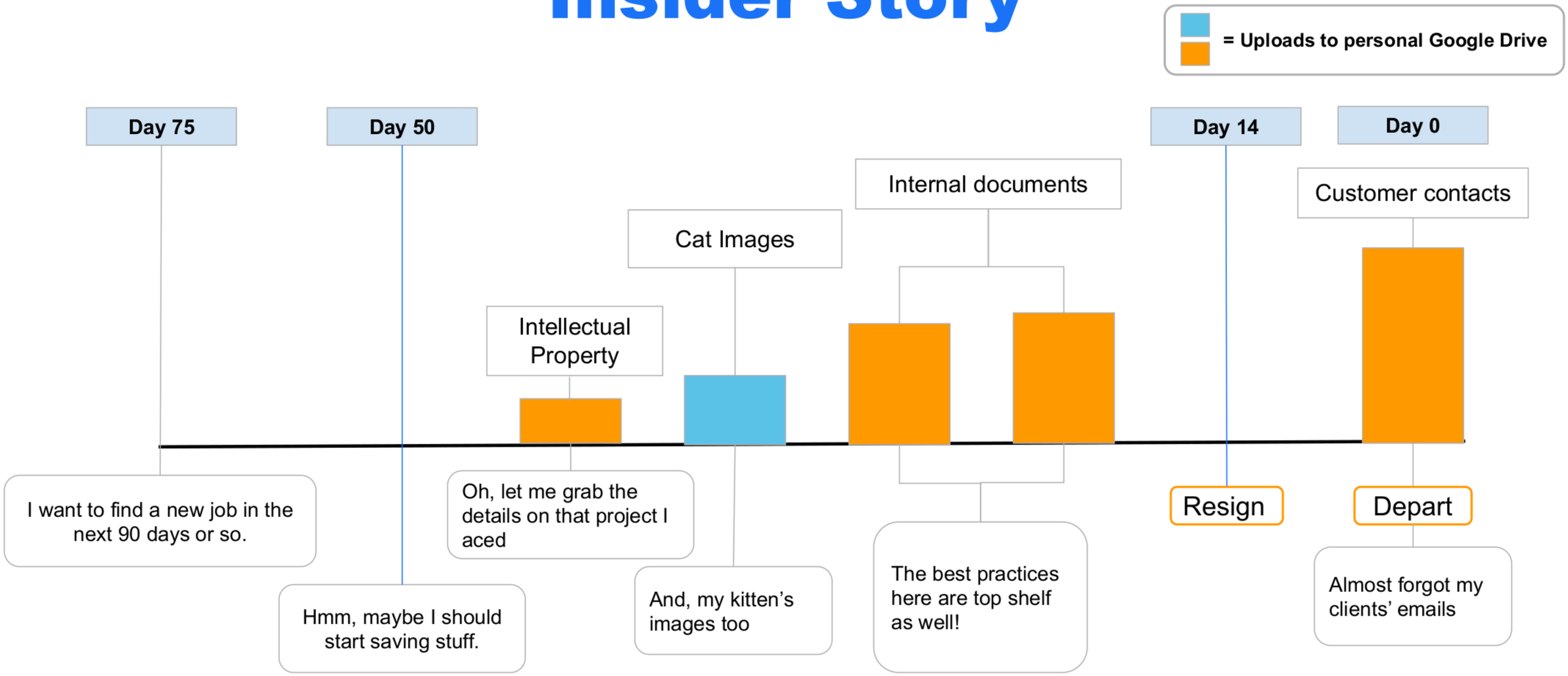
Insider Threats Packing Their Bags With Corporate Data

Dagmawi Mulugeta
Colin Estep

Insider Story



Why should you listen to us?

Our Findings

100% 

Users that left

15% 

Moved data to personal apps

2% 

Mishandled corporate data

207 organisations

4.7M active users

Important data movement starts 50 days prior to exit

Agenda

- The Problem
- Overview of our solution
- Employee Departures
- Data Exfiltration
- Takeaways

Information presented in this talk is based on anonymized usage data collected by the Netskope Security Cloud platform relating to a subset of Netskope customers with prior authorization

The Problem

The problem

A malicious insider who has exfiltrated sensitive corporate data using cloud apps.

