DGAs and Threat Intelligence

John Bambenek – Fidelis Cybersecurity Threat Research Team

HITCON 2015

Intro

- President and Chief Forensic Examiner for Bambenek Consulting
- Adjunct Faculty in CS Department at the University of Illinois at Urbana-Champaign
- Producer of open-source intel feeds
- Work with companies and LE all over the world to address growth in cybercrime

About Threat Intelligence

- Information is a set of unprocessed data that may or may not contain actionable intelligence.
- Intelligence is the art of critically examining information to draw meaningful and actionable conclusions based on observations and information.
- Involves analyzing adversary capabilities, intentions and motivations.

Adversarial objectives

- Here we are generally talking about organized crime, usually financially motivated.

- What we know:
  - Highly rational actors
  - May hire “outside experts” for specific tasks
  - Generally technological sophisticated
  - Desire to remain “quiet” and resilient

My Objectives

- Any good intelligence program needs to also analyze your own objectives.
- I investigate and try to disrupt criminal networks, so my objective is externally focused.
- These efforts are directed toward “criminal” actors, nation-state / APT threats would require a different focus.
- Most people are defensively focused so their information priorities are different.

Malware C2 Network Types

- Static IP / Hostname Lists
- Proxied C2s
- Dynamic DNS
- Fast Flux / Double Flux Networks
- Domain Generation Algorithms
- Tor / i2p hidden services

A History of Malware C2 Networks

- An adversary wants to persist over the long-term and make their networks more resilient against enforcement actions.
- Domains tend to be easier to take down the IPs due to avoidance of jurisdictional issues.
- Development over time will largely show adversaries have acted in ways to ensure increased resiliency.
- We can continue to map forward over time where they are likely to go in the future as a result.

Use of Multiple Techniques

- The most resilient malware C2 use multiple methods of callback.
- Static Lists
- DGAs
- Tor/I2P
- If one or two are blocked, still able to control machine.

- To affect a takedown, need to block all means of communication and updating victim machines.

Domain Generation Algorithms

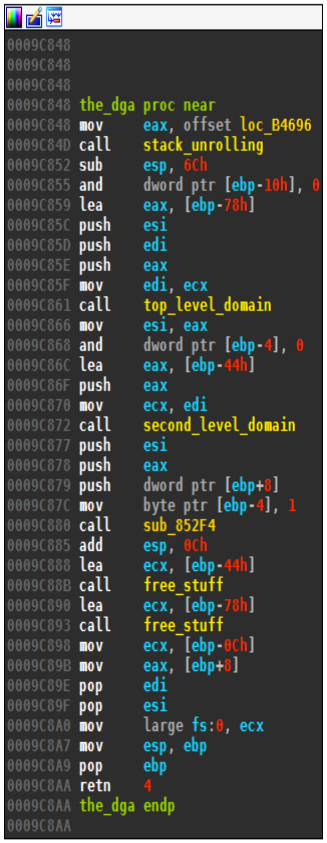
- Usually a complex math algorithm to create pseudo-random but predictable domain names.
- Now instead of a static list, you have a dynamic list of hundreds or thousands of domains and adversary only needs to have a couple registered at a time.
- Can search for “friendly” registrars to avoid suspension.

Reverse Engineering DGAs

- Many blog posts about reversing specific DGAs, Johannes Bader has the most online at his blog:
  - Johannesbader.ch

- No real shortcuts except working through IDA/Debugger and reversing the function.
  - Look for functions that iterate many times.
  - There will be at least a function to generate the domains and a function to connect to all of them to find the C2.
  - As with all reverse engineering, be aware of obfuscation and decoy code meant to deceive you.

