

# Data, Reputation, and Certification Against Spam

John S. Quarterman, Quarterman Creations

Serpil Sayin, Koc University

Jouni Reinikainen, Helsinki Metropolia U.

Eshwaran Vijaya Kumar, U. Texas at Austin

Andrew B. Whinston, PI

McCombs School of Business

IIAR Project, U. Texas at Austin

# Economic Incentives

Spammers, bot herders, phishers, et al.:  
They're all in it for the money.

Email providers, from ISPs to universities:  
Security is an expense, not a profit center  
And outbound spam is an externality.

How do we change this?

# Confusopoly

Nobody wants to be branded a spam haven.  
But how do we know which organizations  
Are sending spam?

Most of them don't mean to  
Many of them don't know themselves  
And Nobody wants to admit to it.  
So we have what Scott Adams calls a  
confusopoly.

# Transparency

Moody's bond ratings

FT business school rankings

Kelley Blue Book for cars

These are all reputation systems.

They endogenize economic externalities

By making comparisons transparent,

Providing economic incentive to do better.

# The Giant Elephant in the Room

What Alex Hutton at Metricon called  
“The Giant Elephant in the room...:  
The necessity of comparative analytics.”

Yes, implement organizational security,  
Yes, measure local effectiveness,  
But to know how well it works you need  
Independent external comparative metrics.

# Need Data

John Moody began by collecting financial data  
On a variety of countries and publishing reports.

IIAR Project at McCombs B. School, U.Texas  
Has been collecting spam blacklist data  
For more than a year, reducing and analysing it.

# Spam Volume per Country

Symantec Apr 2010	Sophos Jan-Apr 2010	Project Honeypot 21 Apr–21 May 2010	IIAR Q3 2009	IIAR Q4 2009	IIAR Q1 2010
US	US	IN	BR	BR	BR
IN	IN	BR	KR	VN	IN
NL	BR	VN	US	KR	US
BR	KR	DE	VN	IN	KR
DE	VN	US	IN	CN	RU
UK	DE	RO	CN	US	VN
FR	UK	RU	RU	RU	CO
PL	RU	UK	PL	CO	UA
VN	IT	IT	CO	PL	AR
IT	FR	PO	AR	AR	DE

# Spamming Countries

**Left 3 table columns** are a few well-known country rankings (there are many more)

No 2 agree on rankings

**Right 3 table columns** are 3 quarters by IIAR

**Each ranking** uses different data and methods

Most don't even use the same time periods

BR, IN, US in all 6 rankings

VN in IIAR and 2 others, etc.



# IIAR Country Rankings Plausible

Which helps validate IIAR data.

Raw data comes from CBL blocklist,  
With custom volume field per blocked address  
collected from 2 CBL spam traps  
(CBL uses more spam traps for their blocklist)

# Top 10 ASNs, Q1 2010

ASN	Owner	Type	CC	%Vol	
7738	T da Bahia	State T	Brazil	4%	
7643	VNPT	Nat. T	Vietnam	3%	
9829	BSNL	Nat. Backbone	India	2%	
8167	T da Santa Catarina	State T	Brazil	2%	
27699	T da Sao Paulo	State T	Brazil	2%	
4766	Korea T	Nat. T	Korea	2%	
24560	Bharti Airtel	Intl T	India	2%	
28573	NET Servicos de Com.	Nat. Cable	Brazil	2%	
17974	PT. T Indonesia	Nat. T	Indonesia	1%	
9050	Romtelecom	Nat. T	Romania	1%	

