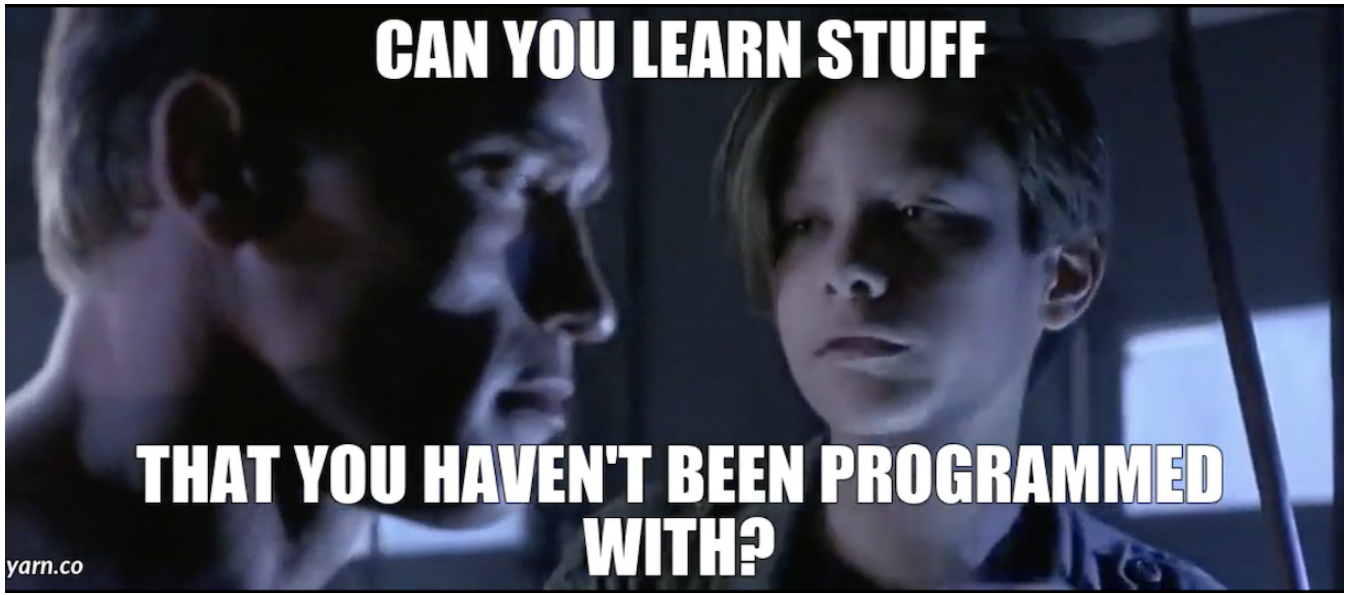
Practical Threat Hunting With Machine Learning

