DiSIEM

73rd
IFIPWG 10.4
Meeting

Building a SIEM in the Cloud

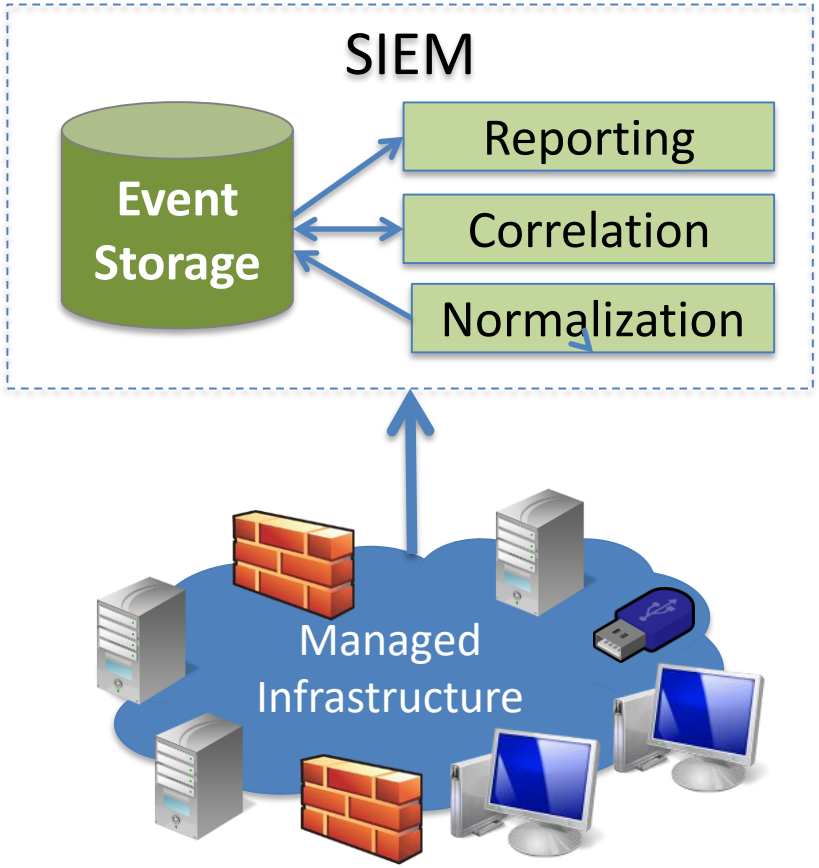
Adriano Serckumencka, Ibéria Medeiros, Alysson Bessani
LaSIGE, Faculty of Sciences, University of Lisboa

Ciências ULisboa

LASIGE
Large-Scale Informatics Systems Laboratory

Security Information and Event Management (SIEM) Systems

- Security Operation Centres: monitor and manage security of organizations infrastructures
- SIEM Systems: distributed tools used to collect, analyse and report
- Why companies spend millions to deploy SIEMs?
  - Compliance
  - Threat complexity