# National Telecom Considered Spammy

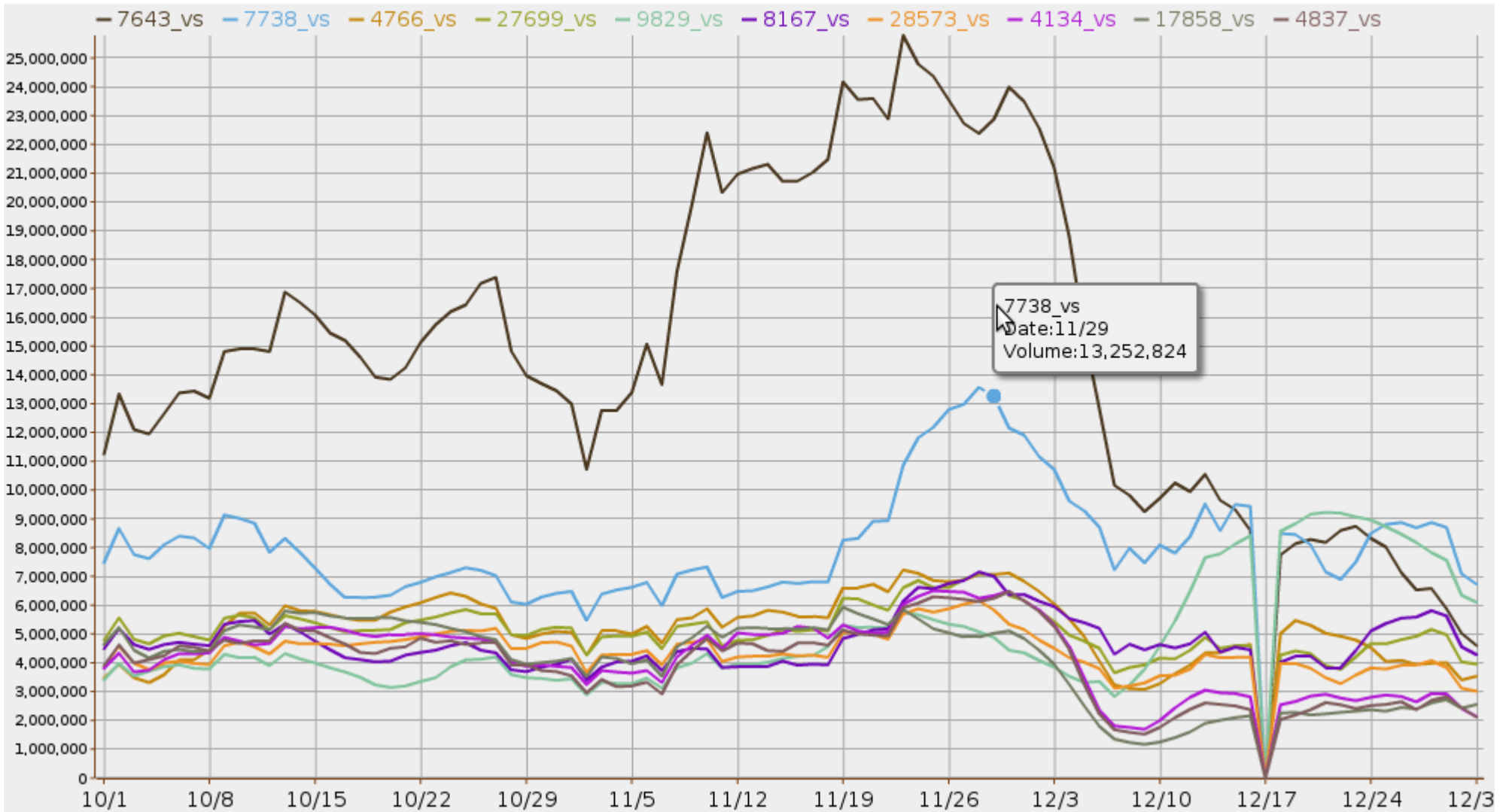
4 of top 10 ASNs are national telecoms.  
Another 3 are telecoms for Brazilian states,  
delegated by national telecom.

Another is a national broadband network.

Gov-controlled telecoms are spammy?

What other patterns can be found?

