Distributed Shuffle Index: Analysis and Implementation in an Industrial Testbed

Enrico Bacis[1], Alan Barnett[2], Andrew Byrne[2], Sabrina De Capitani di Vimercati[3], Sara Foresti[3], Stefano Paraboschi[1], Marco Rosa[1], Pierangela Samarati[3]

[1] Università degli Studi di Bergamo (*Italy*), [2] Dell EMC (*Ireland*), [3] Università degli Studi di Milano (*Italy*)

Objectives

The protection of content confidentiality as well as of access and pattern confidentiality of data moved to the cloud have been recently the subject of several investigations. The distributed shuffle index addresses these issues by randomly partitioning data among three independent cloud providers. We implemented this tecnhique in the high-performance Dell EMC platform, and our experiments confirm the practical applicability of the approach.

Distributed Shuffle Index

The *distributed shuffle index* is a B+-tree index structure that enables efficient key-based data retrieval, while guaranteeing content, access, and pattern confidentiality. It relies on the presence of three independent cloud providers to improve the protection guarantees offered by the single-provider shuffle index.

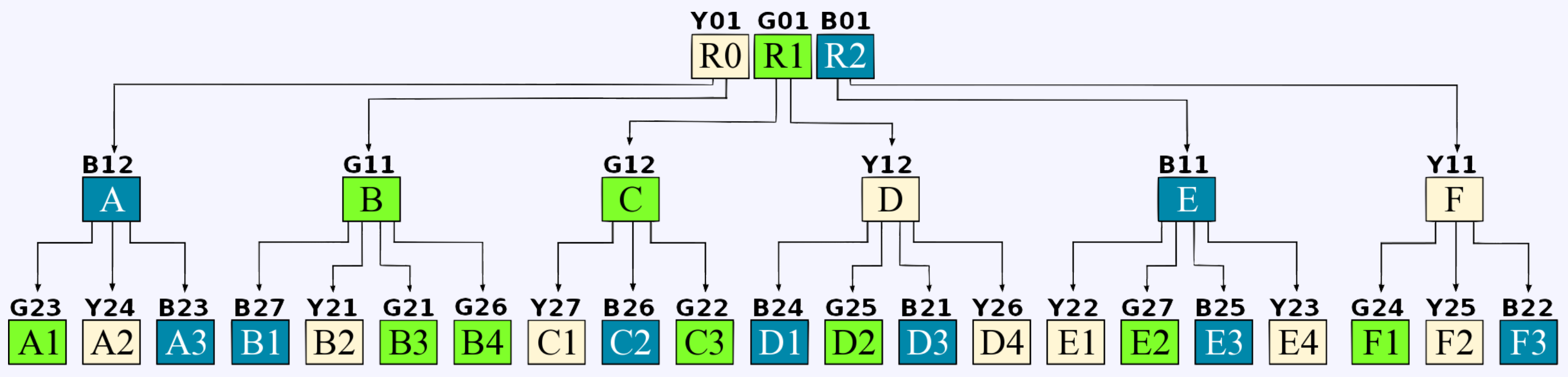


Figure: An example of shuffle index distributed at three cloud providers

Data distribution consists in allocating the nodes composing the shuffle index at three different and independent cloud providers.

Swapping consists in continuously changing the physical allocation of accessed data, which are moved to a different cloud provider after each access.

These techniques guarantee protection of access confidentiality also in case of collusion among the cloud providers.