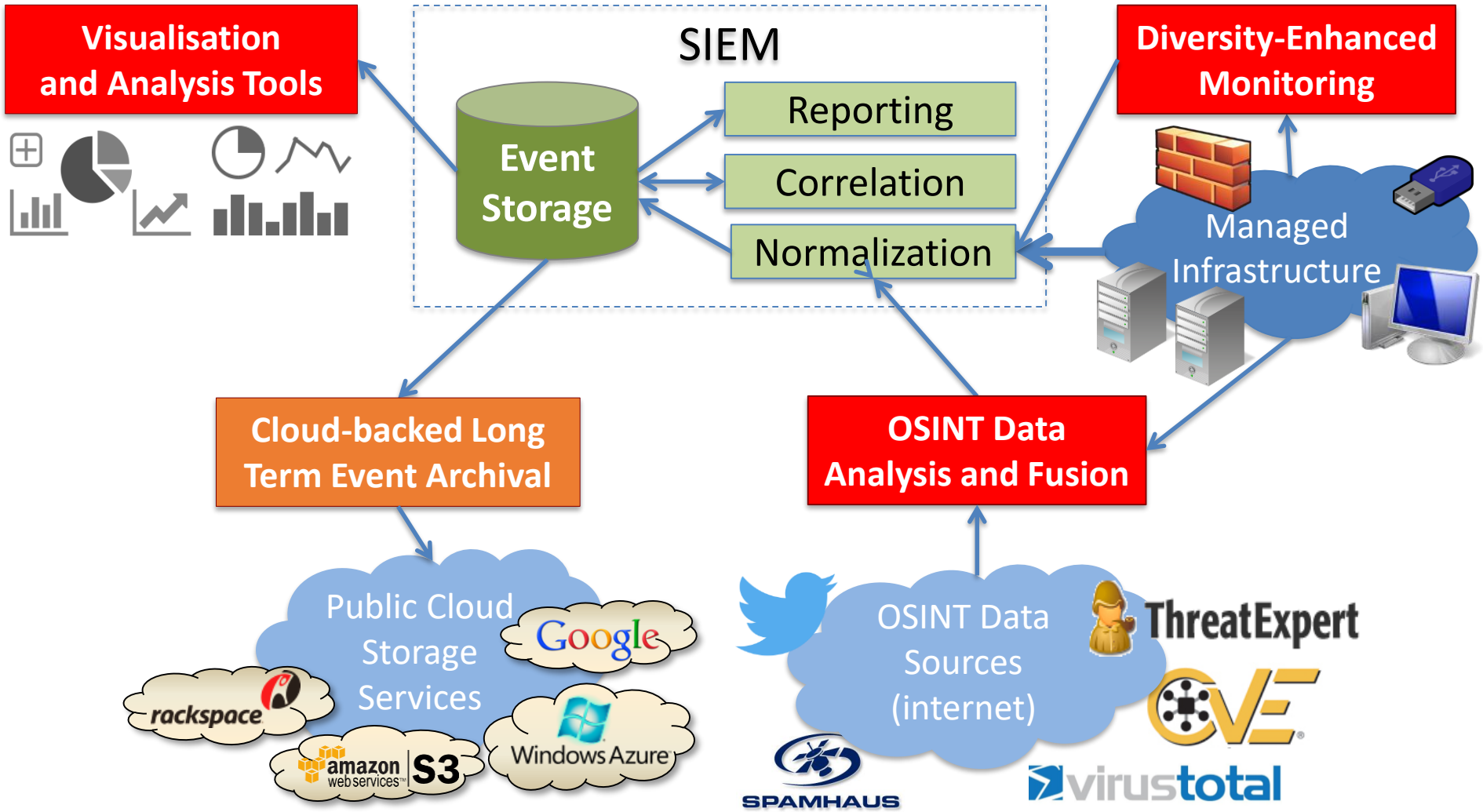
Security Information and Event Management (SIEM) Systems

- Security Operation Centres: monitor and manage security of organizations infrastructures
- SIEM Systems: distributed tools used to collect, analyse and report
- Why companies spend millions to deploy SIEMs?
  - Compliance
  - Threat complexity

Limitations of SIEM Systems

- Threat intelligence (i.e., capability of recognize and rank threats) capacity of SIEMs is still in its infancy
- Current SIEMs can show any “low level” data related with the received events, but they have little “intelligence” to process this data and extract high-level information
- Most data visualisation techniques in current SIEMs are rudimentary
- Event correlation capabilities of SIEMs are as good as the quality of the events fed to it
- SIEMs are incapable of retaining the collected events for a long duration