Craig Chamberlain
@randomuserid

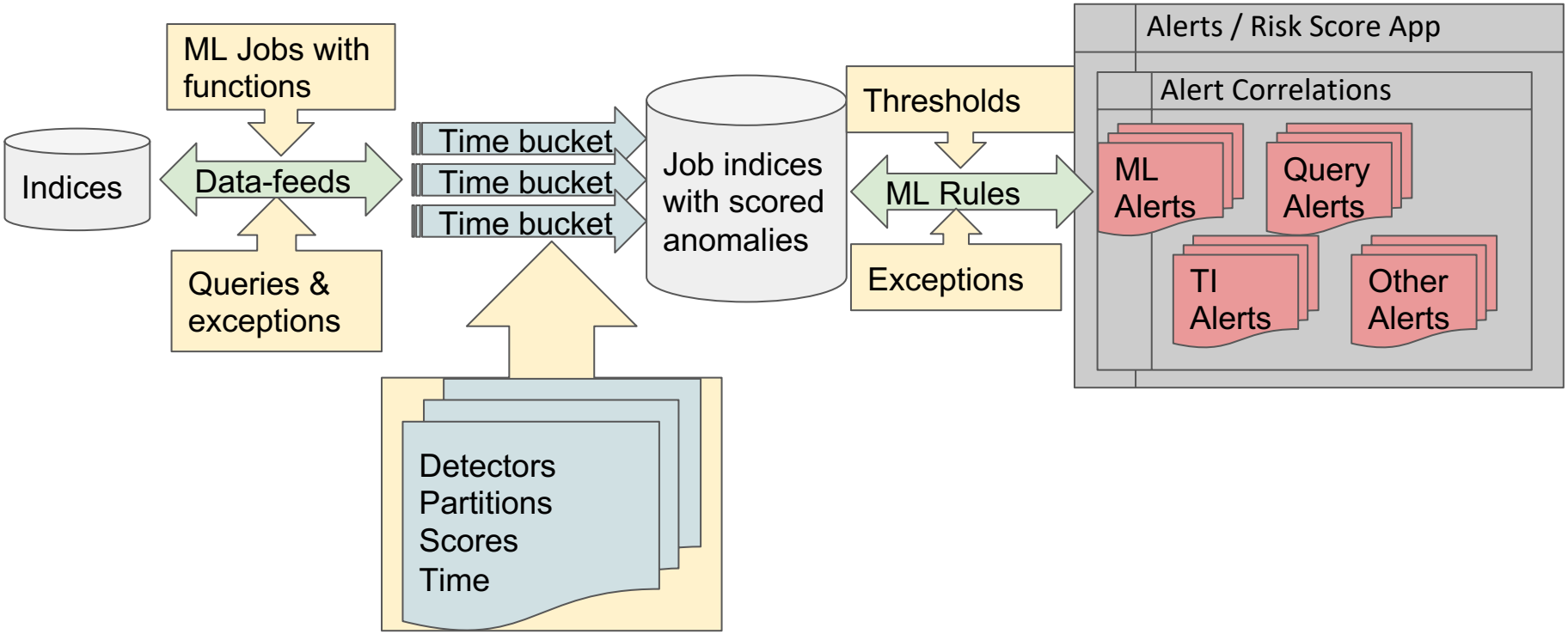
Why hunt with machine learning?

- Add a valuable new layer to the detection stack
- Automate and scale some threat hunting techniques



The Stack: Anomalies to Alerts

https://www.elastic.co/blog/using-elastic-machine-learning-rare-analysis-to-hunt-for-the-unusual



Case Studies: Endpoint Hunting

anomalous process	A rare process name across all hosts
anomalous process, one host	A rare process name for a particular host
anomalous path activity	A rare path (working directory) in a process execution event
anomalous process creation	A rare process name for a particular parent process name
suspicious script	A script with high information content e.g. obfuscation

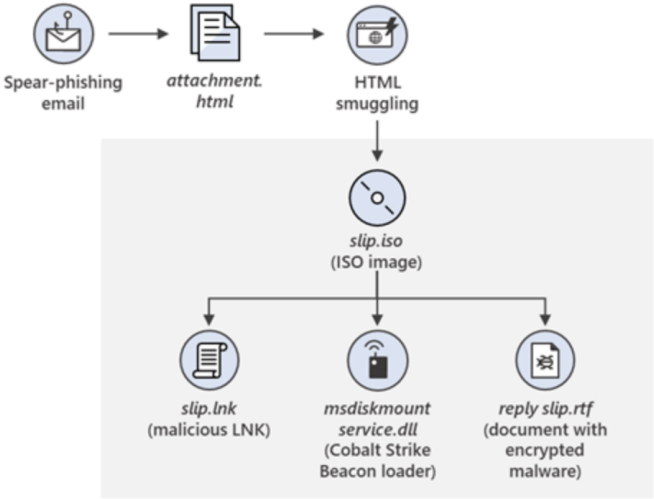
- Rare processes are not always anomalies
  - Managed computer fleets have occasional housekeeping processes
  - These are rare, in that they exist sporadically, but are normal
  - In this case it is more useful to look for rare processes across the host population, which is what the new rare process jobs look for.
- Why look for processes rare to a single host?
  - These jobs sometimes find suspicious and interesting things
  - Some high-value assets (DCs, Exchange) have lower normal variance in process relationships

Microsoft warns SolarWinds hacking group Nobelium is targeting its customers

NOBELIUM Case Study (2021)

APT group active in 1H 2021 (at least)

Detected by unsupervised ML rules with no a priori knowledge - no IOCs, signatures or behavioral searches.