"**Sensitive Data**" refers to data that could hurt the organization if it is exposed externally

The scope of an insider for this presentation is:

- Not using a USB drive
- Not printing out documents and walking out of the building with them
- Not taking pictures of a monitor with their phones

Why is this important?

Insiders

- A 2020 Securonix Insider Threat Report found that 60% of Insider Threats involve "Flight Risk" employees
- Every organization has “flight risk” employees

Data Exfiltration

- More organizations than ever have Personally Identifiable Information (PII) and other sensitive data
- Liability around data breaches are typically on the organization itself

Every organization should have a strategy to address this threat

Defining and Extracting Signals

Volume: Which users are downloading or uploading more than usual?

Nature: What files contain sensitive corporate information?

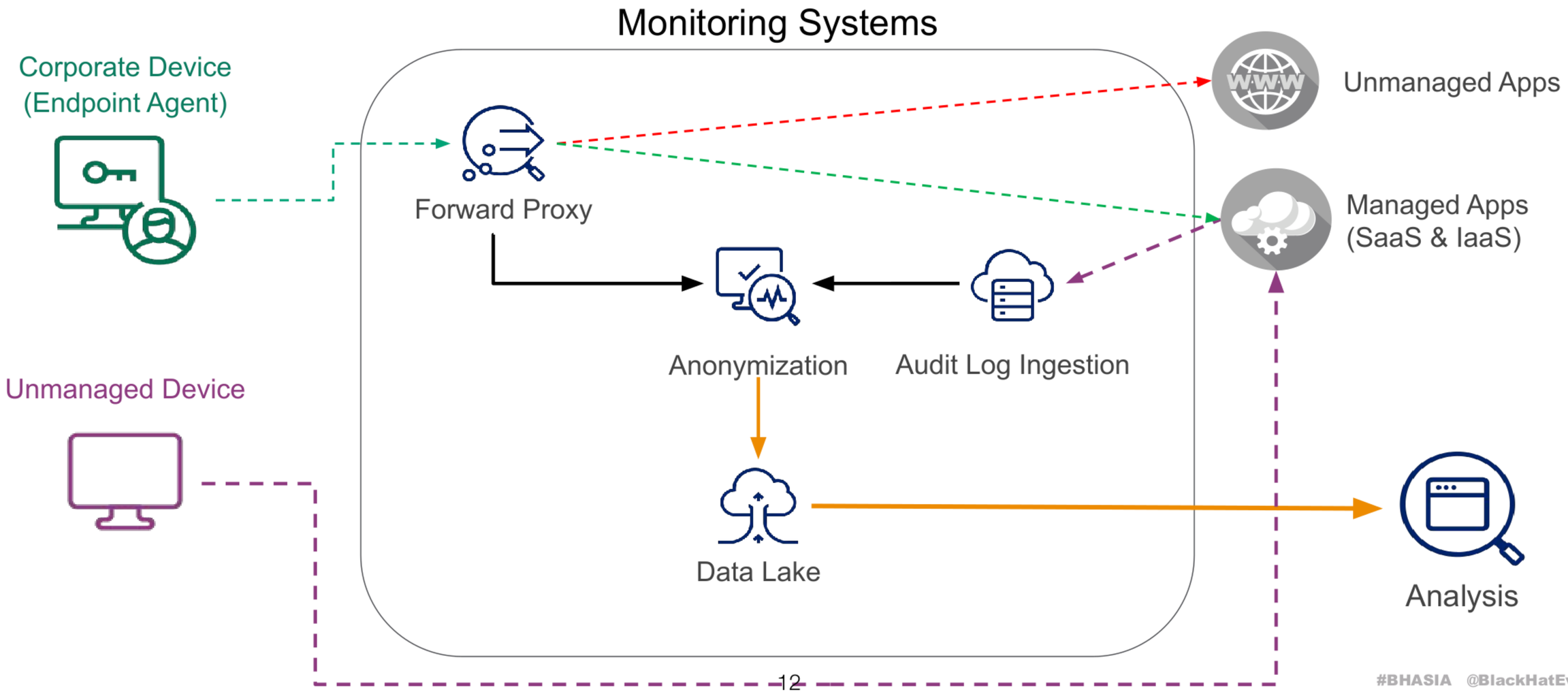
Direction: Are users are saving data to their own personal cloud storage?

