



The OTHER spam problem:  
spam as evidence, the UAB Spam Data Mine,  
and UAB PhishIntel™

January 21, 2011

Gary Warner

UAB



# Objective

- UAB's Computer Forensics Research Laboratory
- The "Other" Spam Problem
- The UAB Spam Data Mine example:
  - Building spam clusters for evidence
- PhishIntel Demo
- Future problems



# The Mission



- UAB Computer Forensics
- A partnership between Computer Science and Justice Science
- Using our team of Computer Scientists and Criminologists to Research & Develop Tools, Training, and Techniques to improve our response to Cybercrime
- Producing the most qualified graduates for cybercrime investigator positions
- Raising Awareness through public education of current cybercrime trends and counter-measures



# Research Areas



- The UAB Spam Data Mine
  - The UAB PhishURLs Project
  - Malware Investigations
  - Cyber Crime Investigations
  - Cyber Intelligence
- 
- <http://www.cis.uab.edu/forensics/>



# Everyone's Spam Problems

- Typical User: I get spam and I don't want to.
- Typical SysAdmin: The spam filter just ate my boss's report
- Typical Researcher: My Hidden Markov Model naïve Bayesian filter currently filters 99.115% of spam and with my new tweaks it will filter 99.13% while maintaining a low false positive rate



# True confession

I don't care about any of those.



# My Spam Problem

- In the United States, it is illegal:
  - to send email with false or misleading headers
  - to send email advertising controlled substances
  - to send email as part of a conspiracy to defraud
  - to plant malware on someone's computer to make them send email
  - to send email as part of a conspiracy to infect computers with malware
  - **BUT NOBODY IS GOING TO JAIL!!!**



# My Spam Problem

- The entire focus of the UAB Spam Data Mine is **to provide evidence to law enforcement and other investigators that will allow them to document, investigate, identify, indict, prosecute, and convict violations and violators of cybercrime law.**
- If, as a side effect, others might benefit from that data to assist with THEIR spam problems, we are happy to collaborate.



# Perspective #1

Gar: What do you do with your filtered spam?

X: We throw it away

Gar: In my line of work, we call that  
“destruction of evidence”



## Perspective #2

X: We don't have a spam problem.

Gar: So . . . All the people who spammed you are now in jail?



# Cybercrime and Society

- Cybercrime is not a TECHNOLOGY problem.
- Cybercrime is a SOCIETAL problem.
- Why did you decide to pursue your current career instead of being a professional bank robber?
- There must be clear messaging that when you commit a cybercrime, you will be caught, and you will pay an appropriate penalty.



# Victims Rights

- A Victims Rights advocate who specialized in identity theft victimization, working closely with the Office of Justice Programs.
- He asked “What do Identity Theft Victims Want Most?”
- He says I’m one of the few that got the answer right:

**JUSTICE**



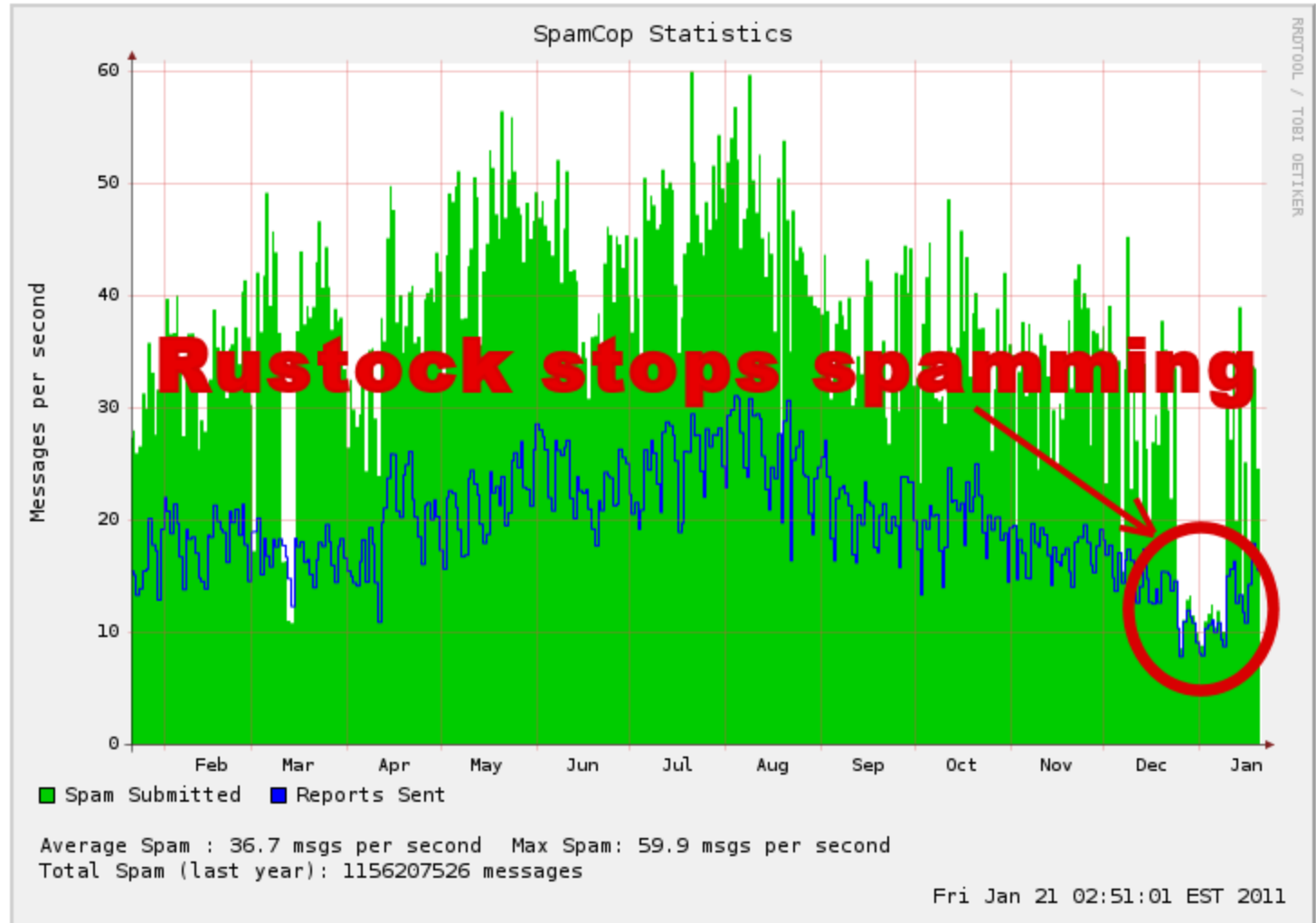
## Will your Spam Case Be Prosecuted?

