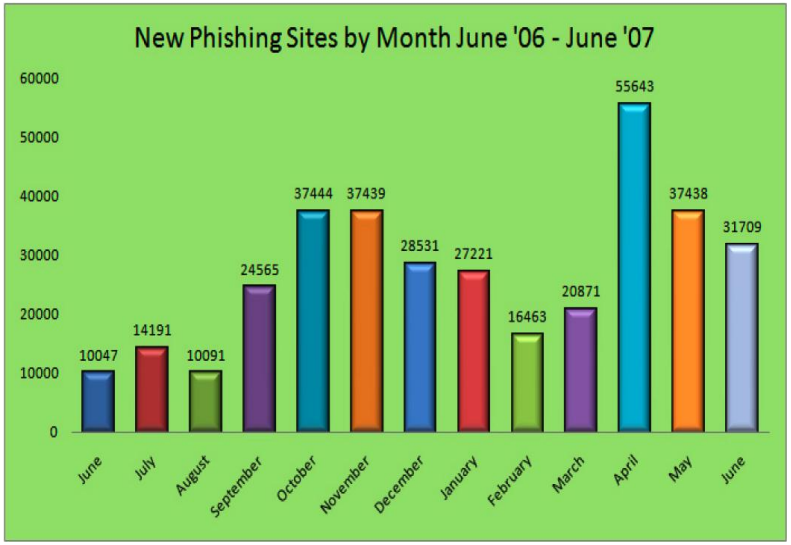
Data Mining the eCriminals: Interesting things lurking in APWG statistics

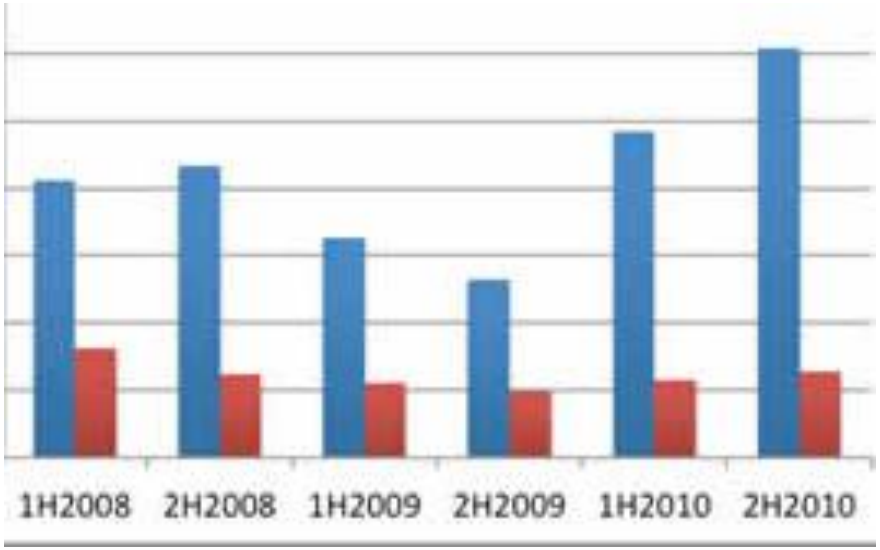
Patrick Cain

APWG – The Cooper-Cain Group, Inc

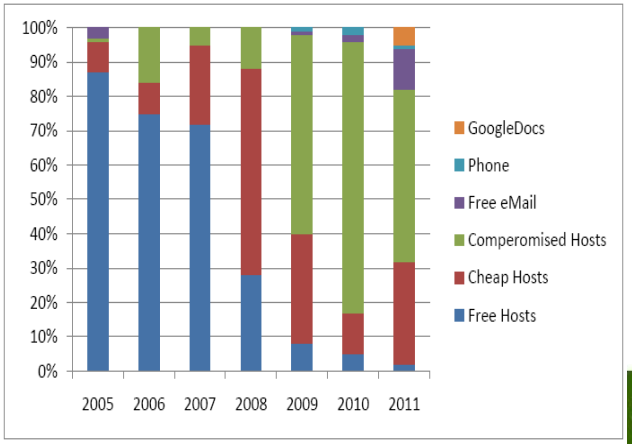
Boston College

We Publish Statistics





RANK	TLD	TLD Location	# Unique Phishing attacks 2H2010	Unique Domain Names used for phishing 2H2010	Domains in registry Oct 2010	Score: Phish per 10,000 domains 2H2010
1	.th	Thailand	125	65	51,438	12.6
2	.ir	Iran	295	169	175,600	9.6
3	.ma	Morocco	73	34	36,669	9.3
4	.ie	Ireland	112	96	151,023	6.4
5	.tk	Tokelau	2,533	2,429	4,030,709	6.0
6 (tie)	.kz	Kazakhstan	49	28	50,534	5.5
6 (tie)	.cc	Cocos Islands	4,963	55	100,000	5.5
7	.in	India	523	421	791,165	5.3
8	.my	Malaysia	68	55	108,211	5.1
9	.hu	Hungary	365	255	542,000	4.7



Anti-Phishing Working Group
APWG

Why Publish Stats?