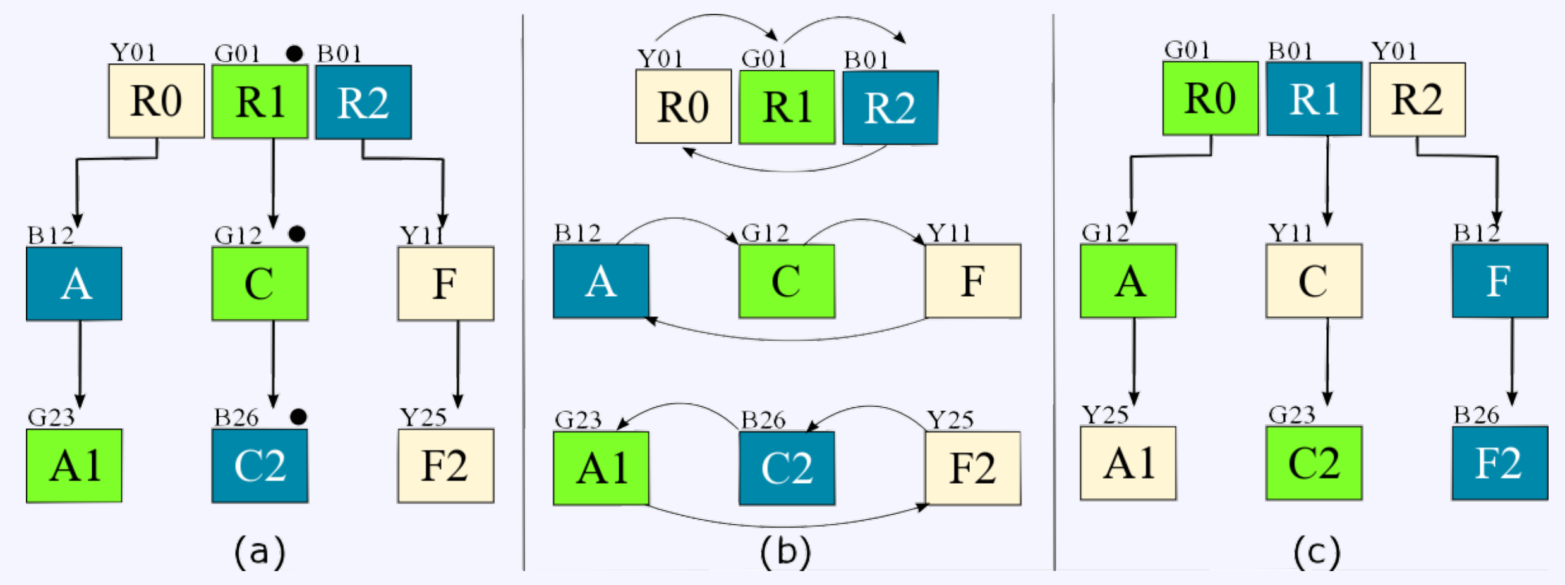
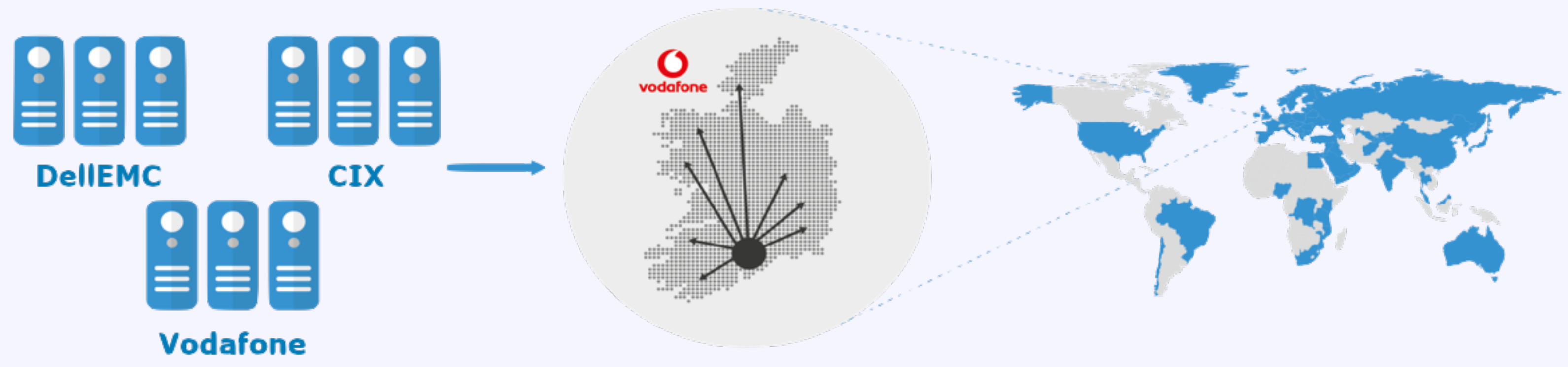


Figure: An example of search for *C2* on the index (a), swapping between accessed nodes (b), and structure of the paths after the access (c)

Industrial Prototype

The distributed shuffle index has been implemented and deployed in the Dell EMC's INFINITE (INternational Future INdustrial Internet TEstbed - www.iotinfinite.org) platform. INFINITE is an IoT innovation platform built for the development of Industrial IoT products and solutions across a wide and diverse range of industries and sectors. It is a strategic initiative led by Dell EMC, Vodafone Ireland and partners, and it is the first-of-its kind in Europe.



The testbed is composed of a full mobile network (2G to LTE) covering the island of Ireland and a cloud infrastructure with a distributed data center architecture. The data center network has 10 Gbps capacity and spans three geographically diverse sites: *i)* *Dell EMC* datacenter with a VMWare cluster and analytics platforms; *ii)* *Vodafone* datacenter with compute, storage, and networking resources; *iii)* *CIX* datacenter with compute, storage, and networking resources.

Deployment Model

Elastic Cloud Storage (ECS) is a turnkey software-defined, cloud-scale, object storage platform selected as a target application for integration with the distributed shuffle index.

