Social Media OSINT Without the Indigestion

Ryan Shaw

Mark Orlando

BSIDES WASHINGTON D.C.

Who Are We?

Ryan Shaw

20 years in Cyber Operations

Managed services, consulting, R&D

Patent holder - Automated Threat Intelligence Platform

Sole Developer - E-mail analysis platform for TSA

Co-founder, Bionic

Mark Orlando

18 years in Cyber Operations

Managed services, consulting, strategy, automation

Former CTO, Raytheon Cyber

Patent holder - Automated Threat Intelligence Platform

Co-Founder, Bionic

About This Talk

This talk is: finding the value where infosec and social media intersect (and yes, influencers)

This talk is not: hard core OSINT methods for intelligence analysts

Social Media as a Resource

Why bother with social media?

- Lots of researchers and other smart people active there
- Don't just focus on collecting all the things
- Understand evolving attacker and campaign TTPs
- Pivot on known IOCs
- Non-obvious impacts to your org
- Changes in targeting and methods

Tweets get sent much faster than blogs or reports get written –

Managing Social Media Overload

All source = noise

Let others do the heavy lifting for IOC scraping

Separate accounts (personal vs. intel sourcing)

Targeted following in support of your role/team

Leverage lists (and don't recreate the wheel)

Get to scripting

Our Good Friend Bias

From Buster Benson's *Cognitive Bias Cheat Sheet:*

1. Too much information
2. Insufficient meaning
3. Insufficient time and resources
4. Insufficient memory

Our Good Friend Bias

When using social media for OSINT or news aggregation, *availability bias* comes into play:

Things that are the most memorable come to mind more quickly, can cause bad assumptions about larger data set

Getting Started

- Org. Threat Modeling: transparency, measurability, alignment
- Shapes everything
  - Tool selection
  - Analysis
  - Data collection
  - Response

Getting Started

Things that can inform your threat model:

- Threats that could impact you/your industry
- Leaks/dox
- New instrumentation, detection methods

Available Solutions

- Commercial threat intel feeds / platforms
- Third party sites/services like Twitonomy
- Public repositories like Slideshare
- Blog/news aggregators

Where We Looked

What's included?

- Twitter
- Vlogs & Blogs

What's not?

- Media
- LinkedIn

There's Gold Out There!

Data-driven answers to key questions

1. Where can I find the most original and timely information?
2. Who are the real influencers in cyber security?
3. Who is flying under the radar?

Case Study: Twitter

Environment

- Three free tier Twitter dev API keys
- Local and AWS hosts *
- Python/shell

Methodology

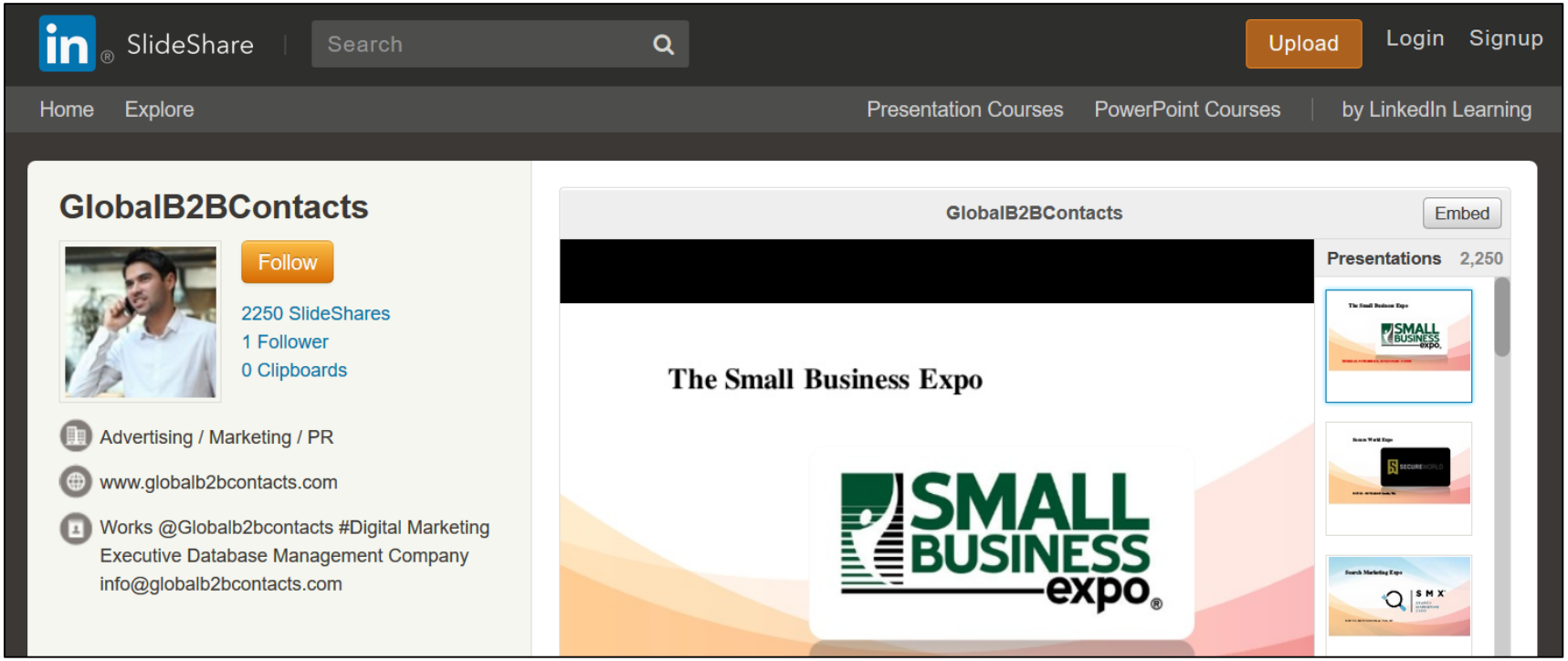
1. Start with known security Twitter accounts, domains, etc… and build from there
2. Profile, tweet, follower, and following calls
3. Domain visits / no scraping **

