Towards Dispersed Cloud Computing

2nd IEEE BlackSeaCom
May 30, 2014, Chișinău, Moldova

Josef Spillner, Alexander Schill
mailto:josef.spillner@tu-dresden.de
xmpp:josef.spillner@jabber.org
Data-Centric Applications in the Cloud

Temperature sensor "INF 3084"

JSON message stream

```json
{
  "sensorevent": {
    "type": "AMBIENT_TEMPERATURE",
    "values": [9.81],
    "timestamp": "2014-03-18T07:52:00+01:00"
  }
}
```

(every 5s to xmpp://inf3084@conference.mobilis-dev.inf.tu-dresden.de/

Function 1: Moving Average

Function 2: Incident Reporting

Mobile Application

Assumptions:
- we trust the sensor admin
- we hardly trust the infra ("honest-but-curious")
- we want precise moving average
- we want incidents with a certain probability (no false negatives)
- all of that in the Cloud
Data Management

<table>
<thead>
<tr>
<th>abstract</th>
<th>files (and sockets)</th>
<th>databases</th>
<th>web services</th>
</tr>
</thead>
<tbody>
<tr>
<td>storage</td>
<td>write()</td>
<td>INSERT</td>
<td>POST, PUT</td>
</tr>
<tr>
<td>retrieval</td>
<td>read()</td>
<td>SELECT</td>
<td>GET</td>
</tr>
<tr>
<td>update</td>
<td>append(), seek()</td>
<td>UPDATE</td>
<td>UPDATE</td>
</tr>
<tr>
<td>removal</td>
<td>unlink()</td>
<td>DELETE</td>
<td>DELETE</td>
</tr>
<tr>
<td>processing</td>
<td>-----</td>
<td>aggregates, procedures</td>
<td>virtual resources, new methods</td>
</tr>
</tbody>
</table>

Are the assumptions met?
Distributed Computing Models

- Distributed hashtables
- Master-master/slave
- Filesharing
- Partitioning
- Distributed databases
- Partition + compose
- Distributed filesystems
- Distributed event processing
- GFS
- RAID
- Distributed web services
- Distributed networks
- Service composition
- Service multiplexing

- Internet
- Federated / distributed with awareness of peers
- Parallel / distributed without awareness of peers (homogeneous, heterogeneous)
- Multipath TCP
Dispersed Data Management

trade-offs: capacity, price, read/write performance, availability, computability

bit splitting (with parity) erasure coding, e.g. CRS secret sharing
Dispersion Examples

(k+m) erasure coding

bitsplitting

replication

secret sharing
Dispersed Computing Models

Storage

Communication/Networking

Computation/Processing

Client

Channel → Target

Channel → Target

Channel

Channel

Client

Channel

Channel

Processor

Processor

Processor

Sender

Channel

Channel

Channel

Receiver
Combined Dispersed Cloud Scenario
Dispersed Cloud Scenario Application

Dispersed Transmission

Dispersed Processing

Dispersed Storage

Function 1: Moving Average

Function 2: Incident Reporting

Mobile Application

Reduction

\[ v_1, v_2 = \text{split}(v) \]

\[ \frac{(x(t-0)+...+x(t-n))}{(n+1)} \]

But what about security?
Combination with:
- homomorphic encryption
- order-preserving encryption
Splitter-NG → lab.nubisave.org

Plugins: Jerasure, JSharing, RAID-1
+ Bitsplitter: Variable-bit stream k+1 splitter in C

Implementation: Dispersion
Implementation: Dispersed Storage

NubiSave → nubisave.org
Implementation: Dispersed Communication

MoreBalance → morebalance.coolprojects.org

Client → Network → Server


dispersing proxy chain

direct proxy

no proxy
Benchmark: Dispersed Communication

Traffic Forwarding Performance

Transmission speed (MB/s)

0 100 200 300 400 500 600 700

1 4 16 64 256 1024 2048 4096

Data size (MB)

Direct

Forward

Dispersion
Implementation: Dispersed Computation

Dispersed Search → lab.nubisave.org

Client

Processor/Reducer: mpolo-concentrator

Network

simple search

Server

Processor 1: mpoloworker

dispersed search

Processor 2: mpoloworker
Benchmark: Dispersed Computation

Search Function Performance

- Simple
- Dispersed
- Dispersed/MT

Compute time (s)

Local  |  LAN  |  WLAN  |  WAN

0      |  50   |  100   |  150   |  200   |  250   |  300   |  350   |
Benchmark: Dispersed Computation

Search Function Performance Distribution

- Simple
- Dispersed
- Dispersed/MT

Compute time (s) vs Measurement run
Summary & Conclusions

Dispersed Storage & Networking
- established topics
- open-source prototypes from our Cloud Storage Lab (lab.nubisave.org)

Dispersed Processing
- hot research topic
- promising algorithms: search, mean average, fixed-point approximation

Dispersed Cloud Computing
- application of Dispersed Computing techniques for complex Cloud applications
- required in order to satisfy growing requirements