Malware Mythbusting: Separating Fact From Fiction

Alex Kirk
Senior Research Analyst

VRT

SOURCEfire®

About the Sourcefire VRT

- Founded in 2001
- 25 team members
  - Core team members based in Columbia, Maryland (USA)
  - Additional offices in Seattle, Calgary, San Francisco, Poland, Italy, Germany
- Mission
  - Provide intelligence and protection to allow our customers to focus on their core business
- Responsibilities:
  - The public face of Sourcefire in the security community
  - Producing and publishing all Sourcefire, Snort, and ClamAV protection profiles
    - SEU, Snort, VDB, ClamAV
  - Threat Intelligence and Monitoring
  - ClamAV Development

Want To Work With Us?

VRT

SOURCEfire®

Lots of Marketing

- McAfee's report on Night Dragon
- Symantec blows up the industry on Duqu
  - Except, of course, for the part where it's actually lame, what with it being dead instantly
- Google and Aurora
- Every news event turns into SEO poison
- Huge claims, huge numbers
- So much to keep up with, what's real and what's not?

Malware Sandbox - Overview

- Sourcefire bought ClamAV in 2007
  - Patent trolls blocked commercializaiton
- What can an IDS company do with all those viruses?
  - Automate the production of real network traffic!
- Simple setup using VMWare ESXi & APIs
  - Yes, I know, not all malware runs in VMWare
  - It's a numbers game – we get plenty of executions

Malware Sandbox - Details

- Samples run for 200 seconds each
  - Partly necessity to deal with all the new files
  - Partly that, if a virus will do something interesting on the network, it will do it fast
- Been running for over a year and a half
  - Over 5 million samples analyzed
  - Over 2TB of network traffic generated
  - Approximately 90% of that traffic is HTTP

Myth #1: Malicious Traffic

- Most of the network traffic generated by malicious programs goes to malicious servers
  - Sure, they might connect to Google, or WhatIsMyIp, but then they go about their merry way
- Makes sense at a prima facae logical level
  - If you're busy doing nasty things, why would you bother with legitimate servers, except to spam, DDoS, or otherwise annoy them?
- Reports always have some evil server in China

Ad Servers, Oh My!

- ad.yieldmanager.com 2,157,795
- ad.foxnetworks.com 1,253,607
- ad.xtendmedia.com 1,208,539
- ad.103092804.com 1,182,779
- ad.adserverplus.com 851,470
- ad.globe7.com 767,009
- ad.scanmedios.com 755,443
- ad.media-servers.net 624,806
- um.simpli.fi is even an ad server!

Whitelisting – How To?

- Keeping a quality whitelist is one of the industry's largest problems
  - Sure, it's easy to add the servers you know
  - But there's plenty of contact with smaller, less known servers/companies
  - Subdomains can be infected, too – see *.dyndns
- 90% solution: use the Alexa Top 1 Million
  - Bigger than any other list out there
  - Free, easy to parse
  - Potential for false negatives, but not too bad

Whitelisting – Some Numbers

- 115,836 unique domain names whitelisted
- 178,331 unique domain names **NOT** whitelisted
- 12,389,762 total “white” hits
- 3,505,162 total “black” hits
- That’s approximately a 4:1 ratio of good:bad
- Plenty of possible theories
  - Click fraud
  - DDoS
  - Trying to look legitimate

URL Problems, Too

- masterbati.com/66c82da6-295b-488b-9229-9a4d90346e8e.ippi?g=66c82da6-295b-488b-9229-9a4d90346e8e
- pharmacymy.com/21d38eb0-e39e-42ff-8898-b9859a621d57.ippi?g=21d38eb0-e39e-42ff-8898-b9859a621d57
- marriedcheater.com/9ad86d88-2ba2-43a3-92c3-232a3063f001.ippi?g=9ad86d88-2ba2-43a3-92c3-232a3063f001
- Looks like a tracking code, certainly not legit
- Over 50,000 hits in the sandbox

Gray Area

- Alerts started swarming in from the field
- With old-school vulnerabilities, it's pretty straightforward
  - Are the triggering conditions met?
  - Is there shellcode?
- With malware traffic, often harder
  - Is the machine infected? Did your AV miss it?
  - How legit is that domain name?
- Matter of critical mass and trusted analysts

