IMPROVING ANALYST WORKFLOW WITH EVENT CLUSTERING

Awalin Sopan

with Adarsh Kyadige, Salma Taoufiq, Tamás Vörös &

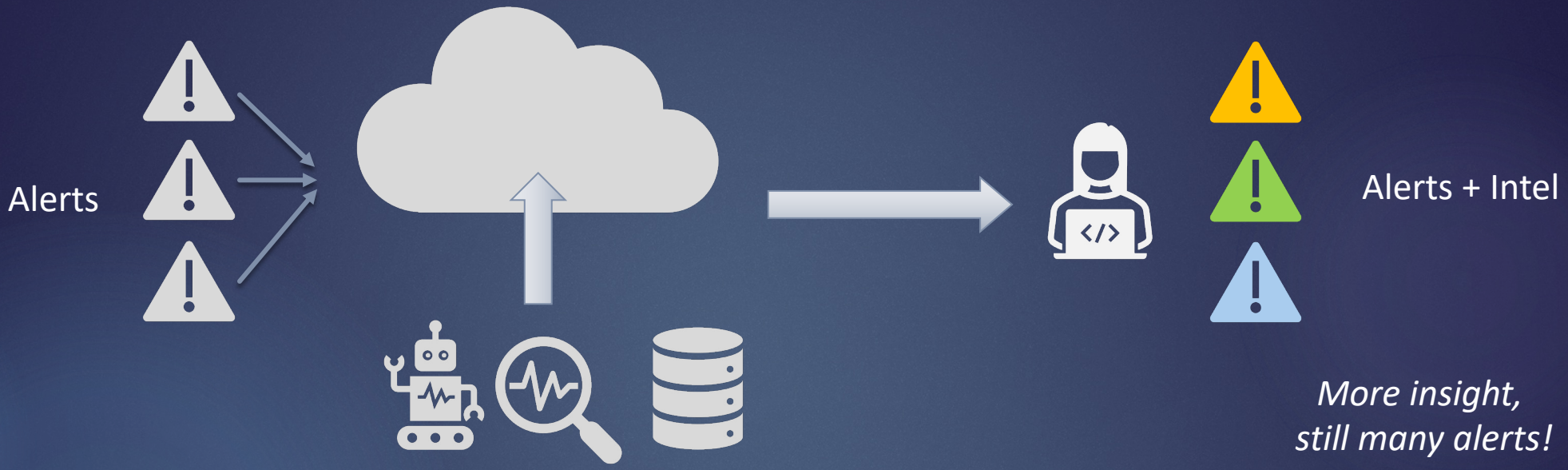
Dr. Konstantin Berlin

Data Scientists/ Engineer/ Director

@SophosAI

SOPHOS

SOC Analyst Workflow: Current State



But …

Similar events in other customers or other times

Now …

Leverage knowledge of how past similar events were resolved

Exact Match Alone Misses Many Similar Alerts

```
Powershell.exe -Noninteractive -Command "& \"C:\Program Files (x86)\PRTG Network Monitor\Notifications\exe\PRTGSlackWebHookNotification.ps1\" -SlackWebHook
'https://hooks.slack.com/services/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' -SlackChannel '#prtgmonitoring' -SiteName 'Production (DMZ)' -Device 'Hay-
ProdAppPU01' -Name 'Ping Jitter (Ping Jitter)' -Status 'Down (before: Warning)' -Down '' -DateTime '04/04/2021 07:28:00' -LinkDevice
'https://Monitoring.corp.loans2go.co.uk/device.htm?id=5297' -Message '0.87 (Jitter) is above the error limit of 0.50 in Jitter' "

Powershell.exe -Noninteractive -Command "& \"C:\Program Files (x86)\PRTG Network Monitor\Notifications\exe\PRTGSlackWebHookNotification.ps1\" -SlackWebHook
'https://hooks.slack.com/services/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' -SlackChannel '#prtgmonitoring' -SiteName 'Z-Team' -Device
'Yourapi.logbookloans.co.uk' -Name 'HTTP Advanced (HTTP Advanced)' -Status 'Down ESCALATION REPEAT' -Down '' -DateTime '12/02/2021 08:33:35' -LinkDevice
'https://Monitoring.corp.loans2go.co.uk/device.htm?id=5974' -Message '3,401 Byte (Bytes received) is below the error limit of 3,419 Byte in Bytes received' "

Powershell.exe -Noninteractive -Command "& \"C:\Program Files (x86)\PRTG Network Monitor\Notifications\exe\PRTGSlackWebHookNotification.ps1\" -SlackWebHook
'https://hooks.slack.com/services/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' -SlackChannel '#prtgmonitoring' -SiteName 'Production (DMZ)' -Device 'Hay-
ProdAppPU01' -Name 'CPU Load (Windows CPU Load)' -Status 'Down (before: Warning)' -Down '' -DateTime '06/03/2021 00:27:00' -LinkDevice
'https://Monitoring.corp.loans2go.co.uk/device.htm?id=5297' -Message '22 % (Total) is above the error limit of 20 % in Total' "
```