Twitter Reality Check

- People don't use templates
- Noise to signal ratio is high – this can be improved, but not eliminated

The high value Twitter accounts flagged for intel discussion shared one Slideshare profile 4x more than any other…

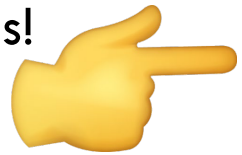
Twitter Reality Check



Lesson: You take the good with the bad (and if you can't sleep, reduce the bad with scripts)

Twitter User Classifications

Community Builder	Great blend of original tweets, re-tweets, and engaging replies
Soapbox	Only posts, never replies. Doesn't follow
Ghost	Inactive
Lurker	Consumes, but doesn't engage
Echo	Primarily re-tweets
Commentator	Quotes others with their two cents
Fraud	Repurposes other's tweets as their own

This! 

Tweet Types

Original	New message from the poster
RetweetOther	Repost someone else's message
ReplyOther	Comment on someone else's tweet
QuoteOther	Retweet someone else with commentary
RetweetSelf	Repost your own message
ReplySelf	Thread building when there is more to be said
QuoteSelf	Retweet your own message with commentary
ManualRetweet	Making other user's tweets look like your own

Our Bias

- Primarily English-speaking tweets / accounts
- Keywords / phrases well known
- Relatively small sample size
- Noise in the dataset

Data Collection & Analysis

Data collection period 10/4/19 - 10/20/19

Total Tweets Collected	25,473,187
Total Profiles Screened	177,362
Keyword / phrase search terms	220
APT spreadsheet terms	848
Matching tweets (all users)	5,142,583
Matching tweets (post noise redux)	100,541
Unique users	34,186
Unique users using 1+ keywords 2 or more times	15,227
Unique users using 2+ keywords	6,677

From those 6,677:

- Extracted all original tweets from last 45 days
- Examined @s, HTs, URLs, and other traits