# Top 10 ASNs, Q4 2009



# About Top 10 ASNs Q4 2009

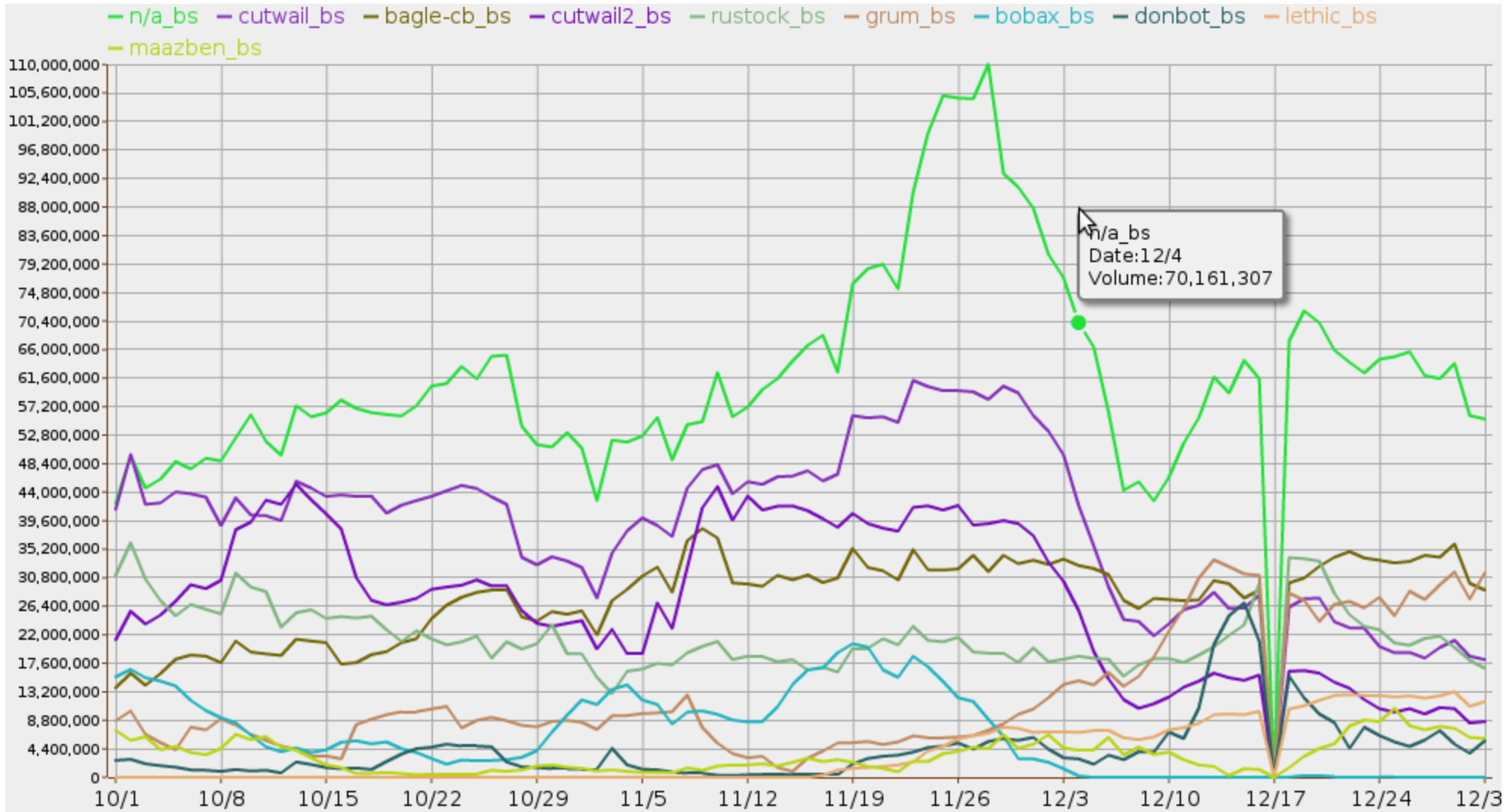
Top ASN for this quarter was AS 7643, VNPT  
Did they do something right end of November?

