Contributions to Session Aware Frameworks for Next Generation Internet Services

PhD Dissertation Defense by

Narjess AYARI

Phd. Prepared at

Orange Labs (SIRP/ASF/INTL) and RESO, LIP – Université de Lyon, France.

Advisors:

Prof. Pascale Vicat-Blanc Primet
Dr. Laurent Lefèvre
Eng. Denis Barbaron

Phd. Advisor (RESO – INRIA - LIP (UMR 5668 CNRS, ENS, UCB))
Phd. Advisor (RESO – INRIA - LIP (UMR 5668 CNRS, ENS, UCB))
Phd. Advisor (Orange Labs)

Tuesday, October 21th, 2008.
- Quantitative growth of the Internet
- New killer applications
  - Voice over IP, video streaming, multi-players games, etc.
- Service success depends on the performance and robustness guarantees over the Internet
- Network operators are concerned with E2E QoS issues while addressing the explosive growth of Internet users
  - Core network contribution in the E2E QoS
  - Edge network contribution in the E2E QoS

This research focuses on the contribution of edge network processing equipments in improving the E2E QoS.
This research focuses on the requirements of scalable and highly available Internet servers subjected to multiple flow-based user sessions.

Service Provider related example

Commercial Web Session
This research focuses on the requirements of scalable and highly available Internet servers subjected to multiple flow-based user sessions.
Operators use resources' redundancy to improve Internet servers' robustness.

What is High Availability?
✔ The capability of a system or a component to continue normal operation despite the occurrence of hardware or software fault.

✔ Availability
✔ Reliability
Limitations of the Existing High Availability Frameworks

Resources
- Overhead of the resource consumption
  - CPU, memory, I/O usage
  - Bandwidth usage
  - Energy usage
  - ...

Fault Tolerance
- Fault detection granularity
- Group communication style
- Fault recovery granularity
- Transparency of the failure recovery
- ...

Performance
- Fault handling procedure
  - Fault detection latency
  - Replica launch latency
  - Fault recovery latency
  - ...

Targeted properties
- True Client server transparency
- Low overhead during failsafe periods
- Good performance across failures

The strength of the fault model
- Fault detection granularity
- Group communication style
- Fault recovery granularity
- Transparency of the failure recovery
  - ...

Context (Cont.)
Key points
Summary of our Contributions
We want to:

- Reduce the number of interrupted sessions
  - Under server overload
  - Across server failures
- Reduce the number of "angry" clients
We advocate Session Awareness as the concept guaranteeing to take into account the characteristics of the offered network traffic to the Internet server while processing it.

- **Session integrity for cluster-based servers**
  - Guarantee the association of the flows pertaining to the same session to the same cluster resource

- **Session duration & volume for cluster-based and standard servers**
  - QoS in terms of the completion of the offered sessions independently of their duration
    - Lower discrimination against long term sessions ➔ close to real case Business Models

We have provided realistic and efficient Session Aware schemes for improving the E2E QoS while maximizing the operator's Profitability.
Proposal of a Session Aware Framework for High Performance Cluster-based Internet Servers

Proposal of an Active Replication-based Framework that provides High Availability capabilities for Internet Servers across failures.

Proposal and Design of Session Aware Admission Control mechanisms that improve the Scalability pattern of Internet Servers under overload condition
Outline

Part I: Contributions to Session Aware Architectures for High Performance Cluster-based IP Services

Part II: Proposal of An Active Replication-based Framework for Highly Available Internet Servers and Stateful Equipments

Part III: Proposal of Session Aware Admission Control Strategies for Scalable Internet Servers

Part IV: Conclusion and Future Works
Part I

Proposal of a Session Aware Framework for High Performance Cluster-based Internet Servers
A Session Aware Architecture for High Performance Cluster-based Internet Servers

Functional Grouping
- Design Space
- General Architecture
- Summary of our Contributions

Summary of our Contributions

Availability Monitoring
- Availability Monitoring
- Load Accounting
- Admission Control
- Failure Recovery
- Server Selection
- Session Identification
- Failure Recovery
- Session Identification
- Server Selection
- Admission Control
- Load Accounting
- Availability Monitoring

Data Plane
- Data Forwarding
- Session Awareness
- Failure Recovery

Control Plane
- Failure Recovery
- Server Selection
- Session Identification
- Admission Control

Session Awareness Plane
- Availability Monitoring
- Load Accounting

Design Space
- Enhanced Reliability
- Coarse Grained
- Fine Grained

LVS'98
BEA'02
Our Proposal
**Functional Grouping**

**Design Space**

**General Architecture**

**Summary of our Contributions**

- **Data Forwarding**
- **Internet**
- **Backup Dispatcher**
- **Primary Dispatcher**
- **Default GW**
- **Outgoing Traffic**
- **Incoming Traffic**
- **Admission Control, Server Selection**
- **Session Identification**
- **Failover**

**Session Identification**

**Admission Control, Server Selection**

**Data Forwarding**
Functional Grouping
Design Space
General Architecture
Summary of our Contributions

Data Forwarding
Incoming Traffic L2 Encap.