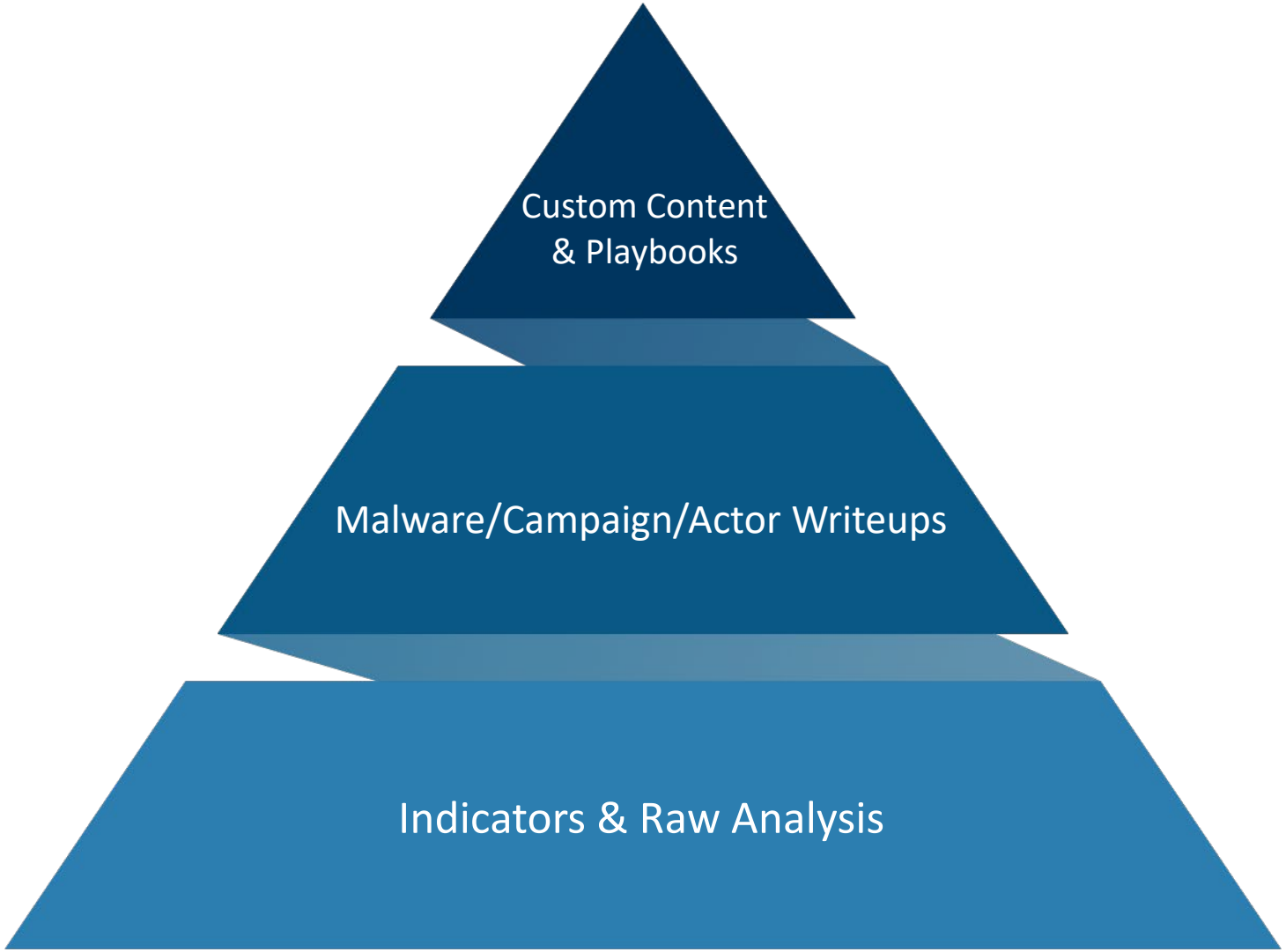
Under the Hood

infoseckitten#~#539990590302474241#~#Original#~#None#~#4#~#None#~#None#~#1#~#Wed Dec 03 03:52:43 +0000 2014#~#Wrote a post tonight on anti-VM malware (http://t.co/ml8ldSg7NK) with free yara rules! (https://t.co/RKVyez67ED) #malware #yara#~#None#~#None#~#None#~#Twitter Web Client#~#None#~#None#~#malware<~>yara#~#http://securitykitten.github.io/vm-checking-and-detecting/<~>https://github.com/securitykitten/public_yara_rules#~#None#~#None#~#None#~#None#~#None#~#None#~#en#~#None#~#None#~#None#~#1570690759

