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	Phd. Advisor (RESO – INRIA - LIP (UMR 5668 CNRS, ENS, UCB)
Dr.	Laurent Lefèvre	Phd. Advisor (RESO – INRIA - LIP (UMR 5668 CNRS, ENS, UCB)
Eng.	Denis Barbaron	Phd. Advisor (Orange Labs)

Tuesday, October 21th, 2008.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

- 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

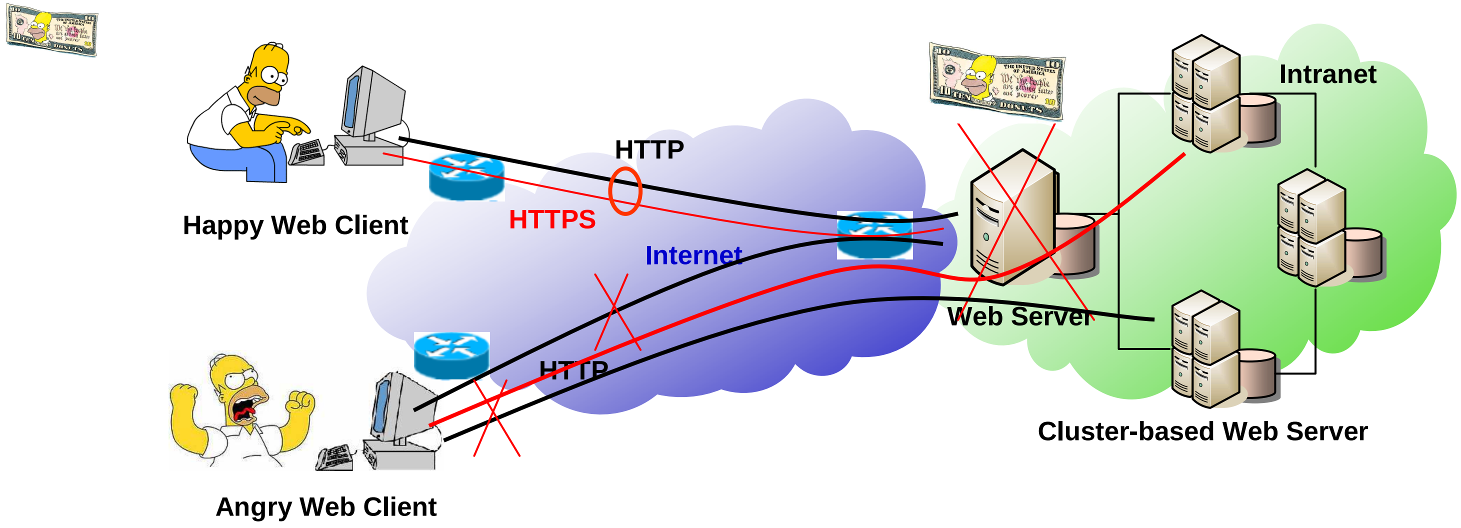
- Network operators are concerned with E2E QoS issues while addressing the explosive growth of Internet users
  - ✓ Core network contribution in the E2E QoS
  - ✓ Edge network contribution in the E2E QoS

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

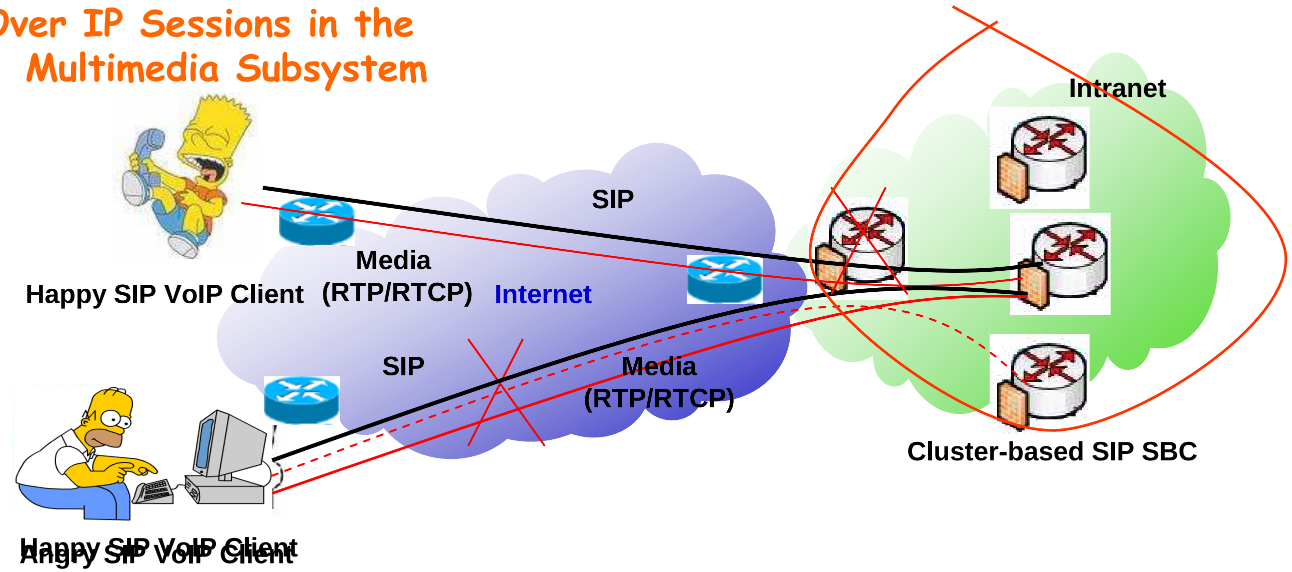
Commercial Web Session



This research focuses on the requirements of scalable and highly available Internet servers subjected to multiple flow-based user sessions.

Network Operator related example

Voice Over IP Sessions in the IP Multimedia Subsystem



SBC: Session Border Controller

Media (RTP/RTCP)

SIP signalling

Operators use resources' redundancy to improve Internet servers' robustness.

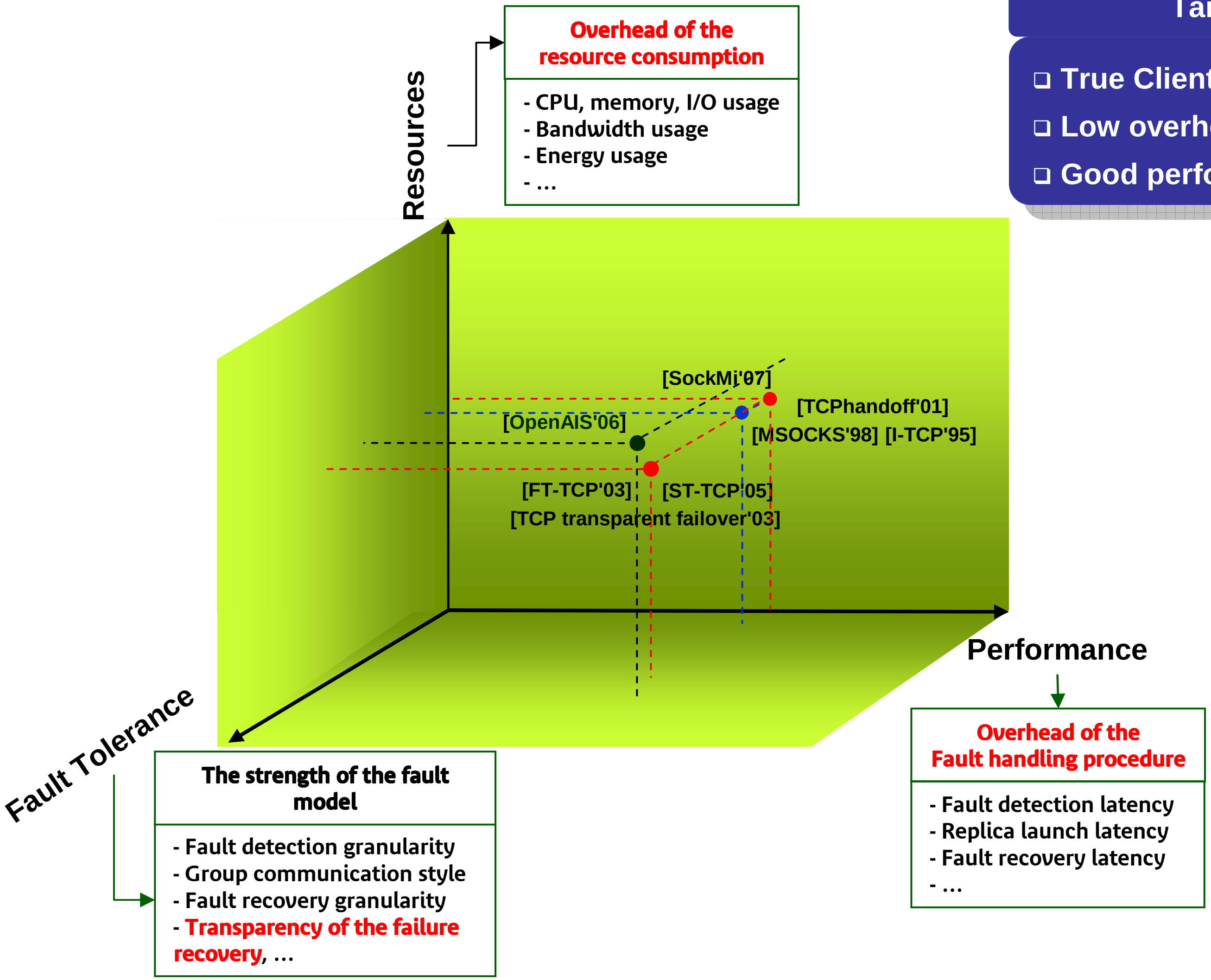
What is High Availability?