- To gauge how bad (or good) things are
- And, we're not trying to sell you something
  - Vendor neutral
- We're not trying to be alarmist
- It does allow for trending
- Can identify obvious areas for improvement
    - Registrars…
- [Everybody has a problem with them]

Anti-Phishing Working Group
APWG

Phishing Terminology

- Phishing – Using social engineering to extract personal data or credentials from a victim.
- A phishing campaign is composed of:
  - Lures – A message used to entice a victim to respond.
    - “I am your bank. Give me your password.”
  - Collector - System used to collect and hold personal data and credentials
  - Credentials
    - Bank or system passwords
    - Tax numbers, birth dates, etc
  - Takedown – Disable collector

Anti-Phishing Working Group
APWG

Total Number of Lures Seen

Total Number of Lures Seen

- Counting the number of (unique) lures and brands and collectors was fun…

  … for a little while ☹
- The goal was to educate banks that phishing was real
  - It worked. Then the stats lost their luster
- Now, the stats are based on domains and TLDs
  - A twice-yearly global phishing domains report is published
  - Use the stats to let registries compare themselves
  - .com & .net account for about 50% of all phish

Anti-Phishing Working Group
APWG

Attacks and Domains for 3 Years

	2H2007	1H2008	2H2008	1H2009	2H2009	1H2010	2H2010
Phishing Domain Names	-	47,342	56,959	55,698	126,697	48,244	67,677
Unique campaigns	28,818	26,678	30,454	30,131	28,775	28,646	42,624
TLDs used	145	155	170	171	173	177	183
IP-based phish	5,217	3,389	2,809	3,563	2,031	2,018	2,318
Malicious reg domains	-	-	5,561	4,382	6,372	4,755	11,769
IDN domains	10	52	10	13	12	10	10