Reversing DGAs Example



```
0009C848
0009C848
0009C848
0009C848 the_dga proc near
0009C848 mov     eax, offset loc_B4696
0009C84D call    stack_unrolling
0009C852 sub     esp, 6Ch
0009C855 and     dword ptr [ebp-10h], 0
0009C859 lea     eax, [ebp-78h]
0009C85C push    esi
0009C85D push    edi
0009C85E push    eax
0009C85F mov     edi, ecx
0009C861 call    top_level_domain
0009C866 mov     esi, eax
0009C868 and     dword ptr [ebp-4], 0
0009C86C lea     eax, [ebp-44h]
0009C86F push    eax
0009C870 mov     ecx, edi
0009C872 call    second_level_domain
0009C877 push    esi
0009C878 push    eax
0009C879 push    dword ptr [ebp+8]
0009C87C mov     byte ptr [ebp-4], 1
0009C880 call    sub_852F4
0009C885 add     esp, 0Ch
0009C888 lea     ecx, [ebp-44h]
0009C88B call    free_stuff
0009C890 lea     ecx, [ebp-78h]
0009C893 call    free_stuff
0009C898 mov     ecx, [ebp-0Ch]
0009C89B mov     eax, [ebp+8]
0009C89E pop     edi
0009C89F pop     esi
0009C8A0 mov     large fs:0, ecx
0009C8A7 mov     esp, ebp
0009C8A9 pop     ebp
0009C8AA retn    4
0009C8AA the_dga endp
0009C8AA
```

From http://johannesbader.ch/2015/05/the-dga-of-ranbyus/

Types of DGAs

- Almost all DGAs use some time of “Seed”.
- Types:
  - Date-based
  - Static seed
  - Dynamic seed
- Seed has to be globally consistent so all victims use the same one at the same time.

Other DGA Hardening Techniques

- Choice of gTLD matters.
  - Some doing have WHOIS protection, make it hard to sinkhole
- Rotation of seeds
- Some malware has rudimentary “sinkhole awareness”
- Adversarial objectives: Maintain control, limit surveillance

Examples of select DGAs - Cryptolocker

- Used 1000 domains a day across 7 gTLDs. Order domains are queries in based on GetTickCount()
- Eerily similar to DGA described in Wikipedia article on DGAs.
- Used previously by Flashback OSX Worm.
- Never changed during the life of the malware campaign.
- Successfully taken down in June 2014.
- Special thanks to Vladimir Kropotov for his help on this!

Examples of select DGAs - Cryptolocker

- Intel conclusions:
  - Likely written by a third party.
  - Went days without a domain registered, actor wanted to get paid but wasn't overly concerned about keeping everything going 24x7.
  - Tended not to shift registrar even after domains were suspended.
  - Likely didn't monitor his own domains because the ratio of malicious to sinkholed domains was about 1:125.
    - Way to go on the OPSEC good guys. ☺D

Examples of select DGAs - Tinba

- Generated 1,000 domains a day, not date-seeded.
- Seeded by an initial hostname and a defined gTLD (one or more).
- Changes seeds often and tends to update already infected machines.
  - At least sinkholing tended to be ineffective for more than a few days.

Examples of select DGAs - Tinba

- Intelligence conclusions:
  - These guys care about their infrastructure.
  - Likely they are actively monitoring to see when their DGA is cracked and adapting accordingly.
  - Likely they wrote DGA with this kind of flexibility in mind.

Examples of select DGAs - Bedep

- Uses a dynamic seed – currency exchange values for foreign currency
  - European Central Bank produces daily feeds of the rates, this is used as source data.

- Impossible to predict in advance even though code to generate the domains is publicly available.
  - http://asert.arbornetworks.com/bedeps-dga-trading-foreign-exchange-for-malware-domains/

Examples of select DGAs - Bedep

- To date, all successful takedowns (and for that matter unsuccessful takedowns) seized malicious DGA domains in advance while simultaneously suspending current domains.
- This would decapitate a botnet if and only if there was no fallback mechanism to reach the C2 (i.e. tor).
- How can you do this for Bedep when you don't know future currency values?
  - Intelligence conclusion: this is obviously an intentional choice.