Second was AS 7738, T. da Bahia.

Pretty impressive: a state ISP  
comes in second worldwide!

(Data loss on 17 Dec 2009.)

# Top 10 Botnets Q4 2009

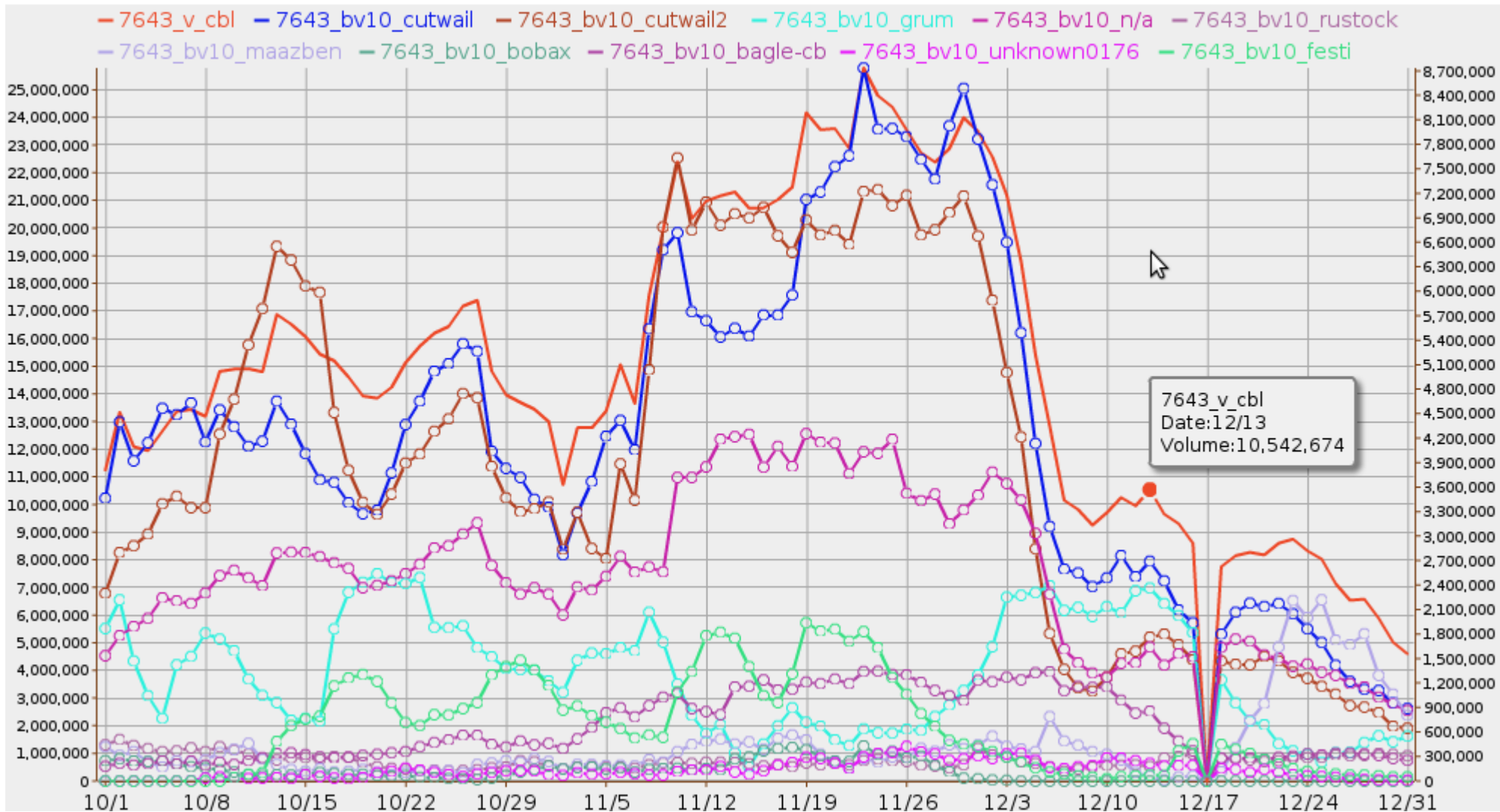


# About Top 10 Botnets Q4 2009

Maybe VNPT didn't do anything good end Nov;  
Maybe cutwail and cutwail2  
finished a spam campaign.