SOC Analyst Workflow: Current State

FPs make the bulk of alerts

Bulk of similar catastrophic FPs cause a lot of noise and alert fatigue

Identify and filter out those noise from workflow

Examples of Similar Commands

Powershell (Get-Command ‘C::\\users\\AwalinSopan\\AppData\\Local||Programs\\Stratasys\\Installs\\GrabCAD-Print-Installer.exe’).Version.Tostring()

Powershell (Get-Command ‘C::\\users\\KonstantinB\\AppData\\Local||Programs\\Stratasys\\Installs\\GrabCAD-Print-Installer.exe’).Version.Tostring()

Powershell (Get-Command ‘C::\\users\\SalmaTaou\\AppData\\Local||Programs\\Stratasys\\Installs\\GrabCAD-Print-Installer.exe’).Version.Tostring()

Our Proposal:
Cluster Alerts
and
Make Decision on Cluster Level

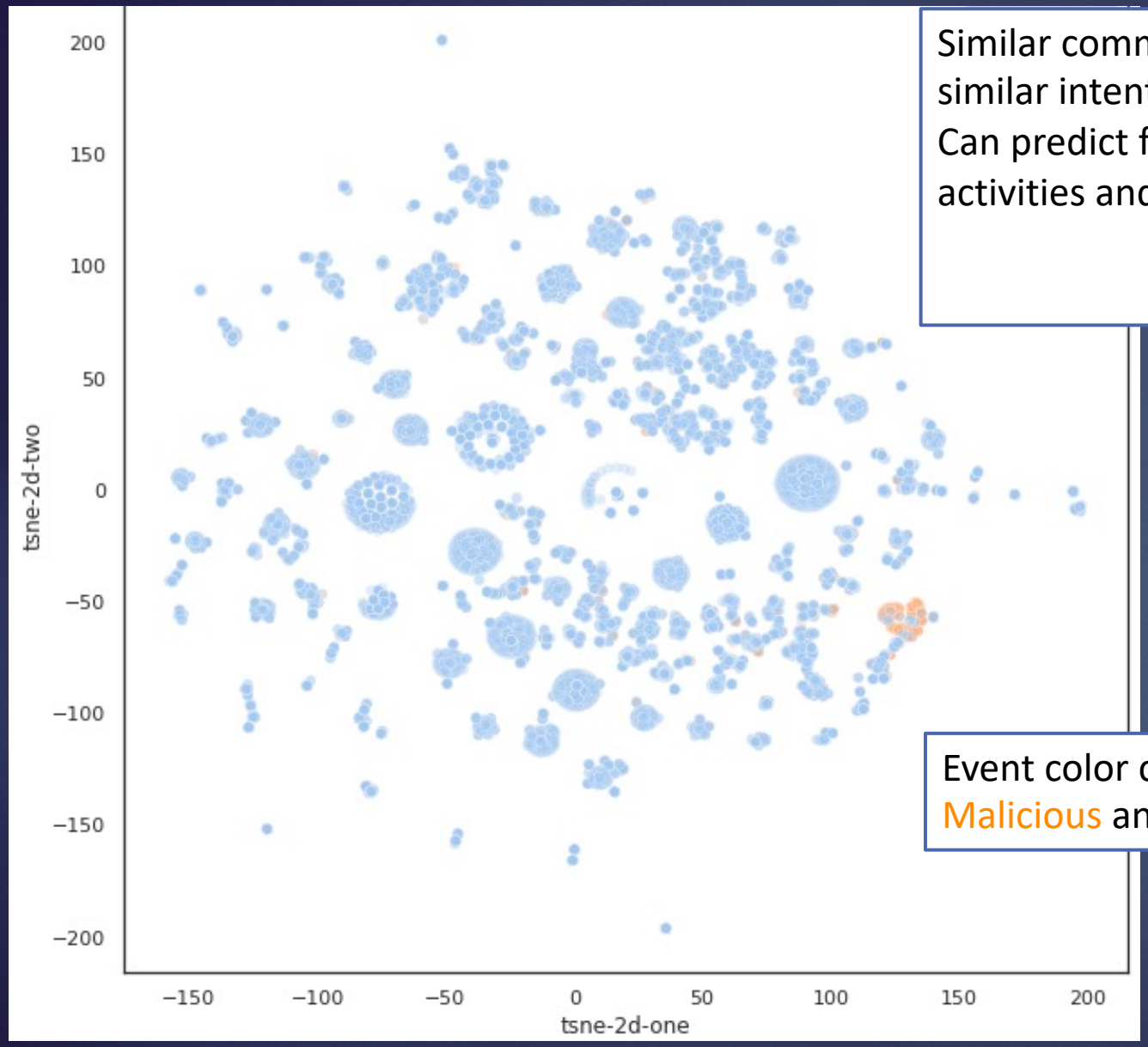
Assumption:

- We can measure similar alerts and quantify similarity
- Scale neighborhood search across millions of alerts

Prototype Human in the Loop UI:

- Speed up workflow
- Enable decision based on prior alerts where IOC/ ML detection may not be perfect

Similar clusters have similar labels

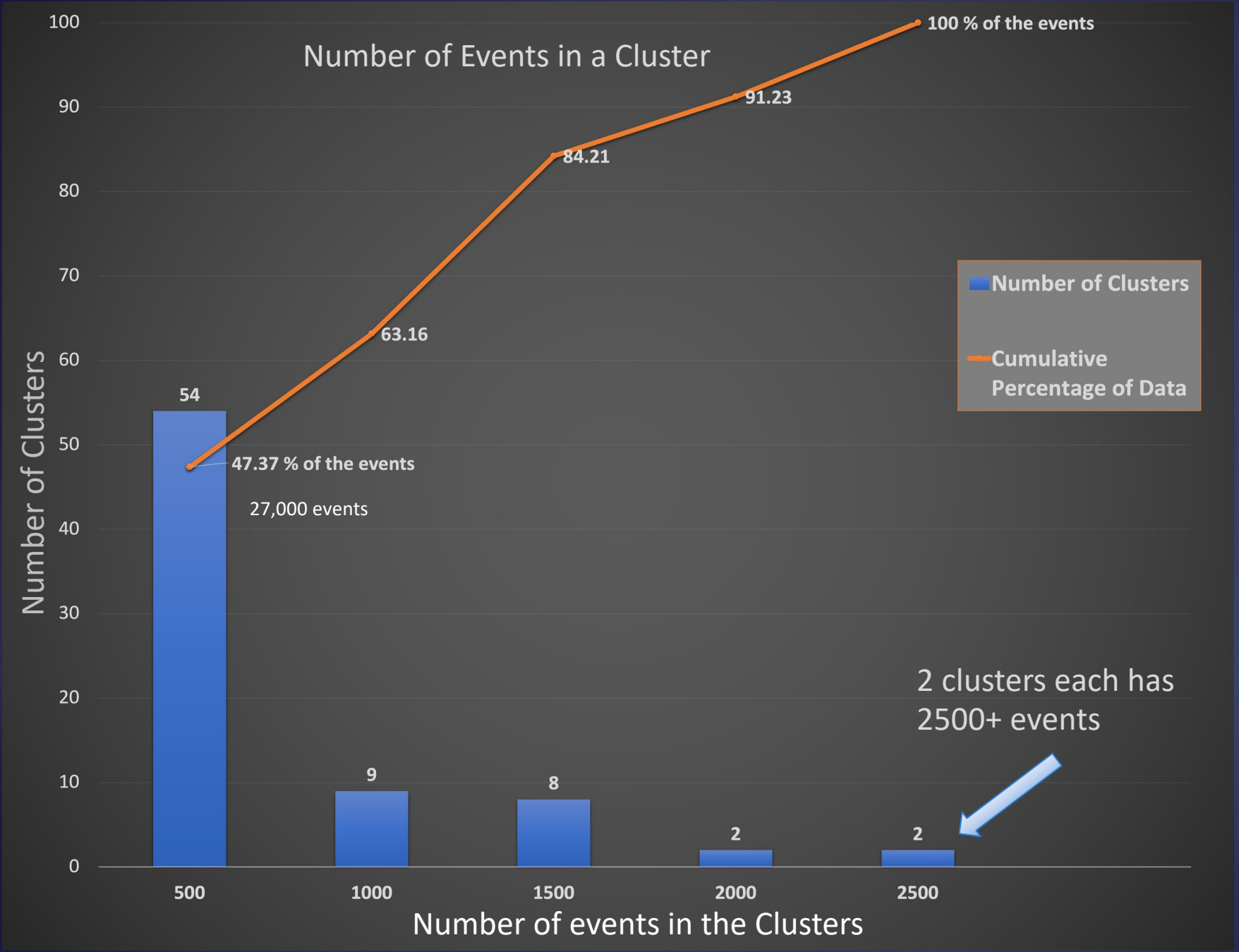


Similar commands indicate similar intent and action.
Can predict from past similar activities and commands.