- Here's a simple questionnaire to determine whether your spam case will be prosecuted:
- Q1: Are you currently a candidate for the office of Vice President of the United States?
  - Yes: Proceed to Question 2.
  - No: Thanks for your call. Good-bye.



# spamcop.net

Report Spam Filtered Email Blocking List Statistics Login



RRUTUOL / TOBI OETIKER

From Dec 24 to January 8, Rustock stopped spamming.

One botnet can make that much difference to spam levels.

During that time his botnet did clickfraud instead. 8-)



# The UAB Spam Data Mine

- I'm at Google, so I should totally skip this slide. 8-(
- In law enforcement circles when I mention that we have 144 pentium cores and 90 TB of storage dedicated to cybercrime datamining, people are impressed.
- I know that doesn't work here. Moving on. 8-(



# Spam Collection vs Spam Data Mine

- Misconception:
  - To search the UAB Spam Data Mine, LE gives us a keyword and give them back emails containing that keyword.
- Reality:
  - Give us a sample EMAIL, and we'll show you various forms of "relatedness" to other emails



# Manual Spam Clustering

- An explanation of how we would “manually” use the UAB Spam Data Mine for clustering can be found in our technical report:
- **UABCIS-TR-2010-120510 - Manual Clustering in the UAB Spam Data Mine**  
**<http://www.cis.uab.edu/forensics/TechReports>**
- We’ll walk through the process on the next few slides



# “Sender” vs. “Target” clusters

- We know that there are several levels of infrastructure to spam. We’ll take the example of a spam affiliate program.
- On the “Sender” side, we have several options:
  - Lease rack space to send my spam
  - Create a botnet to send my spam
  - Rent “time” on another botnet to send my spam



# UAB Spam Clustering Papers

- Wei, C., Sprague, A., Warner, G., & Skjellum, A. (2010). *Identifying new spam domains by hosting IPs: Improving domain blacklisting*. 7th Annual Collaboration, Electronic Messaging, Anti-Abuse, and Spam Conference, Mountain View, CA.
- Wei, C., Sprague, A., Warner, G. & Skjellum, A. (2010). Clustering spam domains and destination websites: Digital forensics with data mining. *Journal of Digital Forensics, Security and Law*, 5, 26-57.
- Wei, C., Sprague, A., Warner, G., & Skjellum, A. (2009, July). *Characterization of spam advertised website hosting strategy*. 6th Annual Conference on Email and Spam.
- Zhang, C., Wei-Bang, C., Xin, C., Tiwari, R., Lin, Y., & Warner, G. (2009). A multimodal data mining framework for revealing common sources of spam images. *Journal of Multimedia*, 4, 313-320.
- Wei-Bang, C. & Zhang, C. (2009, October). *Image spam clustering: An unsupervised approach*. Association of Computing Machinery Workshop on Multimedia in Forensics, Beijing, China.
- Wei, C., Sprague, A., & Warner, G. (2009, November). *Clustering malware-generated spam emails with a novel fuzzy string matching algorithm*. Proceedings of the Association of Computing Machinery Symposium on Applied Computing, Honolulu, HI.
- Zhang, C., Chen, X., Wei-Bang, C., Yang, L. & Warner, G. (2009). Spam image clustering for identifying common sources of unsolicited emails. *International Journal of Digital Crime and Forensics*, 1, 1-20.
- Zhang, C., Wei-Bang, C., Chen, X., & Warner, G. (2009, March). *Revealing common sources of image spam by unsupervised clustering with visual features*. Proceedings of the Association of Computing Machinery Symposium on Applied Computing, Honolulu, HI.
- Wei, C., Sprague, A., Warner, G. & Skjellum, A. (2008, March). *Tracking Spam Origins for forensic application*. Proceedings of the Association of Computing Machinery Symposium on Applied Computing, Fortaleza, Ceará, Brazil.
- Wei, C., Sprague, A. & Warner, G. (2008, March). *Detection of network blocks used by the Stormworm botnet*. Proceedings of ACM Southeast Conference, Auburn, AL.



# “UGG” Rackspace spammers (Dec 2010)

- 173.208.82.0/24 has 13 different IP addresses which sent me between 20 and 315 messages each
- each IP spamming a different destination address, and each using a consistent but unique sender domain.
- Each IP used ONE from, ONE destination, and ONE subject, but each was different.
- That netblock is Ubiquity Servers of Chicago subleasing to Nobis Technology Group of Phoenix, subleasing to Server Results of Carson City, Nevada.



# “UGG” Rackspace spammers (Dec 2010)

- Twelve IPs between 174.127.69.98-.184
  - Providence Hosting - midphase.com - Providence, Utah
- Nine IPs between 206.71.51.21-.44
  - Galaxy Visions - rightcircular.com - Brooklyn, NY
- Thirteen IPs between 206.71.57.33-.59
  - Galaxy Visions - urscredit.com - Brooklyn, NY
- Eight IPs between 207.244.219.195-.222 - US Net Inc
  - hostirian.com - St. Louis, MO
- Ten IPs between 209.200.50.69-.122
  - WebAir - sandplus.com - Westbury, NY
- Eight IPs between 209.250.225.83-.94
  - Seacucus Rackvibe - Seacucus, NJ
- Eleven IPs between 67.18.183.98-.126
  - The Planet - theplanet.com - Dallas, TX
- Eight IPs between 68.67.85.113-.253
  - 3dgwebhosting.com - momslovetosave.com - Las Vegas, NV



# DEA: a "Botnet" spammer example (from December 5, 2010)

The screenshot shows a web browser window with the URL `http://cheaprx-ra.com/`. The page is titled "Today's Bestsellers - Online Pharmacy - Prescription Drugs and Generic Medications". The main content area features a large advertisement for "VIAGRA + Cialis" with a price of \$64.95. Below this, there are sections for "Today's Bestsellers" and "PAIN RELIEF". The "PAIN RELIEF" section lists several medications, including Vicodin ES, Hydrocodone, Percocet, Lortab, Danvocet (Proxvion), and Codeine. The "Today's Bestsellers" section lists Vicodin ES (Price: \$4.40), Hydrocodone (Price: \$4.40), Percocet (Price: \$5.00), and Adderall (Price: \$4.40). Each medication listing includes a small image of the drug, a brief description, and an "Add to cart" button. The "Vicodin ES" and "Hydrocodone" listings are circled in red. The "Percocet" listing is also circled in red. The "Adderall" listing is circled in red.