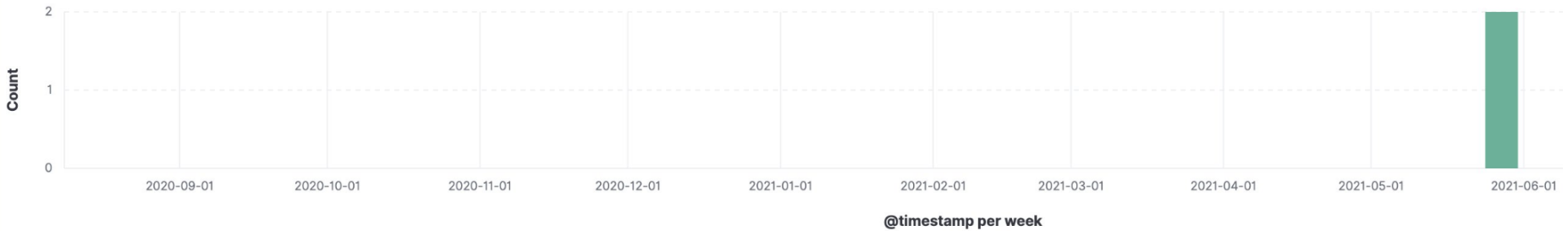


May 28th 2021, 09:00	22	rare by "process.working_directory"	D:\	host.name: nobelium process.name: rundll32.exe user.name: test	v2_windows_anomalous_path_activity_ecs	
May 28th 2021, 12:00	15	rare by "process.name"	rundll32.exe	destination.ip: 12.96.42.215 destination.ip: 192.99.221.77 destination.ip: 34.226.46.235 destination.ip: 34.238.11.122 destination.ip: 54.89.106.200 and 4 more	v2_windows_anomalous_network_activity_ecs	
May 28th 2021, 10:00	14	rare by "process.working_directory"	D:\	host.name: nobelium process.name: rundll32.exe user.name: test	v2_windows_anomalous_path_activity_ecs	

May 28th 2021, 09:00	60	rare by "process.name"	rundll32.exe	destination.ip: 12.96.42.215 destination.ip: 192.99.221.77 destination.ip: 23.32.45.172 destination.ip: 23.35.70.144 destination.ip: 34.226.46.235 destination.ip: 54.89.106.200 destination.ip: 72.21.91.29 destination.ip: 83.171.237.173 destination.ip: 91.199.212.52 host.name: nobelium process.name: rundll32.exe user.name: test show less	v2_windows_anomalous_network_activity_ecs

NOBELIUM: Anomalous DLL Activity

2 hits Reset search Aug 8, 2020 @ 14:03:14.764 - Aug 8, 2021 @ 14:03:14.764 Auto

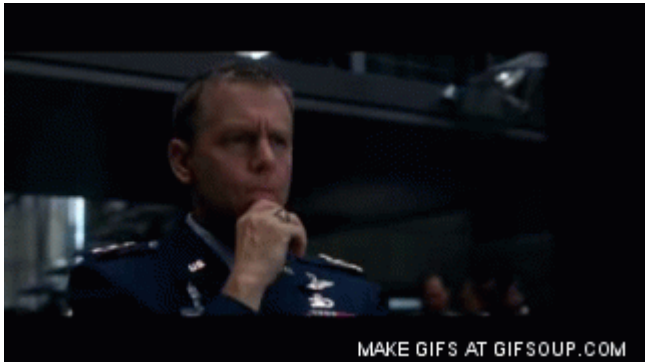


Time	process.name	message	process.command_line
May 28, 2021 @ 10:06:32.093	rundll32.exe	Endpoint process event	"C:\Windows\system32\rundll32.exe" Documents.dll,Open
May 28, 2021 @ 09:57:49.773	rundll32.exe	Endpoint process event	"C:\Windows\system32\rundll32.exe" Documents.dll,Open

Case Studies: Network Hunting

Spike in traffic to a destination country	A relatively large burst of network events to a particular destination country
Rare destination country	Network events with a rare destination country
Spike in network denies	A relatively large burst of denied network connections or events
Spike in network events	A relatively large burst of network connections or events

Case Study
Geographic Detection: Spike in traffic to one destination country



Time	Severity	Detector	Found for	Influenced by	Actual	Typical	Description	Job ID
March 18th 2021	98	high_non_zero_count by "destination.geo.country_name"	Russia	destination.as.organization.name: destination.geo.country_name: Russia source.ip:	2134	4.52	More than 100x higher	high-count-by-destination-country
March 24th 2021	95	high_non_zero_count by "destination.geo.country_name"	Russia	destination.geo.country_name: Russia source.ip:	478	20.5	23x higher	high-count-by-destination-country
March 23rd 2021	89	high_non_zero_count by "destination.geo.country_name"	Russia	destination.geo.country_name: Russia source.ip:	989	14.1	70x higher	high-count-by-destination-country