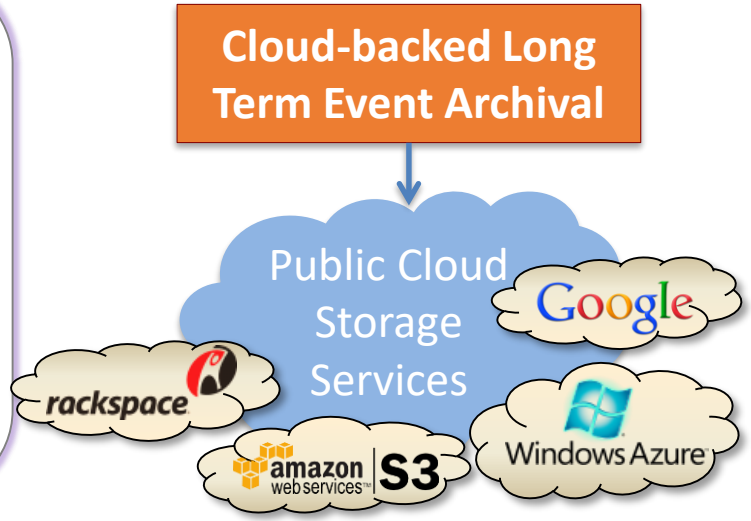
Diversity Enhancements for SIEMs



Ciências ULisboa

SIEMs are incapable of retaining the collected events for a long duration due to storage and event processing constraints

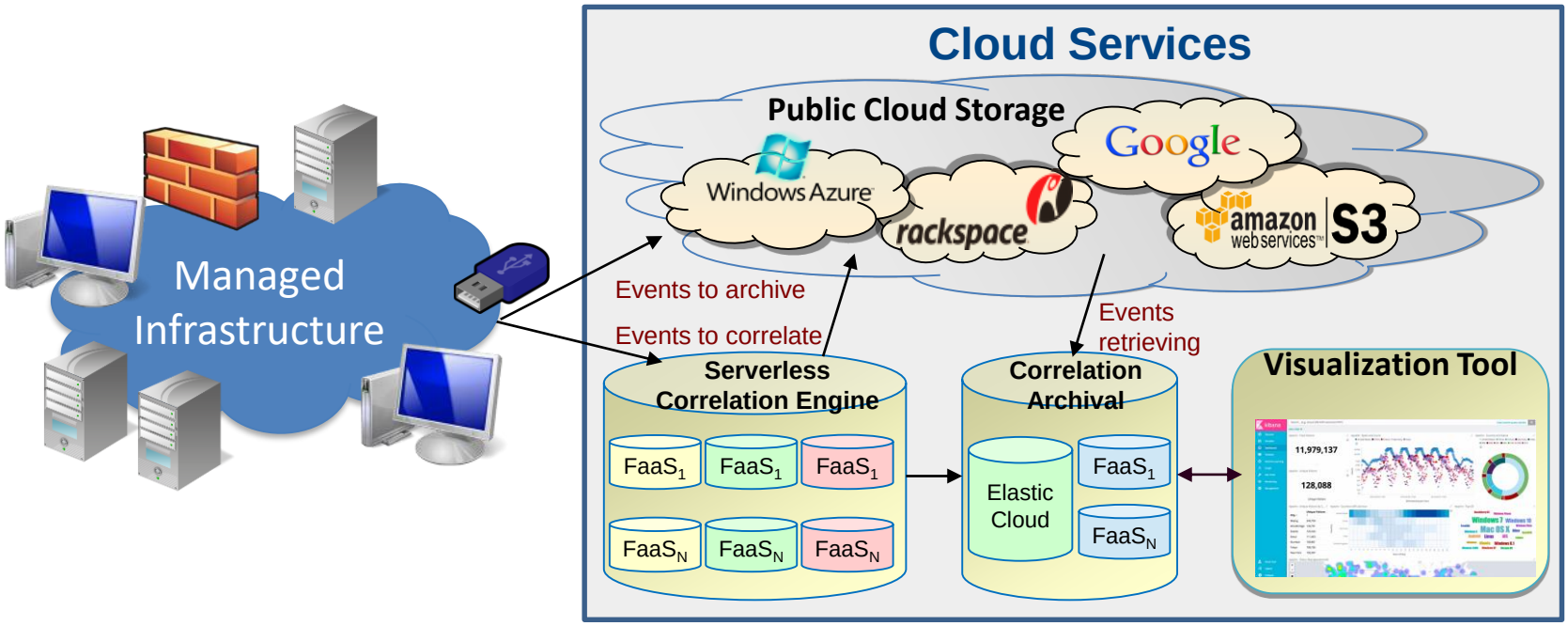
- This limits their forensic investigation capabilities in the long run
- Some zero-day vulnerabilities take up to **320 days** to be discovered
- SIEMs usually keep collected events stored by less than that



A cloud-backed system for storing selected subsets of events for long periods by using cloud storage services

→ **SLiCER** - **S**afe **L**ong-term **C**loud **E**vent a**R**chival

Why not build a SIEM in the cloud ?



