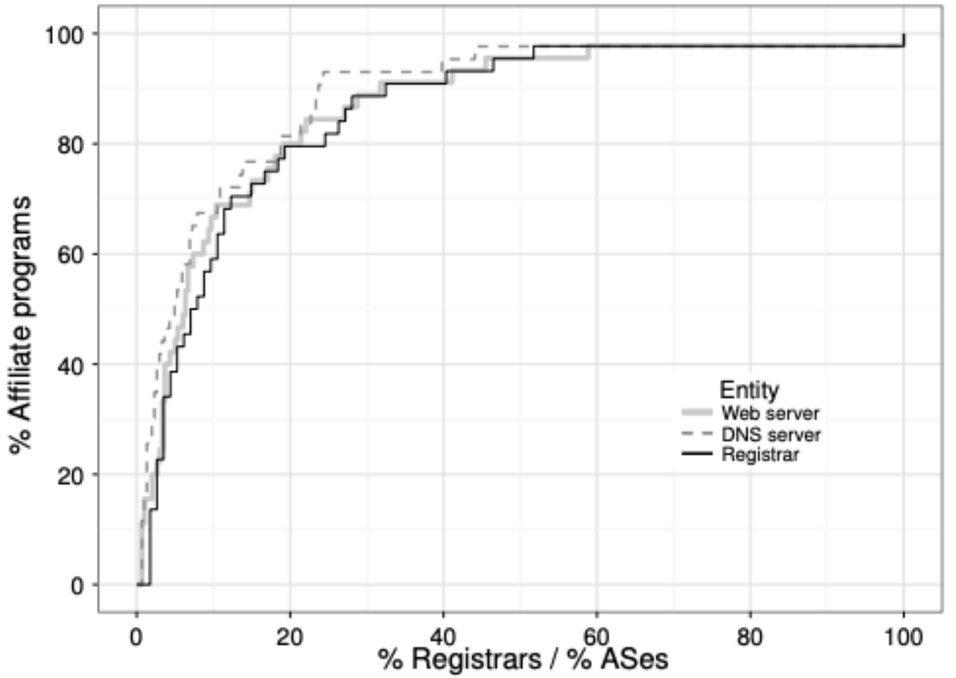


Figure 4: Distribution of infrastructure among affiliate programs. Only a small percentage of programs distribute their registered domain, name server, and Web server infrastructure among many registrars and ASes, respectively.

Where do products come from?

Supplier	*Item*	*Origin*	*Affiliate Programs*
Aracoma Drug	Orange bottle of tablets (pharma)	WV, USA	ClFr
Combitic Global Caplet Pvt. Ltd.	Blister-packed tablets (pharma)	Delhi, India	GlvMd
M.K. Choudhary	Blister-packed tablets (pharma)	Thane, India	OLPh
PPW	Blister-packed tablets (pharma)	Chennai, India	PhEx, Stmul, Trust, ClFr
K. Sekar	Blister-packed tablets (pharma)	Villupuram, India	WldPh
Rhine Inc.	Blister-packed tablets (pharma)	Thane, India	RxPrm, DrgRev
Supreme Suppliers	Blister-packed tablets (pharma)	Mumbai, India	Eva
Chen Hua	Small white plastic bottles (herbal)	Jiangmen, China	Stud
Etech Media Ltd	Novelty-sized supplement (herbal)	Christchurch, NZ	Staln
Herbal Health Fulfillment Warehouse	White plastic bottle (herbal)	MA, USA	Eva
MK Sales	White plastic bottle (herbal)	WA, USA	GlvMd
Riverton, Utah shipper	White plastic bottle (herbal)	UT, USA	DrMax, Grow
Guo Zhonglei	Foam-wrapped replica watch	Baoding, China	Dstn, UltRp

Table VI: List of product suppliers and associated affiliate programs and/or store brands.

Intervening in the chain

- Interventions in…
  - Click support
    - Registrar
    - DNS hosting
    - Web hosting
  - Realization
    - Payments

Payment Interventions seem most effective

- Takedown: acquiring (merchant) bank
  - Add pressure to drop bad customers
  - Challenges: bi-lateral process, potentially slow
- Blacklist: issuing (consumer) banks
  - US Visa / MC could just **refuse** transactions with “bad” acquirers
  - Would this work? Why or why not?

Meta Thoughts on this Paper

- What did we think about this paper? Is this kind of analysis useful, why or why not?
- What surprised you about this paper?
- Why did this paper win an award?
- What did this paper teach you about security, if anything at all?
- How do we reason about the ethics of this type of paper?