Session Identification
Incoming Traffic Inspector

Admission Control
Load Analyser

Load Accounting
Load Manager

Server Selection
Load Analyser

Failure Recovery
Active Replication Proc.

Load Accounting
Load Manager

Availability Manager
availability Manager

Server

Session Identification
Outgoing Traffic Inspector

Server

Load Information

Load Information

Availability Information

Load Information
We have designed the building blocks of a fine grained session aware scalable and robust infrastructure for High Performance Cluster-based Internet Servers

- One-way architecture-based
- Uses proactive techniques to provide the intelligence required by network operator to maximize its profitability, by reducing the number of interrupted sessions either
  - Under overload
  - Across failures

- [FR2007.0756191], [IEEE AICCSA'07], [IEEE ATNAC'07].
Part II

Proposal and Evaluation of an Active Replication-based framework for Highly Available Internet Servers and Stateful Equipments
**Active replication**: Redundant nodes concurrently receive and process the same Internet traffic and provide the same output.

Our Active Replication based Proposal versus Check-pointing- or Message Logging-based Proposals

- Truly client/server transparent framework
- Applies to any stateful equipment
- No overhead to the legitimate conversations during failsafe periods
- Best recovery time
- Deterministic-enough applications ➔ Adding further processing to cope with highly non-deterministic applications
Why Active Replication?

General Architecture and Design Space
An in-depth look at the Framework Building Blocks
Performance Evaluation
Summary of our Contributions

VIP: Virtual IP address of the service
The active replication components are mostly deployed at the backup node.

Three main components:
- State replication component
- Failure detection component
- Failure recovery component

Structure of the Connection Synchronization Message:

|          |          
|----------|----------
| Timestamp|          
| Source IP |          
| Destination IP |          
| Source Port | Destination Port |
|           | Sequence Number |
Consistency ensures that only one node is replying to client requests

- Filtering rules are deployed at replicas to drop the outgoing traffic produced consequently to the active replication.
- Only the management traffic is allowed between replicas.
Implementation for the Linux 2.6.18 Kernel
Performance evaluation is conducted for
Voice over IP Cluster-based Session Border Controller using SIP Stream oriented server
Replicas deploy the same service
Replicas are synchronized through the NTP service
Experiments' methodology

We achieved the following goals:
- Truly client/server transparent
- Low overhead during fail-safe periods
- Good performance across failures

Why Active Replication?
General Architecture and Design Space
An in-depth look at the Framework Building Blocks
Performance Evaluation (Cont.)
Summary of our Contributions

We achieved the following goals:
- CPU, memory, I/O usage
- Bandwidth usage
- Energy usage
- ...
$\Delta = \text{Time required by the kernel space synchronization module to receive the minimal connection state added to the time required to update the fake connection accordingly.}$
Only the "useful" traffic is processed by the backup node (signalling traffic)
Micro-benchmarks are derived to quantify cost of active replication (CPU, memory & Network buffers)

- Depends on the use-case!
- SIP experiment
  - 18% less memory usage overhead at the backup
  - 22% more CPU usage at the backup

We need CPU dimensioning for complex Active/Active redundancy
- Failure recovery latency is the time needed to detect and to recover from a primary's failure.

\[ \text{latency}_{\text{failure recovery}} = \text{latency}_{\text{failure detection}} + \text{latency}_{\text{Takeover}} \]

- Average failure recovery time as a function of the failure detection interval
  - < 1sec. for a FDI = 1 sec.
When only the network level availability is enabled:
- The average recovery rate is improved by 86%.

When the active replication is added:
- The average recovery rate is improved by 94%.

This measure is a cumulative ratio that includes the Core network loss contribution.
• [FR2007.051854], [IEEE Communications Surveys & Tutorials'08], [IEEE AINA'07], [IEEE ARES'07], [IEEE PDCAT'08], [IEEE DSN'08].
Part III

Proposal and Evaluation of Session Aware Admission Control Strategies for Scalable Internet Servers
The objective of the operator is the mean monetary equivalent due to the blocking, the completion and the interruption of the offered sessions over a long time scale.

