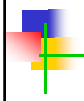


Cluster Computing

Load Balancing Over Network

- Introduction
- Methods
- Common Errors
- Practical Implementations
- Summary

2

Introduction

- Load balancing over a network
  - Use of devices external to the processing nodes in a cluster
  - Distribute workload or network traffic load across the cluster
  - Nodes may be interconnected among themselves
    - must be connected directly or indirectly to the balancing device
- Processing nodes
  - Provide various status information
    - current processor load
    - the application system load
    - number of active users
    - the availability of network protocol buffers
    - other specific resources

3

Introduction

- Balancing device
  - monitors the status of all the processing nodes
  - dictates where to direct the next processing job
  - can be a single unit or a group of units working in parallel or under a tree hierarchy
  - use one or more algorithms or methods together with static or dynamic setting to decide which node gets the next incoming connection request

4

Introduction

- Two ways of network load balancing
  - Network point of view
    - Load balancing system monitors incoming data to a cluster and distributes traffic based upon network protocol and traffic information
  - An application point of view
    - Higher level in the network communications model
    - It is possible to build an application-specific balancing system on top of an existing network-specific balancing system or combine the two into a more complex system

5

Methods

- Implementation of load balancing
  - Through the employment of several basic methods
    - Can be combined to create more advanced system
- Methods
  - Can be looked upon as mathematical functions that work on statistics of network traffic and node status to determine an appropriate target for receiving new load
  - Each of these functions are influenced by several factors
    - define behavior and role of the device

6

Methods

- How new traffic is to be distributed across the nodes of the cluster
  - Factors Affecting Balancing Methods
  - Simple Balancing Methods
  - Advanced Balancing Methods

7

Factors Affecting Balancing Methods

- Define the capabilities and limits of the balancing device
- Influences of the environment that the device works in and have to support
  - The most basic factor: TCP/IP
    - lack of a separate session layer
    - lack of appropriate QoS guarantee system
    - IP, ICMP, TCP, UDP

8

Factors Affecting Balancing Methods

- Network Address Translation (NAT)
  - Converting internal or private network address and routing information into external or public addresses and routes
  - Due to the limited address space of the current version of the IPV4
  - For security reason, NAT as firewall
  - Any balancing device required to perform network address translation must keep separate tables for internal and external representations of computer or host information
  - Cannot be used with VPN (Virtual Private Networks)

9

Factors Affecting Balancing Methods

- Domain Names
  - Form the basis of many balancing methods
  - Mapping Fully Qualified Domain Name (FQDN) to IP address
    - combination of both the host name and the domain name to create a uniquely identifiable name for a system on the Internet
  - Domain Name System (DNS)
    - The standard translation mechanism
    - Mapping names to address and vice versa
    - Map multiple hosts to a single host name
  - As most computer are referenced by their FQDN and not their direct IP address
    - DNS server becomes a crucial aid to the balancing device system to help determine load distribution

10

Factors Affecting Balancing Methods

- Wire-speed Processing
  - Ability to perform network traffic processing and redirection at the full speed of the incoming packets to prevent any traffic bottlenecks at the network device
  - Operating system may be limited in this capacity
  - This can result in slower response or an inability to accept new connections at individual nodes in a cluster

11

Factors Affecting Balancing Methods

- Node Operating System Limitation
  - Some operating systems have limitations
    - the speed at which they can process packets
    - the number of connections they can support
    - the type of traffic they can accept
  - Large number of interrupts as new packets arrive
  - This affects the cluster in much the same way as for wire-speed processing

12

Factors Affecting Balancing Methods

- Balancing Device Limitation
  - All balancing devices have practical limitations incurred by memory and processing speed
  - Balancing methods which work well in small clusters may not be scaleable to large numbers of nodes
    - Keep tables of information on incoming connections and node status
    - Table limit the size of the cluster and the traffic processing rate

13

Factors Affecting Balancing Methods

- Session- and nonsession-based Traffic
  - Session-based traffic
    - Look for IP packet with TCP_SYN and TCP_FIN messages as the start and end of a session
    - Direct all traffic between the source and destination to a specific node in the cluster
  - Nonsession-based traffic
    - Cannot be completely accounted for
    - Created a patchwork system for UDP
      - Keeping track of incoming datagram from a source
      - Establishing a time limit for a 'session'
      - Time interval-based 'UDP session' management