Event color coded by labels:
Malicious and Benign

30,000 Powershell event samples from March, April 2021.

Plotted using t-SNE.



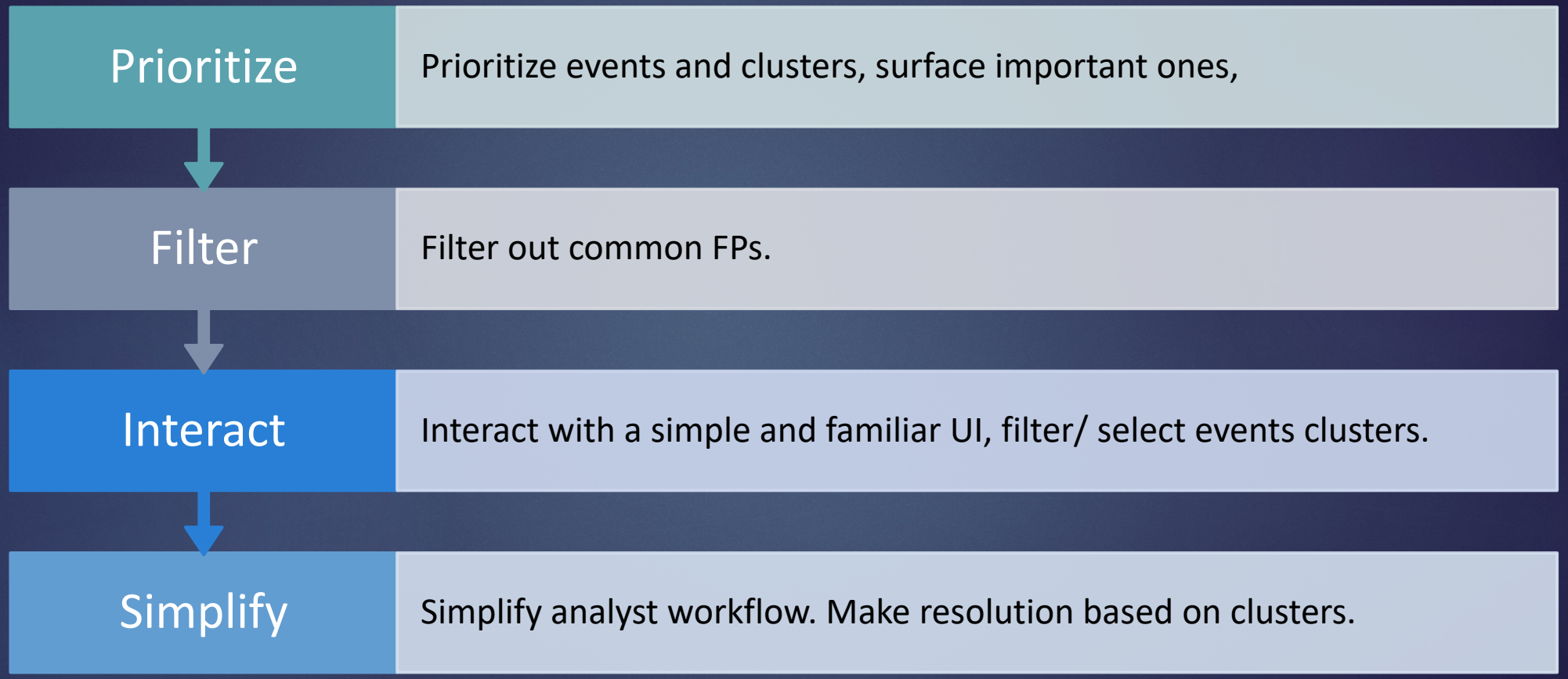
A few clusters capture most events.

April 30, 2021:

- Total 60,000 events.
- Only 75 total clusters!

Number of Clusters = 0.125% of number of Events (Huge data reduction!)