✓ 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



Targeted properties

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

We want to:

- Reduce the number of interrupted sessions
  - Under server overload
  - Across server failures
- Reduce the number of "angry" clients

Concept of Session Awareness

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.

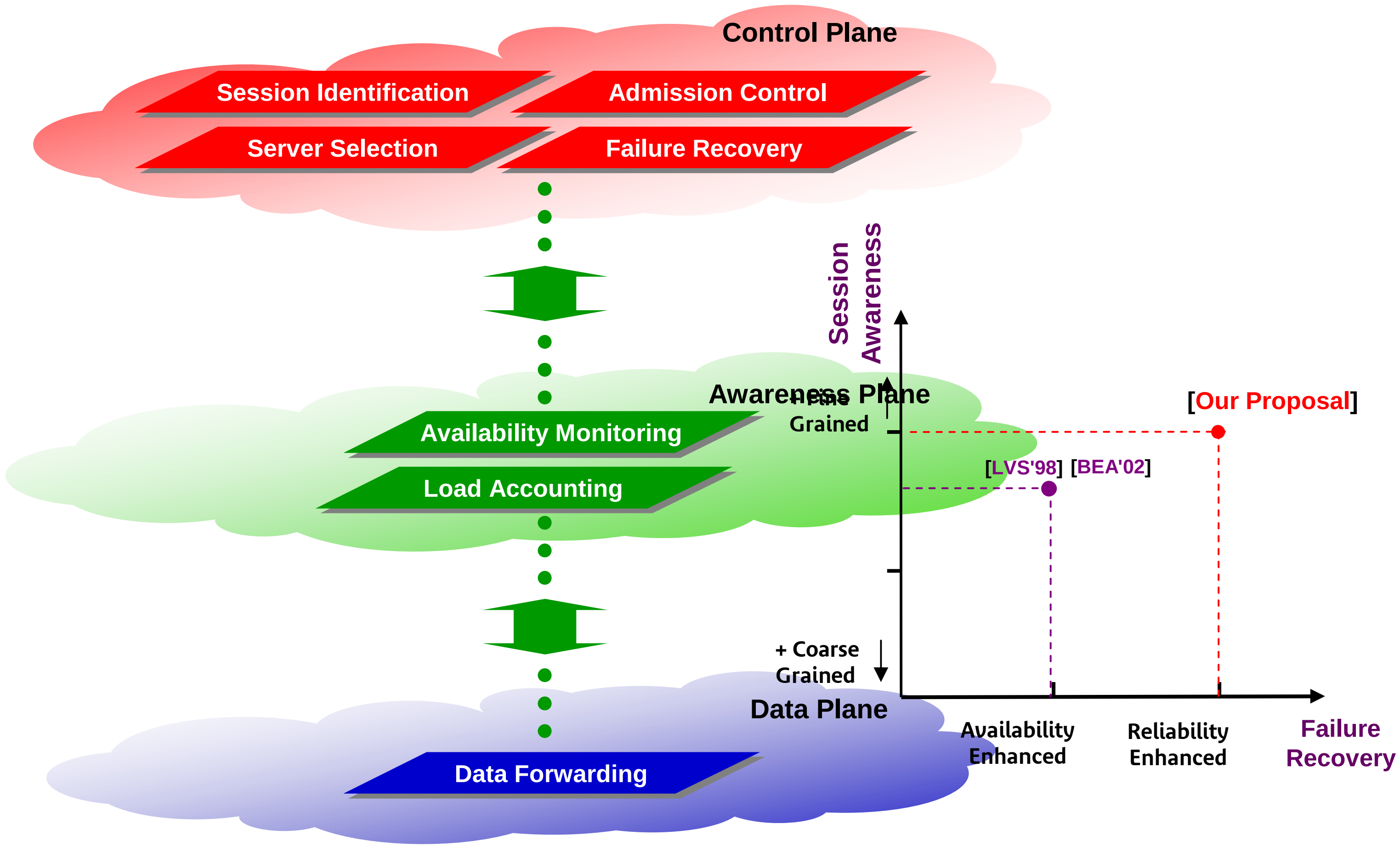
Proposal and Design of Session Aware Admission Control mechanisms that improve the Scalability pattern of Internet Servers under overload condition