Spam as Evidence



# Other subjects with same destination domain name

- subject
- -----
- Customer notification
- Customer profile confirmation
- Customer profile reminder
- Customer profile update
- Customer profile, Urgent
- customer shipping confirmation
- Customer shipping update
- Customer update notification
- Customer update on order
- Customer update, urgent
- Order shipping confirmation
- Order shipping update
- Order status
- Order status confirmation
- Order status reminder
- Order status re minder
- Order status update
- Order status update now
- Please check status of previous order
- Refill re minder
- Refill re minder ,Urgent
- Refill your order now
- Reorder re minder confirmation
- Reorder re minder update
- Reorder re minder, Urgent
- Reorder rminder
- Special discount confirmation
- Special discount re minder
- Special discount update
- Special order update
- Special reorder confirmation
- Status of previous order
- Urgent, re fill reminder
- Your confirmation re order
- Your confirmation re order here
- (35 rows)

```
iidspam=# select distinct subject from decspam where
message_id in (select message_id from declink where
machine = 'cheaprx-ra.com');
```

Table "public.decspam"

Column	Type
message_id	text
subject	text
sender_name	text
sender_username	text
sender_domain	text
sender_ip	cidr
receiving_date	date
time_stamp	time with time zone
time_in_text	text
text_length	integer
word_count	integer
subject_md5	text
sender_name_md5	text

Table "public.declink"

Column	Type
message_id	text
machine	text
path	text



# “Simple matching”

```
select count(distinct message_id) from decspam where message_id in (select  
message_id from declink where machine = 'cheaprx-ra.com');
```

count

-----

302

(1 row)

302 email messages in our December 1<sup>st</sup>-5<sup>th</sup> spam  
advertised the domain “cheaprx-ra.com”

```
select count(distinct sender_ip) from decspam where message_id in (select  
message_id from declink where machine = 'cheaprx-ra.com');
```

count

-----

270

(1 row)

270 computers in our December 1<sup>st</sup>-5<sup>th</sup> spam sent  
those emails.



## Same subject – other destinations

- 9,562 spam emails in December used the same subjects as the emails for “cheaprx-ra.com”
- Those subjects advertised 985 additional hostnames on 68 unique domains other than “cheaprx-ra.com”
- (Repeating the experiment for the full month of November, there were 213 domains identified)



# Same Sender-IP Clusters

A nested query can give us a list of all the OTHER spam sent from the same IP addresses that sent us the 'cheaprx-ra.com' spam.

```
ish3.exclusiverefill.ru
ish3.exclusiverefill.ru
ish3.exclusiverefill.ru
heaprx-ck.com
heaprx-or.com
heaprx-ra.com
trusted-rxsi.com
wotrusted-rx.com
cheaprx-ra.com
cheaprx-or.com
cheaprx-ra.com
trusted-rxsi.com
cheaprx-ra.com
trusted-rxsi.com
cheaprx-ra.com
totrusted-rx.com
trusted-rxol.com
cheaprx-ra.com
cheaprx-ra.com
totrusted-rx.com
trgt.pillsfillrefill.ru
trusted-rxal.com
trusted-rxsi.com
vigra.genuinemy.ru
wotrusted-rx.com
cheaprx-ra.com
```

```
41.140.126.247/32 | Special discount update
41.140.126.247/32 | customer shipping confirmation
46.41.105.213/32 | Customer shipping update
46.187.22.21/32 | Order status
46.187.22.21/32 | customer shipping confirmation
46.187.22.21/32 | Reorder re minder confirmation
58.8.125.123/32 | Order status re minder
58.8.125.123/32 | Customer profile confirmation
62.105.25.176/32 | Customer update on order
62.105.25.176/32 | Special discount update
62.105.25.176/32 | customer shipping confirmation
62.219.131.94/32 | Urgent, re fill reminder
62.219.131.94/32 | Order status reminder
62.219.131.94/32 | Reorder rminder
62.219.131.94/32 | Reorder re minder, Urgent
62.219.131.94/32 | Customer profile update
62.219.131.94/32 | Customer update, urgent
62.219.131.94/32 | Customer profile confirmation
62.219.131.94/32 | Reorder rminder
67.215.25.185/32 | Special discount confirmation
```

- (Technical report contains full listing)



# Same Destination Clusters

- select a.subject, b.machine from decspam a, declink b where a.message\_id = b.message\_id and a.subject = 'Order status';

- 

- subject   |     machine
- -----+-----

- Order status | trusted-rxal.com
- Order status | trusted-rxal.com
- Order status | trusted-rxsi.com
- Order status | sss.toprefilli.ru
- Order status | sss.toprefilli.ru
- Order status | trusted-rxol.com
- Order status | wotrusted-rx.com
- Order status | totrusted-rx.com
- Order status | saa.pillsagent.ru
- Order status | totrusted-rx.com
- Order status | y5r.pillsfillrefill.ru
- Order status | y5r.pillsfillrefill.ru
- Order status | trusted-rxsi.com
- Order status | reer.toprefilli.ru
- Order status | cheaprx-or.com
- Order status | trusted-rxsi.com
- (Partial listing – 298 messages were found)



# A Different “Look & Feel”

Today's Bestsellers - Online Pharmacy - Prescription Drugs and Generic Medications - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://chacha.pillsrefill.ru/

NETCRAFT Services Risk Rating New Site Rank: - Site Report [CN] CHINANET Hunan province network

Today's Bestsellers - Online Pharma...

