Load Balancing - Structure
SYT - 5A HIT

Martin Haidn, Nikolaus Schrack

November 30, 2014

Contents

1   Instruction

The concept of Load Balancing is not new in the server and network space. There are different types of Load Balancing. A Router, for example, distributes the traffic across multiple paths to the same destination. A Server Load Balancer, on the other hand, distributes traffic among server resources rather than network resources.

1.1   The Need and Goals for Load Balancing

The main goal of Load Balancing is to distribute workload across resources. It is suppose to optimize the traffic, maximaize throughput, minimize response time and try not to overload any single recource.

Since the Internet and Intranet have gotten so important for businesses, Load Balancing has become a very essential component for networks and servers. If the network goes down or works poorly, it can critically damage a business. Especially for e-commerce a slow server can make the users experience really bad.

1.2   Use Cases and Examples

This section should pick up the significant points from the "Needs and Goals" and bring them in a relation with specific, real examples.

1.3   Applications

An overview about the common used applications for load distribution.

2   Basic Concepts

2.1   Networking Fundamentals

The OSI model contains seven layers and every single one provides it's own functionality and data.

If we take a closer look on the deeper layers like data link and network, which is representative for layer two and three, we can see that their header information contains IP and MAC addresses. These addresses can be used to decide where a package has to be send when it's revived by a switch.

This basic concept of routing packages builds the fundament for load balancing. It's about making a decision if, and where the data has to go. [1]

2.2   Higher Layered Distribution

Description how load distribution works on OSI-Layers six and seven.

2.3   Load-Distribution Methods

Summary of common load distribution Methods, their benefits and disadvantages.

3   Advanced Concepts

3.1   Session Persistence

Reasons and benefits of using Session Persistence to track and store session data.

3.2   URL Switching

The flexibility of layer seven load balancing and the included url switching.

3.3   Network-Address Translation

Fast Layer 4 load balancing and the appliance as default gateway.

4   Scheduling Algorithms

4.1   Weighted Balance

Ways to guarantee a weighted balance in busy systems.

4.2   Priority

The meaning of priorities concerning the process of load balancing and how to route traffic to a preferred link, as long it's available.

4.3   Overflow

How to prevent traffic flow from slowing down when the connection runs out of available bandwidth.

4.4   Persistance

Eliminate session termination issue for HTTPS, E-banking, and other secure websites.

4.5   Round-Robin

A closer explanation to the scheduling procedure "Round Robin"

5   Caches

5.1   Definition

Define what a cache is for when we talk about load balancing.

5.2   Types

The different types of caches and their usage as well as benefits and disadvantages.

5.3   Deployment

Examples and explanation how to deploy load distribution using caches.

6 Problems

6.1 Mega Proxy Session

Problems triggered through the use of Mega Proxys on the client site.

References

[1] Chandra Kopparapu. *Load Balancing Servers, Firewalls and Caches.* John Wiley and Sons, Inc., New York, 2001.

7 Sources

Titel: Load balancing servers, firewalls, and caches : [timely, practical, reliable]
Autor: Chandra Kopparapu
Jahr: 2002
ISBN: ISBN 0-471-41550-2
TUWS: DAT:964, DAT:224

Titel: Dynamic load balancing : an overview
Autor: Arnold R. Krommer ; Christoph W. Ueberhuber
Jahr: 1992
ISBN: -
TUWS: DAT:351

Titel: Dynamischer Lastausgleich in Parallelrechnersystemen : genetische Algorithmen und eine spezielle Rechnerstruktur
Autor: Michael Witt
Jahr: 1997
ISBN: -
TUWS: -

Titel: Optimal load balancing in distributed computer systems
Autor: Hisao Kameda
Jahr: 1997
ISBN: ISBN 3-540-76130-6
TUWS: -

Titel: Server Load Balancing
Autor: Tony Bourke
Jahr: August 2001
0-596-00050-2, Order Number: 0502
200 pages, 34.95 USD
Link: http://oreilly.com/catalog/serverload/chapter/ch07.html

Onlinequellen:

Name: Optimal Load Balancing in Distributed Computer Systems
Link: http://bookzz.org/book/2092081/f777c1

Name: Spectral Methods for Efficient Load Balancing Strategies
Link: http://cs.emis.de/LNI/Dissertation/Dissertation3/GI-Dissertations.03-3.pdf

Name: Lastverteilung auf dem Konzept des virtuellen Servers
Link: http://www.nm.ifi.lmu.de/pub/Fopras/fikr02/PDF-Version/fikr02.pdf

Name: Dynamic Load Balancing and Scheduling
Link: http://www2.cs.uni-paderborn.de/cs/ag-monien/RESEARCH/LOADBAL/

Please note that this is just an overview about our collected references so far, to find them in the TU-Library. We'll add our last visits during this elaboration.