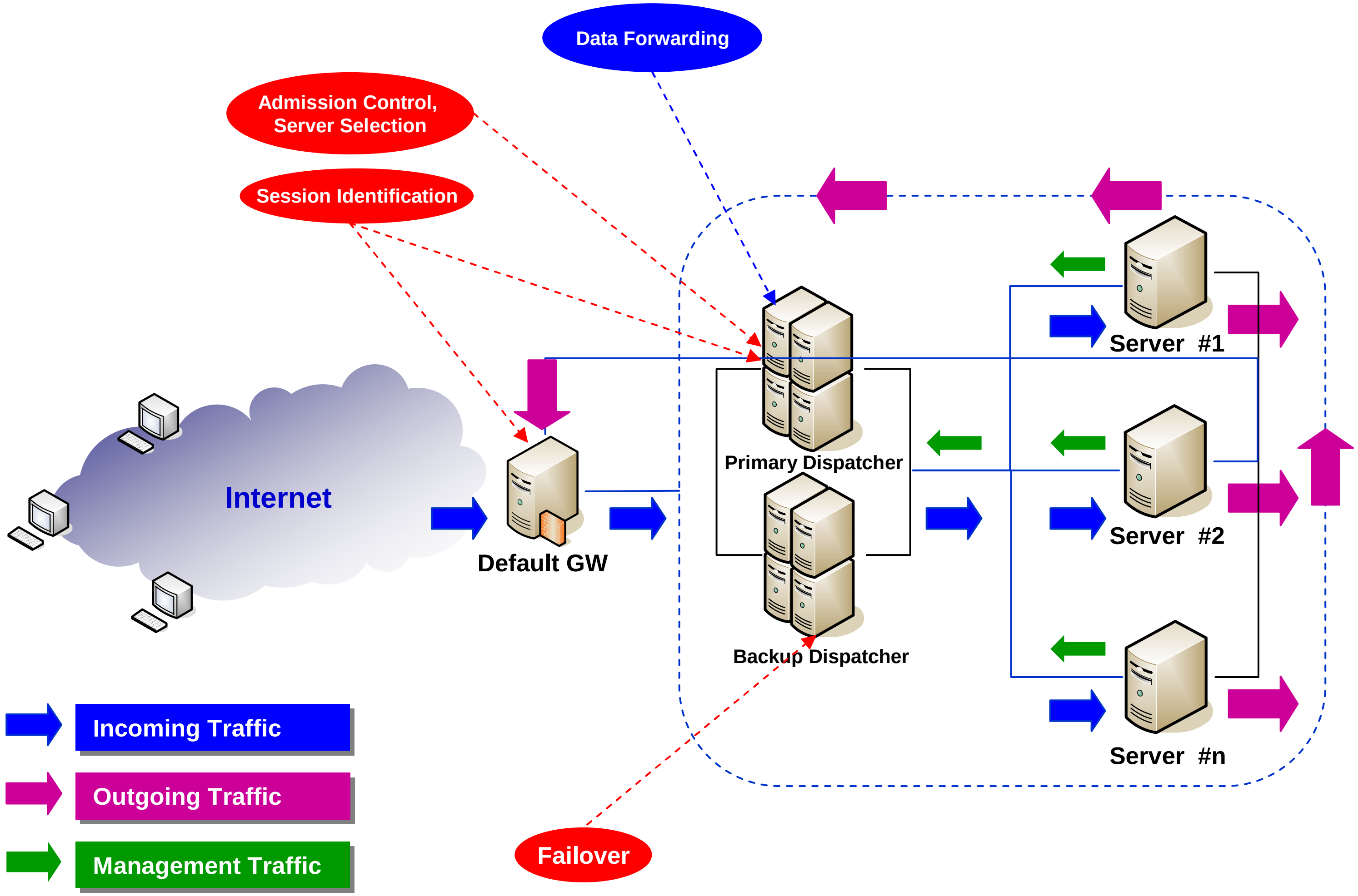
Outline

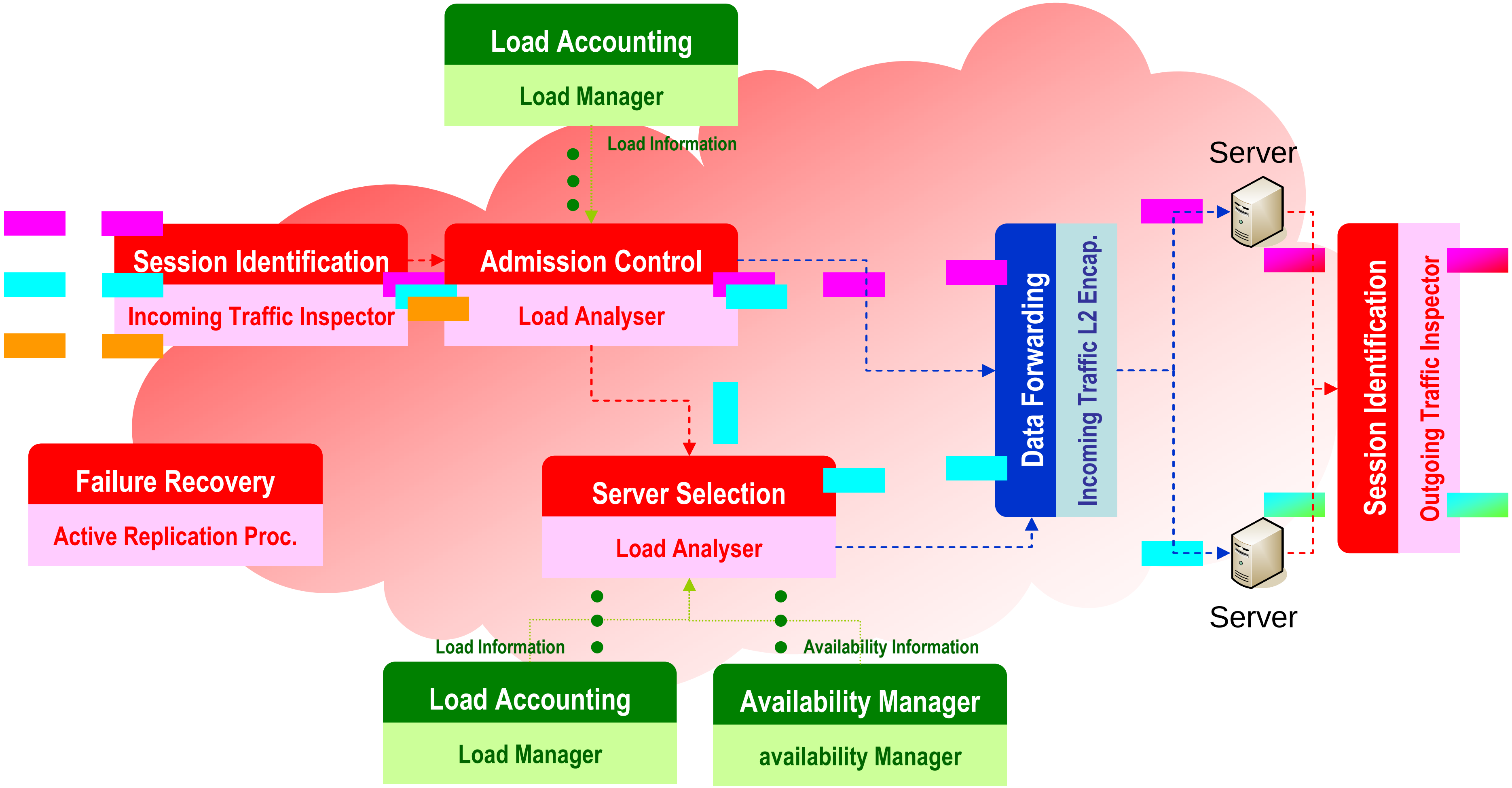
1. Part I: Contributions to Session Aware Architectures for High Performance Cluster-based IP Services
2. Part II: Proposal of An Active Replication-based Framework for Highly Available Internet Servers and Stateful Equipments
3. Part III: Proposal of Session Aware Admission Control Strategies for Scalable Internet Servers
4. Part IV: Conclusion and Future Works

Part I

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


Control Plane
Session Identification
Admission Control
Server Selection
Failure Recovery
Awareness Plane
Availability Monitoring
Load Accounting
Data Plane
Data Forwarding
Session Awareness
+ Fine Grained
+ Coarse Grained
[Our Proposal]
[LVS'98] [BEA'02]
Availability Enhanced
Reliability Enhanced
Failure Recovery


Data Forwarding
Admission Control, Server Selection
Session Identification
Internet
Default GW
Primary Dispatcher
Backup Dispatcher
Server #1
Server #2
Server #n
Incoming Traffic
Outgoing Traffic
Management Traffic
Failover


Load Accounting
Load Manager
Load Information
Session Identification
Incoming Traffic Inspector
Admission Control
Load Analyser
Data Forwarding
Incoming Traffic L2 Encap.
Server
Server
Session Identification
Outgoing Traffic Inspector
Failure Recovery
Active Replication Proc.
Server Selection
Load Analyser
Load Information
Availability Information
Load Accounting
Load Manager
Availability Manager
availability Manager

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].