(The highest curve is for n/a  
because CBL rejects a lot of spam by rules  
that don't require checking which botnet.  
Others: bagle\_cb, rustock, bobax, grum, lethic,  
maazben, donbot)

# Botnets per AS 7643 Q4 2009





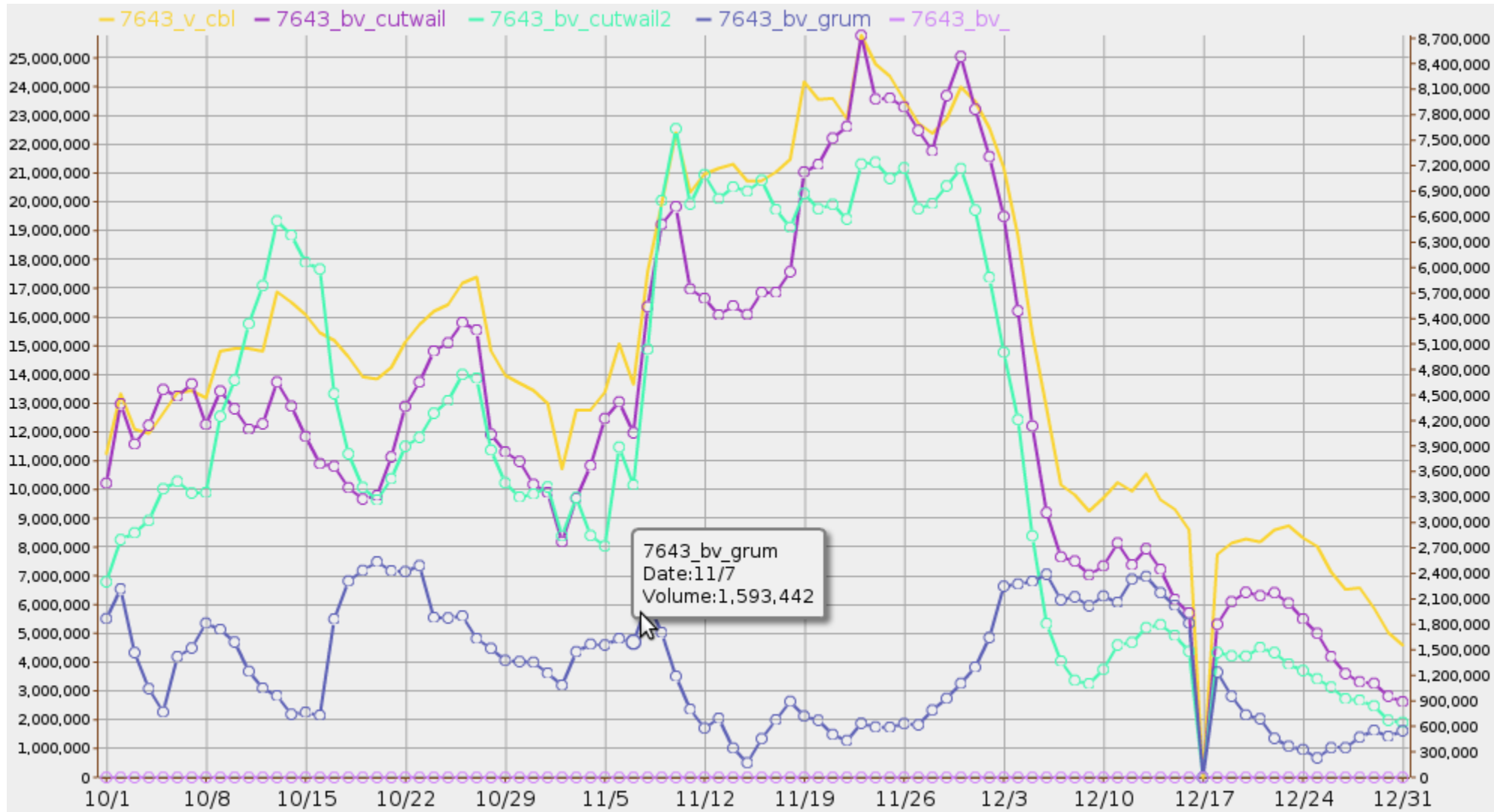
# About Botnets per AS 7643

AS 7643, VNPT, Vietnamese National T

**Left axis:** spam volume this ASN (red line)

**Right axis:** volume top 10 botnets from AS 7643

# Top 3 botnets per AS 7643



# About Top 3 Botnets per AS 7643

For VNPT, total volume tracks

Cutwail + Cutwail2 pretty closely.

