

Load Balancing in Web Server Farms

Αθανάσιος Κουκουλης
M.I.S. Μάιος 2002

Abstract

The need for fast and reliable web services have forced organizations and companies to employ different methods of delivering content to the internet. Load Balancing is one such solution that apart from high levels of reliability, fault tolerance and speed provide scalability and flexibility that are much needed for today's administrators. In this paper we examine how load balancing works different methods available to administrator and the profits gained from employing such policies.

Περίληψη.

Η ανάγκη για τις γρήγορες και αξιόπιστες διαδικτυακές υπηρεσίες έχει αναγκάσει τις οργανώσεις και τις επιχειρήσεις να υιοθετήσουν διαφορετικές μεθόδους για να παρουσιάσουν τις υπηρεσίες τους στο Διαδίκτυο. Η εξισορρόπηση φορτίων είναι μια τέτοια λύση που εκτός από τα υψηλά επίπεδα αξιοπιστίας, την ανοχή βλαβών και την ταχύτητα παρέχει επίσης εξελξιμότητα και την ευελιξία που απαιτούνται από τους σημερινούς διαχειριστές εφαρμογών. Σε αυτό το έγγραφο εξετάζουμε πώς η εξισορρόπηση φορτίων λειτουργεί, τις διαφορετικές διαθέσιμες μεθόδους στον διαχειριστή εφαρμογών και τα κέρδη που αποκομίζονται από την υιοθέτηση μίας τέτοιας πολιτικής.

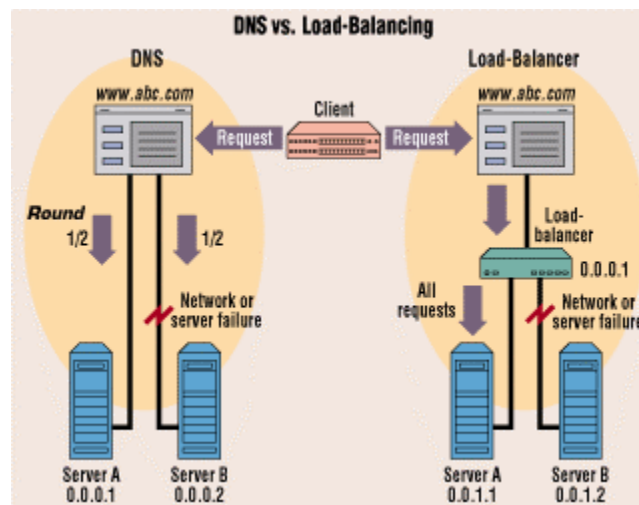
Εισαγωγή

Δεδομένου ότι η βιομηχανία ηλεκτρονικού εμπορίου συνεχίζει να αυξάνεται, περισσότερες επιχειρήσεις στηρίζονται στους ιστοχώρους τους για να επικοινωνήσουν με τους πελάτες τους. Ένας υψηλής απόδοσης ιστοχώρος που γρήγορα και αξιόπιστα εμφανίζει το περιεχόμενο του κερδίζει και διατηρεί πελάτες και είναι κρίσιμος για ένα επιτυχές και ανταγωνιστικό ηλεκτρονικό εμπόριο. Λίγοι πιθανοί πελάτες θα επιστρέψουν σε έναν αργό ιστοχώρο εάν ο αντιμετωπίσουν σημαντικές καθυστερήσεις ή αποτυχία να κατεβάσουν τις πληροφορίες που θέλουν. Κατά συνέπεια, ως τμήμα του προγραμματισμού και της εφαρμογής υποδομής ηλεκτρονικών εφαρμογών μιας εταιρίας, πρέπει να εξεταστεί σοβαρά πώς να βελτιωθεί η απόδοση του ιστοχώρου.

Μπορούν να χρησιμοποιηθούν διάφορες μέθοδοι για να βελτιωθεί η απόδοση του ιστοχώρου: επέκταση του εύρος ζώνης πρόσβασης στο Διαδίκτυο, χρήση γρηγορότερου εξοπλισμού δικτύων, σχεδίαση πιο αποδοτικών εφαρμογών, βελτιστοποίηση και αναβάθμιση του λογισμικού και του υλικού κεντρικών υπολογιστών. Εκτός από αυτές τις δυνατότητες, μπορούμε να βελτιώσουμε την απόδοση του ιστοχώρου με την προσθήκη κεντρικών υπολογιστών (web servers) και να έχουμε το περιεχόμενο του ιστοχώρου μας σε όλους τους κεντρικούς υπολογιστές. Αυτή η μέθοδος μας επιτρέπει να μοιραστεί το φορτίο εξυπηρέτησης των πελατών μεταξύ των κεντρικών υπολογιστών. Επιπλέον, μπορούμε να διατηρήσουμε τους υπάρχοντες κεντρικούς υπολογιστές παρά να τους αποσύρουμε για να τους αντικαταστήσουμε με νέους υπολογιστές.

Το μοίρασμα φορτίου (load sharing), ή ισορρόπηση φορτίου (load balancing), σε πολλαπλούς κεντρικούς υπολογιστές εξασφαλίζει ότι η κυκλοφορία στον ιστοχώρο δεν υπερφορτώνει έναν κεντρικό υπολογιστή ενώ άλλοι κεντρικοί υπολογιστές δεν απασχολούνται. Για να ισορροπήσουμε το φορτίο ανάμεσα στους κεντρικούς υπολογιστές που εξυπηρετούν έναν ιστοχώρο, παραδοσιακά χρησιμοποιείτε το *dns round-robin* χαρακτηριστικό γνώρισμα του οποίου είναι να

διανείμει ομοιόμορφα τις διευθύνσεις IP των κεντρικών υπολογιστών στους χρήστες και κατά συνέπεια οι κεντρικοί υπολογιστές δικτύου σας είναι εξίσου προσιτοί. Εντούτοις, αυτός ο μηχανισμός δεν παρέχει εξισορρόπηση φορτίων σε ένα περιβάλλον στο οποίο οι κεντρικοί υπολογιστές έχουν διαφορετικές ικανότητες υλικού και λογισμικού. Παραδείγματος χάριν, ένα σύστημα κεντρικών υπολογιστών με δύο επεξεργαστές 450MHz Pentium III και 1GB της μνήμης πρέπει να χειριστεί περισσότερο αιτήματα χρηστών και φορτίο σε ένα περιβάλλον όπου υπάρχει πραγματική ισορρόπηση φορτίου από ένα σύστημα υπολογιστών με ένα επεξεργαστή 300MHz Pentium II και 256MB μνήμης. Εντούτοις, η μέθοδος του dns round-robin χρησιμοποιεί εξίσου τους δύο αυτούς υπολογιστές για τη εξυπηρέτηση των επισκεπτών στον ιστοχώρο. Επίσης με αυτήν τη μέθοδο δεν γνωρίζουμε αν κάποιος κεντρικός υπολογιστής είναι διαθέσιμος ή όχι με αποτέλεσμα να υπάρχει ο κίνδυνος σε περίπτωση βλάβης κάποιου υπολογιστή τα αιτήματα που θα έρχονται σε αυτόν να μην εξυπηρετούνται.