14

Factors Affecting Balancing Methods

- Application Dependencies
  - Some applications require that once a source computer has accessed a particular node, they continue to connect to that same node every time in the future
    - continuous service in shared nothing cluster
  - Can be fixed by changing the application code to build a more cluster-aware application
    - this is not always possible

15

Simple Balancing Methods

- A single function that select the node within a cluster to send a new request to
- Some of these methods can be used by themselves
- Used in conjunction with another simple or advanced method

16

Simple Balancing Methods

- Weighting
  - Provides a simple way of conferring load onto the nodes according to the priority value or weight of the node
  - Different weights to the nodes of different capacities
- Randomization
  - Assigns each node with a value generated by a pseudorandom algorithm
  - Works good in identical node environment

17

Simple Balancing Methods

- Round-Robin
  - Assigns the next incoming request to the next node in the list and rotates through the list continuously for further requests
  - Commonly used by itself in DNS
  - DNS servers don't keep track of server load
    - IP caching problem
  - Effective where all the nodes in the cluster are identical in capacity and performance
  - Limitations
    - no knowledge of nodes, address caching

18

Simple Balancing Methods

- Hashing
  - Works similar to the simple weighting system
  - Benefit
    - Packets from the same source address will always get assigned to the same server
- Least Connections
  - Keeps track of all currently active connections assigned to each node in the cluster
  - Assigns the next new incoming connection request to the node which currently has the least connections
  - Differ from actual amount of processing
  - Problem
    - Consume more system resource than others
  - Solution
    - Sets a maximum limit on the number of connections assigned to each node

19

Simple Balancing Methods

- Minimum Misses
  - Keeps long-term track of all incoming requests assignments to the nodes
  - Assign the next incoming request to the nodes which has processed the least number of incoming request in its history
  - Difference with Least Connections
    - this keeps track of the number of current and past connections

20

Simple Balancing Methods

- Fastest Response
  - Keeps track of the network response time between the node and itself
  - Assigns the next incoming connection request to the node with the fastest response
  - Requires active monitoring of the individual nodes
    - Sending ICMP packets with the 'ping' command
    - Proprietary mechanism based upon UDP packets
  - Make little sense except heavy load down
  - Useful in different network segments

21

Advanced Balancing Method

- Primary optimization vectors
  - Network traffic optimization
  - Fair load distribution
  - Network route optimization
  - Response latency minimization
  - Application-specific performance
  - Administrative or network management optimization

22

Primary Optimization Vectors of Advanced Balancing Methods

Advanced Balancing Method	Optimization vectors
Network Traffic-based Balancing	network traffic optimization, network route optimization, response latency minimization
Node Traffic-based Balancing	fair load distribution, response latency minimization
Node Load-based Balancing	fair load distribution, response latency minimization
Load-balancing DNS Resolution	network traffic optimization, fair load distribution
Topology-based Redirection	response latency minimization,network route optimization, network traffic optimization, application-specific performance
Policy-based Redirection	administrative management optimization, network management optimization
Application-specific Redirection	network management optimization, fair load distribution, application-specific performance

23

Advanced Balancing Methods

- Use a combination of the simple systems described earlier
- Network Traffic-based Balancing
  - Requires active monitoring of incoming traffic from different sources and distributing them accordingly to the nodes
  - Focus on predicting the volume of incoming traffic from a source on the network based upon past history
  - Based on a simple weighting function

24

Advanced Balancing Methods

- Node Traffic-based Balancing
  - Converse of the network traffic balancing system
  - Used Least Connections
  - Contact software agent on the node
  - Monitoring of the status of the network buffers
- Node Load-based balancing
  - Software agent
  - CPU or system load in UNIX systems
  - Various system load in Windows NT systems

25

Advanced Balancing Methods

- Load-balancing Domain Name Resolution
  - Involves the Round-Robin method
  - Load-balancing occurs within the DNS server itself
  - Independent of the application that generates the traffic
  - Can create an effective load-balancing system
    - by adding a few algorithms to a standard DNS server application
  - Simple & most popular
  - Best used in a cluster of nodes with identical applications