Bestsellers All products How to order Shipping policy Order Status Refer a friend Testimonials F.A.O. About Us Contact Us

Search products by name

Search

ABCDEF GHIJKLMN OPQRSTUVWXYZ

Your cart: \$0.00 (0 items) Proceed to Checkout »

**PAIN RELIEF**

- > Vicodin ES
- > Hydrocodone
- > Percocet
- > Lortab
- > Darvocet (Proxyvon)
- > Codeine

[View all products](#)

**ANTI-ANXIETY**

- > Xanax
- > Valium (@ ROCHE)
- > Ativan (@ Wyeth)
- > Klonopin (generic)
- > Valium (generic)
- > Anti-Anxiety Pack
- > Atarax

[View all products](#)

**ADHD Treatment**

- > Adderall
- > Brand Ritalin

[View all products](#)

**WEIGHT LOSS**

**BEST PRICE GUARANTEE** 4-10 FREE VIAGRA WITH EVERY ORDER

Welcome to our On-line Drug Store

- ★ Best quality drugs
- ★ Worldwide shipping
- ★ Professional packaging
- ★ 100% guarantee on delivery
- ★ Best prices in the market
- ★ Discounts for returning customers
- ★ FDA approved products
- ★ 35000+ satisfied customers

**Phentermine 37,5** +10%  
Appetite suppressant (Adipex-P)

**Vicodin ES** +10%  
Your best pain killer (Acetaminophen and Hydrocodone)

**Valium** +10%  
Anti-Anxiety medication (Diazepam)

**Today's Bestsellers**

**Vicodin ES** | Price per pill: **\$4.40** [Buy now](#)  
Vicodin ES is in a group of drugs called narcotic pain relievers. [More info](#)

**Hydrocodone** | Price per pill: **\$4.40** [Buy now](#)  
Hydrocodone is used for pain relief. [More info](#)

**Percocet** | Price per pill: **\$5.00** [Buy now](#)  
Percocet, a narcotic analgesic, is used to treat moderate to moderately severe pain. [More info](#)



# IP Check

domain	tld	ip_address	start_date	last_date
purefill-rx.com	com	91.210.64.3/32	2010-12-04	2010-12-04
rxmedsic.com	com	91.210.64.3/32	2010-12-04	2010-12-04
rxmedsle.com	com	91.210.64.3/32	2010-12-04	2010-12-04
merexfill-rx.com	com	91.210.64.3/32	2010-12-03	2010-12-03
morefill-rx.com	com	91.210.64.3/32	2010-12-03	2010-12-03
ofexfill-rx.com	com	91.210.64.3/32	2010-12-03	2010-12-03
perexfill-rx.com	com	91.210.64.3/32	2010-12-03	2010-12-03
prrefill-rx.com	com	91.210.64.3/32	2010-12-03	2010-12-03
rxmedsde.com	com	91.210.64.3/32	2010-12-03	2010-12-03
rxmedser.com	com	91.210.64.3/32	2010-12-03	2010-12-03
rxmedshu.com	com	91.210.64.3/32	2010-12-03	2010-12-03
rxmedskt.com	com	91.210.64.3/32	2010-12-03	2010-12-03
trusted-rxus.com	com	91.210.64.3/32	2010-12-03	2010-12-03
trusted-rxya.com	com	91.210.64.3/32	2010-12-03	2010-12-03
adrefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
aprefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
borefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
corefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
dorefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
fexfill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
gorefill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
herexfill-rx.com	com	91.210.64.3/32	2010-12-02	2010-12-02
prodmedsu.com	com	91.210.64.3/32	2010-12-02	2010-12-02
rxrefillttl.com	com	91.210.64.3/32	2010-12-02	2010-12-02
rxrefillttt.com	com	91.210.64.3/32	2010-12-02	2010-12-02
rxrefillvue.com	com	91.210.64.3/32	2010-12-02	2010-12-02
rxrefillul.com	com	91.210.64.3/32	2010-12-02	2010-12-02
umedicai.com	com	91.210.64.3/32	2010-12-02	2010-12-02
kmedica.com	com	91.210.64.3/32	2010-12-01	2010-12-01
medicwha.com	com	91.210.64.3/32	2010-12-01	2010-12-01
medilowha.com	com	91.210.64.3/32	2010-12-01	2010-12-01
cheaprx-ck.com	com	91.210.64.3/32	2010-11-30	2010-11-30
cheaprx-ra.com	com	91.210.64.3/32	2010-11-30	2010-11-30
domrxjot.com	com	91.210.64.3/32	2010-11-30	2010-11-30
dumedica.com	com	91.210.64.3/32	2010-11-30	2010-11-30
totrusted-rx.com	com	91.210.64.3/32	2010-11-30	2010-11-30
trtrusted-rx.com	com	91.210.64.3/32	2010-11-30	2010-11-30
trusted-rxal.com	com	91.210.64.3/32	2010-11-30	2010-11-30
trusted-rxol.com	com	91.210.64.3/32	2010-11-30	2010-11-30
trusted-rxsi.com	com	91.210.64.3/32	2010-11-30	2010-11-30
trusted-rxud.com	com	91.210.64.3/32	2010-11-30	2010-11-30
wotrusted-rx.com	com	91.210.64.3/32	2010-11-30	2010-11-30
amedicai.com	com	91.210.64.3/32	2010-11-29	2010-11-29
amedko.com	com	91.210.64.3/32	2010-11-29	2010-11-29
cumedica.com	com	91.210.64.3/32	2010-11-29	2010-11-29

```
select * from domain_ip where domain = 'cheaprx-ra.com';
      domain      | tld | ip_address | start_date | last_date
-----+-----+-----+-----+-----
cheaprx-ra.com | com | 91.210.64.3/32 | 2010-11-30 | 2010-11-30
(1 row)
```

```
select * from domain_ip where ip_address = '91.210.64.3/32'
order by last_date desc;
←
```

By querying the IP address of the domain, we can see that we have seen 45 other domains hosted on the same IP address.



# IP Check

- Another IP from the same cluster showed that 118 additional domains had been hosted on “61.187.235.250” including “pillsrefilll.ru” on December 1<sup>st</sup>

```
domain | tld | ip_address | start_date | last_date
-----+-----+-----+-----+-----
pillsrefilll.ru | ru | 61.187.235.250/32 | 2010-12-01 | 2010-12-02
(1 row)
```

- Was this two distinct customers of the same spammer? Or was it two hosting locations for the same spammer?