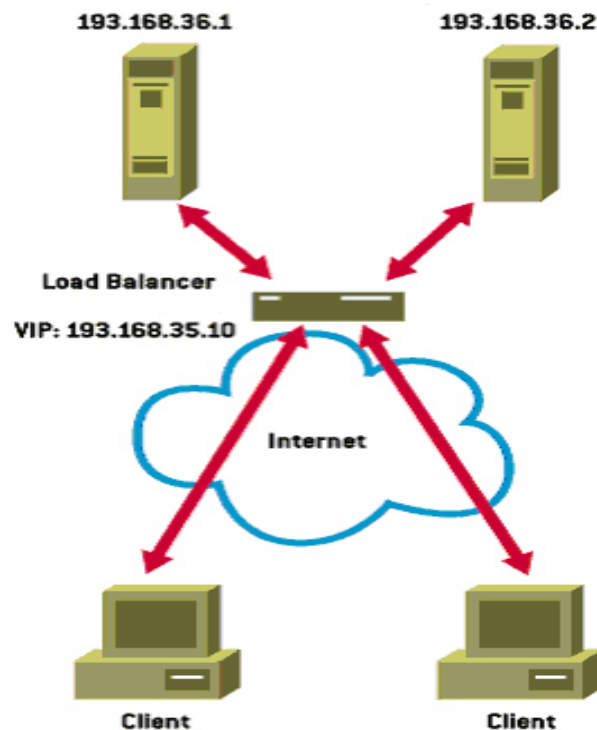


Σύγκριση DNS Round-Robin και Load Balancing

Πρόσφατα, οι προμηθευτές ανέπτυξαν τους εξισορροπητές φορτίων (*load balancers*), τα οποία είναι προϊόντα που μοιράζουν το φορτίο ομοιόμορφα ανάμεσα στους πολλαπλούς κεντρικούς υπολογιστές. Επιπλέον, ο load balancer εξασφαλίζει ανοχή βλαβών των υπολογιστών με τον επαναπροσανατολισμό της

κυκλοφορίας και των χρηστών σε άλλο κεντρικό υπολογιστή σε περίπτωση βλάβης κάποιου υπολογιστή. Επομένως, οι χρήστες αντιμετωπίζουν λιγότερη καθυστέρηση και καμία αποτυχία στις αιτήσεις τους για προβολή ιστοσελίδων.

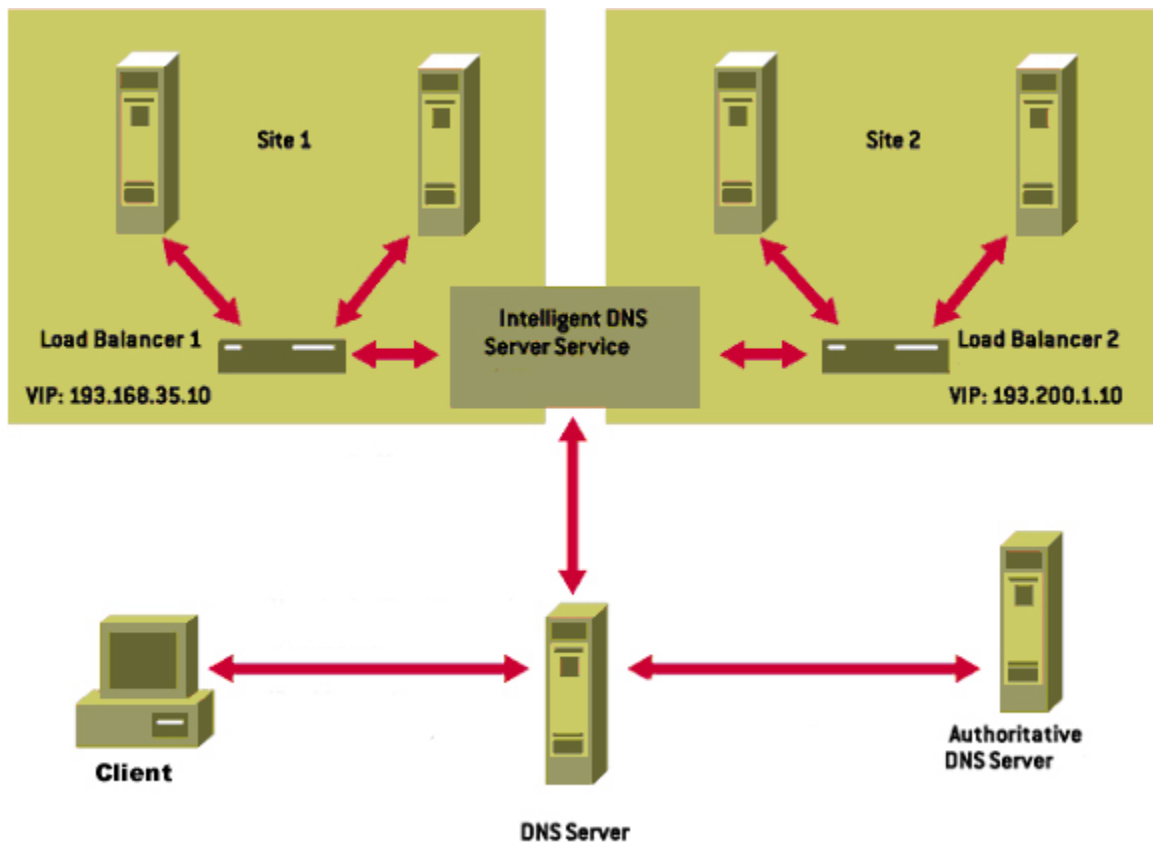
Οι load balancers είναι ένα εργαλείο που κατευθύνει έναν χρήστη στον λιγότερο απασχολημένο ή πιο κατάλληλο υπολογιστή μεταξύ διάφορων κεντρικών υπολογιστών που περιέχουν το ίδιο περιεχόμενο. Ο χρήστης έχει πρόσβαση σε αυτό το σύνολο κεντρικών υπολογιστών ως έναν εικονικό κεντρικό υπολογιστή. Παραδείγματος χάριν, υποθέτουμε ότι έχουμε δύο κεντρικούς υπολογιστές δικτύου σε ένα ενιαίο σενάριο ιστοχώρων: web1.uom.gr με τη διεύθυνση 193.168.36.1 και web2.uom.gr με την IP διεύθυνση 193.168.36.2. Ο load balancer χρησιμοποιεί ένα εικονικό όνομα (π.χ., www.uom.gr) και την εικονική διεύθυνση IP (Virtual IP) (π.χ., 193.168.35.10) για να αντιπροσωπεύσει τον ιστοχώρο. Αντιστοιχούμε το εικονικό όνομα και την αντίστοιχη εικονική διεύθυνση IP με τους δύο κεντρικούς υπολογιστές δικτύου δηλώνοντας το εικονικό όνομα και τη αντίστοιχη εικονική διεύθυνση IP στον DNS (Domain Name Server).