Anti-Phishing Working Group
APWG

Detail from the 2H2010 Report

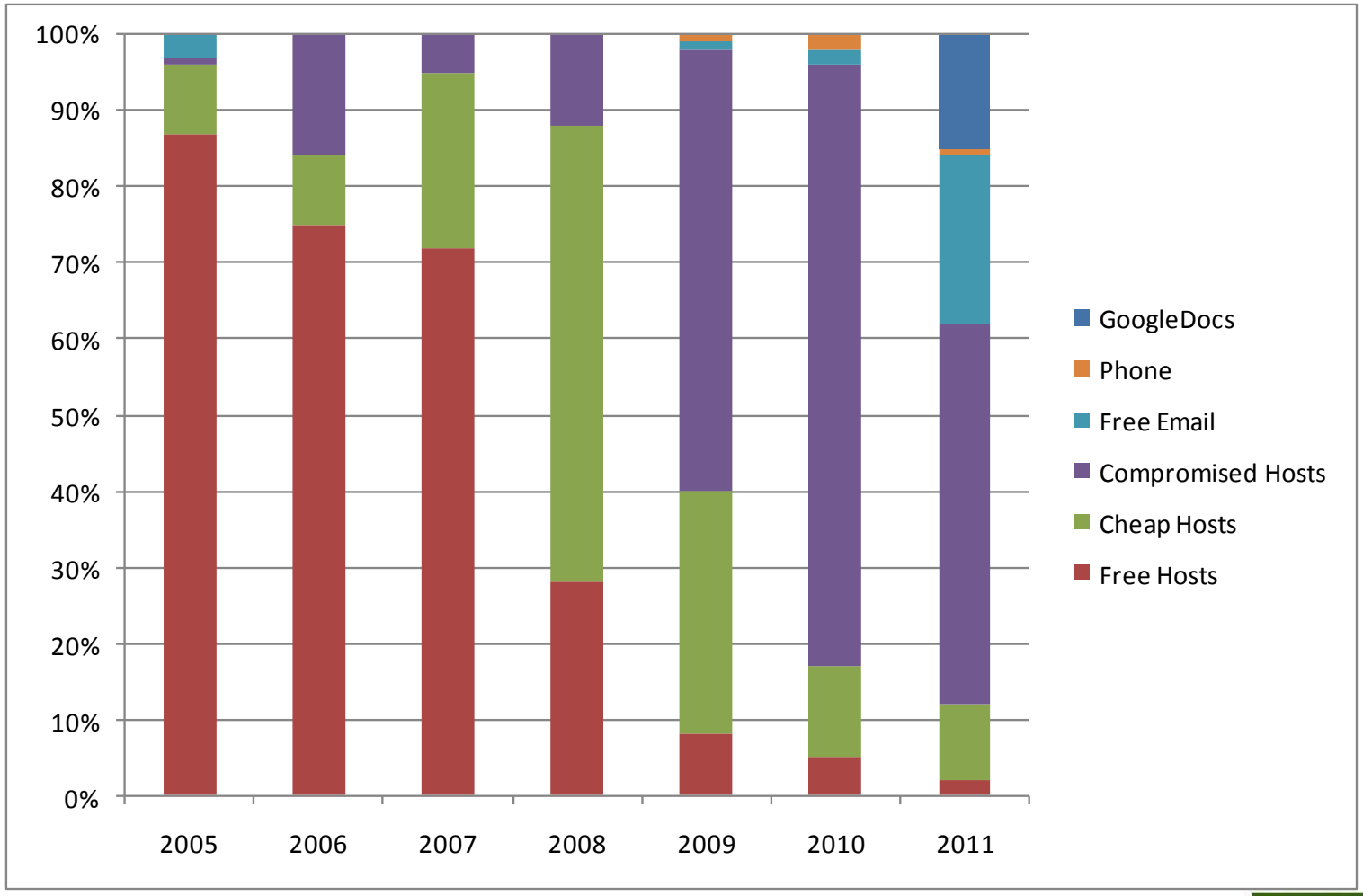
Rank	TLD	TLD Location	# Unique Phishing Attacks 2H2010	Unique Domain Names used for Phishing 2H2010	Domains in Registry 2010	Score: Phish per 10,000 domains
1	.th	Thailand	125	65	51,438	12.6
2	.ir	Iran	295	169	175,600	9.6
3	.ma	Morocco	73	34	36,669	9.3
4	.ie	Ireland	112	96	151,023	6.4
5	.tk	Tokelau	2,533	2,429	4,030,709	6.0
6 (tie)	.kz	Kazakhstan	49	28	50,534	5.5
6 (tie)	.cc	Cocos Islands	4,963	55	100,000	5.5
7	.in	India	523	421	791,165	5.3
8	.my	Malaysia	68	55	108,21	5.1
9	.hu	Hungary	365	255	542,000	4.7

Many Years as a Trend

Year	1H2008	2H2008	1H2009	2H2009	1H2010	2H2010
1	*Hong Kong*	Venezula	Peru	**Thailand**	**Thailand**	**Thailand**
2	**Thailand**	**Thailand**	**Thailand**	Korea	Korea	Iran
3	Belize	Belize	Belize	Ireland	Ireland	Morocco
4	Venezuela	**Soviet Union**	*Belgium*	*Belgium*	Poland	Ireland
5	*Chile*	Romania	Romania	Romania	Chile	Tokelau
6	Romania	*Chile*	Taiwan	Malaysia	Malaysia	Korea
7	**Liechtenstein**	Korea	Korea	.eu	Greece	**Cocos Islands**
8	.name	Vietnam	*Chile*	Iran	Romania	India
9	Taiwan	Russia	Ireland	Poland	Vietnam	Malaysia
10	Korea	Taiwan	Malaysia	Mexico	Czech Rep	Hungary