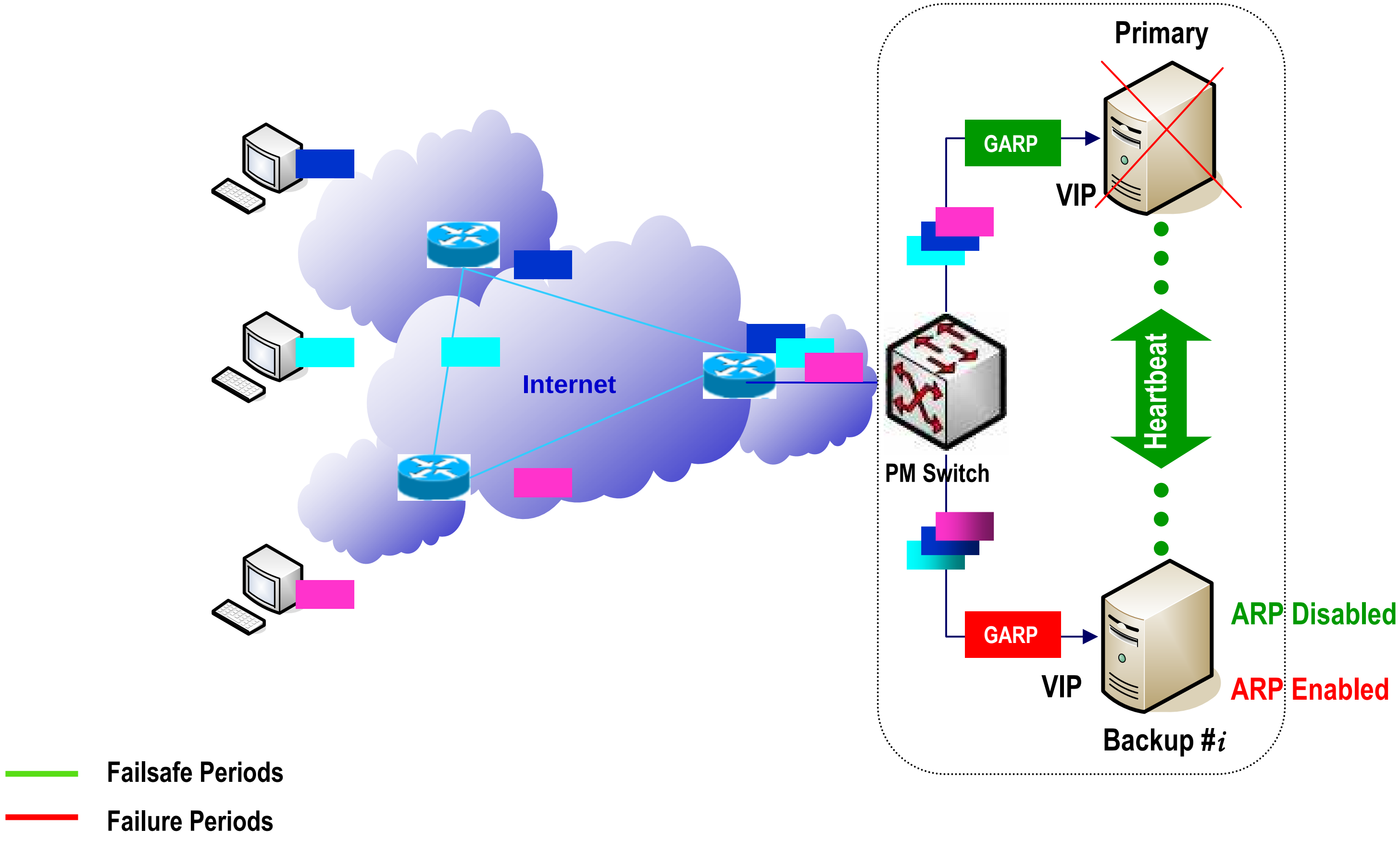
Part II

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

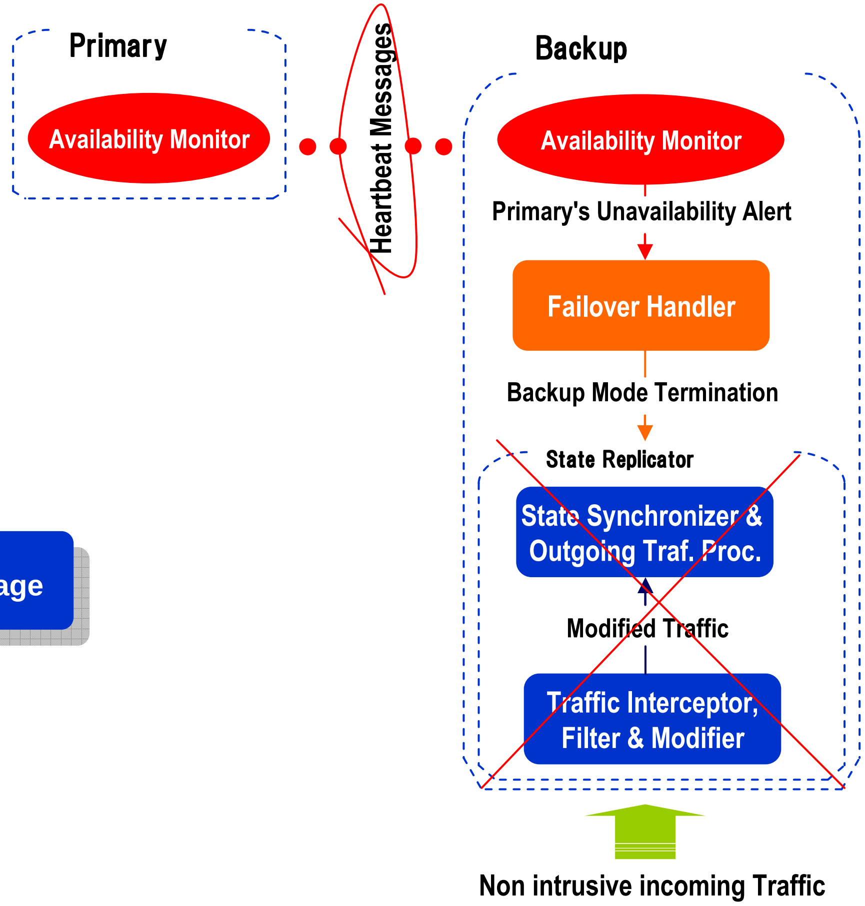


VIP: Virtual IP address of the service

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

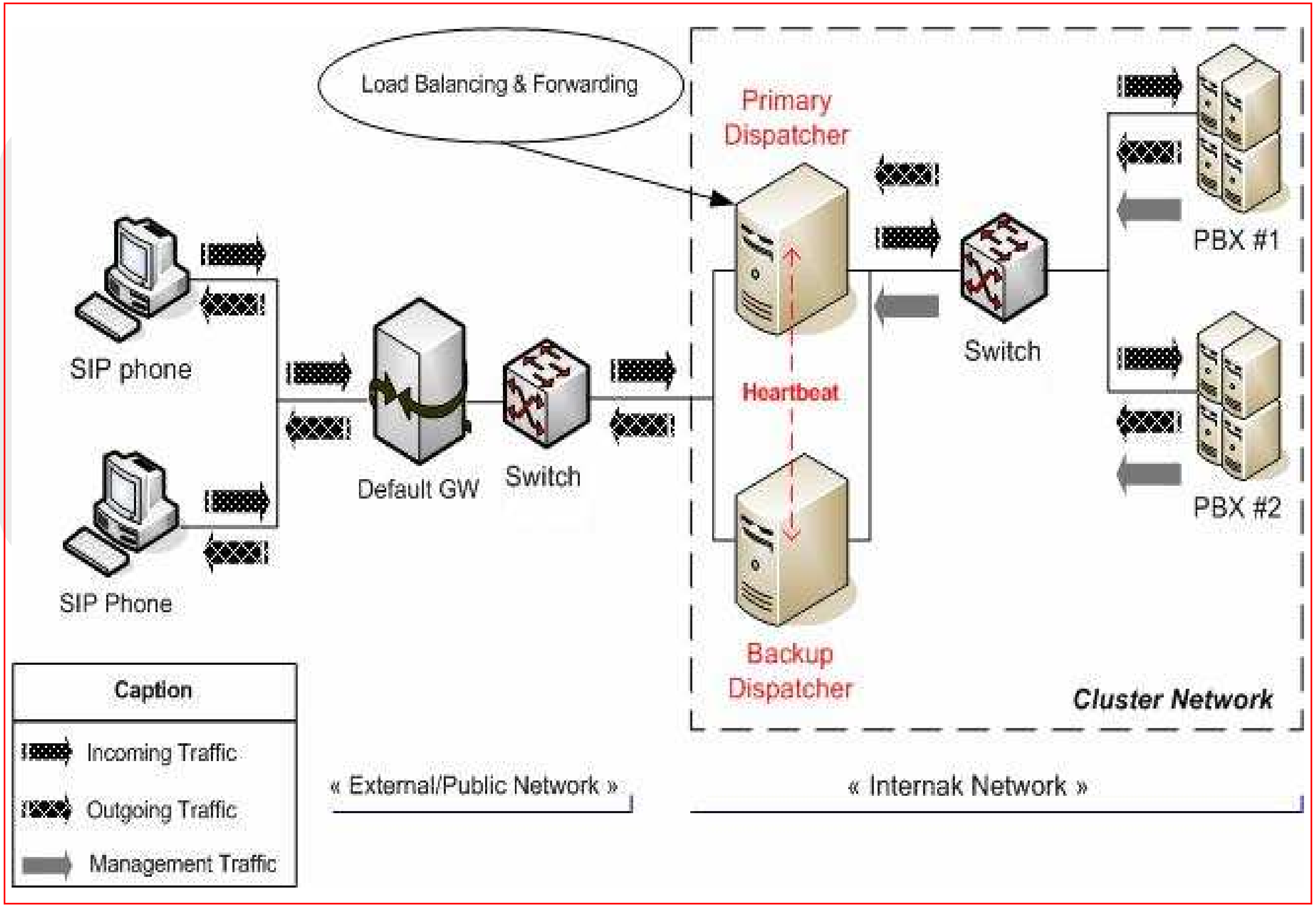
Timestamp	
Source IP @	
Destination IP @	
Source Port	Destination Port
Sequence Number	

Service Consistency

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.