Ο load balancer ελέγχει συνεχώς το φορτίο και τη διαθεσιμότητα κάθε κεντρικού υπολογιστή. Όταν ένας χρήστης ζητήσει πρόσβαση στο www.uom.gr, το αίτημα πηγαίνει στον load balancer αντί ενός από τους κεντρικού υπολογιστή. Με βάση το φορτίο κάθε ελεγχόμενου κεντρικού υπολογιστή και των όρων και των πολιτικών που έχουμε προκαθορίσει, ο load balancer αποφασίζει ποιος κεντρικός υπολογιστής πρέπει να λάβει το αίτημα. Ο load balancer επαναπροσανατολίζει έπειτα το αίτημα από το χρήστη στον κεντρικό υπολογιστή και επαναπροσανατολίζει μετά την απάντηση από τον κεντρικό υπολογιστή στο χρήστη.

Οι load balancers μπορούν επίσης να υποστηρίξουν την εξισορρόπηση φορτίων ανάμεσα σε πολλαπλούς ιστοχώρους. Η εφαρμογή των πολλαπλών ιστοχώρων τοποθετεί τους κεντρικούς υπολογιστές κοντά στους πελάτες και μειώνει την καθυστέρηση απόκρισης μεταξύ του ιστοχώρου και των πελατών μας. Επιπλέον, οι πολλαπλοί ιστοχώροι παρέχουν καλύτερη εξισορρόπηση φορτίων, υψηλή διαθεσιμότητα, και μεγάλη ανοχή βλαβών σε περίπτωση πλήρους αποτυχίας ενός ολόκληρου ιστοχώρου. Σε ένα σενάριο με πολλές τοποθεσίες, κάθε load balancer φορτίων σε κάθε τοποθεσία έχει το ίδιο εικονικό όνομα αλλά μια διαφορετική εικονική διεύθυνση IP. Παραδείγματος χάριν, ο load balancer 1 στη θέση 1 έχει το εικονικό όνομα www.uom.gr και την εικονική διεύθυνση IP 193.168.35.10., ο load balancer 2 στη θέση 2 έχει το ίδιο εικονικό όνομα αλλά μία διαφορετική εικονική διεύθυνση IP 193.200.1.10. Συνδέουμε κάθε load balancer με τους τοπικούς κεντρικούς υπολογιστές χρησιμοποιώντας την ίδια μέθοδο που χρησιμοποιείτε σε ένα σενάριο μίας περιοχής. Εκτός από τον έλεγχο του φορτίου των τοπικών κεντρικών υπολογιστών, οι load balancers φορτίων ανταλλάσσουν πληροφορίες της διαμόρφωσης και των φορτίων των δύο ιστοχώρων και ελέγχουν τη διαθεσιμότητα των ιστοχώρων με τους load balancers των άλλων δικτυακών τόπων. Κατά συνέπεια, κάθε load balancer φορτίων έχει σφαιρικές πληροφορίες φορτίων και διαθεσιμότητας τοπικά. Οι load balancers στα σενάρια με πολλούς ιστοχώρους συχνά λειτουργούν και ως DNS για το εικονικό όνομα. Όταν ο load balancer λαμβάνει ένα αίτημα από κάποιον χρήστη για το εικονικό

όνομα του υπολογιστή, ο load balancer επιστρέφει στον χρήστη την εικονική διεύθυνση IP για το καλύτερο υπολογιστή ανάλογα με το παρών φορτίο στον συγκεκριμένο υπολογιστή, πόσο κοντά βρίσκεται στον χρήστη και άλλους παράγοντες. Έπειτα ο χρήστη έχει πρόσβαση σε εκείνο τον υπολογιστή χωρίς να καταλάβει καμία διαφορά.



Λύσεις Εξισορρόπησης Φορτίου (Load Balancing Solutions)

Υπάρχουν δύο τύποι load balancers: οι συσκευές υλικού (hardware) και λογισμικό.

Hardware Load Balancers

Ένας load balancer φορτίων βασισμένος στο υλικό είναι ένα «κλειστό κουτί» που είναι συνήθως ένα μηχάνημα που τρέχει το λογισμικό εξισορρόπησης φορτίου ενός προμηθευτή. Αυτές οι λύσεις είναι οι ευκολότερες στην υλοποίηση για τους διαχειριστές ιστοχώρων. Το πλεονέκτημα εδώ είναι ότι υπάρχει αφιερωμένο συγκεκριμένο υλικό για μια εργασία. Με άλλα λόγια, οι εφαρμογές των κεντρικών υπολογιστών μπορούν να συνεχίσουν να επεξεργάζονται τα στοιχεία όπως πριν, χωρίς οποιαδήποτε ανάγκη για κάποιο ακόμη κομμάτι λογισμικού, ενώ ένα συγκεκριμένο κομμάτι υλικού χειρίζεται τη διαχείριση κυκλοφορίας. Αυτή η φιλοσοφία οδηγεί σε έναν παράγοντα υψηλότερης απόδοσης για τον τομέα κεντρικών υπολογιστών συνολικά.