Examples of Select DGAs – Matsnu and Rovnix

- Matsnu and Rovnix both use wordlists to generate domains that appear like they would be “reasonable”. Rovnix uses the US Declaration of Independence.
- Problem is that sometimes there is collisions with real domains.

teamroomthing.com,Domain used by matsnu DGA for 16 Aug 2015,2015-08-16
transitionoccur.com,Domain used by matsnu DGA for 16 Aug 2015,2015-08-16
windbearboxreceive.com,Domain used by matsnu DGA for 16 Aug 2015,2015-08-16
winner-care-sir.com,Domain used by matsnu DGA for 16 Aug 2015,2015-08-16

theirtheandaloneinto.com, Domain used by Rovnix DGA
thathistoryformertrial.com, Domain used by Rovnix DGA
tothelayingthatarefor.com, Domain used by Rovnix DGA
definebritainhasforhe.com, Domain used by Rovnix DGA
tosecureonweestablishment.com, Domain used by Rovnix DGA

What the use of DGAs gives the good guys

- Easy ability to sinkhole unused DGA domains to gather additional intelligence.
- Easier ability to do bulk takedowns.
  - *IF* you can predict domains in advance.
- The ability to surveil malicious infrastructure in near real-time.

What the use of DGAs gives the good guys

- The use of DNS in malware severely limits the ability of the adversary to play games.
  - They need the world to be able to find their infrastructure in order to control victim machines.

- Even when DGA changes, the adversary **tends** not to immediately change their infrastructure too.
  - Allows for the use of passive DNS to see the extent of DGA changes.

Sinkholing

- Many security companies do this.
- Many want to hide the fact they do this.
- Most adversaries aren’t stupid enough to not notice.
- Remember, Cryptolocker we had 125 or so sinkholed domain for every 1 malicious domain.

Feed generation on DGAs

sjuemopwhollev.co.uk,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13
meeeqyblgbussq.info,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13
ntjqyqhqwcwost.com,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13,
nvtvqpjmstuvju.net,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13
olyiyhprjuwrsl.biz,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13
sillomslltbgyu.ru,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13
gmqjihgsfulcau.org,Domain used by Cryptolocker - Flashback DGA for 13 Aug 2015,2015-08-13,

From here you could easily feed this into RPZ or other technology to protect your organization. But we want more.

How to set up surveillance on a DGA

- Easy to set up with shell scripting and a non-t1.micro AWS instance.
- Requires GNU parallel and adns-tools to handle bulk DNS queries.

DGA surveillance

- Pre-generate all domains 2 days before to 2 days in future.
- Pipe all those domains into adnshost using parallel to limit the number of lines.
- Able to process over 700,000 domains inside 10 minutes (and I'm not done optimizing).

parallel -j4 --max-lines=3500 --pipe adnshost -a -f < $list-of-domains | fgrep -v nxdomain >> $outputfile

Tinba DGA feed example

bcldleeivfii.com,Domain used by tinba,2015-08-15 04:15
bfoxyvqtolmn.com,Domain used by tinba,2015-08-15 04:15
cniuybkgxelo.com,Domain used by tinba,2015-08-15 04:15
dgscodhlppkk.com,Domain used by tinba,2015-08-15 04:15
djnmllhgwtff.net,Domain used by tinba,2015-08-15 04:15

This is active not-known-sinkhole domains current resolving.

A note on intelligence bias

- How we look at threats and what we tend to do with information will affect how we gather intel and how we process it.
- I tend to be involved in takedowns so I am generally uninterested in sinkholes.
- If you protect an organization, however, you care about your client machines reaching out to sinkholes because they are still infected.

Tinba IP list

5.230.193.215,IP used by tinba C&C,2015-08-15 04:15
54.72.9.51,IP used by tinba C&C,2015-08-15 04:15
95.163.121.201,IP used by tinba C&C,2015-08-15 04:15
104.27.169.12,IP used by tinba C&C,2015-08-15 04:15
104.28.13.180,IP used by tinba C&C,2015-08-15 04:15

Seems like a good list to firewall…
More on that in a moment.