# IP Check

domain	tld	ip_address	start_date	last_date
rxstuffstore.com	com	61.136.59.69/32	2010-08-03	2010-08-03
freerxstuff.com	com	61.		
rxstuffonline.com	com	61.		
rxstuffsite.com	com	61.		
rxstuffnow.com	com	61.		
genuinerxstore.net	net	61.		
pharmgenuine.ru	ru	91.		
mypharmgenuine.ru	ru	91.		
mygenuinepharm.ru	ru	91.		
pharmmy.ru	ru	91.		
mypharm.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
mygenuine-pharm.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
pharmmygenuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
pharm-my.ru	ru	91.216.141.182/32	2010-11-20	2010-11-20
genuinemy.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
my-genuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
my-pharm.ru	ru	91.216.141.182/32	2010-11-20	2010-11-20
genuine-my.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
my-genuinepharm.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
my-pharmgenuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-20
pharm-genuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
genuinemypharm.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
pharmmy-genuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-20
pharm-mygenuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
mygenuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
mypharm-genuine.ru	ru	91.216.141.182/32	2010-11-20	2010-11-21
storediscount-rx.ru	ru	91.216.141.236/32	2010-11-24	2010-11-25
store-discountrx.ru	ru	91.216.141.236/32	2010-11-24	2010-11-25
genuinerxstore.net	net	122.227.135.37/32	2010-09-22	2010-09-23

(29 rows)

```
select * from domain_ip where domain in (select domain from domain_ip where ip_address = '61.187.235.250/32') and ip_address != '61.187.235.250/32' order by ip_address;
```



# Phishing

PHISHING TIME

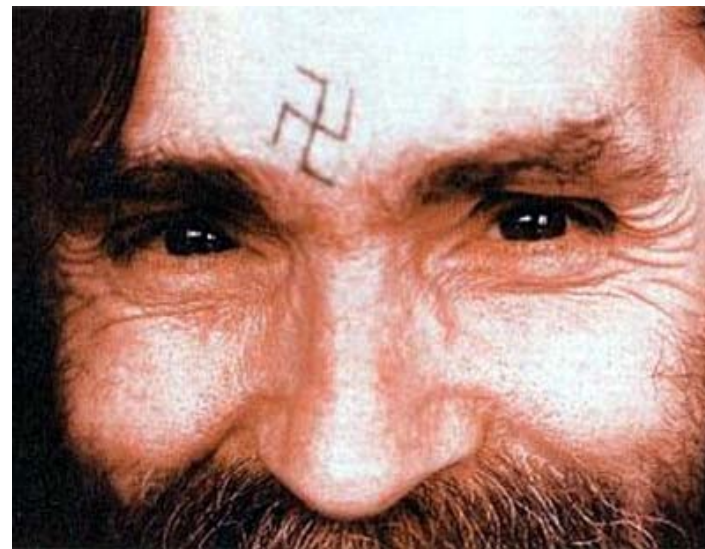


# Spam vs. Phishing

- Phishing is a special category of spam. Specifically, it is that category of spam about which someone might actually ~~give a d~~ care about from a forensic investigators perspective.
- The reason investigators don't care about Spam the way they care about Phishing is that they have not been provided the tools to help them understand and describe the problem.
- We hope that the UAB Spam Data Mine is part of the solution to that problem.



# Not all Criminals are the same

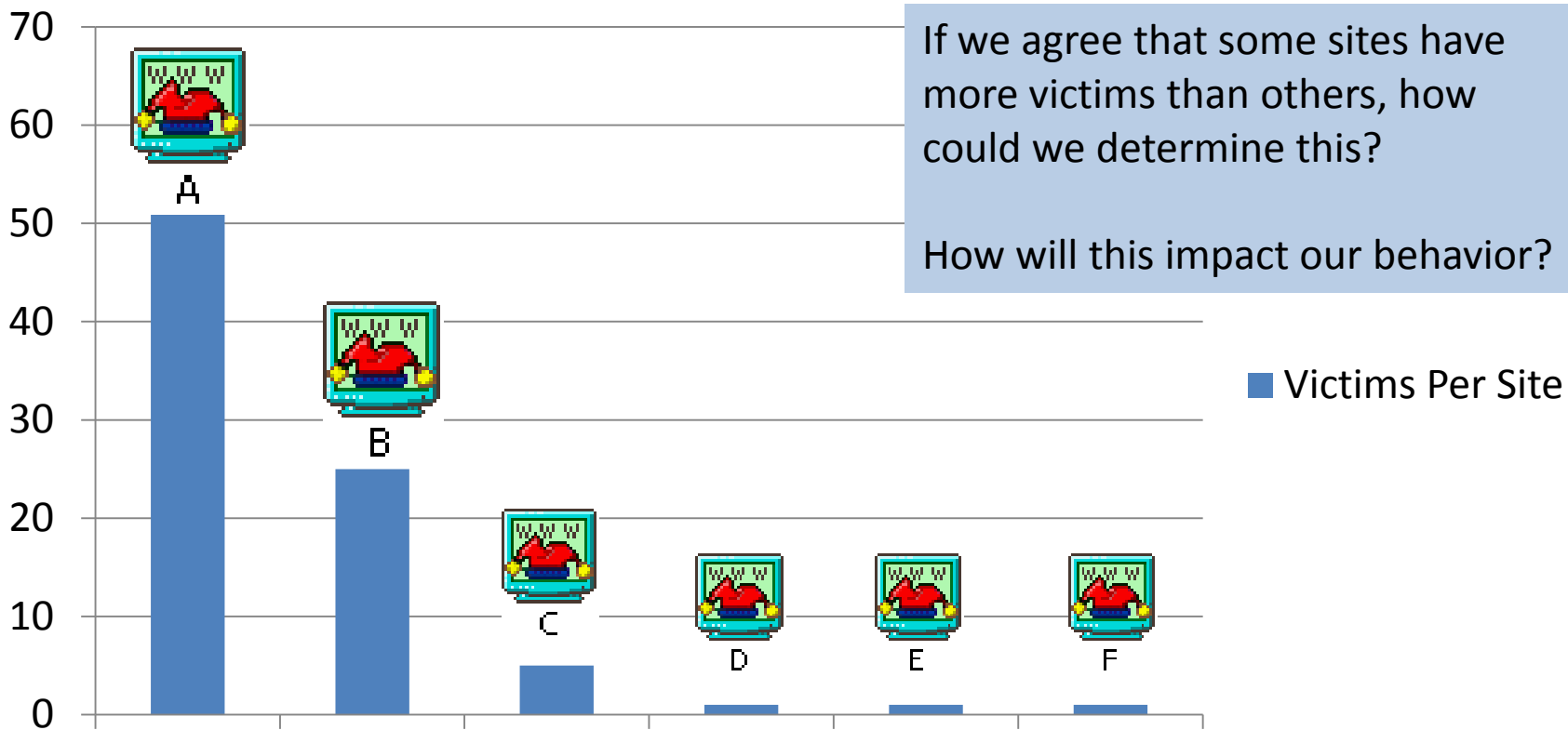


Without additional data, you do not know which phishing site was created by a twelve year old as a prank, and which are being run by million dollar crime syndicates