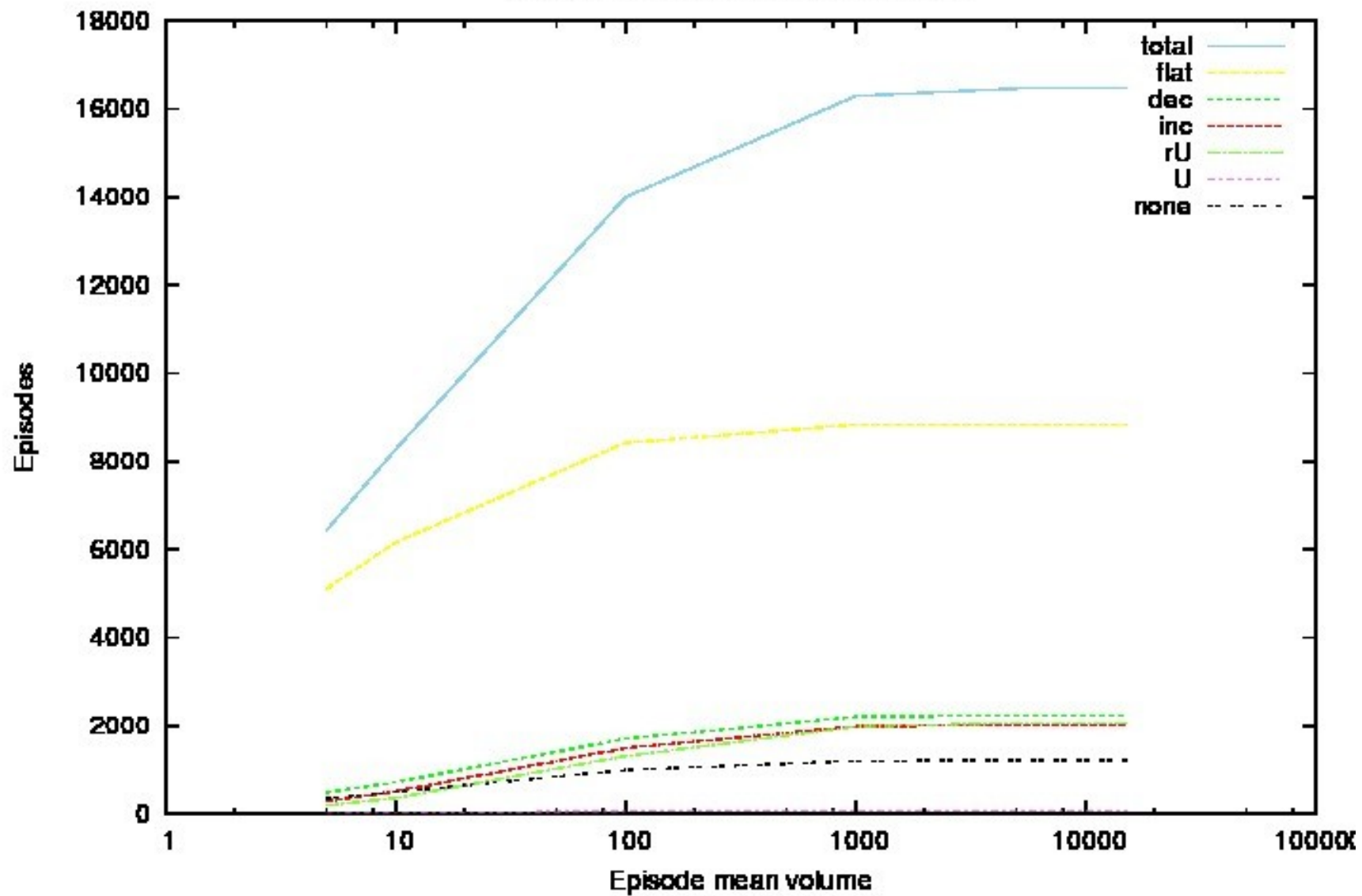
Although on 17 Oct 2009 they decrease

While grum increases keeping total volume up.

We can drill down farther, into specific IP addresses, but you get the idea:

Compare at very high levels, such as countries or ASNs or botnets, then mix and drill down to find clusters and correlations.

Episode mean volume and shape labels



# Proposed Reputation System

Could publish this kind of material as a  
Reputation system (RS)  
providing **market signals**

about security-conscious email providers:

**Economic incentive** for more effective infosec.

A mechanism to **turn the economic externalities**  
Of spam and botnets **into internal incentives.**

(Or for national telecoms, policy incentives.)

# The IIAR Method

**Scope:** the whole Internet

(all spam volume and addresses found by CBL and half a dozen other blocklists, compared to the entire Internet address space and all ASNs)

**Consistency:** daily, with permanent archives

**Variety:** vol/size, %addr, etc. + summary

**Applicability:** can correlate with other information about networks and organizations



# Table, Top 10 ASNs / size Q4 2009

ASN	Country	Description
30792	Ukraine	Luganet Lukansk
48917	Bulgaria	Optinet Ltd
48014	Russia	Interanet Ltd (Voronezh)
38668	Korea	Konkuk University Hospital
45899	Vietnam	VNPT
10030	Malaysia	Celcom ISP
41479	Ukraine	Technoclub, Ltd.
39685	Czechia	Firm Radio Ltd.
34605	Ukraine	Linnet Home Network
10067	Korea	LGNET-China-AS-KR



# About Top 10 ASNs / size Q4 2009

VNPT manages to be in this top 10, too,  
Although it's a different ASN this time.

All the other ASNs are different,  
There's more variety of types,  
And more variety of countries.

# Not Just ISPs

Botnets try to infest every kind of organization that sends email.

Ranking hosting centers for customer choice: providing **economic incentive** for better hosting infosec.

Banks, retailers, NGOs, etc.: **nobody wants a reputation for bad security.**

Each type of organization can be ranked with its peers.

# Rankings by Org Type

Each type of organization  
can be ranked with its peers.

Hosting centers, colos, banks, medical, etc.

Fortune 500: data available to normalize  
By customers, by employees, by market cap...

# Outbound Measures Show Results

**Traditional** application and certification of information security (infosec) techniques, procedures, and policies, **usually about inbound measures**, is great, but doesn't say what works.

**A reputation system using external measures of outbound spam**

Can show which ASNs are actually doing better.