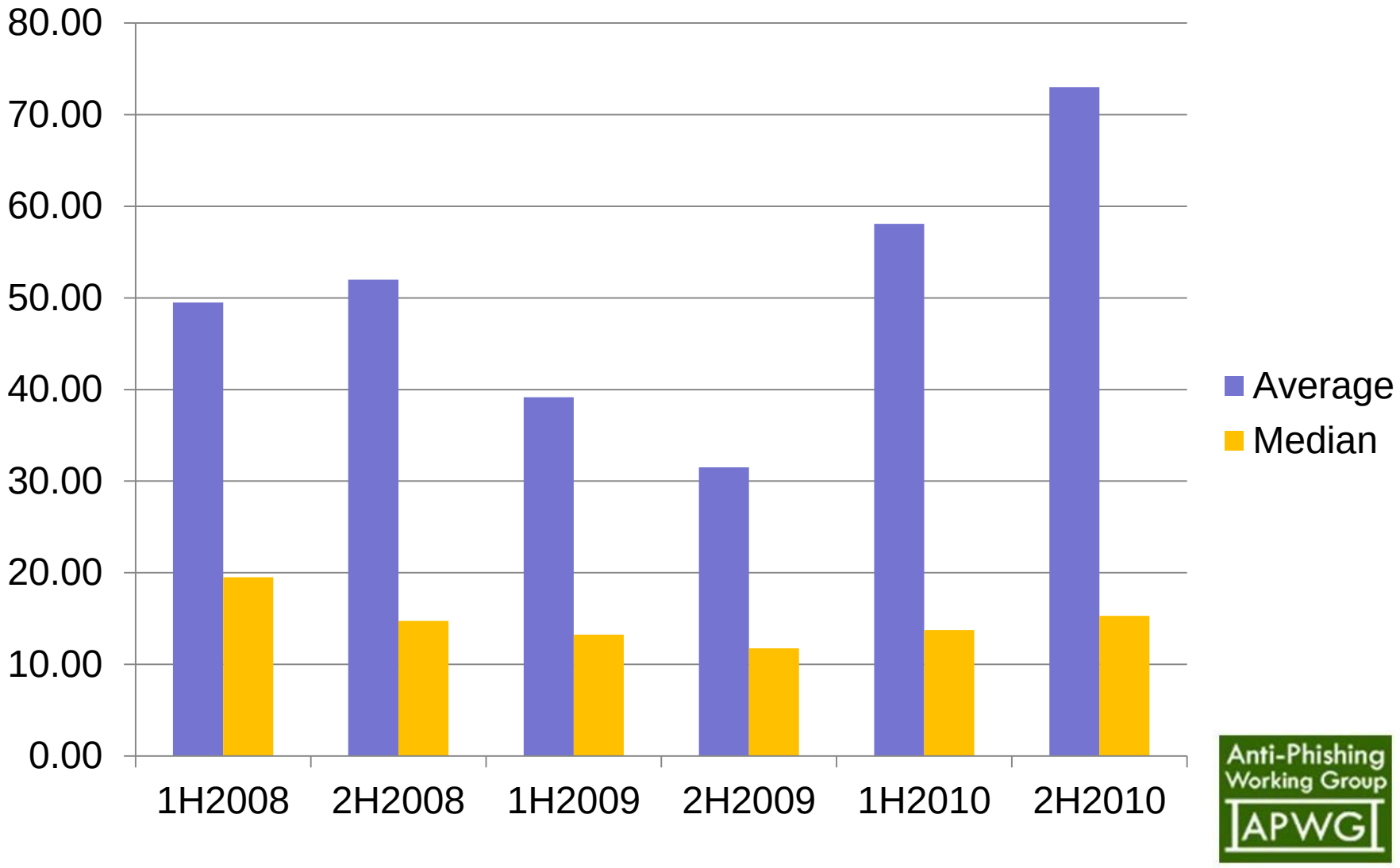
Anti-Phishing Working Group
APWG

Type of Credential Collection Sites



Anti-Phishing Working Group
APWG

Collector Site Uptimes



Anti-Phishing Working Group
APWG

The future of Statistics

- The numbers and pictures are nice….

  …but what are we REALLY trying to do?

Anti-Phishing Working Group
APWG

Adventures in Statistics

- One use of the stats is to convince the banks, governments, polizei, etc, that there is a problem
  - … and to calm down the media hounds
- Phishing, spam, CC fraud, etc used to be distinct
  - Now, organized crime is involved
  - Even minor groups have turned into cooperatives
  - It's now lumped up as Electronic crime (eCrime)
- Everybody knows the numbers are increasing
  - But they're only our numbers
  - How do we get to see a bigger picture?

Anti-Phishing Working Group
APWG

The real purpose of stats… ☺

- The goal it to catch the bad guy
- How do we get countries to devote resources to eCrime?
- How do we get LEA's attention?
  - We need justice's attention
- How do we get Justice's attention?
  - Define risks; education
  - Sounds like a paper.. ☺ (Has it been done before?)

Anti-Phishing Working Group
APWG

What got into Pat?