March 1st 2021	51	rare by "destination.geo.country_name"	Iran		destination.as.organization.name: [redacted] destination.as.organization.name: [redacted] destination.geo.country_name: Iran rare-destination-country-3 destination.ip: [redacted] destination.ip: and 2 more		

Unusual destination country

Unusual spike in firewall denies

Time	Severity	Detector	Influenced by	Actual	Typical	Description	Job ID
March 18th 2021	96	high_count	destination.as.organization.name: (Inc. destination.as.organization. [redacted] destination.geo.country_name: United States destination.port: 443	34841	16838	↑ 2x higher	spike-in-network-acl-denies

Case Study: The *Sunburst* Backdoor / Trojan Horse in the SolarWinds *Orion* Product

SUNBURST C2 detection: *dns_tunneling* looks for a spike in child domain names, for for one parent domain, in DNS events.

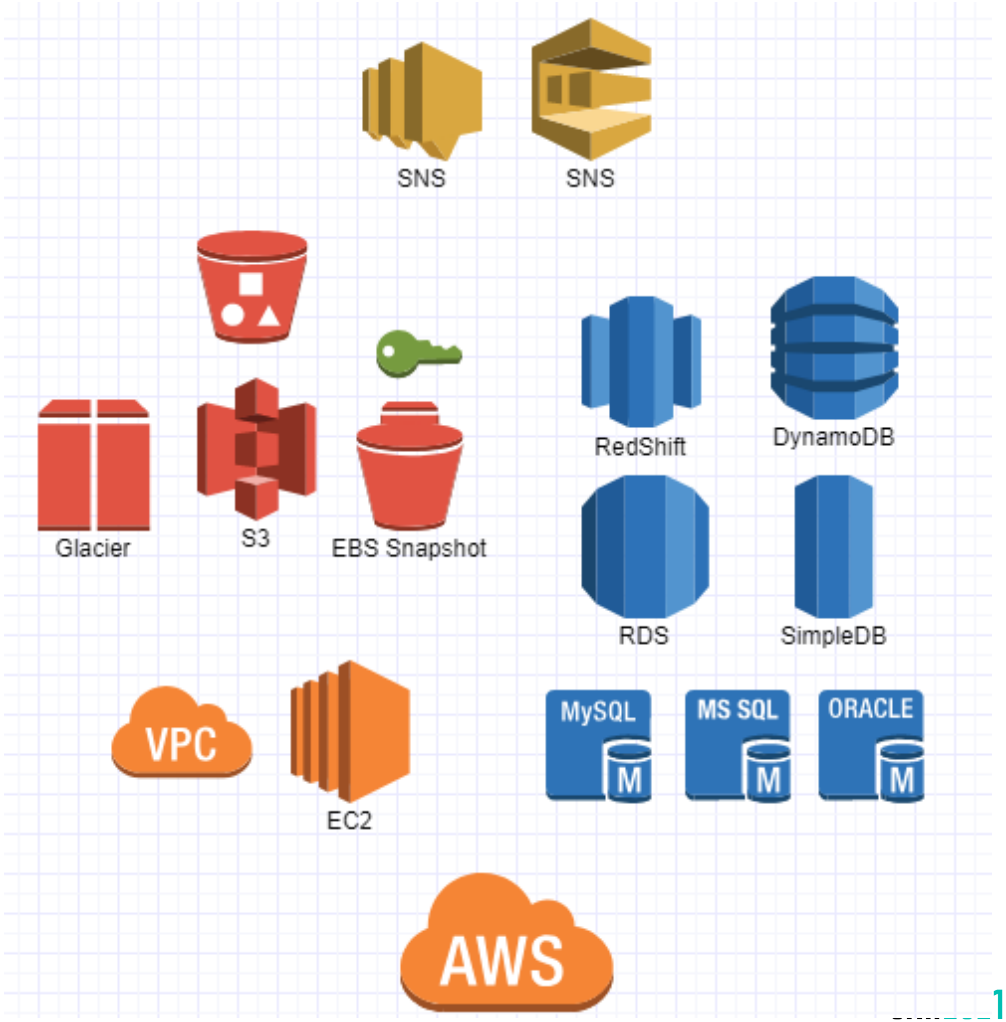
time	severity ↓	detector	found for	influenced by	actual	typical	description	actions
December 16th 2020, 10:00	83	high_info_content("dns.question.name") over "dns.question.etld_plus_one"	avsvmcloud.com	destination.ip: 1.1.1.1 ⊕ ⊖ dns.question.etld_plus_one: avsvmcloud.com ⊕ ⊖	7004	38.44959600784599	↑ More than 100x higher	⚙
December 16th 2020, 09:00	54	high_info_content("dns.question.name") over "dns.question.etld_plus_one"	avsvmcloud.com	destination.ip: 1.1.1.1 ⊕ ⊖ dns.question.etld_plus_one: avsvmcloud.com ⊕ ⊖	21567	222.77166427376523	↑ 97x higher	⚙