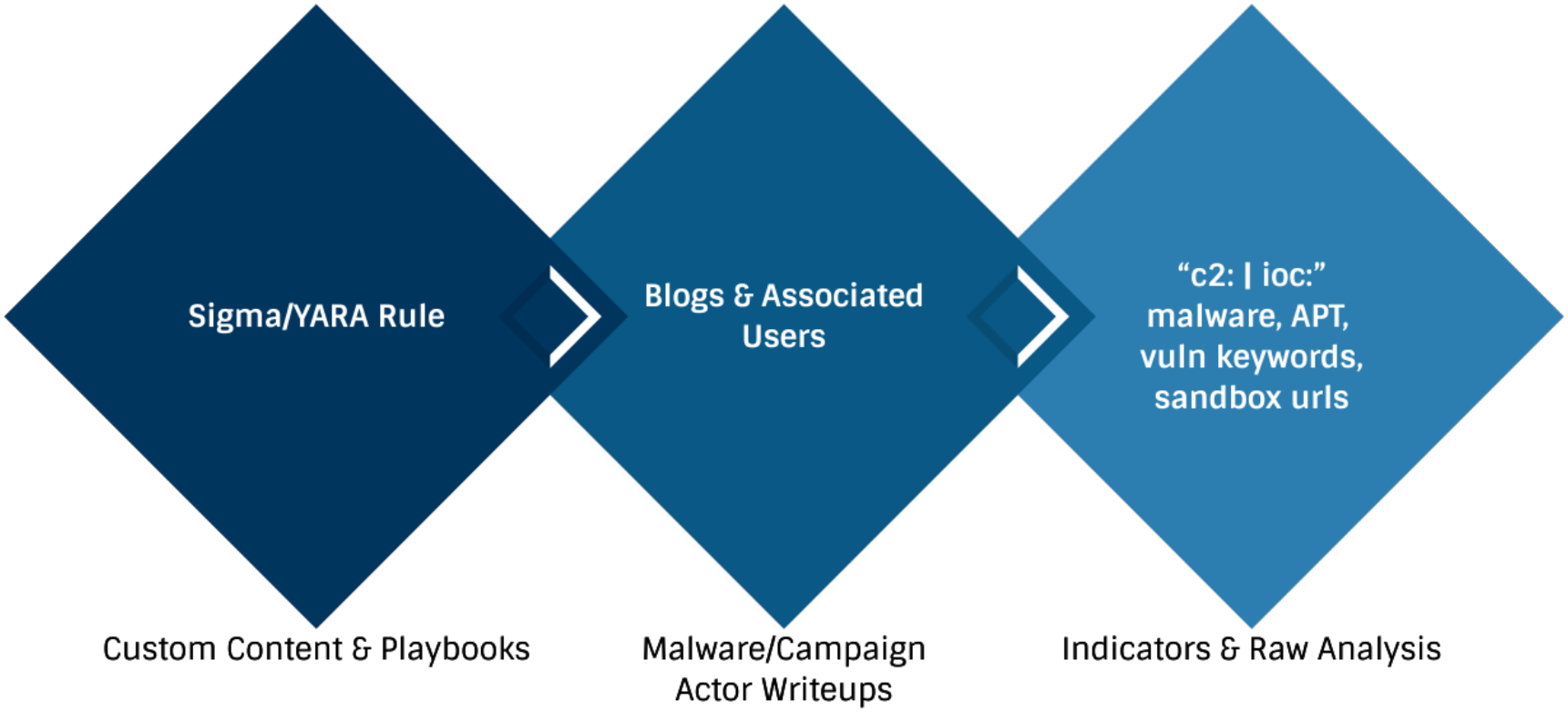
Infosec Twitter (30k ft view)

HANDLE	FOLLOWER COUNT	ORIGINAL TWEETS	2019 ORIG TWEETS	RETWEET OTHER	REPLY OTHER	QUOTE OTHER	RETWEET SELF	REPLY SELF	QUOTE SELF	TIME RANGE (days)
pod2g	404,178	49	1	39	109	2	0	1	0	1989
swiftonsecurity	291,602	23	23	48	59	17	10	37	5	4
briankrebs	263,483	49	49	28	90	16	0	15	2	89
Ih8sn0w	257,521	33	1	32	131	1	0	0	2	999
kevinmitnick	238,648	88	88	33	56	21	0	2	0	184
Mikko	195,794	19	19	51	82	26	0	14	7	29
e_Kaspersky	181,262	121	121	43	17	9	0	10	0	141
I0n1c	141,562	87	87	23	49	31	3	1	5	25
pwnallthethings	139,415	23	23	14	88	9	0	65	1	37
troyhunt	139,246	28	28	42	108	17	4	1	0	17
juliettekayyem	127,460	25	25	106	14	46	0	2	7	19
schneierblog	118,981	200	200	0	0	0	0	0	0	189
evacide	114,241	52	52	9	84	45	0	3	0	63
thegrugq	107,104	5	5	173	16	6	0	0	0	5
hacks4pancakes	100,321	5	5	47	123	2	0	21	0	4
danielmiessler	95,488	45	45	39	93	6	4	12	0	32
matthew_d_green	94,043	17	17	62	84	6	0	29	2	5
gcluley	89,380	102	102	19	64	10	0	5	0	19
dakami	88,299	11	11	9	133	29	0	17	0	4
k8em0	83,450	16	16	56	78	29	1	15	5	14