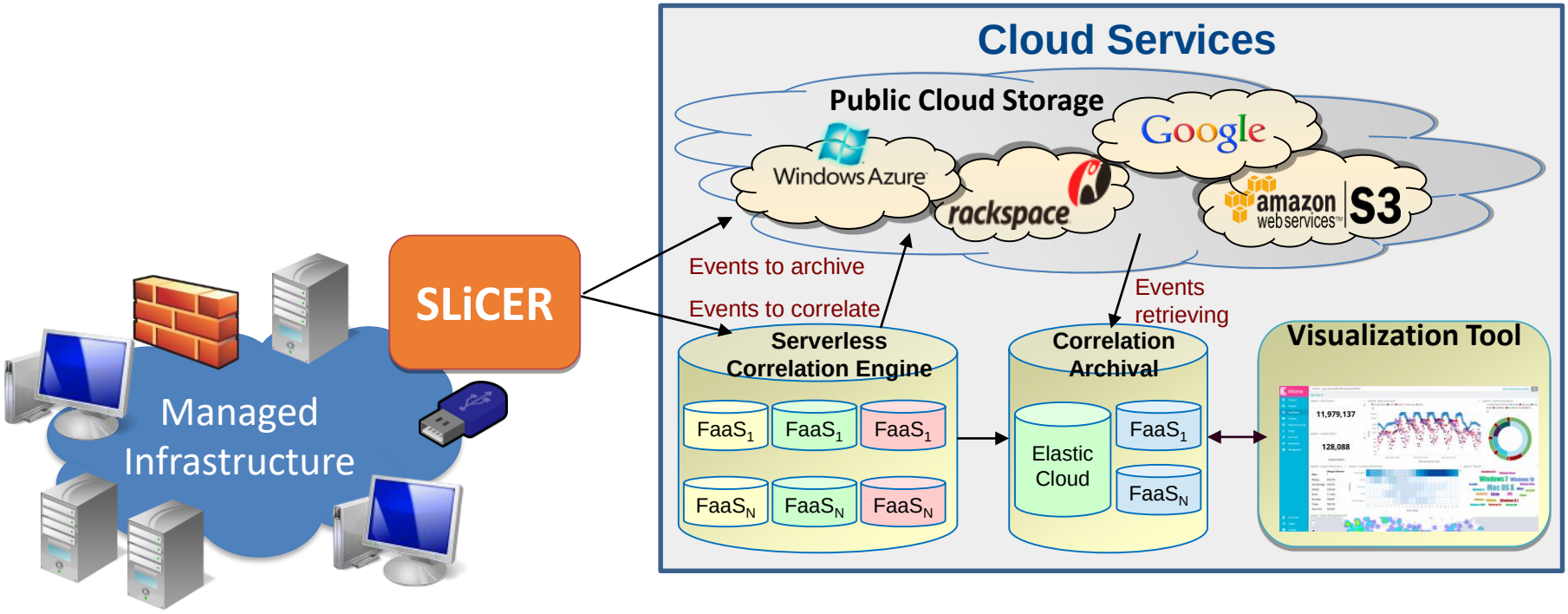
Serverless: focus on application, not on infrastructure -> Function as a Service (FaaS)

removes the need for the traditional 'always on' server system sitting behind an application

AWS Lambda

Ciências ULisboa

Why not build a SIEM in the cloud ?