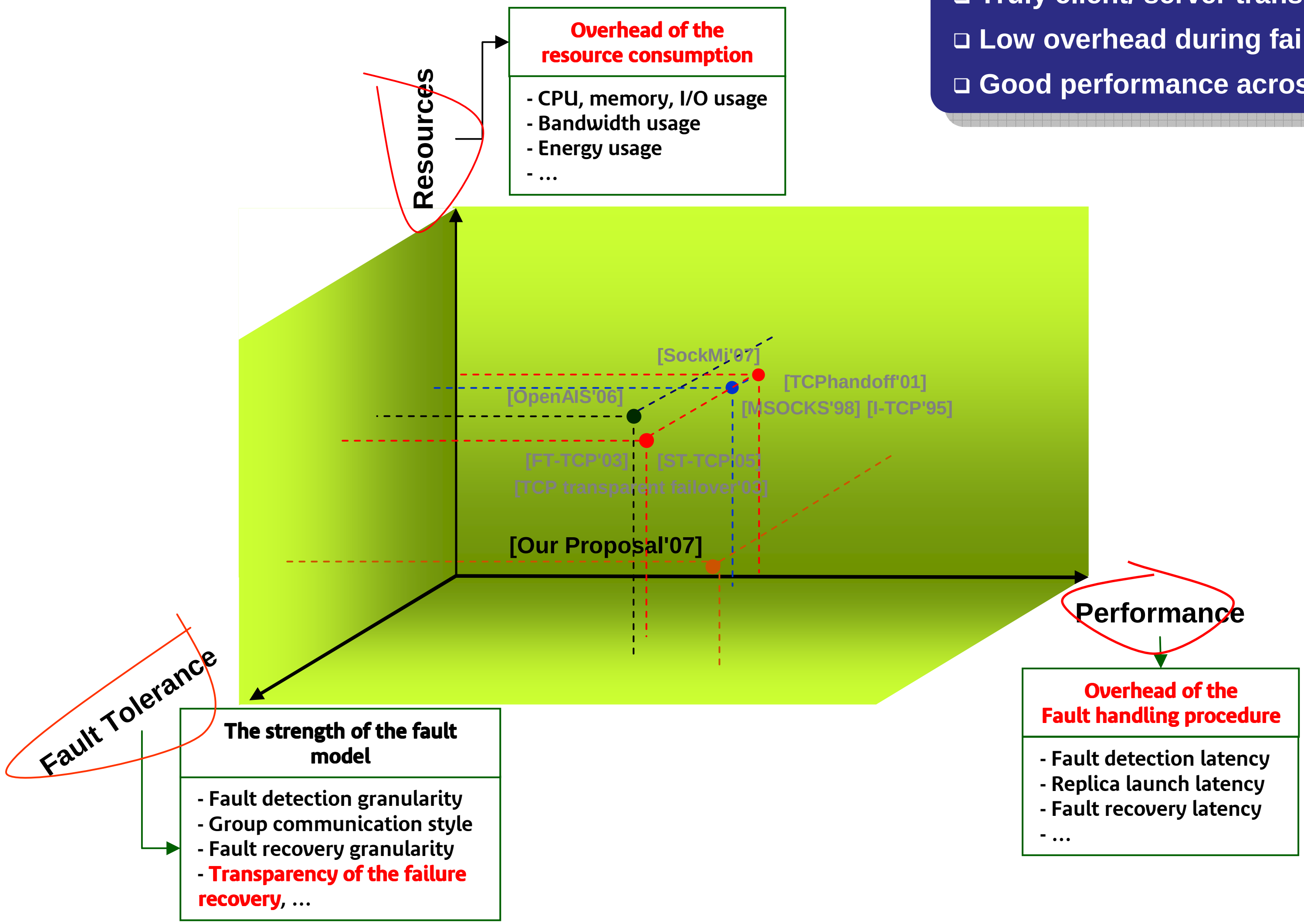
Experiment Setup



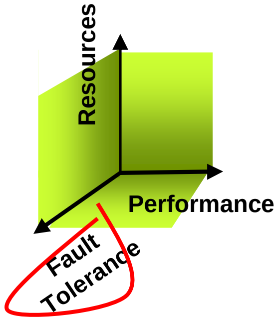
Experiments' methodology

We achieved the following goals:

- Truly client/ server transparent
- Low overhead during failsafe periods
- Good performance across failures

