**LITERATURE SURVEY**

**1) On the duality of resilience and privacy**

**AUTHORS:**  J. Crowcroft

Protecting information has long been an important problem. We would like to protect ourselves from the risk of loss: think of the library of Alexandria; and from unauthorized access: consider the very business of the ‘Scandal Sheets’, going back centuries. This has never been more true than today when vast quantities of data (dare one say lesser quantities of information) are stored on computer systems, and routinely moved around the Internet, at almost no cost. Computer and communication systems are both fragile and vulnerable, and so the risk of catastrophic loss or theft is potentially much higher. A single keystroke can delete a public database, or expose a private dataset to the world. In this paper, I consider the problems of providing resilience against loss, and against unacceptable access as a dual. Here, we see that two apparently different solutions to different technical problems may be transformed into one another, and hence give better insight into both problems.

**2) Depsky: dependable and secure storage in a cloud-of-clouds**

**AUTHORS:** A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa

The increasing popularity of cloud storage services has lead companies that handle critical data to think about using these services for their storage needs. Medical record databases, large biomedical datasets, historical information about power systems and financial data are some examples of critical data that could be moved to the cloud. However, the reliability and security of data stored in the cloud still remain major concerns. In this work we present DepSky, a system that improves the availability, integrity, and confidentiality of information stored in the cloud through the encryption, encoding, and replication of the data on diverse clouds that form a cloud-of-clouds. We deployed our system using four commercial clouds and used PlanetLab to run clients accessing the service from different countries. We observed that our protocols improved the perceived availability, and in most cases, the access latency, when compared with cloud providers individually. Moreover, the monetary costs of using DepSky in this scenario is at most twice the cost of using a single cloud, which is optimal and seems to be a reasonable cost, given the benefits.

**3) Nccloud: A network-coding-based storage system in a cloud-of-clouds**

**AUTHORS:** H. Chen, Y. Hu, P. Lee, and Y. Tang

To provide fault tolerance for cloud storage, recent studies propose to stripe data across multiple cloud vendors. However, if a cloud suffers from a permanent failure and loses all its data, we need to repair the lost data with the help of the other surviving clouds to preserve data redundancy. We present a proxy-based storage system for fault-tolerant multiple-cloud storage called NCCloud, which achieves cost-effective repair for a permanent single-cloud failure. NCCloud is built on top of a network-coding-based storage scheme called the functional minimum-storage regenerating (FMSR) codes, which maintain the same fault tolerance and data redundancy as in traditional erasure codes (e.g., RAID-6), but use less repair traffic and, hence, incur less monetary cost due to data transfer. One key design feature of our FMSR codes is that we relax the encoding requirement of storage nodes during repair, while preserving the benefits of network coding in repair. We implement a proof-of-concept prototype of NCCloud and deploy it atop both local and commercial clouds. We validate that FMSR codes provide significant monetary cost savings in repair over RAID-6 codes, while having comparable response time performance in normal cloud storage operations such as upload/download.

**4) Scalia: an adaptive scheme for efficient multi-cloud storage**

**AUTHORS:** T. G. Papaioannou, N. Bonvin, and K. Aberer

A growing amount of data is produced daily resulting in a growing demand for storage solutions. While cloud storage providers offer a virtually infinite storage capacity, data owners seek geographical and provider diversity in data placement, in order to avoid vendor lock-in and to increase availability and durability. Moreover, depending on the customer data access pattern, a certain cloud provider may be cheaper than another. In this paper, we introduce Scalia, a cloud storage brokerage solution that continuously adapts the placement of data based on its access pattern and subject to optimization objectives, such as storage costs. Scalia efficiently considers repositioning of only selected objects that may significantly lower the storage cost. By extensive simulation experiments, we prove the cost-effectiveness of Scalia against static placements and its proximity to the ideal data placement in various scenarios of data access patterns, of available cloud storage solutions and of failures.

**5) Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services**

**AUTHORS:** Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha

By offering storage services in several geographically distributed data centers, cloud computing platforms enable applications to offer low latency access to user data. However, application developers are left to deal with the complexities associated with choosing the storage services at which any object is replicated and maintaining consistency across these replicas.

In this paper, we present SPANStore, a key-value store that exports a unified view of storage services in geographically distributed data centers. To minimize an application provider's cost, we combine three key principles. First, SPANStore spans multiple cloud providers to increase the geographical density of data centers and to minimize cost by exploiting pricing discrepancies across providers. Second, by estimating application workload at the right granularity, SPANStore judiciously trades off greater geo-distributed replication necessary to satisfy latency goals with the higher storage and data propagation costs this entails in order to satisfy fault tolerance and consistency requirements. Finally, SPANStore minimizes the use of compute resources to implement tasks such as two-phase locking and data propagation, which are necessary to offer a global view of the storage services that it builds upon. Our evaluation of SPANStore shows that it can lower costs by over 10x in several scenarios, in comparison with alternative solutions that either use a single storage provider or replicate every object to every data center from which it is accessed.