Serverless: focus on application, not on infrastructure -> Function as a Service (FaaS)

removes the need for the traditional 'always on' server system sitting behind an application

AWS Lambda

SLiCER Key Features

- Organization and storage of the events in a cloud-of-clouds
  - ✓ ensures security, cloud fault tolerance, and cost efficiency
- A data model in which events are aggregated in blocks before being transferred to the clouds
  - ✓ low costs in storing and retrieving data from the cloud
- A process for indexing events in the blocks considering event properties normally used for performing searches (e.g., the IP address of the event source)
  - ✓ acceptable query performance in the cloud-backed archive

SLiCER challenges

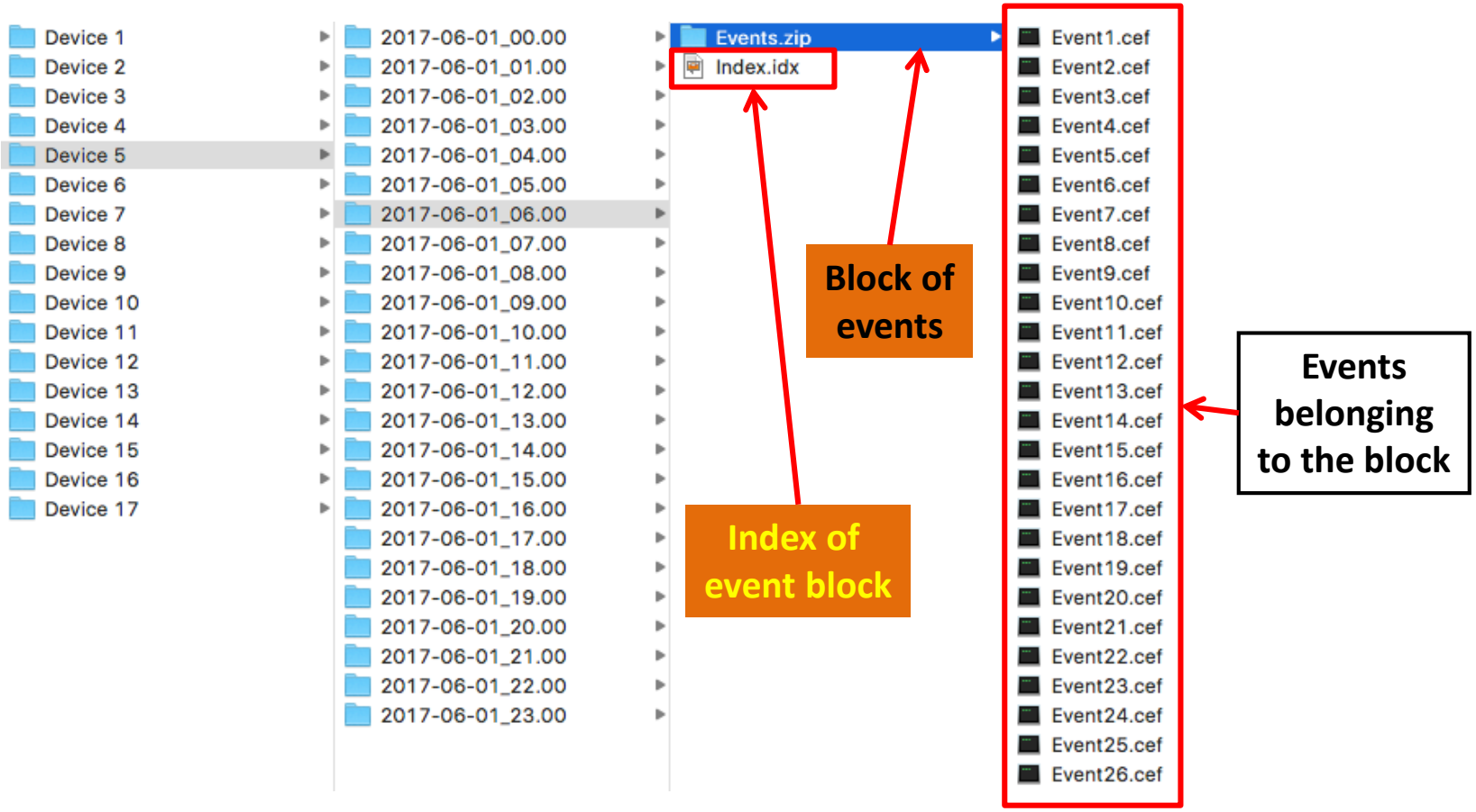
1. Organize the collected events to facilitate the queries to be performed on the archive.