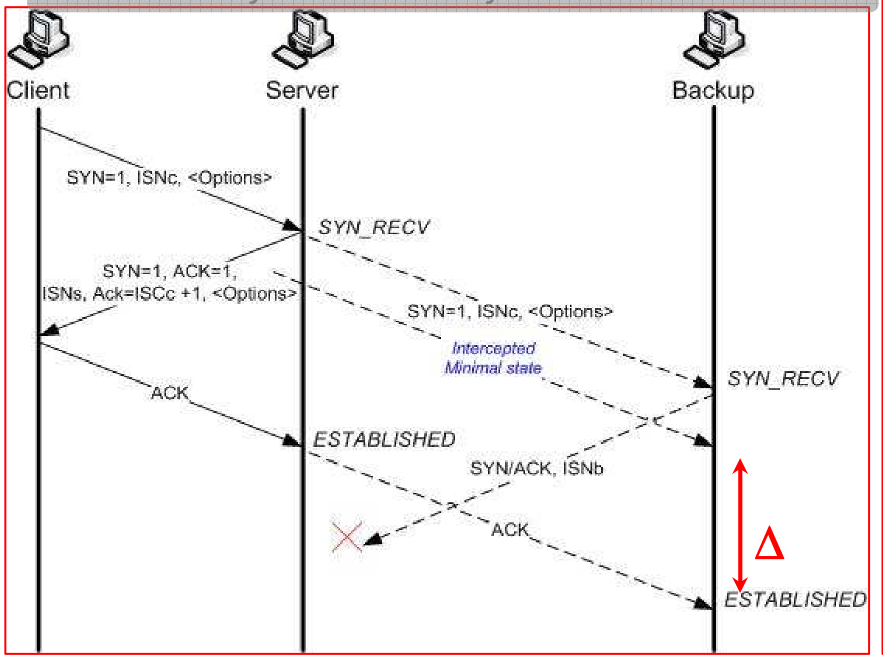


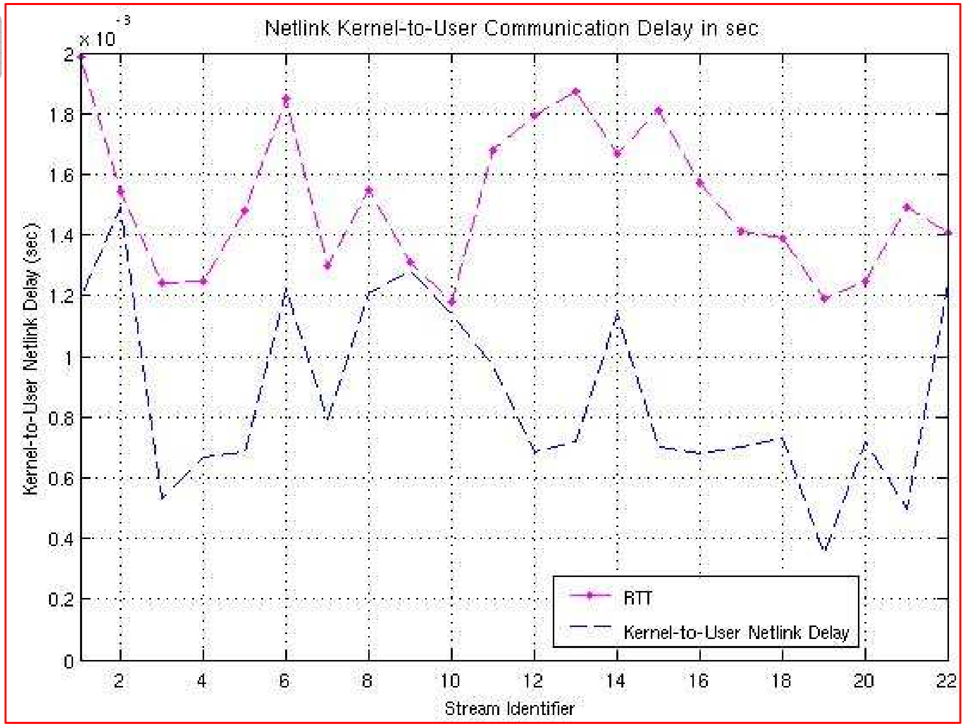
Failsafe Periods' Performance: TCP State Replication Consistency



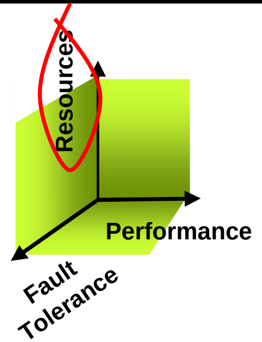
- Δ = Time required by the kernel space synchronization module to receive the minimal connection state added to the time required to update the fake connection accordingly.

Early TCP State Synchronization



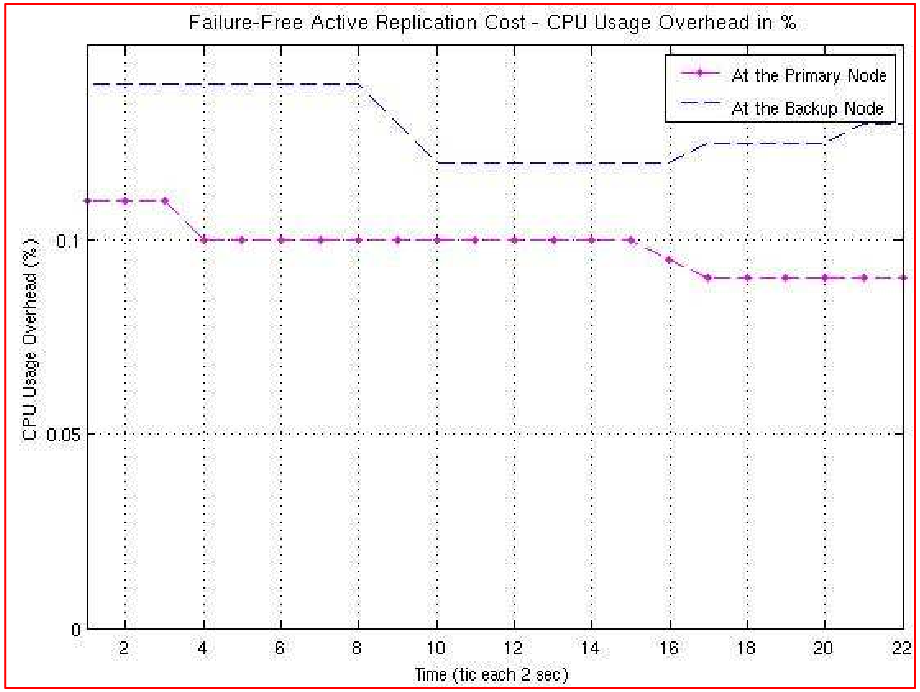


Failsafe Periods' Performance: Resources' Usage



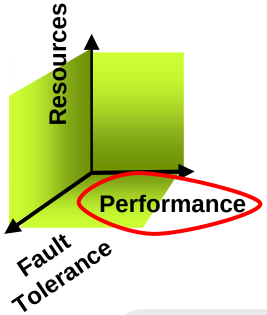
- Only the "useful" traffic is processed by the backup node (signalling traffic)
- Micro-benchmarks are derived to quantify cost of active replication (CPU, memory & Network buffers)

- 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

Performance across Failures : Failure Recovery Time

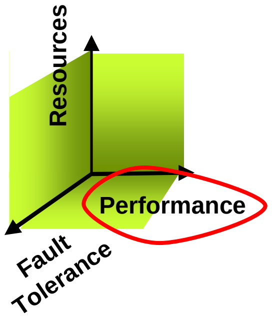


- Failure recovery latency is the time needed to detect and to recover from a primary's failure.

$latency_{failure\ recovery} = latency_{failure\ detection} + latency_{Takeover}$

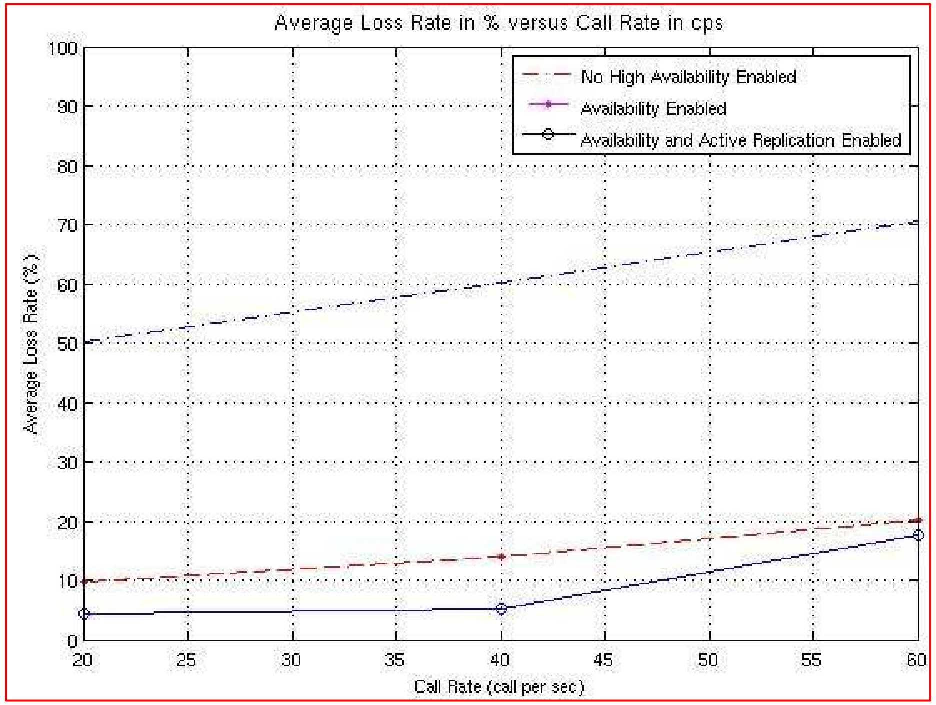
- Average failure recovery time as a function of the failure detection interval
  - < 1sec. for a FDI = 1 sec.