Tweet Value


Custom Content
& Playbooks
Malware/Campaign/Actor Writeups
Indicators & Raw Analysis

Deriving that value



Content is King

“yara rule” hits in Original tweets	331
“sigma rule” hits in Original tweets	70
Unique handles	250
URLs included	374

Top 10 Hashtags Used

104	#yara
55	#malware
38	#dfir
35	#sigma
21	#cybersecurity
16	#infosec
12	#threathunting
10	#apt
7	#siem
6	#security

Top 10 Linked FQDNs

81	github.com
19	fb.me
14	bilgiguvenlik.net
11	virustotal.com
11	bit.ly
10	tdm.socprime.com
8	lnkd.in
6	nextron-systems.com
5	gist.github.com
5	buff.ly

IOCs, because we have to

Fang use [.] [d] hxxp	331
c2:	70
ioc:	250
URLs included	374

Top 10 Handles	
399	scumbots
267	romonlyht
202	noladefense
200	dgafeedalerts
197	phishstats
129	kesagatame0
127	cryptophishing
120	botysrt
103	pennysoc
96	ipnigh

Top 10 Linked FQDNs	
251	pastebin.com
237	app.any.run
191	urlscan.io
127	virustotal.com
116	cc.uec.ac.jp
55	phishtank.com
36	pulsedive.com
10	app.threatconnect.com
7	github.com
7	beta.virusbay.io

Top 10 Hashtags Used	
33	#infosec
30	#cybersecurity
14	#malware
9	#threathunting
9	#malwareanalysis
9	#banker
8	#rat
8	#emotet
6	#lokibot
5	#ursnif
5	#agenttesla
4	#threatintel
3	#gootkit
2	#spelevoek
2	#nanocore
2	#maldoc
2	#keylogger
2	#jasperloader
1	#vidar

Threat Hunting Keywords / Phrases

Sandbox Use

	Last 45 Days		Full Sample Size	
	Tweets	Users	Tweets	Users
TOTAL / UNIQUE	**1456**	**219**	**6920**	**1092**
app.any.run/tasks/	724	104	1972	273
virustotal.com/	511	116	3353	735
otx.alienvault.com/pulse/	100	17	731	115
hybrid-analysis.com/	39	19	469	155
analyze.intezer.com/	36	7	110	15
urlhaus.abuse.ch/	19	6	64	33
virusbay.io/sample	18	7	94	36
joesandbox.com/analysis	5	3	28	15
cape.contextis.com/analysis/	4	4	92	13
apkscan.nviso.be/report	0	0	1	1
sandbox.anlyz.io/	0	0	4	2
sandbox.pikker.ee/analysis	0	0	1	1
sanddroid.xjtu.edu.cn/report	0	0	1	1