Cloud Hunting
Case Studies

Cloud Threat Hunting

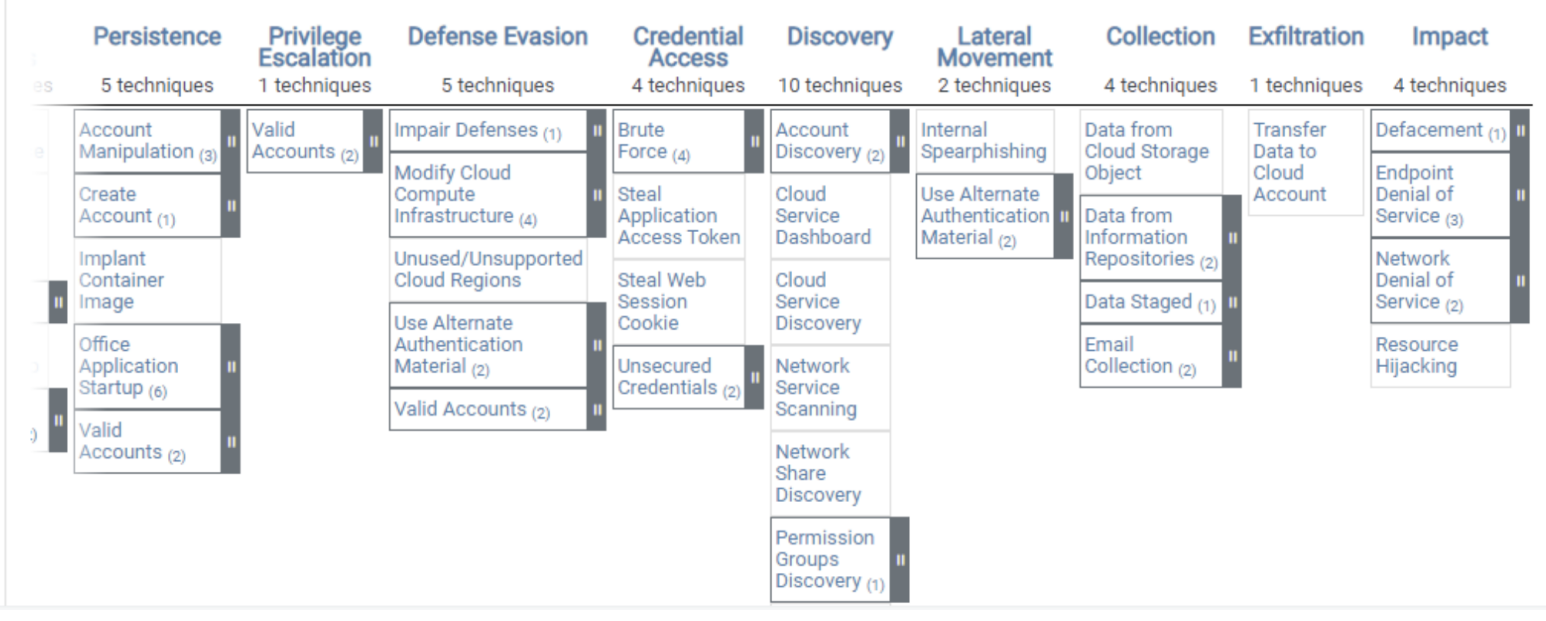
Cloud incidents often lack clear evidence or indicators of misuse or abuse activity in API transaction logs

The difference between normal user activity and credentialed access is a matter of nuance.



212 services

Services. As of 2020, AWS comprises more than **212 services** including computing, storage, networking, database, analytics, application services, deployment, management, mobile, developer tools, and tools for the Internet of Things.