Performance across Failures: Scalability Tests



Loss ratio = (Average number of failed calls) / (Total number calls)

- 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:** $C_i > C_b$

$$\max_p (R_c N_c - C_b N_b - C_i N_i)$$

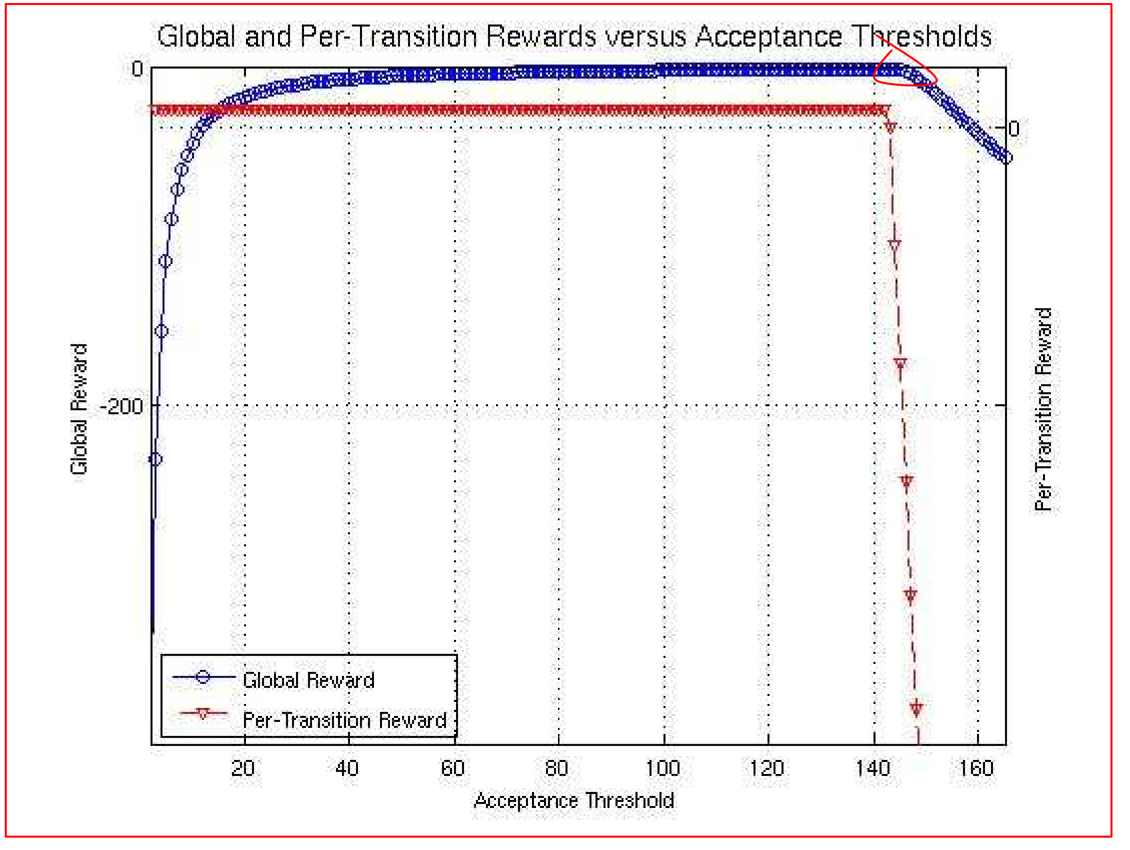
p is the acceptance probability

N_c, N_b and N_i are respectively the mean number of completed, blocked and interrupted sessions per unit of time.

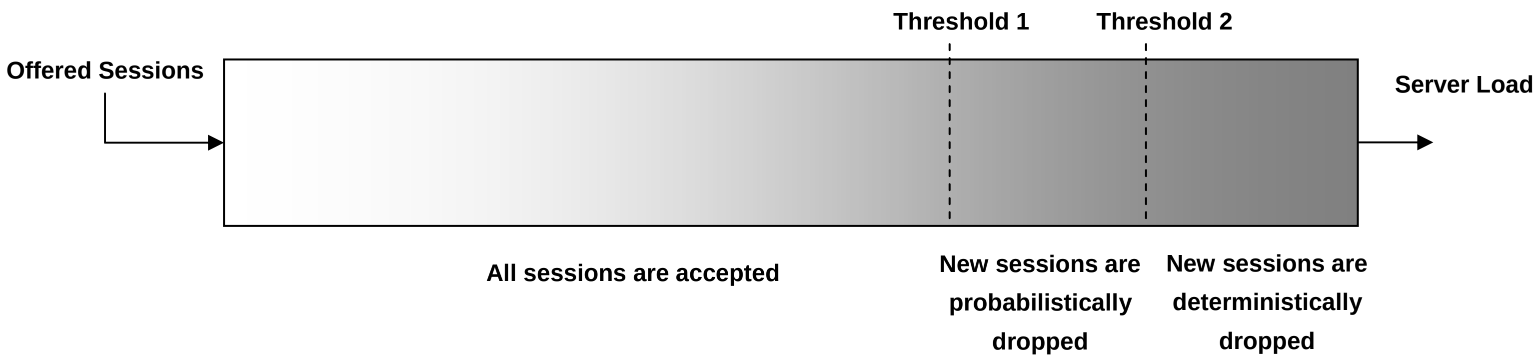
Realistic Traffic Profiles ...

SIP Traffic Profile

Parameter	Value
Maximum Number of Concurrent Sessions	230
Source throughput	14



<u>Our Strategy</u>: Give the traffic pertaining to the active sessions a higher priority under overload.



$$p = \begin{cases} 0, & if\; l_S \leq T_1 \\ f(l_S),\; f(x) = \left\| \frac{x - T_1}{T_2 - T_1} \right\|, & if\; (New\; Session)\; and\; T_1 < l_S \leq T_2 \\ 1, & if\; (New\; Session)\; and\; T_2 < l_S \leq C \\ 1, & if\; l_S > C \end{cases},$$

Simulation Scenario & Methodology

Admission Control Strategies

- **Request Aware (RAC)**
- **Session Aware (SAC)**
  - **Single Threshold-based**
    - ON/OFF Model (ON/OFF SAC)

$$p = \begin{cases} 0, & if\; load \le T_1 \\ 1, & if\; (new\; session)\; and\; load > T_1 \end{cases}$$

    - Responsive Model (RESP SAC)

$$p = \begin{cases} 0, & if\; load \le T_1 \\ \left\| \frac{load - T_1}{capacity - T_1} \right\|, & if\; (new\; session)\; and\; load > T_1 \end{cases}$$

  - **Double Threshold-based**
    - Responsive Model (Double SAC)

$$p = \begin{cases} 0, & if\; l_S \le T_1 \\ f(l_S),\; f(x) = \left\| \frac{x - T_1}{T_2 - T_1} \right\|, & if\; (New\; Session)\; and\; T_1 < l_S \le T_2 \\ 1, & if\; (New\; Session)\; and\; T_2 < l_S \le C \\ 1, & if\; l_S > C \end{cases}$$

Voice CBR Traffic Profile Model

Simulation Scenario & Methodology

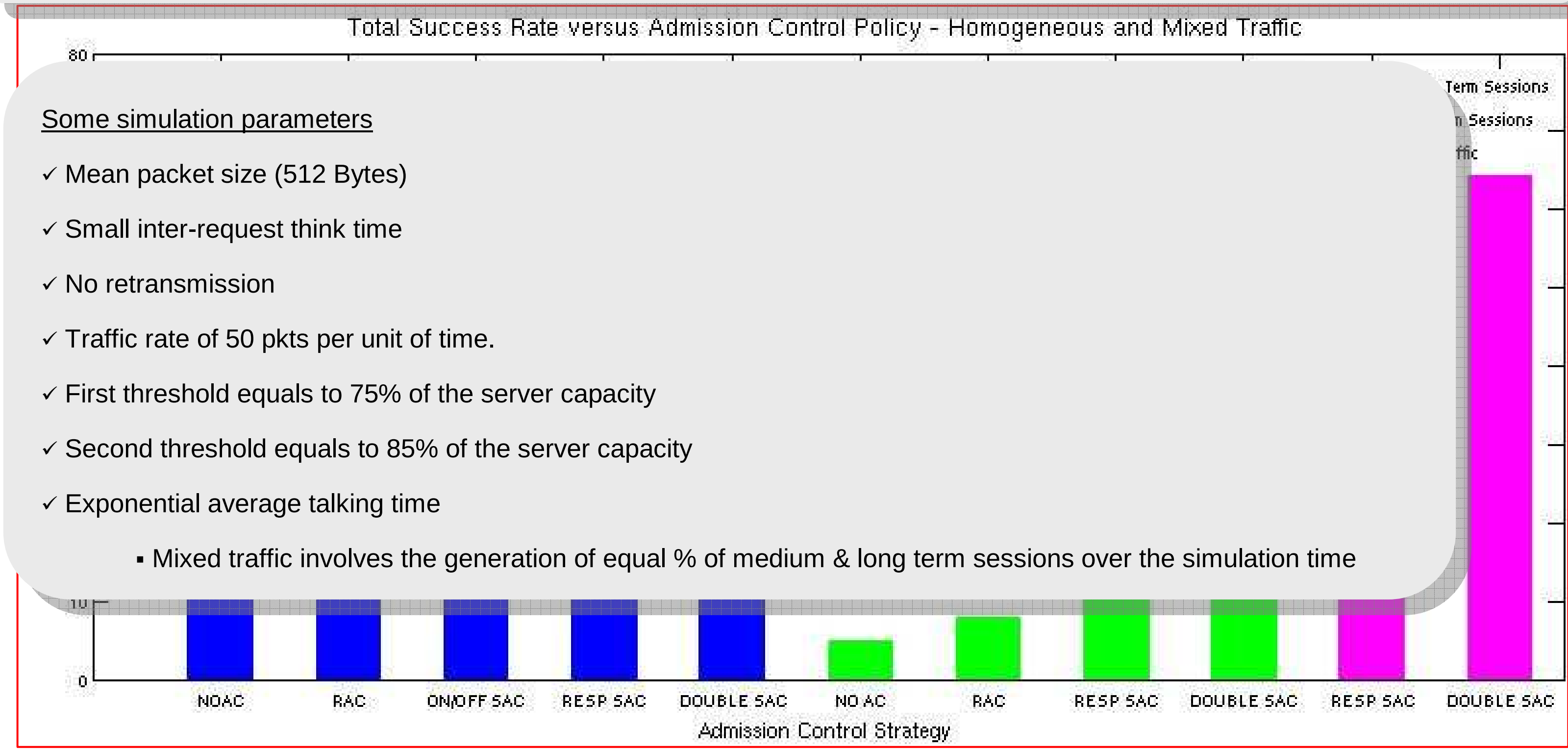
- We consider two simulation scenarios
  - Generating homogeneous sessions over the same simulation run, having a duration that ranges from medium to long term.
  - Generating mixed sessions over the same simulation run, meaning that both medium term and long term conversations are equitably generated over the simulation time