# Infosec per ASN?

**What if we also knew** which infosec techniques, procedures, and policies each ASN uses?

Possible sources: OSSTMM

(Open Source Security Testing Methodology Manual),  
Verizon Business or ICSA Labs (see “Necessary Measures,” Baker, et al., CACM Oct 2007)

Delft U. or MSU or Trend Micro (see “The Role of Internet Service Providers in Botnet Mitigation”, van Eeten et al., WEIS 2010)

# Which Specific Infosec Works?

ASN spam volume  
+ ASN infosec  
= infosec effectiveness

# Which Specific Infosec Works?

If we know which infosec ASNs are using,  
And we see different levels of spam volume that  
correlate with specific infosec,  
That's a clue as to **which** specific technique,  
policy, or procedure **works**.

# Which Specific Infosec Works?

ASN spam volume  
+ ASN infosec  
= infosec effectiveness

+ exploits per botnet  
+ botnets per ASN  
= infosec effectiveness against  
specific exploits



# Exploits per Botnet?

Spam source addresses are proxies for bots.

Which exploits does each botnet use?

Are some exploits used by several botnets?

Many other organizations collect this.

Reputation system may be able to show **which** infosec **works against which exploits.**

# Spatial and Temporal

A reputation system is like a cross-sectional study (rankings compare ASNs),  
but also basically different:  
More detailed, more comprehensive,  
Longitudinal during a long time period,  
And especially ongoing.

# Diverse, Comprehensive, Current

Consider ENISA 2009 spam survey  
(European Network and Information Security Agency)

Lots of good information, but:

Only types of Internet providers vs. all types

Only 92 respondents vs. >10,000 ASNs

Once a year vs. daily

# Temporal or Longitudinal

Use it for experiments:

Try something and watch rankings change.

No need to construct an event chain:

Change infosec and watch rankings.

Additional rankings by

Most improved,

Quickest reacting,

Fastest immunity.

# Potential Infosec ROI

Given how much a measure costs,

How long it lasts,

And how much effect it has  
(according to reputation system)

Could compute ROI for that measure.

# Beyond Loss Reduction to Profit

From the ENISA 2009 spam survey:

“When asked if spam prevention is a factor in the customers' choice of provider, over half said yes, while less than a third said no.”

“...suggesting that generally all providers consider it necessary to have effective anti-spam measures for the sake of attracting and retaining customers.”

# Coordinating Incentives

A reputation system can be used to dig down  
to help determine what security works

But the main point is  
to make the confusopoly transparent  
And thus coordinate economic incentives

# No more Cheap Talk

Cheap talk: providers say they're doing effective security, but how do customers know?

Use reputation and certification to  
Turn cheap talk into effective communication.



# Deployment Organizations

Reputation System (RS)

Produces rankings,  
(High level and detailed),  
Custom drilldowns and analyses

Certification Authority (CA)

Uses rankings  
To certify organizations as in  
Certain classes (like bond ratings)

# Audit and Insurance

Providers could use rankings or certification  
In service level agreements (SLAs),  
Thus in effect self-insuring with external audit.

Insurers could use rankings or certification  
In customer evaluation before writing policies  
And in claims adjustment.

# Elinor Ostrom

Elinor Ostrom (Nobel Prize, Economics, 2009)

Critiques Garrett Hardin's  
myth of the “tragedy of the commons”:

Pure government solutions require perfect government understanding and monitoring.

Pure private solutions require a transparent market or end up in monopoly.

# Effective Commons Management

Ostrom examines many historical and current successful commons. All are hybrids, with much participation by those most affected.

They typically require all participants to know what others are doing:  
That's a reputation system.

# Ack, Merci, Contact

Thanks to CBL for the volume data.

Thanks to Team Cymru for mappings of different data types.

This material is based upon work supported by the  
National Science Foundation under Grant No. 0831338.

Any opinions, findings, and conclusions or recommendations  
expressed in this material are those of the author(s) and do not  
necessarily reflect the views of the National Science Foundation.

Contact: [antispam@quarterman.com](mailto:antispam@quarterman.com)

[iiar@utlists.utexas.edu](mailto:iiar@utlists.utexas.edu)