Case Study: Exfil in AWS

34. In or around September 2016, the Conspirators also successfully gained access to DNC computers hosted on a third-party cloud-computing service. These computers contained test applications related to the DNC's analytics. After conducting reconnaissance, the Conspirators gathered data by creating backups, or "snapshots," of the DNC's cloud-based systems using the cloud provider's own technology. The Conspirators then moved the snapshots to cloud-based accounts they had registered with the same service, thereby stealing the data from the DNC.

Time	eventName
September 27th 2018, 15:46:05.000	DescribeSnapshotAttribute
September 27th 2018, 15:41:15.000	DescribeSnapshotAttribute
September 27th 2018, 15:41:15.000	DescribeTags
September 27th 2018, 15:41:15.000	DescribeSnapshotAttribute
September 26th 2018, 18:27:18.000	DescribeTags
September 26th 2018, 18:27:18.000	DescribeSnapshotAttribute
September 26th 2018, 18:27:18.000	DescribeSnapshotAttribute
September 26th 2018, 17:07:29.000	DescribeSnapshotAttribute
September 26th 2018, 17:07:26.000	DescribeTags
September 26th 2018, 17:07:26.000	DescribeSnapshotAttribute
September 26th 2018, 17:07:26.000	DescribeSnapshotAttribute
September 26th 2018, 17:05:37.000	DescribeSnapshotAttribute
September 26th 2018, 17:05:35.000	DescribeSnapshotAttribute
September 26th 2018, 17:05:35.000	DescribeSnapshotAttribute
September 26th 2018, 17:05:35.000	DescribeTags
September 26th 2018, 17:04:58.000	DescribeSnapshotAttribute
September 26th 2018, 17:03:49.000	DescribeSnapshotAttribute
September 26th 2018, 17:03:44.000	ModifySnapshotAttribute
September 26th 2018, 17:03:19.000	DescribeSnapshotAttribute
September 26th 2018, 17:03:09.000	DescribeTags
September 26th 2018, 17:03:09.000	DescribeSnapshotAttribute
September 26th 2018, 17:03:09.000	DescribeSnapshotAttribute

Time	sourceIPAddress	eventName	requestParameters.snapshotId
September 26th 2018, 17:03:44.000	61.170	ModifySnapshotAttribute	snap-0d5b275c244291f5f

Table JSON

View surrounding documents | View single document

@timestamp		September 26th 2018, 17:03:44.000
t	@version	1
t	_id	UMJ2HGYBWy07VLjza8yf
t	_index	cloudtrails-2018.09.26
#	_score	-
t	_type	doc
t	awsRegion	us-east-1
t	eventID	fb6a2adf-cd02-4964-bbea-fcde1f795138
t	eventName	ModifySnapshotAttribute
t	eventSource	ec2.amazonaws.com
t	eventType	AwsApiCall
t	eventVersion	1.05
t	recipientAccountId	[redacted]
t	requestID	5c10d3f7-b713-4569-bb95-8f7bafe4b2f3
t	requestParameters.attributeType	CREATE_VOLUME_PERMISSION
?	requestParameters.createVolumePermission.add.items	{ "userId": [redacted]1" }
t	requestParameters.snapshotId	snap-0d5b275c244291f5f
t	responseElements._return	true
t	responseElements.requestId	5c10d3f7-b713-4569-bb95-8f7bafe4b2f3
t	sourceIPAddress	192.168.1.121
t	tags	cloudtrail
t	userAgent	console.ec2.amazonaws.com
t	userIdentity.accessKeyId	ASIAZI5AL2TYHMISYE6V
t	userIdentity.accountId	637602092272
t	userIdentity.arn	arn:aws:iam::[redacted]:root
t	userIdentity.principalId	[redacted]2
t	userIdentity.sessionContext.attributes.creationDate	September 26th 2018, 14:08:45.000
t	userIdentity.sessionContext.attributes.mfaAuthenticated	false
t	userIdentity.type	Root

attempted connections to Capital One's server from TOR exit nodes, and a number of connections from IP addresses beginning with 46.246, all of which Capital One believes relate to activity conducted by the same person involved in the April 21, 2019, intrusion, because they involve similar unusual communications through the misconfigured firewall to the server discussed above. Specifically, according to Capital One, the logs show:

- On or about March 12, 2019, IP address 46.246.35.99 attempted to access Capital One's data. I know, from checking publicly-available records, that this IP address is controlled by IPredator, a company that provides VPN services.
- On or about March 22, 2019, the *****-WAF-Role account was used to execute the List Buckets Command several times. These commands were executed from IP addresses that I believe to be TOR exit nodes. According to Capital One, the *****-WAF-Role account does not, in the ordinary course of business, invoke the List Buckets Command.
- Also on or about March 22, 2019, the *****-WAF-Role account was used to execute the Sync Command a number of times to obtain data from certain of Capital One's data folders or buckets, including files that contain credit card application data. A number of those commands

UNITED STATES ATTORNEY
700 STEWART STREET, SUITE 5220
SEATTLE, WASHINGTON 98101
(206) 553-7970

Unusual geolocation / command pair

Unusual user / command pair

Rare method for a user	An unusual CloudTrail operation for a particular user
Rare city for an API method	An unusual CloudTrail operation for a source city name
Rare country for an API method	An unusual CloudTrail operation for a source country name
Rare error code	A rare error code in the CloudTrail logs
Spike in an error message	A relatively large increase in the occurrence of a particular error in the CloudTrail logs

Detecting Anomalous AWS Commands

- The ListBuckets command is used too frequently for conventional search rules.
  - This method is called thousands of times per week, and tens of thousands of times per month, even in a mid-sized AWS environment.
  - Any search rules for the ListBuckets command would be deactivated in a matter of minutes after creating a tsunami of alerts
- What we need is a measurement of unusual command activity
  - A command from a user context that does not normally use the command
  - API command activity from an unusual geolocation

ML Job: Unusual AWS Command for a User

June 30th 2020	< 1	rare by "event.action" partition by "user.name"	ListBuckets	source.geo.city_name: source.ip: user.name: craig.	rare_method_for_a_username	

This user has not called the “ListBuckets” method before!

Unusual AWS Command for a User:
Security Group Change

>	June 20th 2020	< 1	rare by "event.action" partitionfield="user.name"	AuthorizeSecurityGroupIngress	source.geo.city_name: user.name:	rare-method-for-a-username	⚙
>	June 20th 2020	< 1	rare by "event.action" partitionfield="user.name"	CreateSecurityGroup	source.geo.city_name: user.name:	rare-method-for-a-username	⚙

Why is this user making changes to security groups?

Unusual AWS Command for a User:
EC2 Activity

>	June 20th 2020	< 1	rare by "event.action" partitionfield="user.name"	SharedSnapshotVolumeCreated	user.name:	rare-method-for-a-username	
>	June 20th 2020	< 1	rare by "event.action" partitionfield="user.name"	RunInstances	source.geo.city_name: [redacted] user.name:	rare-method-for-a-username	

Why did this user begin starting EC2 instances?