Effect of Session Awareness on the Average Success Rate

Suc

sions)

Let us simulate the behaviour of a processing server over a duration of 100 sec



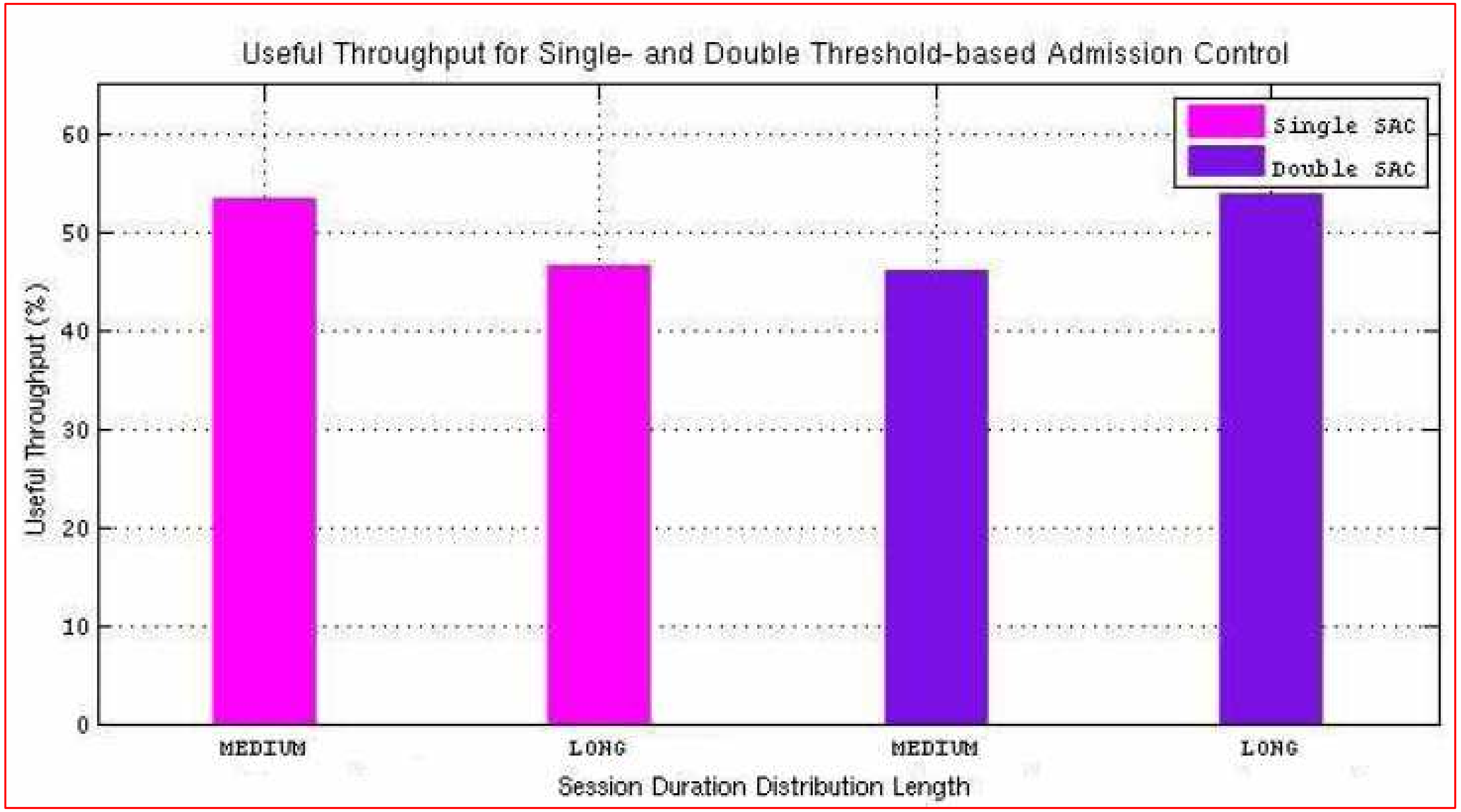
<u>Some simulation parameters</u>

- ✓ Mean packet size (512 Bytes)
- ✓ Small inter-request think time
- ✓ No retransmission
- ✓ Traffic rate of 50 pkts 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

Percentage Histogram of Completed Calls for Different Admission Control Strategies - Mixed Traffic

What About Discrimination Against Long Term Sessions?

Success rate = (Average number of completed sessions) / (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

Summary of the Contributions

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

Future Works

- ❑ Extending 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.

Future Works (Cont.)

- ❑ 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**
  - **"Gestion d'au moins une connexion établie avec au moins un serveur",** France Telecom R&D Patent**,** June **2006.**
  - **"Procédés de gestion de sessions multi-flux entre un terminal et un serveur",** France Telecom R&D Patent**,** June **2007.**
- **Journal Papers**
  - N. Ayari, D. Barbaron, L. Lefèvre and P. Vicat-Blanc Primet, **"Fault Tolerance for Highly Available Internet Services: Concepts, Approaches, and Issues",** IEEE Communications Surveys and Tutorials 2008.
- **Some Research Papers**
  - N. Ayari, P. Neira, L. Lefèvre and D. Barbaron, **"Towards a Dependable Architecture for Highly Available Internet Services",** third IEEE International Conference on Availability, Reliability and Security (IEEE ARES'08), Barcelona, Spain, March 2008.
  - N. Ayari, D. Barbaron, L. Lefèvre and P. Vicat-Blanc Primet, **"A Session Aware Admission Control Scheme for next Generation IP Services",** the Fifth annual IEEE Consumer Communications and Networking Conference (IEEE CCNC 2008), Las Vegas, USA, January 2008.
  - N. Ayari, D. Barbaron, L. Lefèvre and P. Vicat-Blanc Primet**, "T2CP-AR: A system for Transparent TCP Active Replication",** the IEEE 21st International Conference on Advanced Information Networking and Applications (IEEE AINA-07), Niagara Falls, Canada, May 2007.

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	Phd. Advisor (RESO – LIP (UMR 5668 CNRS, ENS, INRIA, UCB)
Dr.	Laurent Lefèvre	Phd. Advisor (RESO – LIP (UMR 5668 CNRS, ENS, INRIA, UCB)
Eng.	Denis Barbaron	Phd. Advisor (Orange Labs)

Tuesday, October 21th, 2008.