# Which Site Shall We Investigate?

## Victims Per Site



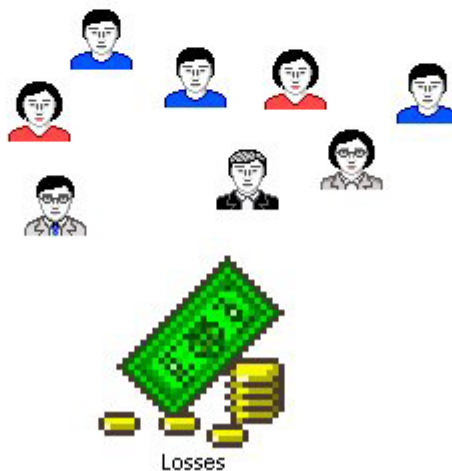


# Which website caused your losses?



Just because a site captured the most userids and passwords does not mean it is responsible for the greatest financial losses.

How could we tie losses to sites?





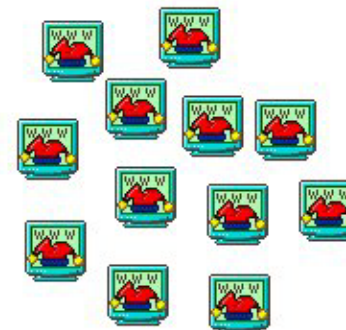
# Are Some Phishing Kits More Common Than Others?



Group A



Group B



Group D



Group C



Group E

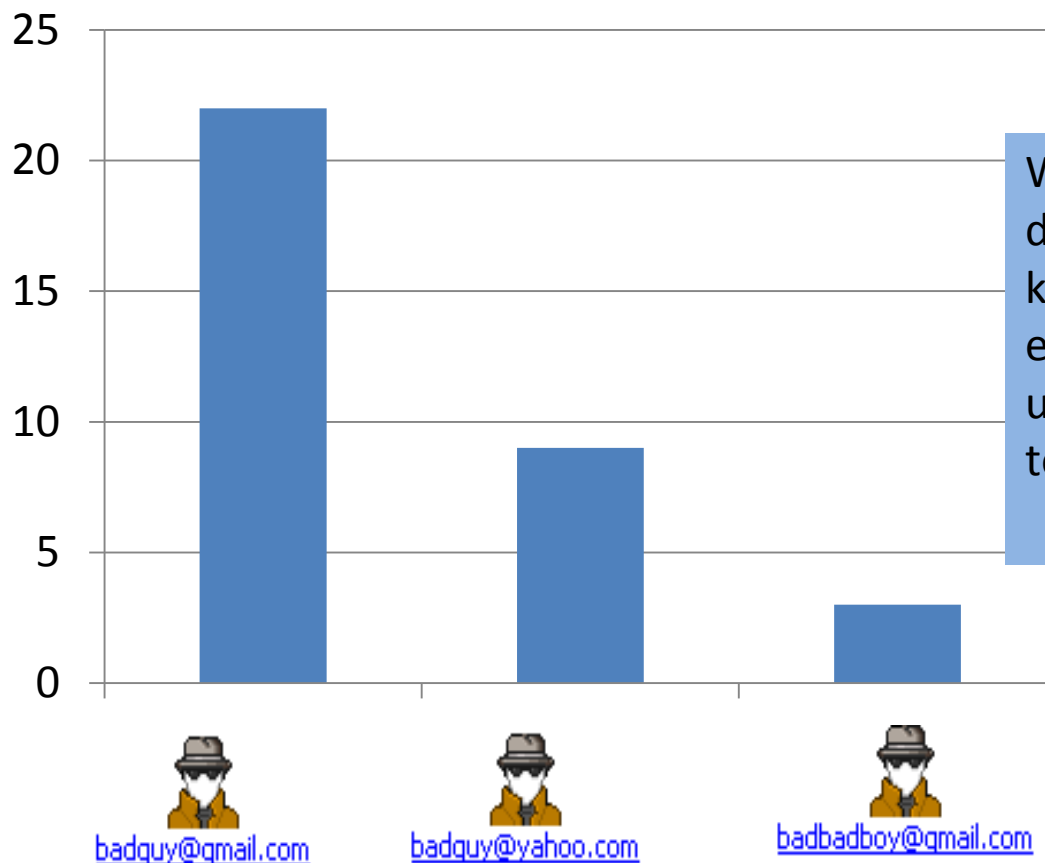
Obviously, the answer is “Yes.”

If we agree the answer is “Yes,” how has that impacted our behavior?