Another Candidate

- GET c.openclick.com/t?c=eJwNybEFwyBQgOG1uaUI1hqPdLipc5aWdpd4UDHYoOb96-R-8Fcn9Rolj-yr5jyuDmtI3QGHvK-YcxFDzlmKiJgM.jVwsA4X6HMhoWf2F7QyDUW22InO23O7xrZ--wq0l60=&requestId=ToS@VQrYakAAAFAsO78AAAAI
- Hundreds of hits in sandbox, appears focused
- Popped up recently, so it looks like a new tactic

Not Exactly Infected

- Immediate flood of false positives again
  - Including customers who had OS X and Linux machines generating alerts
- Events looked completely identical
- …but all of the Referer domains were SEDO link farm crap
- Difference is in the rate of these requests
  - SEDO domains do it once per load
  - Malware does it multiple times per minute

Really, China?

- beacon.sina.com.cn/a.gif?V=2&CI=sz:1024x768|dp:32|ac:Mozilla|an:MSIE|**cpu:x86|pf:Win32**|jv:1.3|ct:lan|**lg:en-us**|tz:4|fv:6&PI=pid:0-9999-0-0-1|st:1|et:1|ref:http%3A//blog.sina.com.cn/waiwaitu|hp:N|PGLS:|ZT:|MT:|keys:&UI=**vid:64.214.53.100**.20737127663002490 2|sid:64.214.53.100.20737127663002490 2|lv::1:1:1|un:|uo:|ae:&EX=ex1:|ex2:&gUid_1276594113000
- Turns out sina.com.cn is one of the biggest Chinese-language web properties on the planet

User-Agent: Install Stub

- User-Agent strings are a remarkable source of hilarity and detection
- If it's undocumented on Google, and it's common in the sandbox, flag it
- Immediately on release, hits came in from the field – on http://stats.norton.com
  - IP address lined up with public queries, so it's not spoofed
  - No idea what these binaries are doing hitting a Norton web server

Myth #2: C&C Locations

- “Most of the malware in the wild comes from shady countries like China, Russia, or Brazil”
- Trend Micro reported last month on Android malware using a Chinese blog as a C&C
- Operation ShadyRAT wasn’t explicitly identified as China, but that’s the consensus
- “Governmental rules and regulations … do not exist or are too weak within countries such as China, Russia, and Brazil, which has in turn led to these countries becoming global centers for Botnet administration.” – Kaspersky

Geolocation Tells The Tale

- Far and away, the leading country for hosting malicious servers is the United States
  - Total of 125,267 distinct, non-whitelisted IPs with geolocation data
  - 51,829 of those – 41.37% – are in the US
  - China comes in a distant 2nd at 14,708 (11.74%)
  - Even combined with Russia and Brazil, total is still a paltry 20,984 (16.75%)
  - Surprise: Germany is 3rd, with 7,086 (5.66%)
  - 9 of top 20 countries are in Europe; 14 of 20 are undisputedly 1st World countries

AS-Path Info

- Upgrading the malware sandbox
  - Props to go Dean Freeman (@wdf_vrt)
- New piece of data being collected: AS-Path
  - For those who may not be familiar, an AS is a collection of routers under common control
- Can span geographic boundaries
  - We all know nation-states are less important today, especially with the Internet
- Makes sense that policies within an AS would be similar, so it's a good way to analyze

AS-Path Results

- Disproportionately skewed towards a small number of AS-Paths
  - 46.76% of all malicious domains carried over just two distinct AS-Paths (13149, 11328)
- Number one hit: Secure Hosting, Ltd., “The Truly Global Offshore Hosting Provider”
- Second-largest provider – SoftLayer Technologies, Inc. – is just a big provider
- APNIC has 5,875 total hosts over 202 total paths (11%)
  - Lines up with other information about Asian threats

Myth #3: Malicious Domains Are Transient

- Everyone talks about Fast Flux domains
  - "[A] growing, sophisticated technique called fast-flux service networks which we are seeing increasingly used in the wild . . . with public DNS records that are constantly changing, in some cases every few minutes." – Honeynet Project
  - "Because the hosts and controllers in a botnet are operating from captured resources, botnets are dynamic and short-lived, which makes them difficult to defeat." – Cisco
- My research looks for generic detection because of the constant "whack-a-mole" game

ilo.brenz.pl 0wns you

- Looking through samples from my Binary C&C over HTTP presentation, one domain kept recurring: ilo.brenz.pl
- First appeared in my sandbox April 7, 2010
  - Which is essentially as soon as it got running
- Last appeared November 14, 2010
- Clearly known within the space now – I've presented, blogged, etc. about it
  - But nobody seems interested in taking it down
  - If you are, let me know!