Infosec Keywords / Phrases

Marketing to the Rescue

· Aug 27

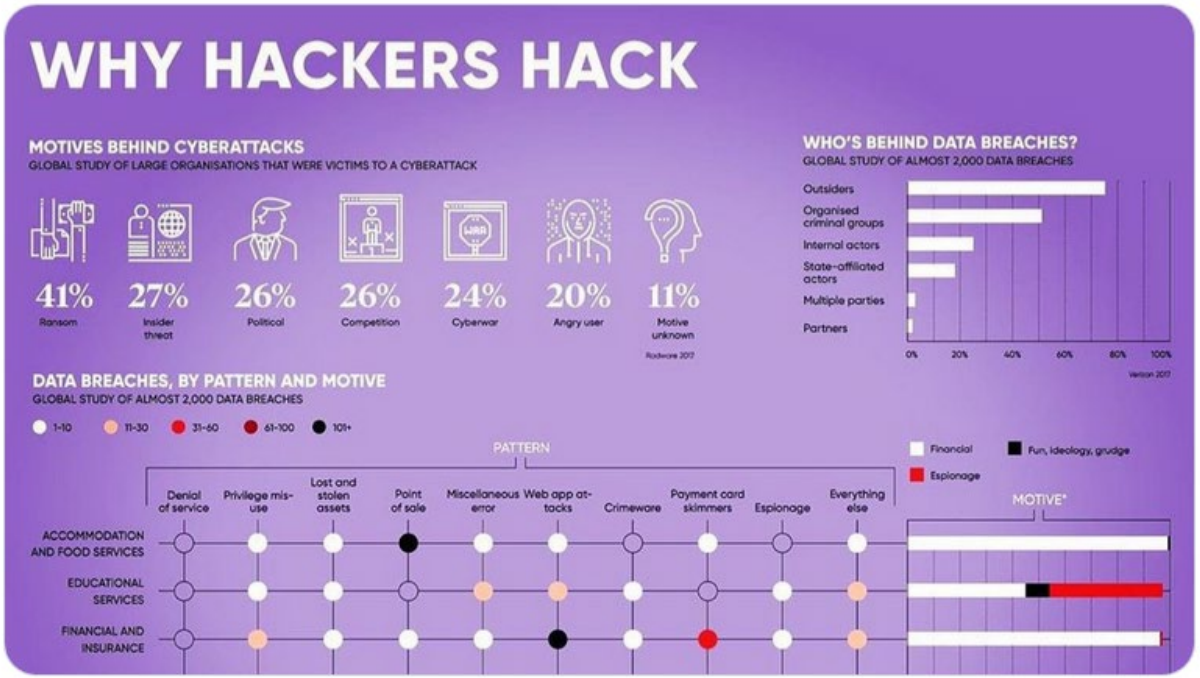
Why hackers hack!!!! 🛡️

•

•

•

#cyber #cybersecurity #vulnerability #infosec #hacker #secure #databreach #cloud #securecode #webdeveloper #data #datasecurity #cybercrime #coder #AWS #programmer #webdevelopment #developers #networks #security



Disinformation Campaign?

How About Our Most Followed?

HANDLE	KEYWORD MATCHES (ALL)	KEYWORD MATCHES (APT)
pod2g	9	
swiftonsecurity	2	
briankrebs	22	
ih8sn0w	3	
kevinmitnick	24	
mikko	8	
e_kaspersky	69	2
i0n1c	16	
pwnallthethings	1	
troyhunt	5	
juliettekayyem	1	
schneierblog	66	3
evacide	8	
thegrugq	1	
hacks4pancakes	1	
danielmiessler	9	
matthew_d_green	2	
gcluley	58	
dakami	1	
k8em0	1	

e_kaspersky

turla

zebrocy

schneierblog

cloud hopper

notpetya

triton

Working as a Community

Korben Dallas @KorbenD_Intel · Sep 18
Replying to @MeltX0R @JAMESWT_MHT and 2 others
FYI, if you mask your tweet IOCs like www[.]halimatoudi[.]com or 103.13.222[.]31 others can copy/paste those directly into @virustotal, and it will remove the brackets.

MeltX0R
@MeltX0R

Possible #Ke3chang #APT ?

C2 is timing out, so unable to obtain further payload.

MD5: 7c91d69ee49394ab960d8695a1866ec5
C2: www[d]halimatoudi[d]com:443

@thor_scanner @cyb3rops it looks like Thor scanner flagged on this file. Would it be possible to be provided the yara rule?

“I’ve tuned my sources, now what?”

1. Prioritize the intel
2. Contextualize & pivot
3. Create content, playbooks, and rules for your SIEM, SOAR, and other platforms
4. Share back

Door Prizes

Github with raw data and analyses: https://github.com/bioniccyber/socialmediaintel

What else is in there?

Other social media

- Blogs
- Podcasts

Door Prizes

Social scraper for domains

Twitter list(s)

https://twitter.com/secdatanoms/lists

That's a Lot About Twitter, But…

Slack 

Peers / Sector / Communities of interest

News

https://danielmiessler.com/newsletter/

https://thecyberwire.com/

Twitch 

https://www.twitch.tv/GrumpyHackers

Future Plans

New to the field/discipline training

- Conference talk recordings / slide decks
- Twitter lists, blogs, and podcasts by discipline

Code refinements and how-to

- Twitter dev API search (local/AWS)
- NLP

Thank You!

Ryan: @secdatanoms

Mark: @markaorlando