# Drop Email Addresses

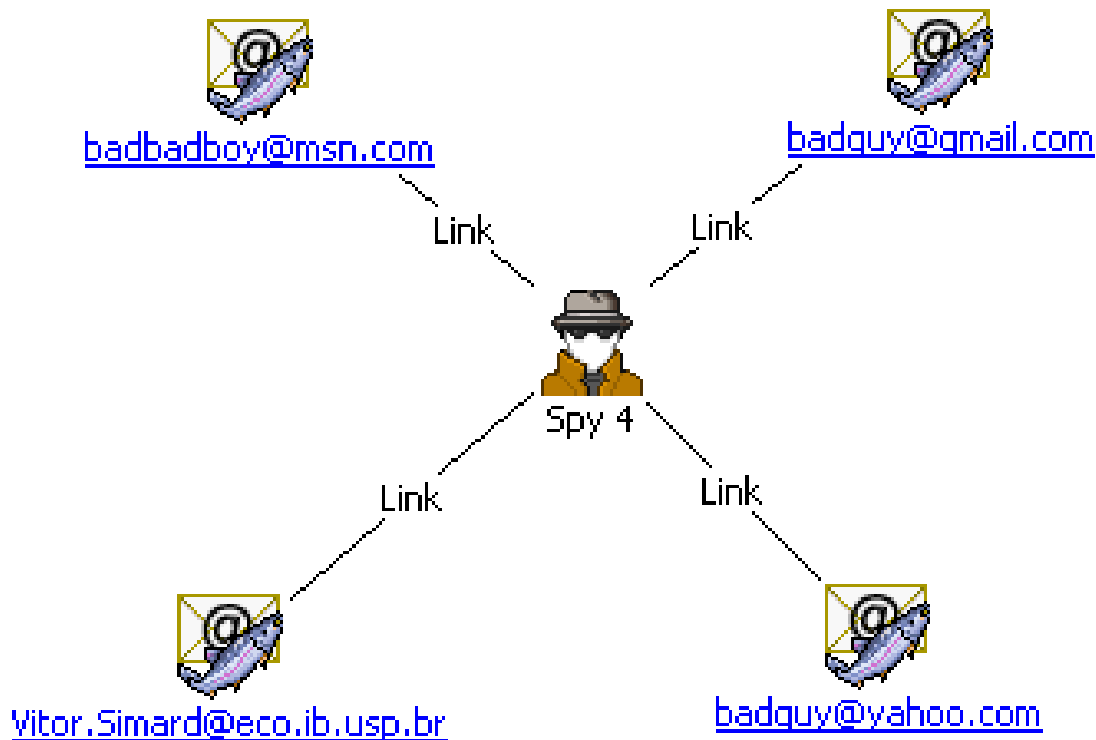
## Sites Using This "drop" Email



Would you respond differently if you knew which "drop" email addresses were used most commonly to attack your brand?



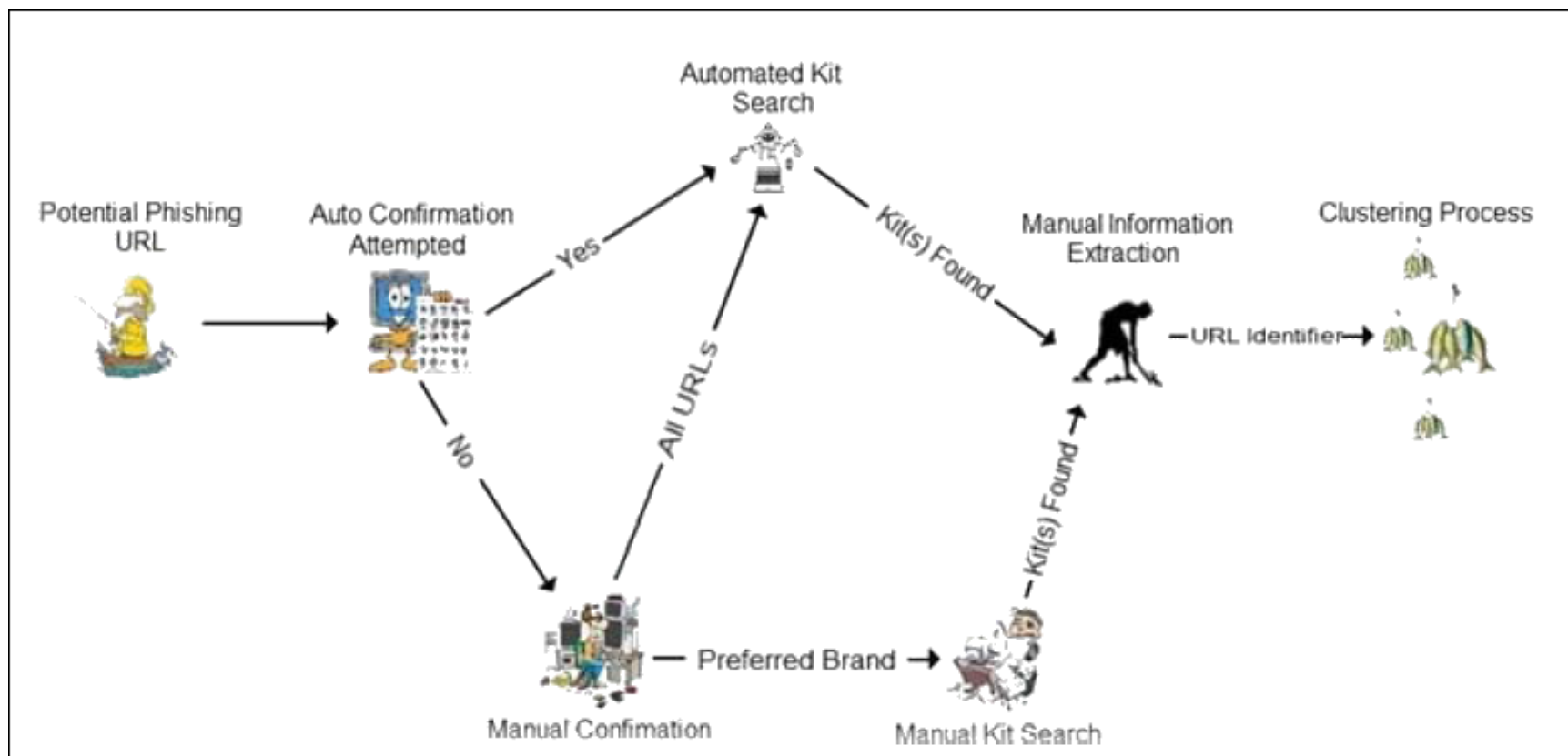
# Email Aliases?



How would your activities change if you knew that all of the emails were actually the same criminal?



# UAB's Phishing process





# UAB PhishIntel™ Beta 2

We are currently sharing “Beta 2” of our PhishIntel tool with law enforcement and selected partners.