Plenty of Persistent Domains

- irc.zief.pl: 2010-04-07 - 2011-11-05 (577 days)
- play.unionsky.com: 2010-04-07 - 2011-10-22 (563 days)
- adsfac.us: 2010-04-07 - 2011-10-17 (558 days)
- ktr.t134.net: 2010-04-08 - 2011-08-16 (495 days)
- tooldawn.com: 2010-04-07 - 2011-05-25 (413 days)
- ayb.host127-0-0-1.com: 2010-04-07 - 2011-05-24 (412 days)
- a.95622.com: 2011-01-09 - 2011-11-06 (301 days)
- kzqinferno.com: 2010-04-08 - 2010-09-04 (149 days)
- 113128url.cptgt.com: 2010-04-07 - 2010-07-12 (96 days)

Exceptions, Not The Rule

- Overall picture is quite distinct
- 261,923 unique domain names
  - 105,308 (~40%) appear only one time!
  - 37,926 (~14.5%) appear only twice
  - Less than half of the domains are really multi-use
- What's the threshold for caring about a specific domain, as a defender?
  - If it's 50 hits, you only have 18,733 hosts (~7%)
  - If it's 100 hits, you only have 10,900 hosts (~4%)

Time Frames Agree

- Discarding one- and two-time appearances, we see a median “TTL” of 12.89 days for domains appearing in my sandbox
- For domains we care about (50+ hits), number climbs to ~118 days
- Shoots up to ~296 days for hosts with 1,000+
- Overall lesson: malicious domains are **both** transient and persistent
  - Once you cross a certain threshold of activity, the likelihood that a domain is here for good increases drastically

IP Addresses vs. Domain Names

- IP addresses are considerably more transient than domain names
  - Median TTL for IP addresses on domains hit at least 50 times: 11.82 hours
  - Range is very broad: hundreds of days on the high end down to one-time-use IPs on the low end
    - Saw 158,855 IP addresses only once for all domains with at least 5 hits in the sandbox
  - Behavior varies wildly even for a given domain name
    - ilo.brenz.pl, for example, has 299 unique IPs associated
    - 24 one-time uses, but high end is 492 days

Myth #4: Porn Is Full Of Malware

- “Honey, I swear, I was only looking at that page for research purposes!”
- Everybody knows somebody who went to a porn site and got owned six ways to Sunday
- Decided to investigate this after a customer inquiry
  - “Can you guys provide coverage for the Sifiref malware we found with MD5 71b11fa18b3099c35b7c96c5a2e1e00a?”
  - Original filename: dog-doing-girl.avi.exe

File Names

- 30,275,252 files in the ClamAV database came in with an in-the-wild filename attached
  - Astonishingly small 674 (0.0002%) have “fuck”
    - _how_to_fuck_cool_girls_-_tutorial.exe_
    - Obama_caught_fucking_17yr-old-November-13-2008.scr
  - Similarly small 1,874 (0.0006%) have “sex”
    - Sex_Drugs_and_Cocoa_Puffs_split_000.html (WTF?!?!?)
    - False Positive: WindowsExplorer.adml
  - Tiny numbers for media types: 1,111 “.avi”, 572 “.wmv”, 263 “.mpg/.mpeg”, 3,689 “.mp3”
  - Compare to “.exe” – 7,760,179 (25.6%), and my personal favorite “.rar” - 1,034,050 (3.4%)

Domain Names

- Similarly small numbers for domain names
  - 586 contain “fuck” (0.22%)
  - 2,660 contain “sex” (1.01%)
  - 315 contain “dick” (0.12%) vs. 159 “pussy” (0.06%)
- Real humor is if you look for the word “porn”
  - 1,136 domains with “porn” are not whitelisted
  - 1,135 domains with “porn” **are** whitelisted
  - Makes you wonder what people are doing with their botnets
- Personal favorite: customer asks if hit for “smellypussy.info” is a false positive