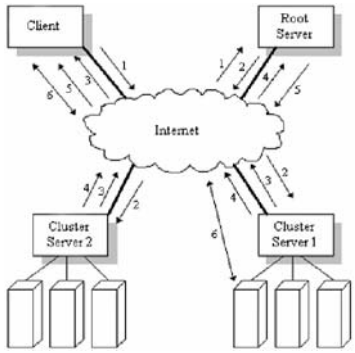
26

Advanced Balancing Methods

- Topology-based Redirection
  - Redirect traffic to the cluster nearest to the user's computer in terms of network topology
    - Hop count (static) and network latency (dynamic)
  - Hop count is the number of routers the packets have to traverse to reach the destination
  - Network latency is the amount of time taken for a network packet to travel between the client and the cluster balancing device
    - fastest response
    - a top level node in a particular domain
  - Effective in several clusters deployed across a network

27

Ping Triangulation in Topology-based Redirection



28

Advanced Balancing Methods

- Policy-based Redirection
  - Application of a mathematical or functional set of rules that define the balancing behavior of the cluster
  - Bandwidth Allocation Policies
    - higher priority on network administration & security control
  - Administrative Policies
    - specific to the needs of each message in network environment
  - Security Policies
    - proper access right for access the resource

29

Advanced Balancing Methods

- Application-specific Redirection
  - Provides load-balancing features dependent upon the type application or resource the client trying to access
  - Support for application level of sessions
  - Database & Web load-balancing

30

Common Errors

- There are four common errors
  - Overflow
  - Underflow
  - Routing errors
  - Induced network errors
- That can be destabilize efficient network clustering

31

Common Errors

- Overflow
  - Occur when too much network traffic to process
  - Occur at the balancing device or at individual nodes
  - Result
    - lost packets or throttling of packets intended for a destination node
    - loss of data and processing
  - The balancing device
    - Usually much greater than that of individual cluster node
    - But it possible to be overflow
    - Result in throttling or deleting some data streams to the nodes (leaving an adequate level of traffic to the node)
  - In TCP connections
    - There is an idle timeout clock for receiving an acknowledge
    - In an overflow situation, the acknowledge can't be send back
    - Retries from the client to deliver the same packet again until the timeout limits or connection dropped

32

Common Errors

- Underflow
  - A problem within the cluster itself
    - where one node is not getting enough traffic as compared to the other member nodes
  - Result
    - The node is underutilized or starved while others are getting loaded down
    - Indicating an inefficient distribution of traffic
  - This is typically a problem
    - with the algorithm itself or
    - with the improper use of the system
  - Problem of Non symmetric nodes
    - where nodes in the cluster are not identical in power and one or more member nodes have far more computing resources than other

33

Common Errors

- Routing Errors
  - It occur
    - between a balancing device and the cluster node
    - between the source client and the cluster nodes
  - Typically, it occurs from misconfiguration or a disconnected link

34

Common Errors

- Induced Network Errors
  - Errors generated by
    - normal use of the network
    - not an incorrect or unstable network state
  - Is not really errors
    - but results from delays in the propagation of packets along a network route
  - Too much traffic can result in
    - a bottleneck in the network route in network route
    - appear as errors
  - These errors are temporary, but can last for hours
  - In particular, the Fastest Response method and Topology-based redirection are the most affected by these errors

35

Practical Implementations

- A number of vendors have different approached, but arrived with similar solutions
- There is no commonly accepted standard
  - Most vendor implementations are proprietary and work with only other products from the same vendor

36

Simple Balancing Methods in Vendor Implementations

Vendor	Product	Weighted	Random	Round-robin	Hashed	Least Connects	Minimum Misses	Fastest Response
Alteon	ACEdirector	Y	N	Y	N	Y	N	N
Check Point	FireWall-1	N	Y	Y	N	N	N	Y
Check Point	FloodGate-1	Y	N	Y	N	Y	N	Y
Check Point	VPN-1	N	N	N	N	N	N	N
Cisco	Distributed Director	Y	Y	Y	N	Y	N	Y
Cisco	Local Director	Y	N	Y	N	Y	N	Y
F5 Labs	3DNS & BIG/ip	Y	N	Y	N	Y	N	Y
HolenTech	HyperFlow	Y	N	Y	Y	N	N	N
HydraWEB	HYDRA	Y	N	Y	N	N	N	Y
Resonate	Central Dispatch	N	N	Y	N	Y	N	Y
Resonate	Global Dispatch	N	N	Y	N	Y	Y	Y
RND Networks	Web Server Director	Y	N	Y	N	Y	N	Y
RND Networks	WSD for Distrib. Sites	Y	N	Y	N	Y	N	Y
Sun	StorEdge	Y	N	N	N	N	N	Y