2. Ensure the security of the events stored in the cloud

3. Retrieve events from the clouds in a cost-efficient way

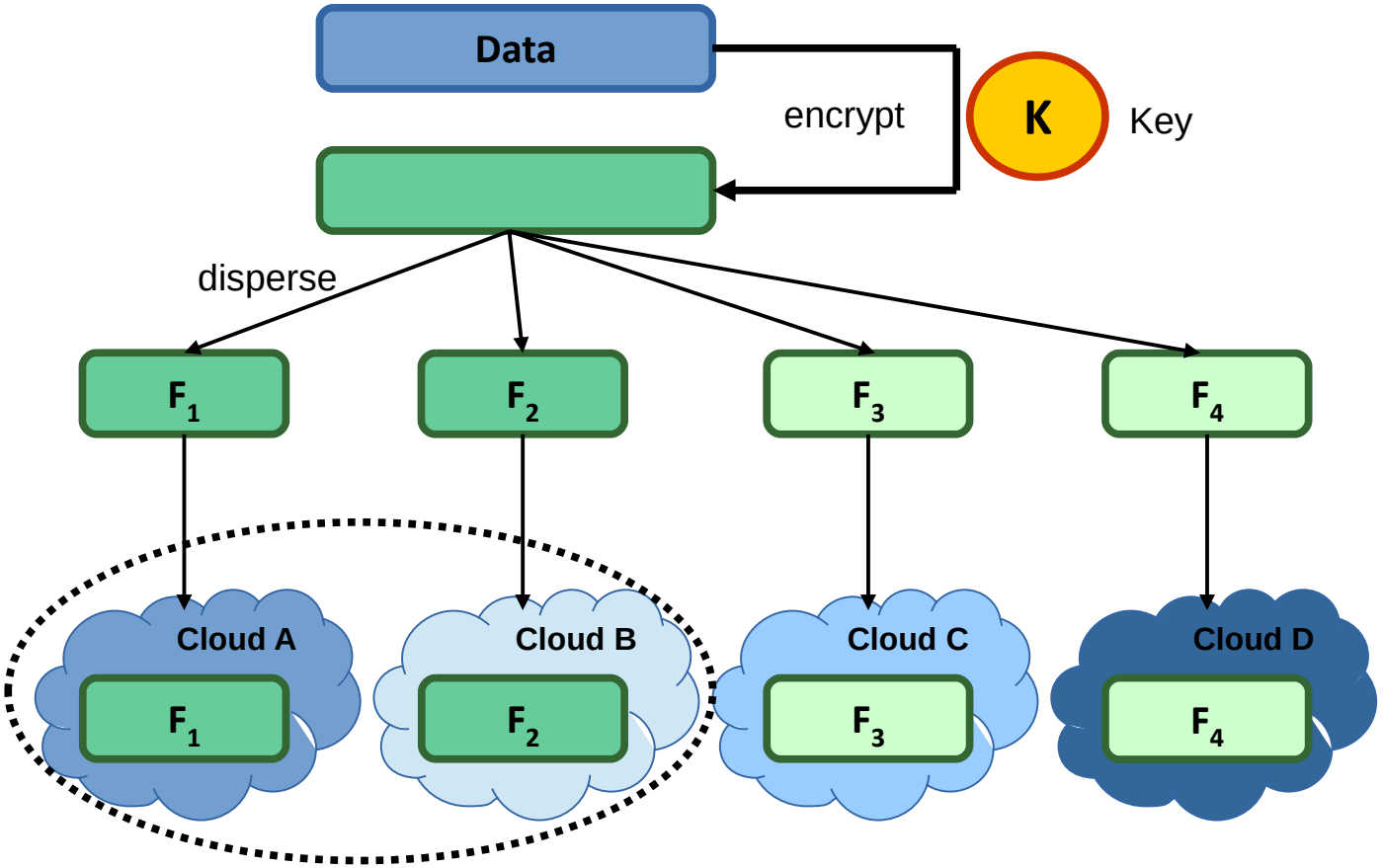
C1. Organize the events to facilitate queries on the archive