Overview of our solution

Elements of our solution

- Architecture to monitor cloud traffic
- Applying labels to the cloud traffic
- What the events look like
- Analysis with Anomaly Detection

Architecture



Applying Labels: Application Instances

The domain associated with a cloud application, which indicates who controls that particular application, is an instance. We use some heuristics to label the instances as data comes in for analysis.

Application	Domain	Label	Percentage of Traffic
Google Drive	netskope.com	Business	50%
Google Drive	gmail.com	Personal	15%
Google Drive	foobar.com	Unknown	35%

Applying Labels: DLP

We need a way to label the files that contain an organization's sensitive information.

DLP policies should alert when something contains the following:

- Intellectual Property
- Secrets
- Data in scope for compliance (PCI-DSS, GDPR, etc.)

We set policies in DLP to tell us when something sensitive has been accessed.

What the events look like

User	App	App Instance label	Activity	File Name	DLP Violation
dagmawi@gmail.com	Google Drive	personal	upload	black_project.docx	Secret project code names

Analyzing the Data

- Use Anomaly Detection to find changes in behavior
- Focus on data movement with DLP
- Correlation between sensitive data movement and anomalous behavior are key

The approach above produces very useful results to find data exfiltration by insiders

Employee Departures

Our Data

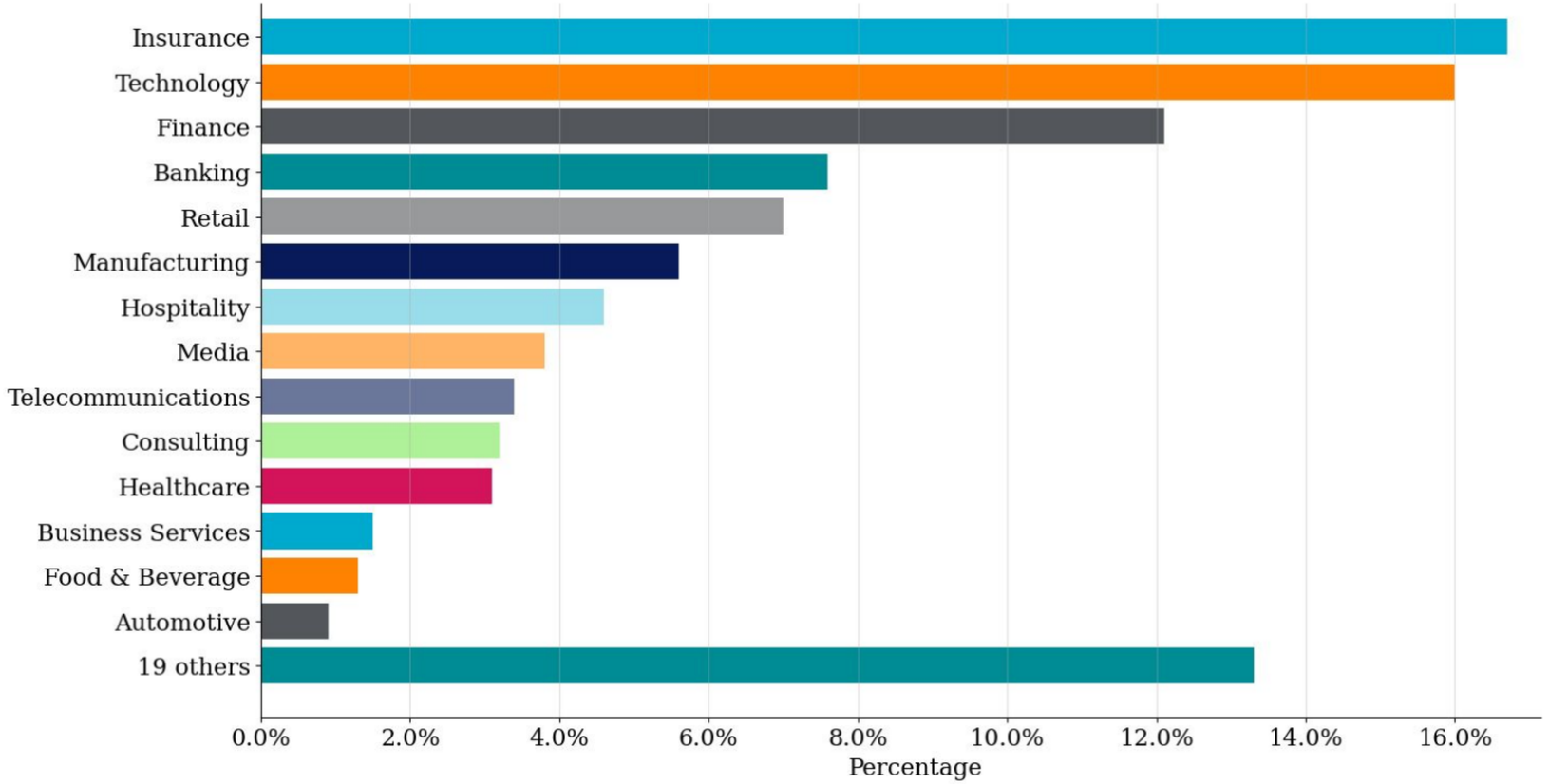
Timeline:
July 2022 to April 2023

207 organisations

4.7M active users

58,314 individuals left their employment

Industry breakdown for Departures



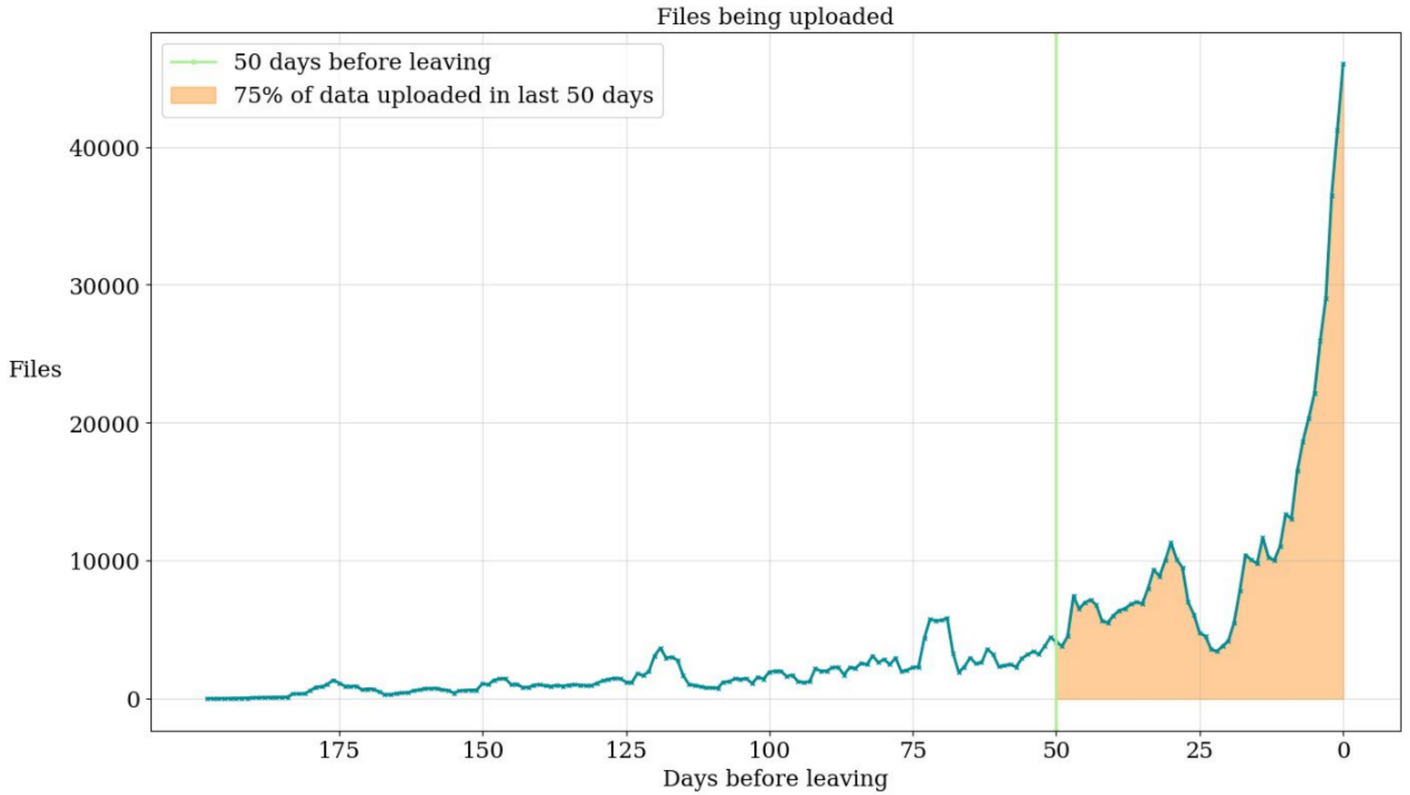
How many people move data to personal apps?