- We hang out internationally
  - We try and get countries to take eCrime seriously
    - How do we get cops/gov’ts actionable?
- Lots of people use our stats as a driver for change
  - But get/give different conclusions

  are the current stats meeting the ‘mission’?
  - I wondered if we were looking at the stats ‘big picture’ wrong

Anti-Phishing Working Group
APWG

A Diversion

- Interaction with the UN eCrime Commission convinced us that some organizations, companies, and member-states will never report any type of specific eCrime statistics.
- This is bad
  - Stats help countries prioritize response
  - Stats help plan response actions
  - Our stats won't help (non-country specific) you!
- It will get worse
  - APT, night dragon, cheese slider, etc
- What's a crime fighter to do?

Anti-Phishing Working Group
APWG

Modify Our Current Stats?

- Define the risks to an organization from the internet
  - Kind of like what ISO/IEC 27032 may do
- Refine some (general) threats from those risks
- Identify threat-specific malicious behaviour

- Report stats as ‘threats and risks’ based
  - We’ll need new types of reporting
  - And more people to report things
  - Or not. Use it ‘internally’, too

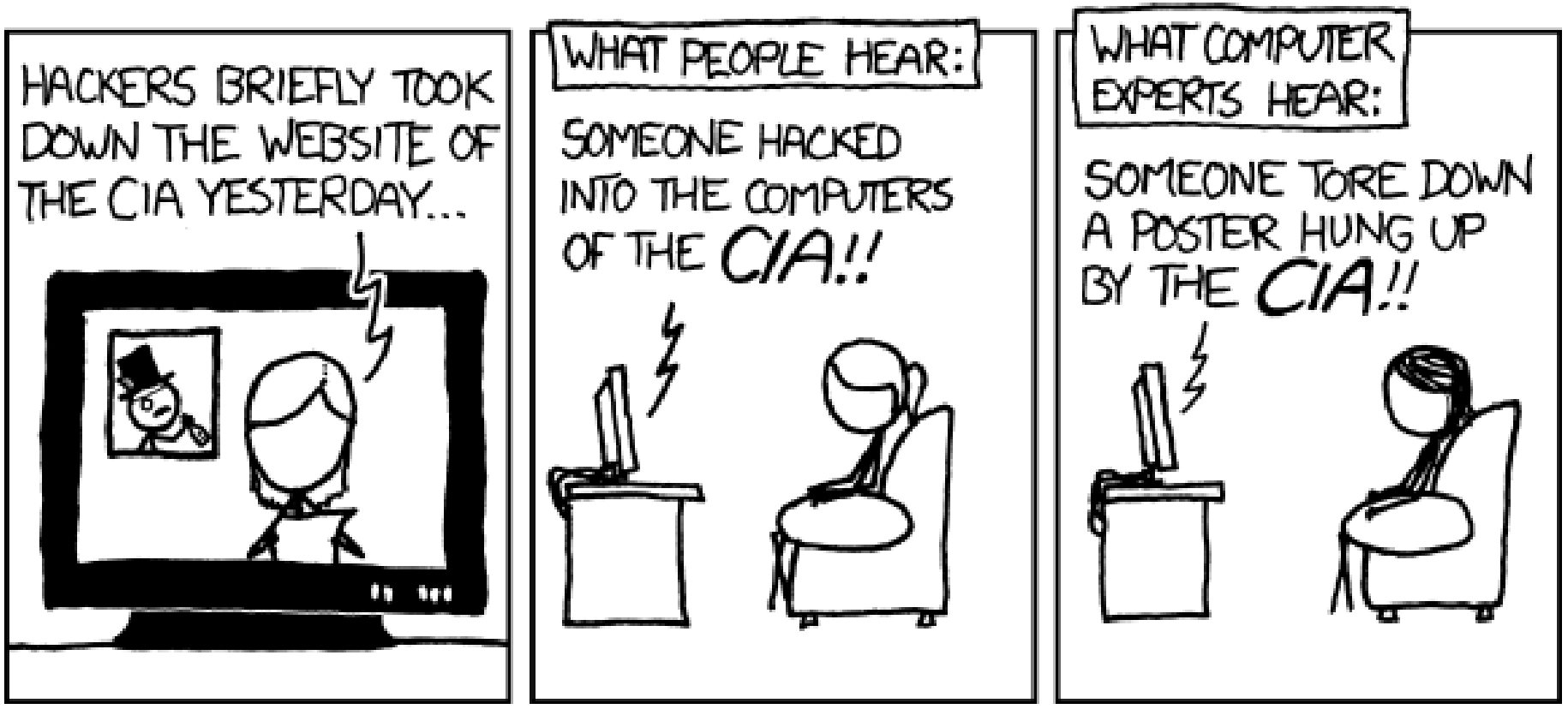
Anti-Phishing Working Group
APWG

So how could this be useful?