Break Time + Attendance

https://tinyurl.com/cse227-attend

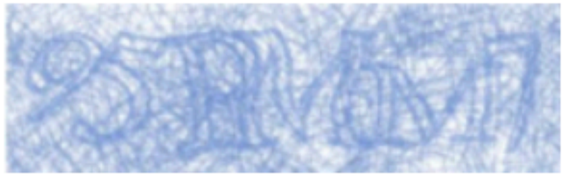
Codeword:
Click-Here!

Re: CAPTCHAs – Understanding CAPTCHA-Solving Services in an Economic Context

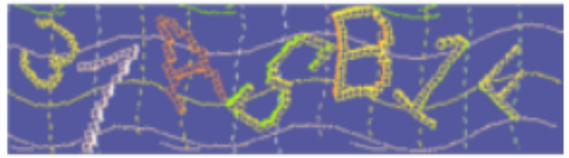
What's a CAPTCHA?

What's a CAPTCHA?

(a) Aol.



(b) mail.ru



(c) phpBB 3.0

(d) Simple Machines Forum

(e) Yahoo!

(f) youku

Why do we have CAPTCHAs?

Why do we have CAPTCHAs?

Limit the abilities of attackers to scale their activities via automated means.

Why do we have CAPTCHAs?

Limit the abilities of attackers to scale their activities via automated means.

How do we measure CAPTCHA abilities?

Measuring security is hard…

- How do you reason about how “well” a security mechanism works?
  - Metrics are a rabbit hole… what are the “units” of security?
- Big claim in this paper: Follow the money
  - Solving CAPTCHAs has acquired value; **retail market**
- Price of a solved CAPTCHA becomes a **cost** on a balance sheet of a scammer; *does increasing that cost make the scams go away?*

Scenario: Breaking CAPTCHAs

- Let’s say you’re an attacker and you want to break CAPTCHAs in practice for some purpose
  - How might you go about doing this?

Scenario: Breaking CAPTCHAs

- Let’s say you’re an attacker and you want to break CAPTCHAs in practice for some purpose
  - How might you go about doing this?
- Two strategies
  - Technical mechanisms
  - Human mechanisms

Technical Mechanisms

- Automated image processing
  - OCR + segmentation + heuristics
  - Almost as old as CAPTCHAs
  - Today: AI? Yeah.
- **Successful technically** (i.e., can be built… with errors)
- Paper claims this is **unsuccessful as a business**
  - Why?

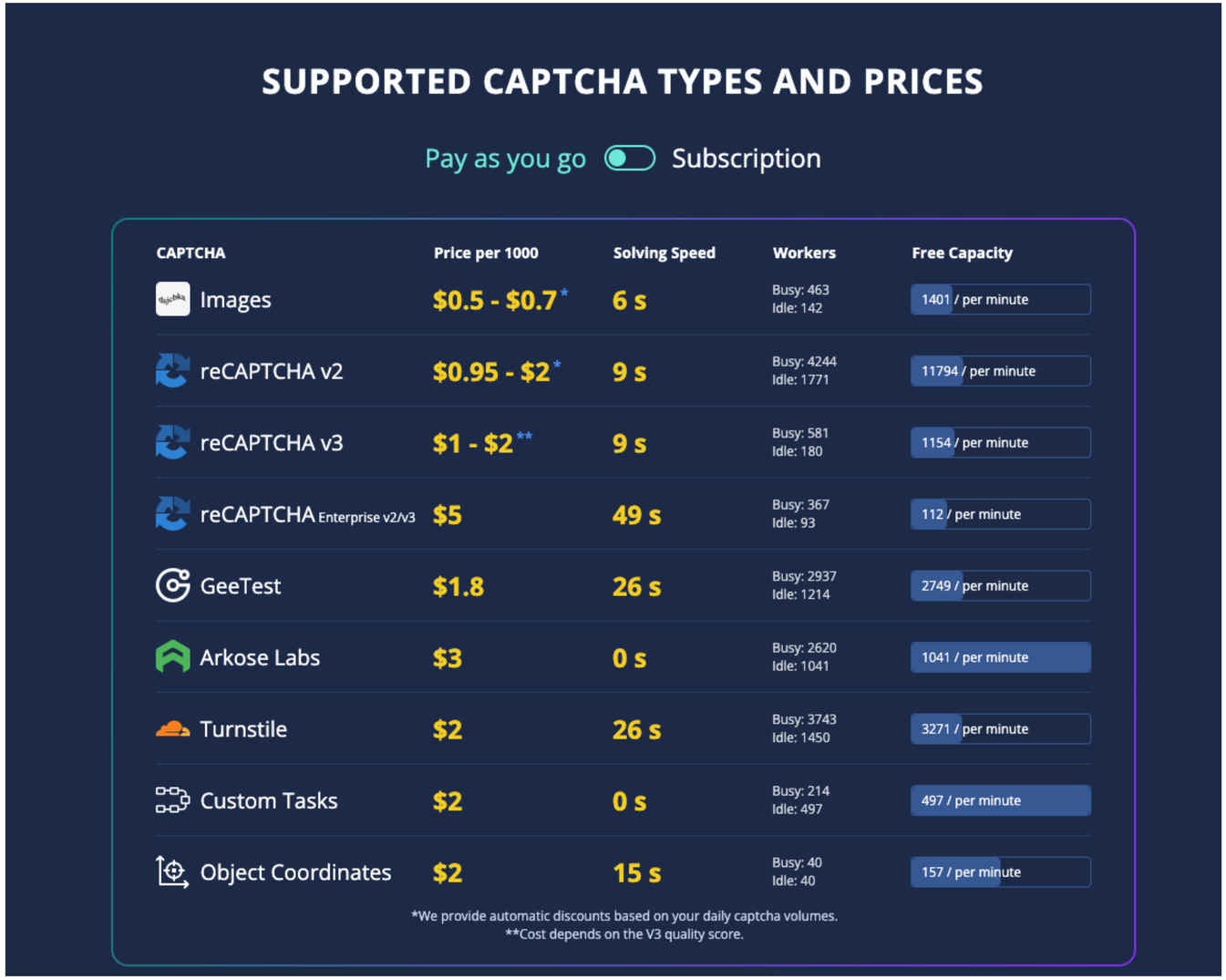
Why do software solvers fail economically?