85% of flight risks **did not move data** to their personal apps

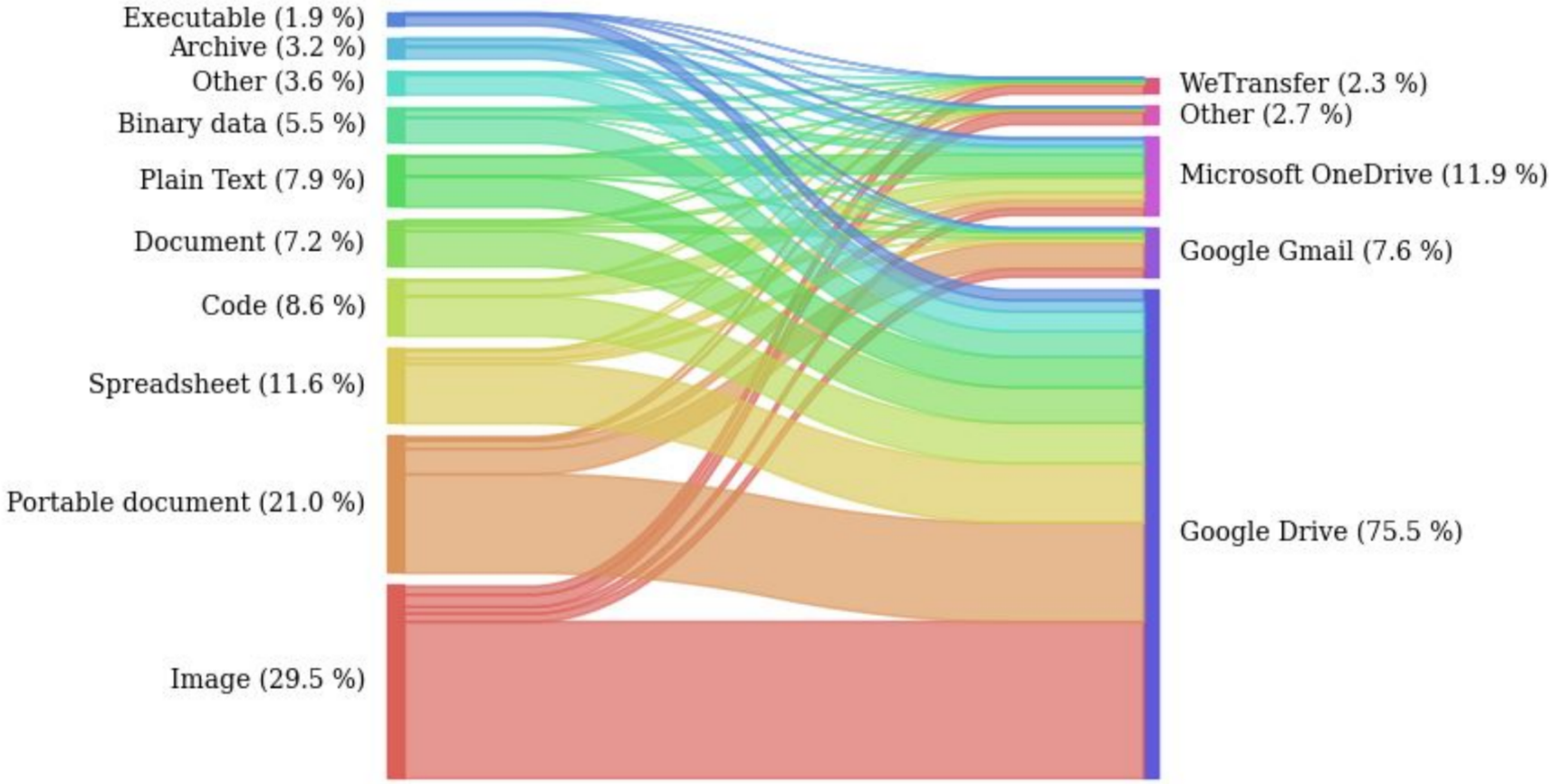
15% of flight risks **moved data** to their personal apps

When is the data moved to personal apps?