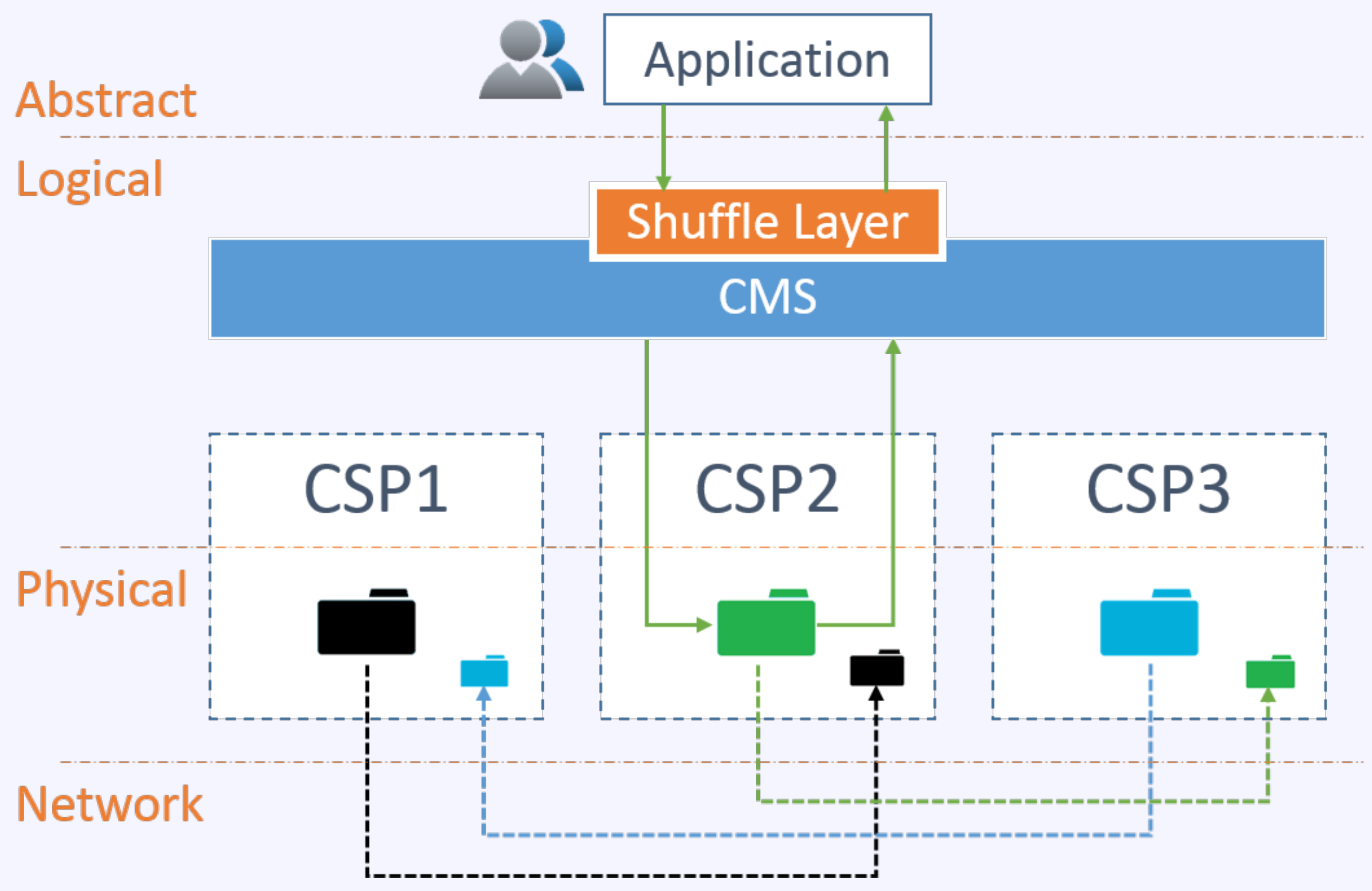


Figure: Shuffle index deployment model

The shuffle index was deployed on a VM running a Content Management Service responsible for the distribution of data across ECS nodes. ECS nodes were deployed to the testbed in three installations not 'aware' of each other. Once operational, each single-node configuration of ECS was migrated to a different physical location of the testbed.

Experimental Results

The data structure used for the experiments is a 2-level index with fan-out 27 and we performed 100 accesses over the index. The table below shows a comparison between the distributed shuffle index and a plain encrypted index with the same static structure.

	Plain encrypted index	*Distributed shuffle index*
ECS	$0.04210s$ $\sigma = 0.01322s$	$0.19674s$ $\sigma = 0.08075s$
Commercial Providers	$0.56777s$ $\sigma = 0.25588s$	$1.07609s$ $\sigma = 0.42817s$

Figure: Access times and their standard deviation σ

The 10 Gbps connection speed provided by INFINITE testbed enables the rapid transmission of data during swapping operations, which effectively counteracts the potential negative impacts of the physical distribution of the experiments.

References

- E. Bacis, S. De Capitani di Vimercati, S. Foresti, S. Paraboschi, M. Rosa, and P. Samarati. Distributed shuffle index in the cloud: Implementation and evaluation. In *Proc. of IEEE CSCloud*, New York, USA, June 2017.
- S. De Capitani di Vimercati, S. Foresti, S. Paraboschi, G. Pelosi, and P. Samarati. Three-server swapping for access confidentiality. *IEEE TCC*, 2017.
- S. De Capitani di Vimercati, S. Foresti, S. Paraboschi, G. Pelosi, and P. Samarati. Shuffle index: Efficient and private access to outsourced data. *ACM Transactions on Storage (TOS)*, 11(4):1–55, October 2015.

Acknowledgment: This work was supported by the EC within the H2020 under grant agreement 644579 (ESCUDO-CLOUD)