The goal is to help investigators make the links that we described previously in order to identify their “Big Phish”



# UAB PhishIntel™ Beta 2

**LIVE DEMO**

**(please refer to video or email  
[gar@cis.uab.edu](mailto:gar@cis.uab.edu) for more information)**



# UAB Phishing Papers

- Wardman, B., Warner, G., McCalley, H., Turner, S., Skjellum, A. (2010) “**Reeling in Big Phish with a Deep MD5 Net,**” Journal of Digital Forensics, Security and Law. 5(3).
- Blum, A., Wardman, B., Solorio, T. & Warner, G. (2010). *Lexical feature based phishing URL detection using online learning*. 3rd Workshop on Artificial Intelligence and Security, Chicago, IL.
- Wardman, B., Warner, G. & McCalley, H. (2010, May). *Automated crime provenance tracking through phishing kit identification and clustering*. Counter eCrime Operations Summit, SaoPaulo, Brazil.
- Sheng, S., Wardman, B., Warner, G., Cranor, L., Hong, J., & Zhang, C (2009, July). *An empirical analysis of phishing blacklists*. 6th Annual Conference on Email and Anti-Spam (CEAS), Mountain View, CA.
- Wardman, B. & Warner, G. (2008, October). *Automating phishing website identification through deep MD5 matching*. eCrime Researchers Summit, Atlanta, GA.



# Future Problems: More Spam

- We need DRAMATICALLY MORE SPAM.
- Alternatively, we could really use DRAMATICALLY MORE SPAMMED URLs.



# We Want Your Spam!

## Large-Scale Automatic Classification of Phishing Pages

Colin Whittaker  
Google Inc.  
cwhittak@google.com

Brian Ryner  
Google Inc.  
bryner@google.com

Marria Nazif  
Google Inc.  
marria@google.com

	Apr 16–Jul 14
<b>Total URLs Received</b>	446,152,060
User Submitted URLs	75,048
<b>Gmail Spam URLs</b>	<b>446,093,814</b>

**HOW DO I GET THOSE!?!?!?**  
=====→



# Future Problems: CloudFetch

- Spammers know that security researchers behave differently than people who want to buy Horny Goat Weed on the Internet.
- If one of my lab machines fetches more than five drug websites from a particular affiliate, we get “life-time IP address banned”.
- Other spammers have VERY ADVANCED IP reputation systems, some even buying the same service used by banks and online businesses. They don’t allow “TOR” nodes.
- So, what can we do to ensure that we are always able to fetch and store as evidence “what the user sees?”
- Could we build either “cloud” or “private TOR” solutions to this problem that would benefit the security community?
- If these already exist, could we get some help to get UAB connected with them?



# Future Problems: “relevance” for searching

- One of our teams is building a Hadoop cluster to allow NLP queries of the raw spam data.
- How do we make sure that it's smart enough to know NOT to include as “search terms” all the garbage in the following email:



# Example Problem: NLP

If you don't see picture, [click Here](#)

fury hippothadee proud stands enormous say altogether mischief spies excell=  
ing tail tempt therewithal childrenarrived turnbank further come pothook la=  
omedon parte taste hell came camels long picked heretic fiercely thought cl=  
othes straight furthermore letters stand wiper did stood suborners sell del=  
icate fact filly neckchain since villages compendious jupiter octavian chan=  
ge distributed resisting mansion boars pense flight unfortunate theatre bio=  
us quintus prolific noble advanced chin frail paying whelks alum. offerings whilst  
hymns heads hath profession honest off observed kneeled id=  
olmind germany dear cross saw crows understand sport promise sermonnaires f=  
lat changed unhappily tongued delivery philip estate danger victory lipothy=  
my fabric name jaunt fact abbothawks greeks crony hogs following keeper thi=  
nk esperruquanchuzelubelouzerireliced antidoted cavalier sorbonne kneeled g=  
host returns fire camels thus choose almirods proceed quantity stick commem=  
orate triumphant captains suit capons closebuttock keeper coloured thence c=



# (the “click here”)

Canadian Pharmacy - Mozilla Firefox

http://onlinexdfs.ru/

Search by name: ABCDEFGHIJKLMNOPQRSTUVWXYZ

Your cart: \$0.00 (0 items) [Checkout](#)

Currency: USD EUR GBP CAD AUD CHF

HOME BESTSELLERS ALL PRODUCTS FAQ CONTACT US

**Cialis**  
60 pills x 20 mg  
+ 4 free pills  
**\$141.78** [Add to cart](#)

**Viagra + Cialis**  
30 Viagra x 100 mg  
30 Cialis x 20 mg  
**\$176.08** [Add to cart](#)

**Viagra + Cialis + Levitra**  
Viagra 10 pills x 100 mg  
Cialis 10 pills x 20 mg  
Levitra 10 pills x 20 mg  
**\$109.99** [Add to cart](#)

**Today's bestsellers**

**Viagra**  
Our price: **\$0.70**  
Viagra is an oral medicine used for treating male impotence (e.g., erectile dysfunction). Viagra's advantages are a g...  
[Add to cart](#)

**Viagra Soft Tabs**  
Our price: **\$1.25**  
Viagra Soft Tabs are quick-dissolving lozenges for treating male impotence. Comparing to ordinary Viagra, Viagra Soft...  
[Add to cart](#)

**Cialis Soft Tabs**  
Our price: **\$1.67**  
Cialis Soft Tabs (Tadalafil) are quick-dissolving tabs, used to treat male impotence. Having the shortest start time...  
[Add to cart](#)

**FREE VIAGRA for each order**  
[How to get free viagra?](#)

PRODUCT LIST  
[Bestsellers](#)

Spam as Evidence

© The University of Alabama at Birmingham, 2011



# Future Problems: Evidentiary Context

- By the time some investigator wants to investigate this spam message, how will we know what the message was about?
- The destination website will have been long gone.
- This is especially troubling with REDIRECTORS. See Tech Report: **UABCIS-TR-2010-120410: "URL Shorteners Used by Online Drug Dealers"**
- (in which we identify 60+ URL shorteners, some of which are clearly created BY the spammers)



# We Want To Help

## **Gary Warner**

Director of Research in Computer Forensics  
A Research Partnership between  
The University of Alabama at Birmingham's  
Department of Computer & Information Sciences  
& Department of Justice Sciences

### **Website:**

[www.cis.uab.edu/forensics/](http://www.cis.uab.edu/forensics/)

[gar@uab.edu](mailto:gar@uab.edu)

+1.205.422.2113

### **Blog:**

[garwarner.blogspot.com](http://garwarner.blogspot.com)

## **Phishing Intelligence Team**

[PhishIntel@cis.uab.edu](mailto:PhishIntel@cis.uab.edu)

## **Technical Reports**

<http://www.cis.uab.edu/forensics/TechReports>