- Traditional security arms race
  - Spam vs. anti-spam, virus vs. anti-virus, etc.
  - **Bad guy obfuscates**, and the **good guy recognizes**
  - In this case… CAPTCHAs are the obfuscation, and bad guys want to recognize… which is *hard*
- Lifetime of a software solver is low
  - CAPTCHAs evolve all the time, and new software takes time and is expensive
  - At what point does it make financial sense to do this?

Hypothetical in the paper

- Let’s say it costs $10,000 to develop a new CAPTCHA breaking system (with ~20% success rate)
- Must solve 20 million CAPTCHAs (100 million attempts at 20% accuracy) before being noticed and the CAPTCHA gets changes…
  - Just to break even at 50c for every 1000

…or you could just pay someone



History of human solving

- First starts to emerge in late 2006
  - How does this side-step the assumptions set out by CAPTCHA makers?
- **Retail service** market emerges
  - Market pressure (free market wins)
    - 2007: ~$10/1000 solves
    - 2008: ~$1.5/1000 solves
    - 2009: ~$1/1000 solves
    - 2010: ~$0.75/1000 solves

This paper: Active measurements of CAPTCHA solving services

- How did the authors measure CAPTCHA services?

This paper: Active measurements of CAPTCHA solving services

- How did the authors measure CAPTCHA services?
  - They purchased the service and saw what happened!

Service	$/1K Bulk	Dates (2009–2010)	Requests	Responses
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%)

Table 1: Summary of the customer workload to the CAPTCHA-solving services.

This paper: Active measurements of CAPTCHA solving services

- How did the authors measure CAPTCHA services?
  - They purchased the service and saw what happened
- Authors learned many things by direct measurement
  - Accuracy, speed, pricing, etc.

This paper: Active measurements of CAPTCHA solving services

- How did the authors measure CAPTCHA services?
  - They purchased the service and saw what happened
- Authors learned many things by direct measurement
  - Accuracy, speed, pricing, etc.
- Authors were able to identify many things via *side channels*
  - Worker wages
  - Geolocations
  - How did the authors corroborate these findings without ground truth?

Mr. E Anecdotes (ha-ha)

- Paper was able to corroborate some empirical findings with an interview of a person who ran a CAPTCHA solving services
  - "Dabbled with automated solving but new solvers stopped working too quickly."
  - 50% of his revenue is *profit*!
  - 5 – 10% error rate in CAPTCHA solving
  - Labor markets in China, India, Bangladesh, Vietnam
  - *75% of traffic was generated by 5 – 10 clients!*

Meta Thoughts on this Paper

- So… do CAPTCHAs work? Why or why not?
- What is the economic theory behind employing CAPTCHAs (since they're still employed today)
- What *surprised* you about this paper?
- What do we think about this type of analysis? Has it changed the CAPTCHA ecosystem at all?

Next time…

- Usability and human factors… getting higher and higher level now
  - Now getting into *my* territory as a researcher
- Keep up with your projects. Things are due soon :)