Advanced Balancing Methods in Vendor Implementations

Vendor	Product	Network Traffic	Node Traffic	Node Load	Load-Balanced DNS	Topology Redirect	Policy Redirect	Application Redirect
Alteon	ACEdirector	Y	N	N	N	N	N	Y
CheckPoint	FireWall-1	N	N	N	N	N	Y	Y
CheckPoint	FloodGate-1	Y	N	N	N	N	Y	Y
CheckPoint	VPN-1	N	N	N	N	N	Y	Y
Cisco	Distributed Director	Y	N	N	N	Y	Y	N
Cisco	LocalDirector	Y	N	N	N	N	Y	N
F5 Labs	3DNS & BIG/ip	Y	N	N	Y	N	N	Y
HolenTech	HyperFlow	Y	N	N	N	N	N	N
HydraWEB	HYDRA	N	N	Y	Y	N	N	Y
Resonate	Central Dispatch	N	Y	Y	Y	N	Y	Y
Resonate	Global Dispatch	N	Y	Y	Y	Y	Y	Y
RND Networks	Web Server Director	Y	N	N	Y	N	N	Y
RND Networks	WSD for Distributed Sites	Y	Y	Y	Y	N	N	Y
Sun	StorEdge	N	N	Y	N	N	N	Y

General Network Traffic Implementations

- Independent of the software application using the network and transport layers
  - IP balancing
  - TCP session load-balancing only
  - UDP session

General Network Traffic Implementations

- HolenTech HyperFlow
  - Load balancing at the IP network level
    - independent of the TCP and UDP
    - not be functionally useful or efficient as balancing TCP sessions
  - Weighting & round-robin in initial load balancing
  - Two level hashing as the basic method for mapping
    - one-to-one, many-source-to-one
    - multiple balancing devices

General Network Traffic Implementations

- Cisco LocalDirector
  - LAN-based system originally based on NAT
  - CIP (Channel interface processor)
  - Least Connections (default), Weighted Percentage, Fastest Response, Round-Robin
- Cisco DistributedDirector
  - WAN-based system based on DNS
  - Topology-based redirection
  - UDP-based Director Response Protocol (DRP)

General Network Traffic Implementations

- Resonate Central Dispatch
  - Primary scheduler communicates with the agent to determine server and network traffic load
  - Round-Robin, Fastest Response, Least Connection, Node Load-based, Node Traffic-based
- Resonate Global Dispatch
  - Topology-based Redirection server that works with RCD
- Alteon Networks ACEdirector
  - Ethernet switches with load balancing
- F5 Labs BIG/ip and 3DNS
  - Load balancing, DNS, firewall

Web-specific Implementations

- HydraWEB Load Manager
  - Web content level clustering
  - Portions of URL may be distributed across several nodes for asymmetric balancing
  - Agents on nodes to monitor
- RND Network Web Server Director and Director Pro
  - LAN-based cluster WSN, WSN Pro
  - WSN-DS (Distributed Sites) for distributed environment
  - Dynamically reassigns nodes from other clusters to become part of the loaded system

43

Other Application Specific Implementations

- Sun Microsystems StorEdge
  - expansion of RAID to two-node cluster
  - remote mirroring (replication)
  - high-bandwidth direct connection between the two end-points
- Check Point FireWall-1
  - network access security monitors or firewalls
- Check Point VPN-1
  - IP-gateway providing certificate-based authentication
- Check Point FloodGate-1
  - bandwidth can be assignment via domain names, IP address, or user information

44

Summary

- Separate balancing device
  - in a network load balancing system
  - monitor traffic
  - execute a method of distributing traffic to a cluster of nodes
- Balancing methods
  - implemented independently, but very similar
- DNS as a crucial part in many load-balancing method
- Network layer (IP) & transport layer (TCP, UDP) implementation
- Instead of QoS, best-guess and proprietary method

45