1. Create event blocks by device and interval of time
2. Create an index for each block of events



C2. Ensure the security of the events stored in the cloud

Use cloud-of-clouds storage approach for security & dependability



C3. Retrieve events from the cloud in a cost-efficient way

Storage Pricing (varies by region)

Region: EU (Ireland)

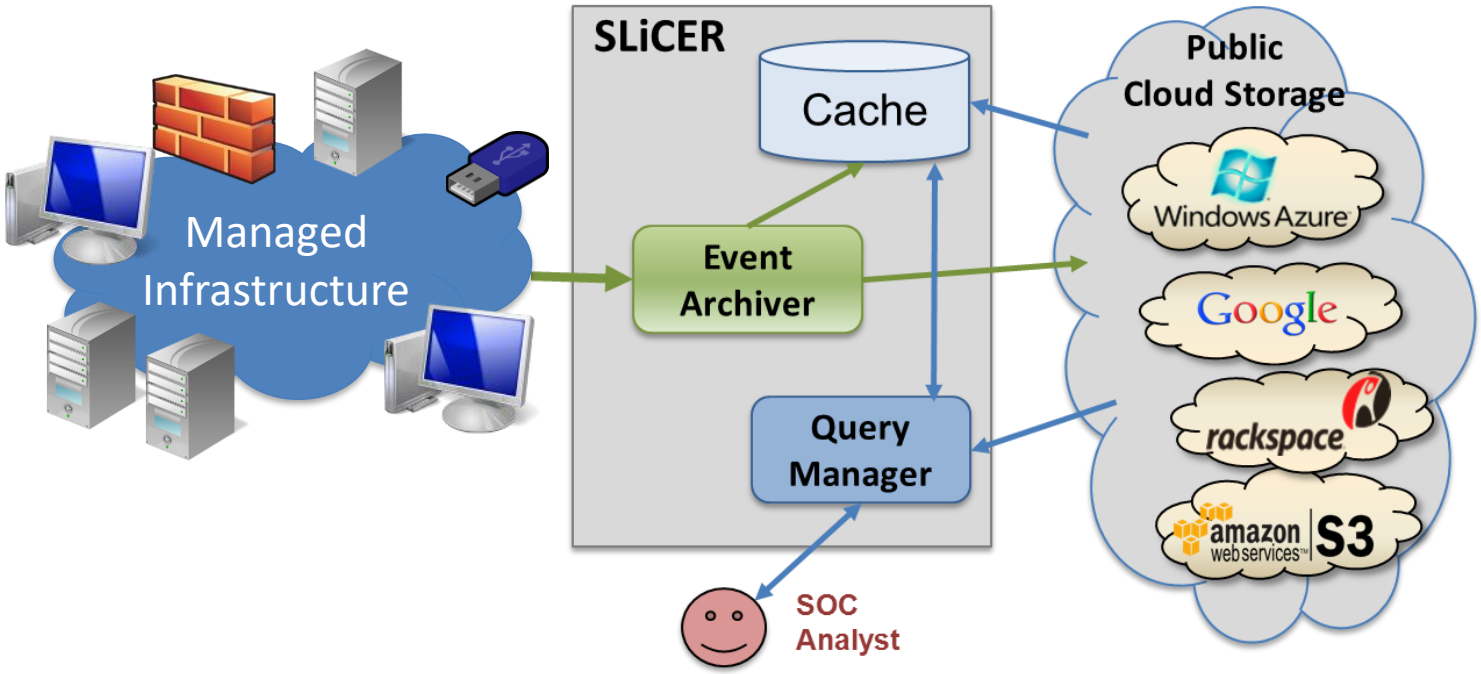
amazon web services S3

	Standard Storage	Standard - Infrequent Access Storage †
First 50 TB / month	$0.023 per GB	$0.0125 per GB
Next 450 TB / month	$0.022 per GB	$0.0125 per GB
Over 500 TB / month	$0.021 per GB	$0.0125 per GB
Data Transfer OUT From Amazon S3 To Internet		
First 1 GB / month		$0.000 per GB
Up to 10 TB / month		$0.090 per GB