- I volunteered to lead an effort to write an “Internet Threat Assessment” to help our friends and us come up with useable stats, understand the risks, and educate justice ministries.

- This is live research; views welcome
  - ‘Live’ as in still changing

Anti-Phishing Working Group
APWG


HACKERS BRIEFLY TOOK DOWN THE WEBSITE OF THE CIA YESTERDAY...
WHAT PEOPLE HEAR:
SOMEONE HACKED INTO THE COMPUTERS OF THE CIA!!
WHAT COMPUTER EXPERTS HEAR:
SOMEONE TORE DOWN A POSTER HUNG UP BY THE CIA!!

The Top-Level Risks

- Financial Loss
- Data Misuse
  - Proprietary
  - Personal
- Content Controls
  - Content Restrictions
  - Access to Prohibited Content
- Business Interference
- Loss of Network Control
- Distribution of Prohibited Speech
- Loss of Privacy
- (Reputation)
- (People/Knowledge)

Digging into the Risks/Threats

- Financial Loss
  - Fraudulent transactions
  - Improper Credential Use
  - Laundering Activities
  - Extortion
- Proprietary Data Misuse
  - Possession
  - Corruption, Deletion
  - Misuse
  - Cyber Stalking
- Personal Data Misuse
  - Possession
  - Alteration
  - Misuse/Trafficing?
  - Falsification
- (Controlling Content)
- Access to Prohibited Content
  - Illegal porn
  - Pirated artistic works
- Distribution of Prohibited Speech
  - Hate speech
  - Death threats
  - Cyber-bullying
- Business Interference
  - DOS
- Loss of Network Control
  - Network Service Unavail – (DOS)
  - Network Compromised
- Loss of Privacy
  - Data Aggregation

Down to the Details

- Map the Risks to likely attacks
  - Using CAPEC mappings (initially)
- Describe how to determine, collect, report those attacks
  - Let people do it themselves
  - Maybe convince some collusion to get area statistics

Anti-Phishing Working Group
APWG

Risks vs Participants

Risk	Company	Government	Person	Alien
Financial Loss	✓	✓	✓	
Data Misuse	✓	✓		
Proprietary	✓	✓		
Personal	✓	✓	✓	
Controlling Content				
Access to Prohibited Content	✓	✓	✓	
Restrictions	✓	✓	✓	
Distribution of Prohibited Speech	✓	✓	✓	
Business Interference	✓	✓		
Loss of Network Control	✓	✓		
Personal Data Misuse		✓	✓	
Loss of Privacy	✓	✓	✓	

The Path Forward

- Flush out a document
  - Humorously called: Internet Risk Assessment
  - Why do a doc? Set the tone; define vocabulary
  - Use it as a tool to educate our ‘friends’
- Longer-term
  - Get more data (from others) into the stats
  - Provide our squishy-stats in a more general form so we track evolution.

Anti-Phishing Working Group
APWG

IEEE Stop eCrime Effort

- Run as a joint APWG-IEEE Industry Connection Program
- First Phase Deliverables
  - eCrime Glossary
  - Initial Guidance to Responders
  - Gap Analysis
  - List of Relevant Publications
- Try to get the ‘community’ to give us data.

Anti-Phishing Working Group
APWG

Our overall next steps

- Run an eCrime IODEF Pilot this fall to see if this all works
  - Multi-country, multi-language, multi-grief
  - Can we report and understand set scenarios
  - See if we can collect the new types of stats
- (unrelated) Figure out how to measure eCrime

Anti-Phishing Working Group
APWG

Other Event Info

- CrimeFighters want more data in our stats
  - Collect more data items
- As we slop data around, there's more to agree on…
  - Data Sharing Restrictions
  - The attack 'method'
  - The 'impact' of the attack
- LEO guidance on data to put in a report
- Watch ITU-related and other efforts

Anti-Phishing Working Group
APWG

Additional Information

- Special thanks to
  - Greg Aaron of Afilias
  - Rod Rasmussen of Internet Identity
- For the Global Phishing Report

- All reports are available on
  - http://apwg.org/resources.html

Anti-Phishing Working Group
APWG

Thank you

Pat Cain
Resident Research Fellow
APWG
pcain@antiphishing.org