Should also check NS info too

5.230.193.215,Nameserver IP used by tinba C&C,2015-08-15 04:21
5.45.69.31,Nameserver IP used by tinba C&C,2015-08-15 04:21
46.166.189.99,Nameserver IP used by tinba C&C,2015-08-15 04:21
50.7.230.28,Nameserver IP used by tinba C&C,2015-08-15 04:21
54.75.226.194,Nameserver IP used by tinba C&C,2015-08-15 04:21

Should also check NS info too

ns3.freedns.ws,Nameserver used by tinba C&C,2015-08-15 04:21
ns4.freedns.ws,Nameserver used by tinba C&C,2015-08-15 04:21
ns-canada.topdns.com,Nameserver used by tinba C&C,2015-08-15 04:21
ns-uk.topdns.com,Nameserver used by tinba C&C,2015-08-15 04:21
ns-usa.topdns.com,Nameserver used by tinba C&C,2015-08-15 04:21

With these two data points you can usually quickly validate what is a sinkhole and what is likely malicious and bears further investigation.

DGA Surveillance

- Looking at those four data points you now have solid information to make decisions based on the data.
- You could block domains/IPs.
- You could block nameservers (some times).

Adversarial Response

- Adversaries know we are doing this.

- In response:
  - They change seeds frequently
  - They have non-DGA communication mechanisms
  - They engage in counterintelligence

Counterintelligence

- The tactics by which an adversary thwarts attempts to gather information on itself.
- Remember the domain and IP lists before?
- What if an adversary registers domains that they aren’t using?

Counterintelligence – or worse version

- What if adversary knows you pump these IP lists directly into your firewall (and I know people do this with my feeds)?
- Anyone recognize these IP addresses? They are the DNS Root Servers

198.41.0.4
192.228.79.201
192.33.4.12
199.7.91.13
192.203.230.10
192.5.5.241
192.112.36.4
128.63.2.53
192.36.148.17
192.58.128.30
193.0.14.129
199.7.83.42
202.12.27.33

Counterintelligence – or worse version

- Taking action on information without analysis is generally a bad idea, especially when the information is under the complete control of the adversary.
- This is why intelligence analysis is so important.
- (I whitelisted the root servers after I noticed an adversary tried to do an attack similar to this.)

Whois Registrar Intel

- Often actors may re-use registrant information across different campaigns. There may be other indicators too.
- Sometimes **even with WHOIS privacy protection** it may be possible to correlate domains and by extension the actor.
- Most criminal prosecution in cybercrime is due to an OPSEC fail and the ability to map backwards in time of what the actor did to find that fail that exposes them.

Whois Info

- Many actors will use WHOIS protection… some just use fake information.
- “David Bowers” is common for Bedep.

ubuntu$ grep "David Bowers" *.txt | grep Registrant

whois-bfzflqejohxmq.com.txt:**Registrant** Name: David Bowers
whois-demoqmfritwektsd.com.txt:**Registrant** Name: David Bowers
whois-eulletnyrxagvokz.com.txt:**Registrant** Name: David Bowers
whois-lepnzsiqowk94.com.txt:**Registrant** Name: David Bowers
whois-mhqfmrapcgphff4y.com.txt:**Registrant** Name: David Bowers
whois-natrhkylqoxjtqt45.com.txt:**Registrant** Name: David Bowers
whois-nrqagzfcsnneozu.com.txt:**Registrant** Name: David Bowers
whois-ofkjmtvsnmy1k.com.txt:**Registrant** Name: David Bowers

David Bowers

bfzflqejohxmq.com,Domain used by bedep (-4 days to today),2015-08-16
eulletnyrxagvokz.com,Domain used by bedep (-4 days to today),2015-08-16
natrhkylqoxjtqt45.com,Domain used by bedep (-4 days to today),2015-08-16
nrqagzfcsnneozu.com,Domain used by bedep (-4 days to today),2015-08-16

But why stop with just known DGAs, what other domains are associated with “David Bowers”?