75% of all files uploaded to personal apps were uploaded in the last 50 days



What sort of data gets moved?



Files moved in the last 50 days

Data Exfiltration

What kind of data exfiltration?

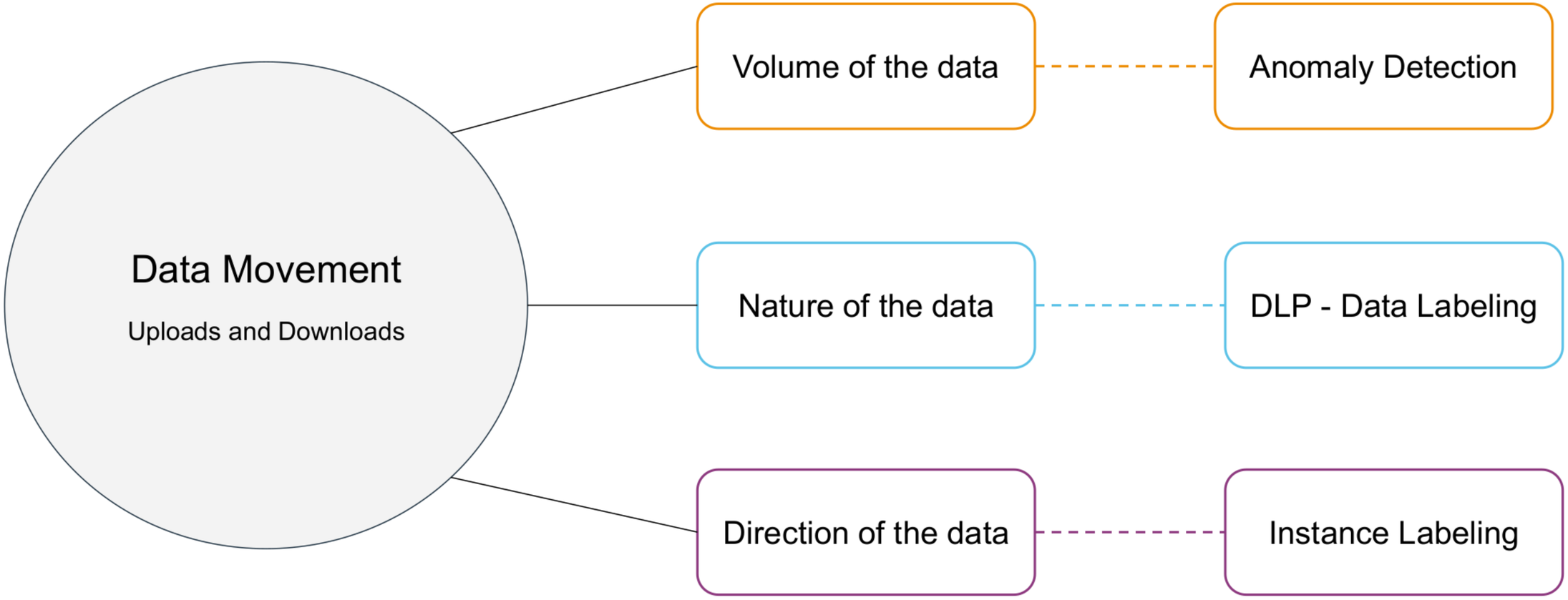
A malicious insider who has exfiltrated sensitive corporate data using cloud apps.