All About Economics

- Filename counts start going up when you offer something
  - 7,397 have “free”
  - 809 have “money”
- Domain names are similar
  - 1,408 have “cash” (0.54%)
  - 962 have “money” (0.37%)
  - 5,708 have “free” (2.18%)
  - 12,931 have “download” (4.94%)
    - 9,321 of these are whitelisted – backs up the theory that lots of malware is just installing apps they get paid for

Myth #5: Spambots, Spambots, Everywhere

- Everybody's heard of Rustock
  - Averaged 192 spams per machine per minute
- Sending spam is a good paying gig – goes along well with the economics of malware
- Sourcefire customer several years ago who was able to stop plans for a massive network upgrade when we helped them find 12 compromised spambots

Most Malware Doesn't Spam

- Recently started tracking stats on port connections – data on 733,875 samples
  - Only 26,654 of them try to talk to port 25 (3.6%)
- Huge range in the amount of mail they attempt to send
  - One had 10,548 connections in 121 seconds
  - Had 2,788 with fewer than 5 connections in the 200-second recording period
  - Lines up nicely with the concept of High Volume Spammers vs. Low Volume Spammers (see http://www.eecs.umich.edu/~zmao/Papers/leet08.pdf)

Spam As A Secondary Behavior

- Many of the samples sending spam seemed to do so as an afterthought
- Even sample with the most outbound connections started with a query returning

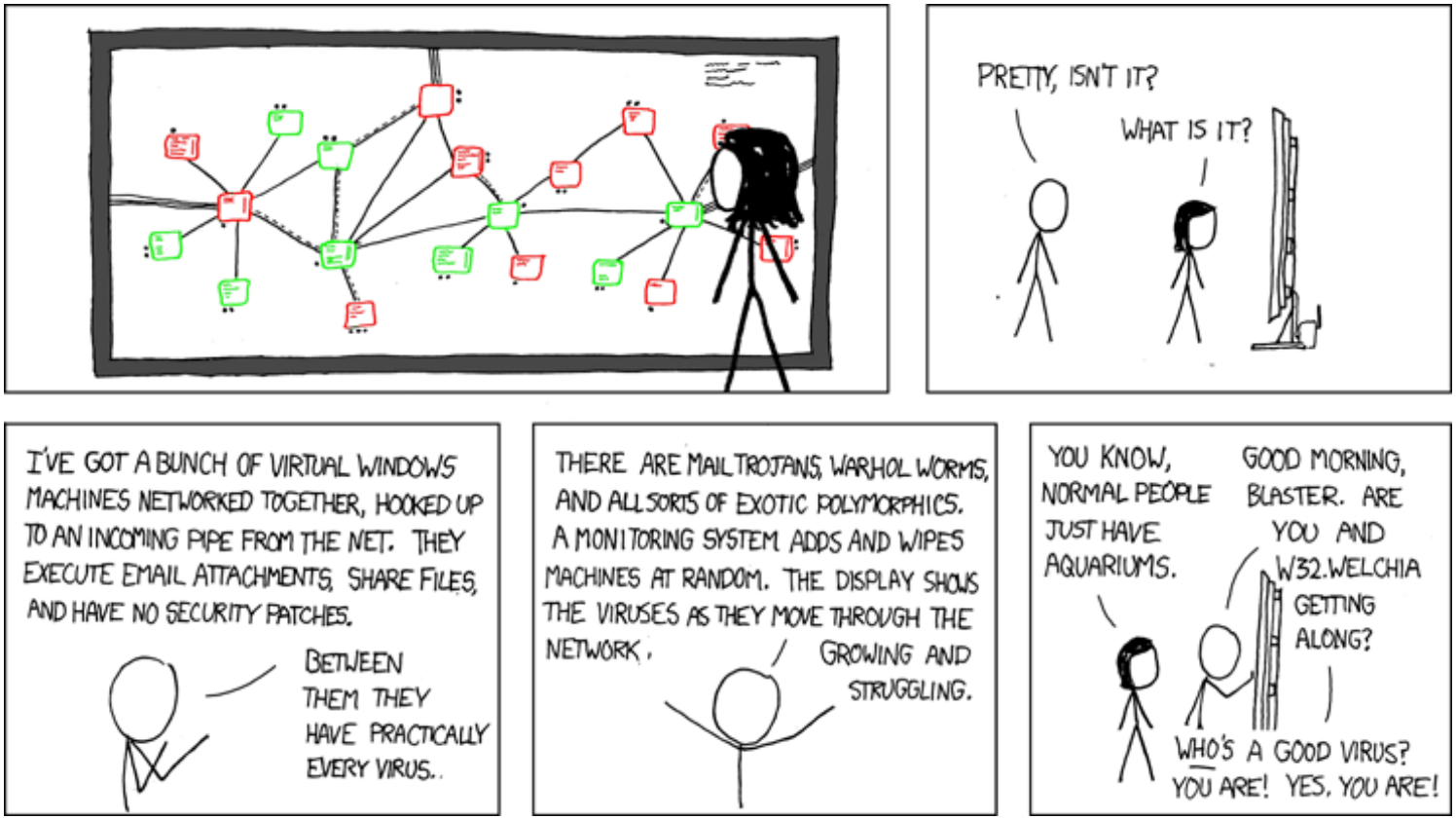
```
GET
/md5.php?hdd=&pc_adi=&fic=taktuk&tim
e=7/13/2011%205:25:59%20AM HTTP/1.1


#komut#udpflood 82.165.137.159 25
10000 300#komutson#
```