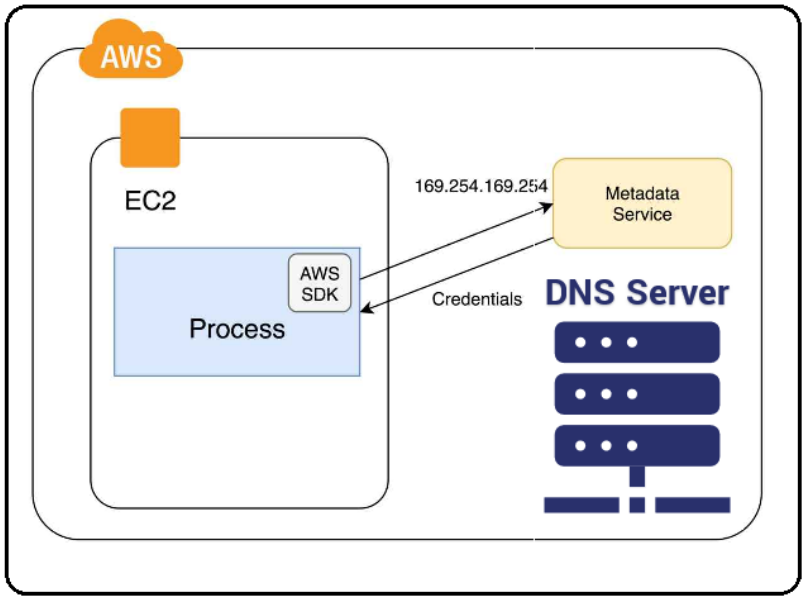
Rare Method for a Country

June 17th 2020	< 1	rare by "event.action" partition by "source.geo.country_iso_code"	AssumeRole	aws.cloudtrail.user_identity.arn: arn:aws:iam::144492464627:user/cloudtrail-ingest source.geo.country_iso_code: US source.ip: 18.189.184.12	rare_method_for_a_country	
June 17th 2020	< 1	rare by "event.action" partition by "source.geo.country_iso_code"	UpdateAssumeRolePolicy	aws.cloudtrail.user_identity.arn: arn:aws:iam::144492464627:user/cloudtrail-ingest source.geo.country_iso_code: US source.ip: 18.189.184.12	rare_method_for_a_country	

These privileged methods should not be coming from unusual geolocations!

Rare Method for a Country

The Metadata Service



Four ML jobs for the metadata service;

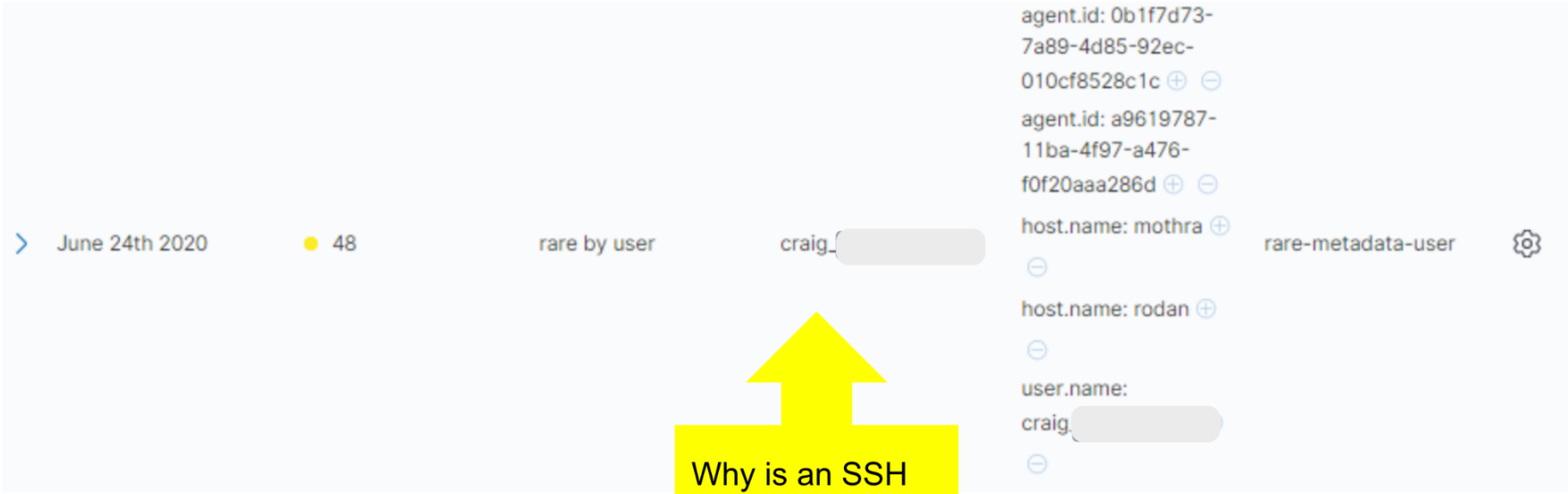
- Linux
  - Rare process name calling the metadata service
  - Rare username calling the service
- Windows
  - Rare process name calling the metadata service
  - Rare username calling the service

Rare Metadata Process

June 24th 2020	43	rare by "process.name"	<unknown process>	agent.id: 0b1f7d73-7a89-4d85-92ec-010cf8528c1c host.name: rodan	rare-metadata-process
June 24th 2020	43	rare by "process.name"	curl	agent.id: a9619787-11ba-4f97-a476-f0f20aaa286d host.name: mothra user.name: craig	rare-metadata-process

Why is this process interrogating the metadata service?

Rare Metadata User



- 740 unsupervised ML jobs in development
- Supervised models
  - Suspicious URLs, in addition to LOLBins, DGA activity
- Risk score clustering & correlation of ML and conventional alerts
  - Experimental dash available now

The Road Ahead / Q&A
Craig Chamberlain
@randomuserid


Practical Threat Hunting With Machine Learning
Craig Chamberlain
@randomuserid