"**Sensitive Data**" refers to data that could hurt the organization if it is exposed externally

The scope of an insider for this presentation is:

- Not using a USB drive
- Not printing out documents and walking out of the building with them
- Not taking pictures of a monitor with their phones

Finding Data Exfiltration



Anomaly Detection

Looking for **spikes in certain activities**:

- Different from the user's own patterns
- Different from the rest of the organization

Examples:

- You uploaded 2 TB to Google Drive in one day, which is more than anyone else
- You generated 500 DLP alerts on your last day, when you normally generate 10

Detection Categories

	Heuristic	Anomaly Detection	Anomaly Detection + DLP
Baselines	X	✅	✅
Instance awareness	X	✅	✅
Data Loss Prevention	X	X	✅
Example	Alert me if anyone uploads more than 5 files to Google Drive	Alert me if someone uploads more than they usually do to their personal Google Drive	Alert me if someone uploads corporate secrets in large amount to their personal Google Drive

Detection efficacy

What is the relative signal strength of each type of detection to find someone who is going to leave?

Data Movement Detection	Improvement
Heuristic	Baseline
Anomaly Detection*	15.6 x
Anomaly Detection* + DLP	43.0 x

*Monitoring uploads are vital to get these improvements

Exfiltration by departing employees

2% exfiltrated corporate data via cloud apps

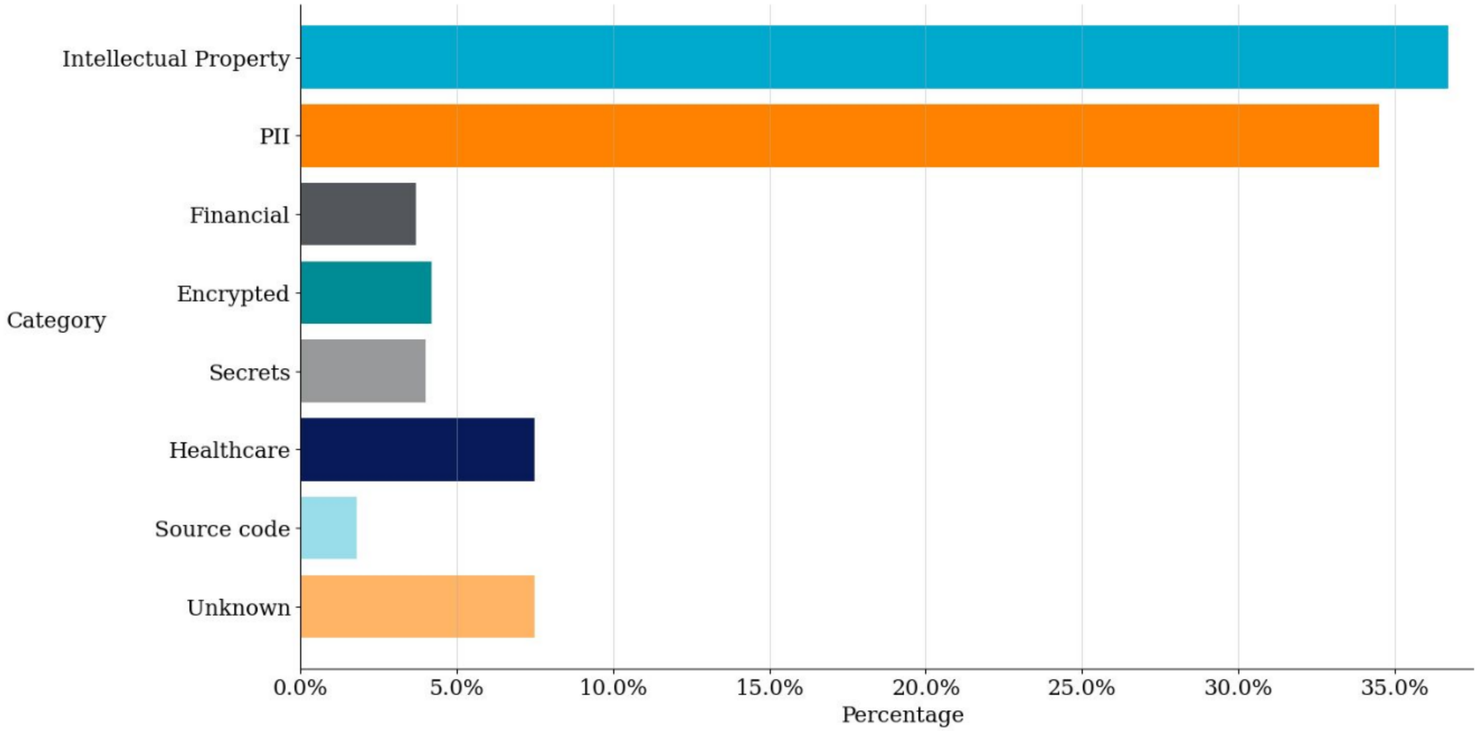
Timeline before departure

- 94% of the files exfiltrated in the last 91 days
- 84% of the files exfiltrated in the last 49 days
- 74% of the files exfiltrated in the last 28 days