Επιπλέον, η συντήρηση του υλικού είναι απλή και μπορεί να γίνει με την ελάχιστη προσπάθεια. Για τις αναβαθμίσεις ή και τις περιοδικές απαιτήσεις συντήρησης, μόνο μια συσκευή στο δίκτυο πρέπει να πειραχτεί, όχι κάθε κεντρικός υπολογιστής. Εάν μια ακόμη μονάδα είναι διαθέσιμη, αυτό μπορεί να γίνει χωρίς καμία διακοπή στην λειτουργία του συστήματος. Οι λύσεις υλικού έχουν τα μειονεκτήματά τους. Παραδείγματος χάριν, μια λύση hardware δεν μπορεί σχεδόν ποτέ να ξέρει τα πάντα για όλους τους κεντρικούς υπολογιστές. Είναι ουσιαστικά αδύνατο να συγκεντρωθούν πληροφορίες όπως η χρήση της ΚΜΕ (Κεντρικής Μονάδας Επεξεργασίας) ή η χρήση μνήμης από τους κεντρικούς υπολογιστές, εκτός αν χρησιμοποιούνται και προγράμματα λογισμικού στους κεντρικούς υπολογιστές. Επίσης, μια λύση υλικού μπορεί να θεωρηθεί ως ένα συγκεκριμένο σημείο αποτυχίας ή βλάβης του όλου συστήματος. Όλη η κυκλοφορία από και προς τους κεντρικούς υπολογιστές περνά από ένα μονό

κομμάτι υλικού, το οποίο μπορεί (ανάλογα με τον παράγοντα αξιοπιστίας του υλικού) να εισαγάγει έναν πιθανό κίνδυνο. Εντούτοις, σε όλες αυτές τις περιπτώσεις υπάρχει μια ακόμη μονάδα διαμορφωμένη για να ενεργοποιηθεί σε περίπτωση που η βασική μονάδα σταματήσει να λειτουργεί.

Software Load Balancing

Μία λύση εξισορρόπησης φορτίου βασισμένη στο λογισμικό δεν απαιτεί τροποποίηση στην συνδεσμολογία ή τον εξοπλισμό δικτύων όταν εισάγεται η υπηρεσία εξισορρόπησης φορτίου στο αγρόκτημα κεντρικών υπολογιστών του δικτύου. Μπορεί να εγκατασταθεί το λογισμικό στους υπάρχοντες κεντρικούς υπολογιστές ή τους ειδικά αφιερωμένους υπολογιστές εξισορρόπησης φορτίου.

Το λογισμικό σε κάθε κεντρικό υπολογιστή επιτρέπει τη σε βάθος ανάλυση του λειτουργικού συστήματος μέσα στον κεντρικό υπολογιστή. Αυτό μπορεί να δώσει στη εφαρμογή την εντυπωσιακή δυνατότητα να εξετάσει τις ζωτικής σημασίας στατιστικές όσον αφορά τον κεντρικό υπολογιστή όπως η χρησιμοποίηση της ΚΜΕ ή μνήμης. Μερικές λύσεις επιτρέπουν ακόμη και το συγχρονισμό των στοιχείων μεταξύ των κεντρικών υπολογιστών σε έναν τομέα. Εντούτοις, οι λύσεις λογισμικού παρουσιάζουν μερικά μειονεκτήματα. Παραδείγματος χάριν, ένα πρόσθετο κομμάτι λογισμικού σε κάθε κεντρικό υπολογιστή απορροφάει ένα άγνωστο ποσό επεξεργαστικής ισχύς σε εκείνο τον κεντρικό υπολογιστή. Όσο περισσότερο η εφαρμογή έχει ανάγκη για επεξεργαστική ισχύ τόσο περισσότερο μειώνεται η ισχύ των κεντρικών υπολογιστών για να εξυπηρετήσει τις ανάγκες της εξισορρόπησης φορτίου. Ένα νέο κομμάτι του λογισμικού στους κεντρικούς υπολογιστές παρουσιάζει επίσης ένα πρόσθετο σημείο αποτυχίας ή προβλήματος ανά κεντρικό υπολογιστή. Θα ήταν ειρωνικό ότι το κομμάτι του λογισμικού που εγκαθίσταται για να εξετάσει την κατάσταση ενός κεντρικού υπολογιστή να καταλήγει να είναι αυτό το σημείο αποτυχίας.

Λειτουργίες Load Balancers

Ανεξάρτητα σε ποια κατηγορία προϊόντων ανήκει ο load balancer, εκπληρώνει τις ακόλουθες τρεις λειτουργίες: έλεγχος φορτίου και κατάστασης των κεντρικών υπολογιστών, επιλογή του σωστού κεντρικού υπολογιστή για έναν χρήστη, και επαναπροσανατολισμός της κυκλοφορίας μεταξύ του χρήστη και του κεντρικού υπολογιστή.

Έλεγχος Φορτίου και Κατάστασης Υπολογιστών

Ο load balancer ελέγχει συνεχώς το φορτίο και την κατάσταση των διοικούμενων κεντρικών υπολογιστών έτσι ώστε μπορεί να χρησιμοποιήσει τις πληροφορίες φορτίων και κατάστασης για να επιλέξει τον καλύτερο διαθέσιμο κεντρικό υπολογιστή για να ανταποκριθεί σε ένα αίτημα χρηστών. Οι load balancers χρησιμοποιούν δύο μεθόδους για να ελέγξουν τους κεντρικούς υπολογιστές: εξωτερικός έλεγχος και εσωτερικός έλεγχος