Our Prototype



Demo of UI Prototype

- The system observed ~1.5 million total security events and ~3,500 of these events triggered alerts.
- Clustered new (unresolved) and previous (resolved) events.
- Accumulated group-level prediction (ML Score) and priority based on the cluster.
- Similarity Metric used for Clustering: Jaccard Similarity of MinHash representations of two commands to determine similarity. Locality Sensitive Hashing for faster computation.

UI Features

- Clusters of alerts in tabular format:
  - Cluster level aggregate stats
  - Timeline of events in the cluster showing pattern and trend of similar alerts
  - ML Score based on all alerts in a cluster
- Show nearest neighbor clusters of a selected cluster and the resolution for those neighbors
- Find interesting clusters based on filtering / sorting
- Escalate or Suppress all alerts in a cluster

User Feedback

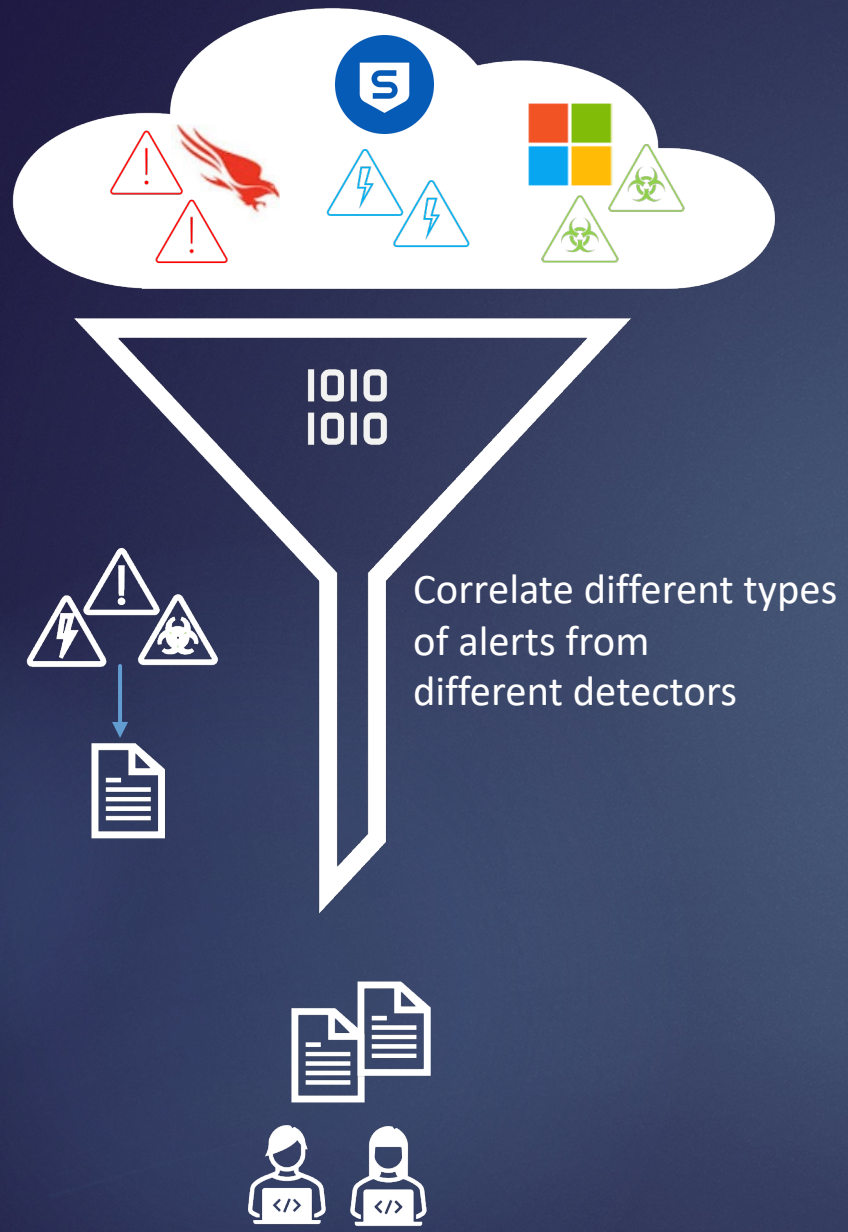
Understand analyst workflow

Understand analyst pain points

What information are useful to them

Take Away

- Clustering compresses information, reduces data load
- Reuses previous information guides new resolution
- Recycles analysts' knowledge and saves time (good for the planet)



Future Work

Generalize and Extend work beyond PowerShell to find Similar alerts across detectors and devices

Thank you!

@SophosAI

Catch our Research Director **Konstantin Berlin** and Senior Data Scientist **Adarsh Kyadige** with more questions!