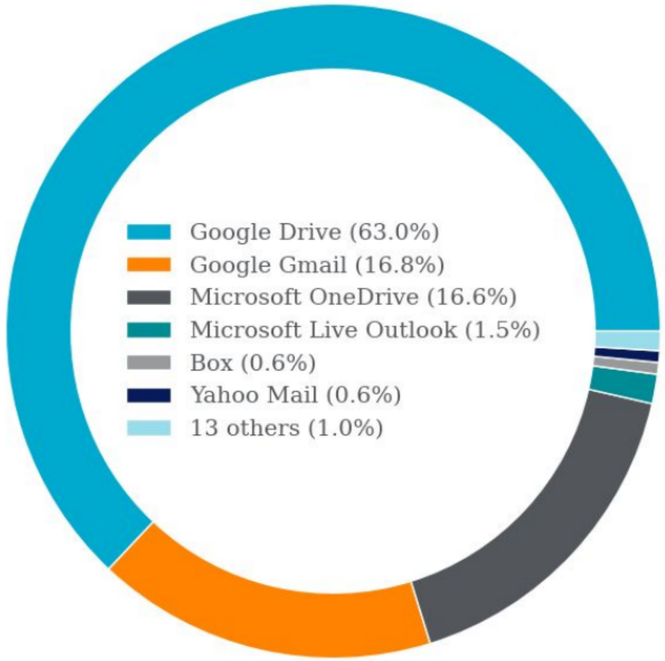
If you monitor the last 30 days of employment, you may get around 75% of the files being mishandled before someone leaves.

In order to catch the 2% of people leaving and doing this, you need proactive analysis

Data Targeted



Policies violated



Apps used

Current Limitations

Current Limitations

- Analysed a finite set of data movement sources
- Scope was insiders that end up leaving the organization, but there are ones that do not
- Unknown traffic (neither personal or business) was primarily excluded from our analysis

Future improvements

- Expanding the set of apps we analyze
- Developing other flight risk signals like uploads of resumes
- Reduction in business activities (saw ~10% reduction on overall business activities)

Takeaways

The problem

100% 

Users that left

15% 

Moved data to personal apps

2% 

Mishandled corporate data

2% is not a lot, most users are alright!

The problem

100% 

Users that left

15% 

Moved data to personal apps

2% 

Mishandled corporate data

2% is not a lot, most users are alright

- Incorrect!

~70% of the data targeted was Intellectual Property and PII

Worst case scenario

Trade Secret Theft

Investigation Into Theft of Intellectual Property from GE Leads to Two Guilty Pleas

"He thought he was the smartest guy in the room." That's how FBI Albany Special Agent Vin Manglavil described Jean Patrice Delia, who pleaded guilty to conspiring to steal trade secrets from General Electric Compa that he believed he could download th secret—and launch a company to cor what he was up to.

Tesla filed a lawsuit against a former employee this week after it learned he made changes to company source code and exported gigabytes of proprietary data to unknown third parties.

"Sernas was traveling on company business, carrying a company laptop that had the GE trade secret files on it," Murphy said. The investigation also uncovered evidence that Sernas and Delia had sent the calculations over email and uploaded them to cloud storage accounts.

Black Hat Sound Bytes

- **2%** of flight risks take sensitive data with them
- **75%** of data is uploaded in the last 50 days, before the typical 14 day notice
- Monitoring the **nature, volume,** and **direction** of data moved will allow you to detect these cases

Following up…

Twitter: Dagmawi (@dagmulu)

Linkedin: Colin (colinestep) Dagmawi (dmulugeta)

Future updates on our Netskope Threat Labs Blog

netskope
THREAT LABS

Thank you!

Questions?