Εξωτερικός Έλεγχος

Για να ελέγξει εξωτερικά έναν κεντρικό υπολογιστή, ο load balancer υπολογίζει το χρόνο απόκρισης ενός κεντρικού υπολογιστή με την αποστολή ενός αιτήματος στον κεντρικό υπολογιστή και την αναμονή για μια απάντηση. Η χρησιμοποίηση ενός ping (Internet Control Message Protocol (ICMP) ping) είναι ο απλούστερος τρόπος για τον load balancer να ελέγχει εξωτερικά έναν κεντρικό υπολογιστή. Το ping εξετάζει τη διαθεσιμότητα ενός κεντρικού υπολογιστή και το χρόνο επιστροφής μεταξύ του κεντρικού υπολογιστή και του load balancer. Εάν ο load balancer δεν λαμβάνει μια απάντηση από τον κεντρικό υπολογιστή μετά από διαδοχικά ping, ο load balancer υποθέτει ότι ο κεντρικός υπολογιστής δεν είναι διαθέσιμος. Οι διαχειριστές συνδέουν συνήθως τους κεντρικούς υπολογιστές δικτύου άμεσα με τον load balancer, έτσι εάν ο χρόνος απόκρισης είναι μεγάλος, ο load balancer ξέρει ότι ο κεντρικός υπολογιστής είναι απασχολημένος

Εντούτοις, το ICMP ping εξετάζει μόνο τη στοίβα IP ενός κεντρικού υπολογιστή αλλά δεν μπορεί να ελέγξει την κατάσταση της στοίβας TCP που το HTTP χρησιμοποιεί. Για να ελέγξει ότι η στοίβα TCP ενός κεντρικού υπολογιστή λειτουργεί, ο load balancer προσπαθεί να εγκαταστήσει μια σύνδεση TCP, που απαιτεί μια τριπλή «χειραψία», με τον κεντρικό υπολογιστή. Σε μια τριπλή χειραψία, ο load balancer στέλνει στον κεντρικό υπολογιστή ένα πακέτο TCP με το δυαδικό ψηφίο SYN 1. Εάν balancer φορτίων λάβει πίσω από τον κεντρικό υπολογιστή ένα πακέτο TCP με το δυαδικό ψηφίο SYN 1 και το δυαδικό ψηφίο ack 1, ο load balancer στέλνει ένα άλλο πακέτο TCP με το δυαδικό ψηφίο SYN 0 και το δυαδικό ψηφίο ack 1. Μια ολοκληρωμένη χειραψία σημαίνει ότι η στοίβα TCP του κεντρικού υπολογιστή είναι υγιής. Μετά από την ολοκλήρωση της χειραψίας, ο load balancer διακόπτει αμέσως τη σύνδεση για να κάνει οικονομία στους πότους του δικτύου. Ο load Balancer μπορεί να υπολογίσει την απόδοση σύνδεσης TCP ενός κεντρικού υπολογιστή βασισμένος στο χρόνο που παίρνει για να ολοκληρωθεί η παραπάνω «χειραψία».

Εκτός από τη παραπάνω δοκιμή, ο load balancer μπορεί να ελέγξει το χρόνο απόκρισης και τη διαθεσιμότητα ενός κεντρικού υπολογιστή με την υποβολή ενός αιτήματος HTTP για το περιεχόμενο μίας ιστοσελίδας ή ένα URL. Παραδείγματος χάριν, αν η αρχική σελίδα του web1.uom.gr είναι η index.htm, ο load balancer μπορεί να κάνει μία εντολή HTTP Get ζητώντας το περιεχόμενο index.htm από τον web1.uom.gr. Εάν ο load balancer λάβει από τον κεντρικό υπολογιστή δικτύου τον κώδικα 200, η αρχική σελίδα στον web1.uom.gr είναι διαθέσιμη. Ο load balancer μετρά το χρόνο απόκρισης με τη μέτρηση του χρόνου μεταξύ της αποστολής του αιτήματος και της λήψης του κώδικα επιστροφής.

Εσωτερικός Έλεγχος

Αν και ο εξωτερικός έλεγχος μας δίνει την δυνατότητα να εξακριβώσουμε χρήσιμες πληροφορίες, παρέχει περιορισμένους ή καθόλου πληροφορίες για διάφορες σημαντικές πτυχές της κατάστασης ενός κεντρικού υπολογιστή, συμπεριλαμβανομένης της ΚΜΕ, τη μνήμη και άλλα στοιχεία συμπεριφοράς των συστημάτων και της εφαρμογής. Μόνο ο εσωτερικός έλεγχος μπορεί να παρέχει τέτοιες λεπτομερείς πληροφορίες για το φορτίο των κεντρικών υπολογιστών. Για να ελέγξει εσωτερικά έναν κεντρικό υπολογιστή, ο load balancer χρησιμοποιεί ένα εσωτερικό ελεγκτικό πρόγραμμα, το οποίο τρέχει φυσικά σε κάθε κεντρικό υπολογιστή, ελέγχει τη κατάσταση του υπολογιστή, και ενημερώνει τον load balancer. Ο εσωτερικός έλεγχος είναι κοινός για εφαρμογές βασισμένες στο λογισμικό, αλλά load balancers που βασίζονται στο hardware χρησιμοποιούν τον εξωτερικό έλεγχο.

Επιλογή Υπολογιστή για εξυπηρέτηση

Ο load balancer μπορεί να χρησιμοποιήσει τις πληροφορίες από τον εξωτερικό και εσωτερικό έλεγχο ενός κεντρικού υπολογιστή για να επιλέξει ποιος κεντρικός υπολογιστής είναι καλύτερος για το χειρισμό ενός αιτήματος από κάποιον χρήστη. Εάν όλοι οι κεντρικοί υπολογιστές έχουν την ίδια ποιότητα υλικού και λογισμικού, μπορούμε να διαμορφώσουμε τον load balancer για να χρησιμοποιήσει ένα round-robin σύστημα για να επιλέξει έναν κεντρικό υπολογιστή βασισμένο στη θέση των κεντρικών υπολογιστών. Εντούτοις, εάν ο load balancer διαχειρίζεται έναν κεντρικό υπολογιστή με ένα επεξεργαστή Pentium III και έναν κεντρικό υπολογιστή με έναν υπέρ επεξεργαστή Pentium, μπορούμε να διαμορφώσουμε τον load balancer για να επαναπροσανατολίζει

περισσότερη κυκλοφορία στον ισχυρότερο κεντρικό υπολογιστή. Αυτή η οργάνωση είναι μια σταθμισμένη round-robin διαμόρφωση.

Πολλές εφαρμογές load balancing επαναπροσανατολίζουν τις αιτήσεις των χρηστών τον κεντρικό υπολογιστή που αυτή την στιγμή έχει τον λιγότερο αριθμό συνδέσεων TCP (least connections algorithm). Παρόλο που αυτό εισάγει ένα στοιχείο πραγματικής εξισορρόπησης του φορτίου ανάμεσα στους υπολογιστές δεν εξαλείφει το πρόβλημα που υπάρχει και με την μέθοδο του DNS round-robin.