> 7x more

When querying the cloud archive, we have to avoid downloading data -> indexing

SLiCER Architecture



Query Model

- A query received by SLiCER must to be represented following the query model that follows the structure of blocks defined on the data model
- A query is composed of 3 components: *devices*, *time period* and *terms*

query = (devices; startTime, endTime; terms)

devices that generated the events	range of time to search the events	set of words we are looking for in an event

Events Search

For each specified **devices**:

1. Retrieve the index file for each block containing events generated between **startTime** and **endTime**
2. Search for the **terms** in the index files
3. For each index file in which the some **term** was found, retrieve the corresponding block of events
4. Search for events satisfying **startTime, endTime** and **terms** in each block

Query Example

Get all actions made by Attacker address 192.168.1.22, at Feb.13.2017, between 00:00:00 and 00:06:00, registered by FTP and Check Point

({FTP, Check Point}; <Feb.13.2017, 00:00:00, 00:06:00>; {Attacker Address=192.168.1.22})

End Time	Name	Attacker Address	Attacker Host Name	Attacker Port	Target Address	Target Port	Target Host Name
Feb 13 2017 00:00:42	Successful Network Logon	192.168.1.22	CHAT	0	192.168.1.23		FTP
Feb 13 2017 00:00:42	User Logoff				192.168.1.23		
Feb 13 2017 00:00:42	Successful Network Logon	192.168.1.22	CHAT	0	192.168.1.23		
Feb 13 2017 00:00:42	User Logoff				192.168.1.23		
Feb 13 2017 00:01:41	Successful Network Logon	192.168.1.21	TICKET	0	192.168.1.23		
Feb 13 2017 00:01:41	User Logoff				192.168.1.23		
Feb 13 2017 00:01:41	Successful Network Logon	192.168.1.21	TICKET	0	192.168.1.23		FTP
Feb 13 2017 00:01:41	User Logoff				192.168.1.23		FTP
Feb 13 2017 00:02:21	Connector Raw Event Statistics						
Feb 13 2017 00:07:21	Connector Raw Event Statistics						

FTP server device

End Time	Name	Attacker Address	Attacker Host Name	Attacker Port	Target Address	Target Port	Target Host Name
Feb 13 2017 00:03:03	accept	172.16.24.100		61862	10.100.2.101	8443	
Feb 13 2017 00:05:08	accept	172.16.27.150		58403	255.255.255.255		
Feb 13 2017 00:05:08	accept	10.100.1.100	DC	59306	10.10.0.21		
Feb 13 2017 00:05:08	accept	10.100.1.100	DC	58925	10.10.0.21		
Feb 13 2017 00:05:08	accept	192.168.1.25	MR	52868	10.100.2.101		
Feb 13 2017 00:05:08	accept	10.100.1.100	DC	59733	204.110.15.178	53	
Feb 13 2017 00:05:08	accept	192.168.1.22	CHAT	4639	10.100.2.101	8443	
Feb 13 2017 00:05:08	accept	10.100.1.106	COLEC	137	192.102.248.135	137	
Feb 13 2017 00:05:08	accept	10.100.1.100	DC	59554	10.10.0.21	53	
Feb 13 2017 00:05:09	accept	192.168.1.25	MR	61033	10.100.1.100	53	DC
Feb 13 2017 00:05:09	accept	172.16.22.103		39141	10.100.2.41	8443	
Feb 13 2017 00:05:09	accept	192.168.1.23	FTP	1395	10.100.2.101	8443	

Check Point device

Indexing

No Index

- no indexes are used, terms are assumed to be present in the blocks

Bloom Filters

- probabilistic data structures that represent a set of elements of the same type (e.g., source IP address, source port)
- fields to be indexed need to be configured by the SOC
- very small, fast to download and query, but generates false positives

Text-based Indexing (Apache Lucene)

- provides indexing and searching over documents with any type of content (formatted or not)
- larger (~20% of the original data), reasonably fast to query
- can be configured to index only certain event fields

Building a SIEM in the Cloud

Adriano Serckumencka, Ibéria Medeiros, Alysson Bessani
LaSIGE, Faculty of Sciences, University of Lisboa

Thank you!