Spam As A Secondary Behavior (con't)

- Some samples fall back on other behavior if spamming isn't working
  - After two connections got RSTs, fell back onto a binary protocol of some sort on port 8000
  - Received large stream from remote host, then started making Gnutella-related HTTP requests, followed by file transfers over Gnutella
- Of the 26,654 that talk on port 25, 26,123 talk on other ports (98%)
  - 22,924 of them talk on port 80 (86%)
  - Huge number of other ports hit

Myth #6: If You've Got One Virus…

- Raise your hand if you've got an older relative whose machine had multiple viruses at once



Immunet Stats To The Rescue

- Sourcefire bought a cloud-based AV company, Immunet, in January
- Lots of cool real-time stats available
- During a sample period in July…
  - 557,285 total machines talking to the cloud
  - 469,601 of them were clean (84.27%)
  - Of the infected ones, though…
    - 26,277 had a single infection (29.91%)
    - 15,498 had exactly two infections (17.67%)
    - 45,909 had three or more infections (52.42%)

Lots of Droppers/Downloaders

- During the same sample time frame, out of the top 40 most frequently detected infections, 3 were droppers or downloaders
- Made up a significant portion of the volume of overall infections
  - Clam.Heuristic.HTML.dropper was #1 sample
  - 31,804 of 140,200 total infections in the top 40 were from these three pieces of malware
- Makes perfect sense from an operational standpoint – you don't need your own malware, you just drop someone else's

Myth #7: Too Smart To Be Obvious

- All of the malware reports you hear about today are “APT”
  - Dissecting run-of-the-mill malware is boring
  - Directed, state-sponsored attacks are sexy
- Average person doesn’t understand computers – let alone exploitation, botnets, etc.
- Besides, if it were ridiculously easy to find malware, wouldn’t we not have the problem we face today?

Yeah…Not so much

- Personal favorite: using a dumb, easily detectable channel to spread dumb, easily detectable Facebook-related scams:

```
PRIVMSG #!hot! :[HTTP]: Updated HTTP
spread message to "did you see this?
OMG!!
http://www.barackobama.ir/facebook-
profile-pic-18262-JPEG"

PRIVMSG #!hot! :[MSN]: Updated MSN
spread message to "LOL
http://www.barackobama.ir/facebook-
profile-pic-63695-JPEG"
```

User-Agent Strings

- Hey, let's announce our malicious intent directly in the requests we send to the Internet!

```
POST /adm/index.php HTTP/1.1
Content-Type: application/x-www-form-urlencoded
User-Agent: malware
Host: prellerstay.co.za
Content-Length: 5
Connection: Keep-Alive
Cache-Control: no-cache

logs=
```

More User-Agent Strings

- The list is spectacular once you start looking
  - `GenericHttp/VER_STR_COMMA`
  - `-`
  - `hello`
  - `Trololo`
  - `microsoft`
  - `123232710181`
  - `privacyscan_agency`
  - `super browser`
  - `PopRocks`
  - `User Agent`