- We associate equivalent monetary values to:
  - the **good completion** of a session ($R_c > 0$),
  - the **rejection/blocking** of a session ($C_b > 0$),
  - the **interruption** of a session ($C_i > 0$),

- **Economic Basis**: $C_i > C_b$

\[
\max_p \left( R_c N_c - C_b N_b - C_i N_i \right)
\]

- $p$ is the acceptance probability
- $N_c$, $N_b$ and $N_i$ are respectively the mean number of completed, blocked and interrupted sessions per unit of time.
SIP Traffic Profile

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Number of Concurrent Sessions</td>
<td>230</td>
</tr>
<tr>
<td>Source throughput</td>
<td>14</td>
</tr>
</tbody>
</table>

Global and Per-Transition Rewards versus Acceptance Thresholds
**Our Strategy:** Give the traffic pertaining to the active sessions a higher priority under overload.

\[
p = \begin{cases} 
0, & \text{if } l_s \leq T_1 \\
\frac{x-T_1}{T_2-T_1}, & \text{if (New Session) and } T_1 < l_s \leq T_2 \\
1, & \text{if (New Session) and } T_2 < l_s \leq C \\
1, & \text{if } l_s > C
\end{cases}
\]
Simulation Scenario & Methodology

Motivation and Optimization Problem Statement
Numerical Study of the Operator’s Reward
Proposal of a Session Aware Admission Control strategy for Internet Servers
Performance Evaluation
Summary of our Contributions

Admission Control Strategies

Request Aware (RAC)

Session Aware (SAC)

Single Threshold-based

Double Threshold-based

Voice CBR Traffic Profile Model

ON/OFF Model (ON/OFF SAC)

Responsive Model (RESP SAC)

Responsive Model (Double SAC)

Voice CBR Traffic Profile Model

\[ p = \begin{cases} 
0, & \text{if load} \leq T_1 \\
1, & \text{if } (\text{new session}) \text{ and load} > T_1 
\end{cases} \]

\[ p = \begin{cases} 
0, & \text{if load} \leq T_1 \\
\frac{\text{load} - T_1}{\text{capacity} - T_1}, & \text{if } (\text{new session}) \text{ and load} > T_1 
\end{cases} \]

\[ p = \begin{cases} 
0, & \text{if } l_s \leq T_1 \\
\frac{x_1 - T_1}{T_2 - T_1}, & \text{if } (\text{New Session}) \text{ and } T_1 < l_s \leq T_2 \\
1, & \text{if } (\text{New Session}) \text{ and } T_2 < l_s \leq C \\
1, & \text{if } l_s > C 
\end{cases} \]
We consider two simulation scenarios.

- Generating **homogeneous** sessions over the same simulation run, having a duration that ranges from **medium** to **long** term.

- Generating **mixed** sessions over the same simulation run, meaning that both medium term and long term conversations are **equitably** generated over the simulation time.
Let us simulate the behavior of a processing server over a duration of 100 sec.

Some simulation parameters:
- Mean packet size (512 bytes)
- Small inter-request think time
- No retransmission
- Traffic rate of 50 packets per unit of time.
- First threshold equals to 75% of the server capacity
- Second threshold equals to 85% of the server capacity
- Exponential average talking time

- Mixed traffic involves the generation of equal % of medium & long term sessions over the simulation time
Success rate = \( \frac{\text{Average number of completed sessions}}{\text{Total number of generated sessions}} \)

Percentage Histogram of Completed Calls for Different Admission Control Strategies - Mixed Traffic.
Session aware admission control as the means to efficiently prevent a server overload while maximizing the operator profitability.

Responsive session aware admission control is beneficial to increase the performance of a server subjected to long lived sessions

- If we define QoS as the completion of sessions independently of their duration, we can say that double threshold-based session aware admission control improves the QoS provided to subscribers by decreasing discrimination against long lived sessions

[IEEE CCNC'08] and ongoing submissions.
Part IV

Conclusion & Future Works
Service aware network management is a key issue for the current and future Next Generation network equipments.

We have proposed a methodology for reliably improving service scalability.

We have proposed and evaluated a proactive operational service aware active replication-based framework that reduces the number of interrupted sessions across failures.

We have proposed and evaluated implicit session aware admission control mechanisms that contribute to providing improved E2E QoS to subscribers.
- Extend the framework to support
  - More redundancy models
  - Highly non deterministic applications

- Push standardization work on transparent frameworks for service reliability across failures

- Investigate complex & realistic analytical model to derive the optimal session aware admission control strategy
  - Session duration, session volume, etc.
- Extend the advocated session aware admission control model to handle more QoS metrics
  - Client category, etc.

- Enhance the proposed session aware admission control model with the means to address the QoS of highly variable Internet traffic
  - Use forecasting techniques to improve stability
Thanks 😊!

Any Questions?
Some Achievements …

• **Patents**

• **Journal Papers**

• **Some Research Papers**
Contributions to Session Aware Frameworks for Next Generation Internet Services

PhD Dissertation Defense by

Narjess AYARI

Phd. Prepared at

Orange Labs (SIRP/ASF/INTL) and LIP/RESO (INRIA/UCBL/ENS de Lyon) – Université de Lyon, France.

Advisors:

Prof. Pascale Vicat-Blanc Primet
Dr. Laurent Lefèvre
Eng. Denis Barbaron

Phd. Advisor (RESO – LIP (UMR 5668 CNRS, ENS, INRIA, UCB))
Phd. Advisor (RESO – LIP (UMR 5668 CNRS, ENS, INRIA, UCB))
Phd. Advisor (Orange Labs)

Tuesday, October 21th, 2008.