David Bowers

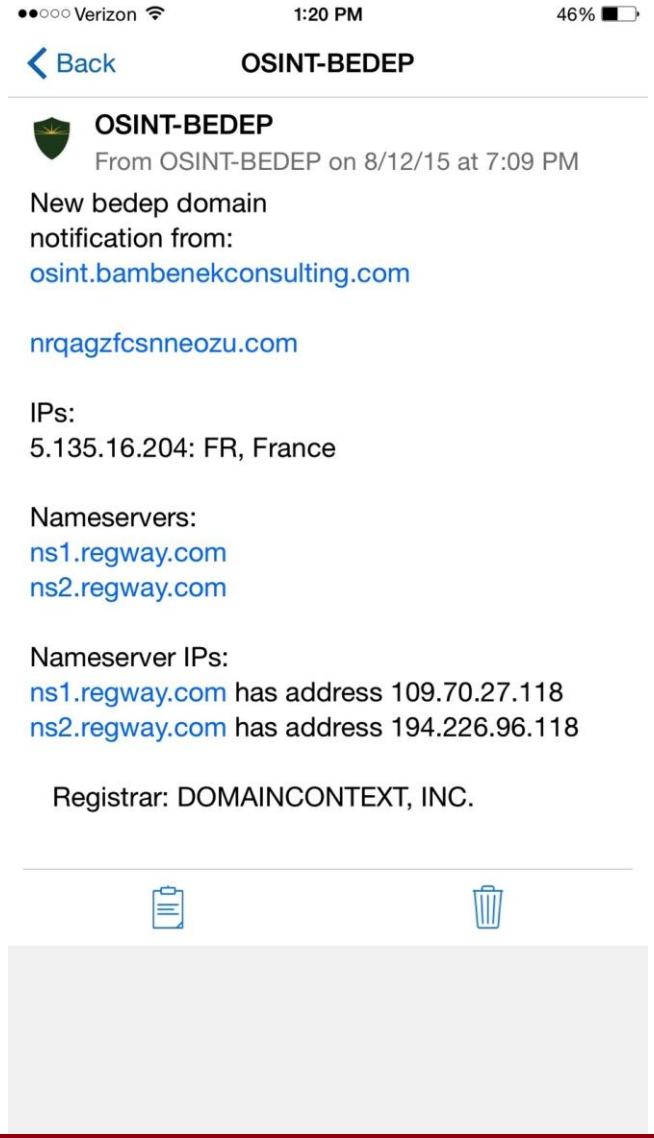
029uhbsdfisjdj4.in	2015-02-25	--
298dkoaldjfiow-yets.in	2015-03-18	--
37aodjdopeoi.in	2015-03-17	--
37kdospwmeop.in	2015-03-25	--
3875jncioeprk.us	2015-03-31	--
394iopwekmcopw.com	2015-01-19	DOMAINCONTEXT, INC.
78i2jpaosieu.in	2015-05-07	--
7u2yopwjh.in	2015-05-07	--
82hasyqtwq.in	2015-05-13	--
82kolesan.in	--	--
a4egjph0jy.us	2015-07-25	--
aachurill.com	2015-04-30	DOMAINCONTEXT, INC.
aachurill.in	2015-04-22	--
abloovoades.com	2015-03-04	DOMAINCONTEXT, INC.
abozpkdiowe28a9.in	2014-12-08	--
absuawpcphiwkkhj8.com	2015-04-19	DOMAINCONTEXT, INC.
ac38vplik8p.com	2015-07-10	DOMAINCONTEXT, INC.
accident-muscle.com	2015-03-05	DOMAINCONTEXT, INC.
ace-nate-rade.in	2015-03-24	--
aderradpow.in	2014-10-13	--
adgeziklopas.ws	2015-02-27	PDR Ltd. d/b/a PublicDomainRegistry.com
adoncorst.com	2015-04-29	DOMAINCONTEXT, INC.

Surveillance is nice, what about notification?