Παραλλαγές των παραπάνω αλγορίθμων είναι το Round Robin με βάρη (*weighted round-robin*) και οι λιγότερες συνδέσεις με βάρη (*weighted least connections*).

Στον αλγόριθμο *weighted round-robin* σε κάθε πραγματικό κεντρικό υπολογιστή ανατίθεται ένα βάρος, v , που αντιπροσωπεύει την ικανότητά του να χειριστεί τις συνδέσεις, σε σύγκριση με τους άλλους πραγματικούς κεντρικούς υπολογιστές που συνδέονται με τον εικονικό κεντρικό υπολογιστή. Δηλαδή οι νέες συνδέσεις ανατίθενται στους πραγματικούς κεντρικούς υπολογιστές v φορές προτού να επιλεχτεί ο επόμενος πραγματικός κεντρικός υπολογιστής στο αγρόκτημα κεντρικών υπολογιστών.

Στον αλγόριθμο *weighted least connections* προσδιορίζει ότι ο επόμενος κεντρικός υπολογιστής που επιλέγεται από ένα αγρόκτημα κεντρικών υπολογιστών για μια νέα σύνδεση στον εικονικό κεντρικό υπολογιστή είναι ο κεντρικός υπολογιστής με τις λιγότερες ενεργές συνδέσεις. Σε κάθε κεντρικό υπολογιστή ανατίθεται επίσης ένα βάρος για αυτόν τον αλγόριθμο. Όταν τα βάρη ανατίθενται, ο κεντρικός υπολογιστής με τις λιγότερες συνδέσεις είναι βασισμένος στον αριθμό ενεργών συνδέσεων σε κάθε κεντρικό υπολογιστή, και στη σχετική ικανότητα κάθε κεντρικού υπολογιστή. Η ικανότητα ενός δεδομένου κεντρικού υπολογιστή υπολογίζεται ως το βάρος εκείνου του κεντρικού υπολογιστή που διαιρείται με το ποσό των βαρών όλων των κεντρικών υπολογιστών που συνδέονται με εκείνο τον εικονικό κεντρικό υπολογιστή.

Ένας πιο περίπλοκος load balancer έχει την δυνατότητα να μας επιτρέψει να προσδιορίσουμε μια πιο πολύπλοκη πολιτική επιλογής κεντρικών υπολογιστών. Παραδείγματος χάριν, μπορούμε να διαμορφώσουμε μία πολιτική για να λαμβάνει υπ'όψιν την χρήση της ΚΜΕ, τη χρησιμοποίηση μνήμης, τον αριθμό ανοικτών συνδέσεων TCP, και τον αριθμό πακέτων που μεταφέρονται στο κεντρικό υπολογιστή. Ο τύπος εξισορρόπησης φορτίων μπορεί να έχει τη εξής μορφή $(5 * \text{χρήση της ΚΜΕ}) + (3 * \text{χρησιμοποίηση μνήμης}) + (6 * \text{ο αριθμός ανοικτών συνδέσεων TCP}) + (5 * \text{ο αριθμός μεταφερμένων πακέτων}) = \text{το φορτίο ενός υπολογιστή}$. Όταν λαμβάνει ένα αίτημα χρηστών, ο load balancer υπολογίζει το φορτίο για κάθε κεντρικό υπολογιστή σύμφωνα με τον τύπο και επαναπροσανατολίζει το αίτημα στον κεντρικό υπολογιστή με το ελαφρύτερο φορτίο.

Μερικές εφαρμογές εξισορρόπησης φορτίων επιτρέπουν την εξισορρόπηση που βασίζεται στο περιεχόμενο. Παραδείγματος χάριν, διάφορα webserver θα μπορούσαν να είναι οργανωμένα για να προσφέρουν στον χρήστη αποκλειστικά τις εικόνες, ενώ άλλες μηχανές επικεντρώνονται στο δυναμικό περιεχόμενο. Άλλοι κεντρικοί υπολογιστές χειρίζονται το βίντεο ροής, και τα λοιπά. Οι διαφορετικοί κεντρικοί υπολογιστές, κάθε ένας που βελτιστοποιείται για να παρέχει έναν συγκεκριμένο τύπο αρχείου ή εγγράφου, μπορούν να παρέχουν την καλύτερη γενική απόδοση.

Σε μερικές περιπτώσεις, αφού ο load balancer αναθέτει έναν κεντρικό υπολογιστή σε έναν χρήστη και ο κεντρικός υπολογιστής και ο χρήστης κάνουν μια αρχική σύνδεση, η εφαρμογή απαιτεί ότι ο load balancer να στέλνει διαρκώς τις κινήσεις του χρήστη σε εκείνο τον κεντρικό υπολογιστή. Αυτή η σύνδεση είναι μια επίμονη σύνδεση (persistent or sticky connection). Παραδείγματος χάριν, ένας χρήστης ψωνίζει σε ένα on-line βιβλιοπωλείο και βάζει τρία βιβλία στο

καλάθι αγορών. Εάν ο κεντρικός υπολογιστής που επεξεργάζεται το καλάθι αγορών αποθηκεύει τις πληροφορίες τοπικά, ο load balancer δεν μπορεί να αποστείλει τη νέα αίτηση του χρήστη σε έναν άλλο κεντρικό υπολογιστή ακόμα κι αν το φορτίο δεν είναι ισορροπημένο ανάμεσα στους κεντρικούς υπολογιστές. Διαφορετικά, τα τρία βιβλία στο καλάθι αγορών του χρήστη θα χαθούν επειδή ο νέος κεντρικός υπολογιστής δεν έχει τις πληροφορίες αγορών του χρήστη. Επομένως, ο load balancer πρέπει να θυμηθεί ποιος χρήστης έχει πρόσβαση σε ποιο κεντρικό υπολογιστή για ένα ορισμένο χρονικό διάστημα που καθορίζετε βάση της συμπεριφοράς των πελατών και το είδος της εφαρμογής. Εάν ενεργοποιήσουμε στον load balancer το χαρακτηριστικό της επίμονης σύνδεσης, αυτό το χαρακτηριστικό γνώρισμα θα υπερισχύει πάντα σε σχέση με άλλες πολιτικές εξισορρόπησης φορτίου.