URLs Full of Dumb

- /data/img?mt_id=107042&mt_dcid=**<IN SERT_DATA_CAMPAIGN_ID>**&v1=&v2=&v3= &s1=&s2=&s3=&mm_bnc HTTP/1.0
- 174.139.2.234/Go.ashx?**Mac=00:0C:29 :FC:3F:0F**&UserId=54&Bate=2.0
- 62.90.136.249/in.php?affid=64100&u rl=5&**win=Windows%20XP+2.0**&sts=
- b.whataboutadog.com/132/checkin.ph p?cid=45799573&aid=10279&time=C:\\ WINDOWS\\TEMP\\\\1318040819.dat&fw =0&v=132&m=0&vm=0

URLs Hiding In Plain Sight

- Plenty of malware uses the same URL scheme over and over and over again
  - `ishlseek.cn/stat2.php?w=170&i=0728c2492c74c2490dd14f3e86d2d863&a=21` – Siferef (9,616 hits)
  - `quadroprivate.net/admin/gate.php?hwid=3821060361&pc=VIRUSCLONE28&localip=192.168.10.165&winver=Windows%20XP%20Professional%20x32` – Zeus (16,116 hits)
  - `motuh.com/borders.php` – Zeus (45,253 hits)

Close, But No Cigar

- POST to `http://gahyob.com/login.php`
- Body: `C9 97 A2 F3 7E 37 CB 7E 27`
- Super-sneaky – you can't just block login.php, and that binary data would probably change over time or otherwise be hard to catch
- Except the part where it sends 20 requests like this in 2 seconds, plus hundreds of DNS queries and NetBIOS name queries
  - Useful piece of malware detection: if a host makes more than, ~25 DNS queries in a second

Content-Type Mismatches

- GET amazzamboni.sites.uol.com.br/modula.ico
  - Executable file is returned without any encoding
- GET jkury.sites.uol.com.br/md30.css
  - `......dd.*.......r:\nova versao do sistema\imagens\barrafis.bmpw..>EC2.. ...x...1..P...., 0 ...`g.......%0`
- POST os.downloadapi.com/v1.0.1/?v=2.0…
  - `Conent-Type: text/html`
  - `.l--..CT..N ..ME.o_C.."-.sUS.l-- ..EC..ON.`>`
  - `.q!-..EC..ON..AM.p"I..T_..DE.`->.G`

I Promise That's Executable

- Some attackers have apparently figured out that it's easy to find a PE file coming down
- GET transfersakk.com/logo.png?tq=gKZEtzo…
  - Content-Type: image/jpeg
  - $..`... m..`... ...`... )..`... i..`... i..`... i..`... i..`... g..n....H<.,.!TH.. .rogR.. .annO..b. ruNI.n@DOS...d..
- GET cdn.dli.trymedia.com//r/release/valusoft/60m_i_1f/jddgsetup.exe
  - MZ.....!..L.!Win32 reqd.$....ACTIVEMARK_MAGIC_TO_CHECK_INSTALLER

The Purloined Config File

- Sometimes it's just easy to hide in plain sight
- GET dl.dropbox.com/u/32472005/index.html
  - `{---CONFIG - INICIO - GENERATION---}`
  - `[Tabela]=jhdjffjiejgxjhejghjgxjfejghjfhjgi)=[fim]`
  - `[DATA]=jaxijigjaxhjaxajabxjigjaaxjaxxjaadjaxajdi)=[fim]`
  - `[Senha]=jgajhbjfejghjfhjfijdijexjeajebjecjed)=[fim]`
  - `[Endereco]=jaxxjihjaaejaacjdhjdhjdhjeejdfjaaijaxdjaaejaxajaadjaahjaxejaxxjaaajaadjdfjiijaaajaxi)=[fim]`

“Malware” Is As Varied As “Software”

- If there’s anything we can take away from studying malware, it’s that it runs the gamut
  - Sometimes it’s incredibly well-done and hard to detect…sometimes it’s clumsy and obvious
  - End goal of any given infection can range from spam to DDoS to… “WTF are they doing there?”
- This amount of variety is what keeps defenders busy – there’s always some novel technique
- As long as there’s profit to be made, malware volume will keep increasing

Contact/Follow Us

- The VRT Blog
  - http://vrt-blog.snort.org
  - Technical and policy analysis
- Twitter
  - ~2000 followers (VRT_Sourcefire)
  - Personal account (alexgkirk)
- Labs
  - http://labs.snort.org
  - All the VRT cool stuff
- Email: alex.kirk@sourcefire.com