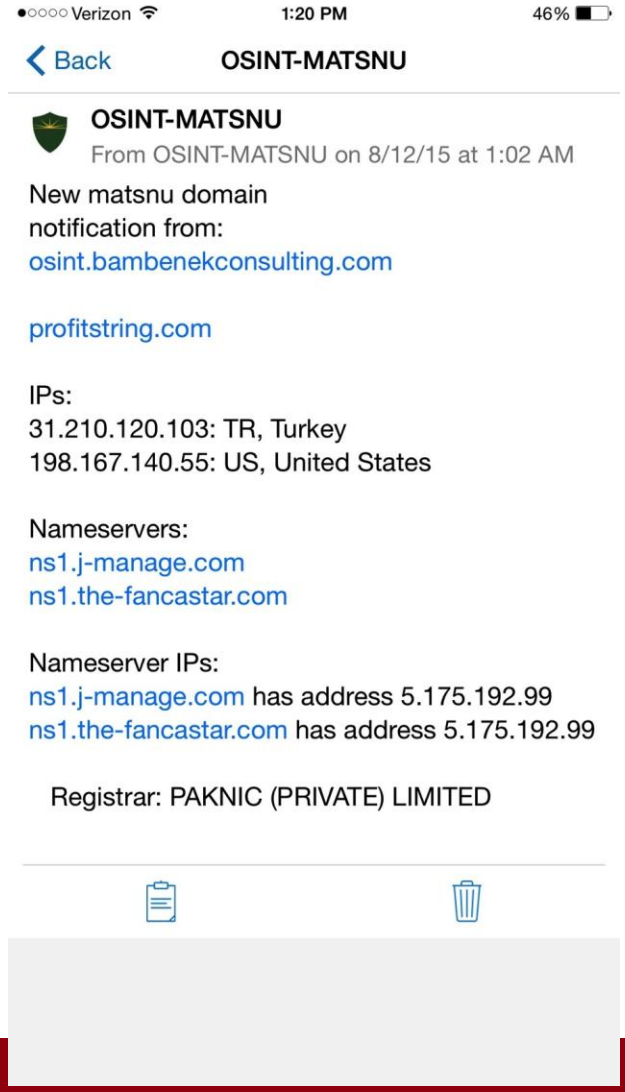
- Creation of feeds and intake is still a passive tactic.
- It is all possible to automate notifications when key changes happen to allow for more near-time actions.
- This uses the Pushover application (Apple and Google stores) which has a very simple API.