Το κλειδί για τη διατήρηση μιας επίμονης σύνδεσης είναι να ανακαλυφθεί η ταυτότητα ενός χρήστη και να δεσμευθεί αυτή η ταυτότητα σε έναν κεντρικό υπολογιστή προορισμού. Ο load balancer χρησιμοποιεί συνήθως τη διεύθυνση IP του χρήστη ως ταυτότητα του χρήστη. Εντούτοις, η διεύθυνση προέλευσης του χρήστη να μην είναι η πραγματική διεύθυνση IP του χρήστη. Πολλά επιχειρήσεις και ISPs χρησιμοποιούν Proxy Servers για να ελέγξουν την κυκλοφορία στο διαδίκτυο και να κρύψουν τις διευθύνσεις IP των χρηστών τους. Κατά συνέπεια, εάν 500 χρήστες έχουν πρόσβαση στον ιστοχώρο από μία επιχείρηση και 10 χρήστες έχουν πρόσβαση στον ιστοχώρο από μια άλλη επιχείρηση, το φορτίο κεντρικών υπολογιστών δεν θα μοιραστεί σωστά επειδή ο load balancer θα δεσμεύσει και τους 500 χρήστες της πρώτης επιχείρησης που έχουν την ίδια διεύθυνση προέλευσης σε έναν κεντρικό υπολογιστή και τους άλλους 10 χρήστες σε έναν άλλο κεντρικό υπολογιστή. Για να υπερνικήσει αυτό το μειονέκτημα, ο load balancer που υποστηρίζει τη διεύθυνση IP και το συσχετισμό αριθμού θυρών TCP μπορεί να διακρίνει τους χρήστες ακόμα κι αν οι χρήστες χρησιμοποιούν τον ίδιο Proxy server. Ο load balancer μπορεί να κάνει αυτήν την διάκριση επειδή κάθε σύνδεση TCP έχει μια μοναδική διεύθυνση πηγής IP και έναν αριθμό θύρας TCP. Ένας άλλος τρόπος να προσδιοριστεί ένας χρήστης

είναι, εάν ο χρήστης χρησιμοποιεί μια ασφαλή μέθοδο επικοινωνίας HTTP, είναι να ελεγχθεί η ασφαλής ταυτότητα επικοινωνίας (SSL Secure Socket Layer Session ID).

Επαναπροσανατολισμός της κυκλοφορίας

Ο load balancer μπορεί να χρησιμοποιήσει διάφορες μεθόδους για να επαναπροσανατολίσει τα αιτήματα των χρηστών στον επιλεγμένο κεντρικό υπολογιστή: media access control (MAC) address translation (MAT) και Network Address Translation (NAT).

MAT. Ο load balancer που χρησιμοποιεί αυτήν την μέθοδο απαιτεί από κάθε κεντρικό υπολογιστή να χρησιμοποιεί και την εικονική διεύθυνση IP (VIP) του load balancer, εκτός από τη φυσική διεύθυνση IP του κεντρικού υπολογιστή. Όταν ο load balancer λαμβάνει ένα πακέτο ενός χρήστη και κάνει μια επιλογή κάποιου κεντρικού υπολογιστή, ο load balancer αντικαθιστά τη διεύθυνση MAC προορισμού στο πακέτο του χρήστη με την επιλεγμένη διεύθυνση της MAC του κεντρικού υπολογιστή και στέλνει το πακέτο στον κεντρικό υπολογιστή. Το πακέτο περιέχει τη διεύθυνση IP του χρήστη, έτσι για να απαντήσει άμεσα στο χρήστη, ο υπολογιστής χρησιμοποιεί την αρχική διεύθυνση IP του χρήστη ως διεύθυνση προορισμού IP. Εντούτοις, ο κεντρικός υπολογιστής χρησιμοποιεί τη εικονική διεύθυνση VIP του load balancer ως διεύθυνση πηγής IP, έτσι ώστε να φαίνεται στο χρήστη ότι η επικοινωνία είναι από τον load balancer. Κατ' αυτό τον τρόπο, το επόμενο πακέτο του χρήστη πηγαίνει στον load balancer παρά στον κεντρικό υπολογιστή που απάντησε στο χρήστη.

NAT. Χρησιμοποιώντας την μέθοδο NAT, ο load balancer αντικαθιστά τη διεύθυνση προορισμού ενός λαμβανόμενου πακέτου (δηλαδή την εικονική διεύθυνση IP του load balancer) με την επιλεγμένη διεύθυνση IP του κεντρικού υπολογιστή και τη διεύθυνση πηγής IP με τη VIP διεύθυνση του load balancer προτού να επαναπροσανατολίσει το πακέτο στον επιλεγμένο κεντρικό

υπολογιστή. Όταν ο load balancer επαναπροσανατολίζει ένα πακέτο του κεντρικού υπολογιστή στο χρήστη, αντικαθιστά τη διεύθυνση προορισμού IP με τη διεύθυνση IP του χρήστη και τη διεύθυνση πηγής IP με τη εικονική διεύθυνση IP του load balancer. Αυτή η μέθοδος κρύβει τις διευθύνσεις IP των κεντρικών υπολογιστών από τους χρήστες, έτσι οι κεντρικοί υπολογιστές δικτύου μπορούν να χρησιμοποιήσουν οποιοσδήποτε διευθύνσεις IP, συμπεριλαμβανομένων των ιδιωτικών διευθύνσεων. Οι κεντρικοί υπολογιστές δικτύου δεν χρειάζεται να συνδέσουν άμεσα με τον load balancer (δηλ., να χρησιμοποιήσουν το ίδιο τμήμα του τοπικού LAN) εφ' όσον μπορούν οι κεντρικοί υπολογιστές και load balancer μπορούν να επικοινωνήσουν ο ένας με τον άλλο μέσω ενός πρωτοκόλλου δρομολόγησης.