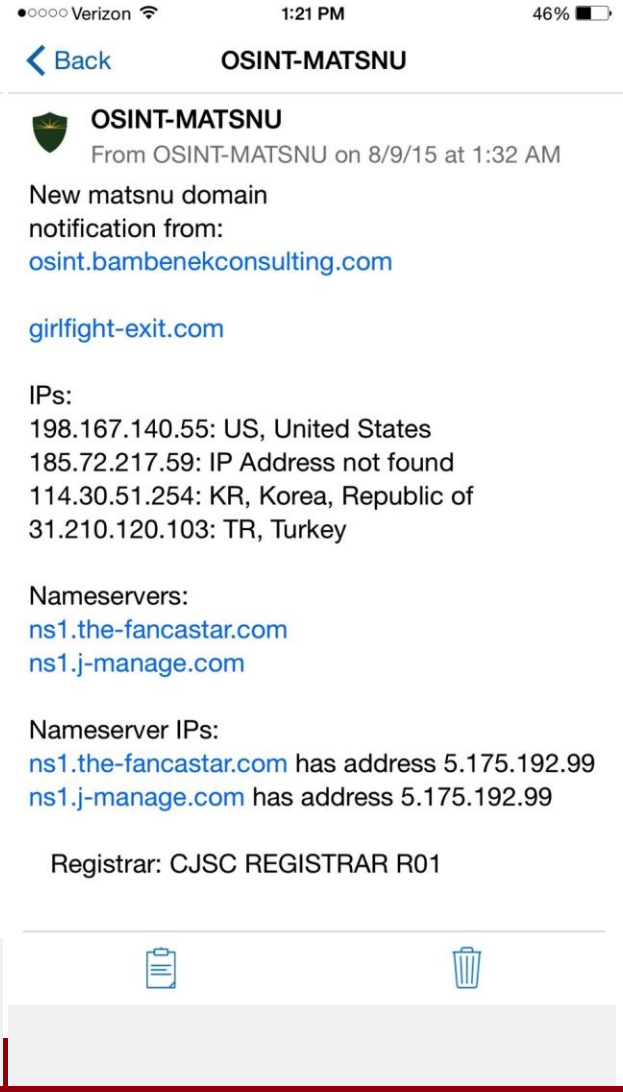
New Dyre domain registered

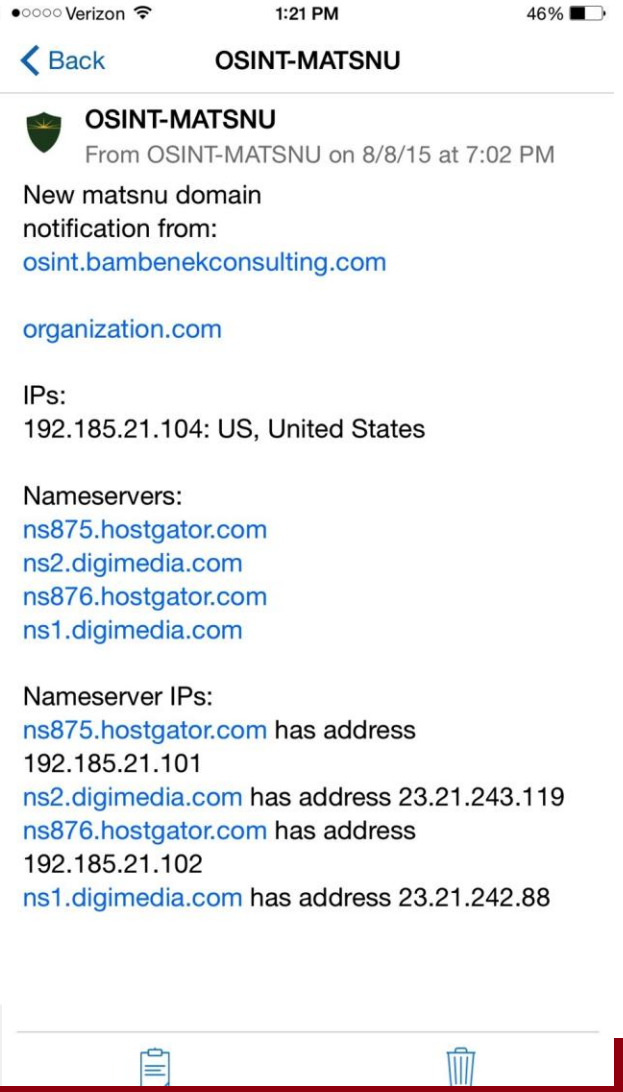
New Bedep Domain Registered



New Matsnu domains registered







Pivoting

- Now that I know the-fancastar.com and j-manage.com serve NS for Matsnu, I can see what else is served by those nameservers to find additional intelligence.
- *As of 24 Aug, this has switched to nausoccer.net and kanesth.com*
- Caution is due, this may not always yield results and may yield false positives. Always correlate with something else before making a final judgement.

Pivoting

Using IP from Matsnu 31.210.120.103

hostkale.com. IN A 31.210.120.103
ns1.hostkale.com. IN A 31.210.120.103
ns2.hostkale.com. IN A 31.210.120.103
linuxtr.hostkale.com. IN A 31.210.120.103
mobiluzman.com. IN A 31.210.120.103
habertemasi.com. IN A 31.210.120.103
kinghackerz.com. IN A 31.210.120.103
eglencekeyfi.com. IN A 31.210.120.103
ns1.eglencekeyfi.com. IN A 31.210.120.103
nejdetkuafor.com. IN A 31.210.120.103
profitstring.com. IN A 31.210.120.103
sirketrehber.com. IN A 31.210.120.103
actstudy-meat.com. IN A 31.210.120.103
….

Last adversarial response

- Starting to see sinkhole-aware malware.
- Some malware always authenticated the C2, but sinkholes still could gather intel.
- Now malware is being written to attempt to bypass sinkholes altogether.

The Future?

- DGAs will be around for awhile as part of several methods of communication to victim machines.
- Tor/I2P will continue to be used because of its advantages but DGAs still needed due to ease of blocking tor.
- Increase in the use of “interesting” dynamic seeds.

Questions?

Thanks Daniel Plohmann, April Lorenzen, Andrew Abakumov, Anubis Networks, many others.

And thanks HITCON!

My feeds: osint.bambenekconsulting.com/feeds/

jcb@bambenekconsulting.com
www.bambenekconsulting.com
+1 312 425 7225