Απαραίτητα χαρακτηριστικά

Εξελξιμότητα – κάθε εφαρμογή load balancing πρέπει να επιτρέπει τη δυναμική προσθήκη και την αφαίρεση των στοιχείων του δικτύου με έναν τρόπο που είναι συνολικά διαφανής στον τελικό χρήστη. Επιπλέον αυτή η διαδικασία πρέπει να είναι σχετικά ανώδυνη για το προσωπικό συντήρησης των δικτύων για να εγκαταστήσει και να διαχειριστεί το υπάρχον υλικό. Οι συσκευές πρέπει να είναι σε θέση να βγουν από το δίκτυο με ένα ελεγχόμενο τρόπο και να επανεισαχθούν σε αυτό βαθμιαία. Μια λύση πρέπει να είναι σε θέση να εργαστεί ετερογενώς μεταξύ των λειτουργικών συστημάτων και των πλατφόρμων υλικού και λογισμικού. Οι συσκευές και εφαρμογές load balancing πρέπει να μας επιτρέπουν να διατηρήσουμε τον υπάρχον εξοπλισμό και να συνεχίσουν να χρησιμοποιούν μία συσκευή που ήταν το πιο εξελιγμένο κομμάτι τεχνολογίας πριν 18 μήνες αλλά είναι τώρα το πιο αργό κομμάτι του δικτύου μας. Με αυτό τον τρόπο τα κόστη των αναβαθμίσεων γίνονται σημαντικότερα λιγότερα ενώ ταυτόχρονα διατηρείται μία συνέχεια σε όλο μας το δίκτυο.

Ευελιξία - μια λύση πρέπει επίσης να προσφέρει πολλά και εύκαμπτα "σχέδια" εξισορρόπησης φορτίων για να χρησιμοποιούνται μέσα στα αγροκτήματα

υπολογιστών. Ένας διαχειριστής χρειάζεται πολλαπλές προαιρετικές δυνατότητες διαθέσιμες για να επιλέξει πώς ο load balancer θα αποφασίζει σε ποια συσκευή να στείλει το κάθε αίτημα των χρηστών. Επιπλέον, ένας διαχειριστής πρέπει να έχει την πολυτέλεια της επιλογής των διαφορετικών σχεδίων για τα διαφορετικά αγροκτήματα συσκευών. Ένα αγρόκτημα πρέπει να είναι σε θέση να χρησιμοποιήσει το σχέδιο A, ενώ ένα δεύτερο αγρόκτημα (που συντηρείται από την ίδια ισορροπώντας λύση φορτίων) χρησιμοποιεί το σχέδιο B. Αυτό είναι μια ανάγκη σε περίπτωση που τα διαφορετικά αγροκτήματα υπολογιστών έχουν διαφορετικές ανάγκες και απαιτήσεις.

Συμπεράσματα

Η ανάγκη για την βέλτιστη εξυπηρέτηση των πελατών μιας επιχείρησης στο Διαδίκτυο επιβάλλει στην επιχείρηση να είναι σε θέση να προσφέρει γρήγορες και αξιόπιστες διαδικτυακές υπηρεσίες. Η λύση της εφαρμογής εξισορρόπησης φορτίων (load balancing) είναι σε θέση να προσφέρει μεγάλο ποσοστό αξιοπιστίας στον ιστοχώρο μας επιτρέποντας την συνεχή παρουσία του στο διαδίκτυο με μεγάλο βαθμό ανοχής βλαβών. Επίσης είναι σε θέση να προσφέρει μεγάλη ταχύτητα στην εξυπηρέτηση των αιτημάτων των πελατών επιλέγοντας τον καλύτερο δυνατό υπολογιστή για την εξυπηρέτηση. Τέλος είναι εύκολα επεκτάσιμο και συντηρήσιμο καθώς κανένα από τα κομμάτια του δικτύου δεν μπορεί από μόνο του να θέσει όλη την εφαρμογή εκτός λειτουργίας.

Συνεπώς οι λύσεις με την χρήση πολλαπλών κεντρικών υπολογιστών (server farms) και χρήση εξισορρόπησης φορτίου (load balancing) είναι οι οικονομικά ευνοϊκότερες και προσφέρουν τα μέγιστα δυνατά πλεονεκτήματα για την δημιουργία ενός εταιρικού ιστοχώρου για την ανάπτυξη της ηλεκτρονικής επιχειρηματικότητας.

Βιβλιογραφία

- 1) A Comparative Analysis of Web Switching Architectures - **Arrow Point Communications**
- 2) Planning for Load Balancing
<http://www.ibm.com>
- 3) Internet Traffic Management Switches – **Fountry Networks Inc.**
- 4) Web site load balancing Solutions *Understanding and evaluating availability systems for Internet presence* - **Networkshop, Inc**
- 5) Achieving Scalability and High Availability for E-Business Clustering in BEA WebLogic® Server - **BEA Systems, Inc**
- 6) Improving Web Server Performance With Zeus Load Balancer - **Zeus Technology Ltd**
- 7) Combining Global Load Balancing and Geo-location with Emissary™ - **Coyote Point Systems, Inc. and Digital Envoy.**
- 8) Bullet Proof Sites with Radware. Building Bullet Proof Internet/Intranet Sites with IP Load Balancing, Scalability Optimization, Fault Tolerance and Availability with Radware - www.radware.com
- 9) Options for load balancing - **IT Week (February 12th 2001)**
- 10) Growing Your E- Business with IBM Server Load Balancing and Caching embedded solutions - www.ibm.com
- 11) “Always There for You” E-commerce requires a highly available Web architecture - and clustering may be the answer - www.iemagazine.com
- 12) Load balancing serves the farms – www.eetimes.com
- 13) Building a Highly Available and Scalable Web Farm
Paul Johns and Aaron Ching - **Microsoft Developer Network**
www.microsoft.com
- 14) Network Load Balancing Technical Overview
Microsoft Developer Network – www.microsoft.com
- 15) Server Load Balancing: The Key to a Consistent Customer Experience
www.riverstonenet.com
- 16) Load Balancing Web Applications
www.onjava.com
- 17) The Tao of Network Load Balancing
www.winntmag.com
- 18) Web-Site Problem-Solving, Load-Balancing Style continued
www.networkcomputing.com
- 19) Server Load Balancing
www.cisco.com
- 20) Web farming 101--NT Load Balancing Service
by **Scot Johnson**
www.elementkjournals.com
- 21) Cisco IOS Server Load Balancing and the Catalyst 6000 Family of Switches
www.cisco.com
- 22) Load-Balancing Switches: Balancing Act
by Logan G. Harbaugh
www.internetwk.com
- 23) Betting the Farm
by Kamesh Namuduri & Palani Ram
www.intelligententerprise.com