

Aktives Load-Balancing in Wireless LAN Hotspots

Robil Daher, Heiko Kopp, Djamshid Tavangarian

Lehrstuhl für Rechnerarchitektur

Institut für Informatik

Universität Rostock

[vorname.nachname]@uni-rostock.de

Abstract: Drahtlose Netzwerke nach dem Wireless LAN Standard IEEE 802.11 sind heute in immer größeren Bereichen im Einsatz. Trotz der vorhandenen Handover-Fähigkeit zeigt sich, dass in derartigen Netzwerken eine zuverlässige Lastverteilung auf erreichbare Basis-Stationen gerade für echtzeitbasierte Anwendungen entscheidende Vorteile hat. Im Rahmen dieses Artikels wird ein Verfahren zum aktiven Load-Balancing in Wireless LAN vorgestellt. Neben der Erläuterung des zu Grunde liegenden Protokolls zeigen erste Messungen die dadurch erreichbare Effizienz. Den Abschluss bildet ein Ausblick über weitere Untersuchungen auf diesem Gebiet.

1 Wireless LAN und Load-Balancing

Wireless LAN nach dem IEEE 802.11 Standard ist nicht zuletzt auf Grund der heute verfügbaren Erweiterungen IEEE 802.11b, g und a in unterschiedlichsten Ausprägungen verfügbar. So werden häufig gerade Firmengebäude, Hochschulen und öffentliche, stark frequentierte Bereiche, wie Flughäfen und Bahnhöfe mit einem drahtlosen Netzwerk ausgestattet. Auf Grund geringer Reichweiten einzelner WLANZellen (Access Point - AP), findet oftmals eine so genannte multizellulare Architektur Anwendung, in der ein Handover zwischen den einzelnen Zellen durch entsprechende Protokolle im Standard geregelt wurde.

Der Algorithmus des Roamings im WLAN, nach dem Standard IEEE 802.11, beginnt mit der Suche benachbarter APs in Reichweite des Clients sobald das Signal-Rausch-Verhältnis zum assoziierten AP einen bestimmten Grenzwert unterschreitet. Existieren mehrere APs, wählt der Client einen dieser aus und sendet diesem ein REASSOCIATION-Request. Ist der ausgewählte AP in der Lage den neuen Client zu verwalten, sendet er eine REASSOCIATION-Response zurück um den Vorgang abzuschließen. Ab diesem Zeitpunkt ist der Client mit dem neuen AP assoziiert[IE98]. Die Zeit dieses Prozesses wird als Handover- oder Roaming-Zeit bezeichnet. Diese Zeit ist grundsätzlich abhängig von Herstellern sowohl der Clients, als auch der APs[Jü04]. Da für den Einsatz von Voice-over-IP eine Latenz von nicht mehr als 50 ms im Worst-Case empfohlen ist, kann dies jedoch zu Problemen führen[Jü04]. Der Begriff Load-Balancing wird als Algorithmus zur ausgewogenen Verteilung von Last auf Knoten (APs) des Systems innerhalb einer multizellularen Organisationsstruktur, mit sich überlappenden Bereichen, verwendet. Die Last eines Wireless LANs nach IEEE 802.11 ist jedoch nicht nur von der Anzahl aktiver Clients pro Zelle, wie in Netzwerken mit fester Kanalzuweisung [FZ02], abhängig, son-

dern wichtige Parameter, z. B. die noch zur Verfügung stehende Bandbreite pro AP mit einbeziehen.

Zwei unterschiedliche Mechanismen können dabei angewandt werden. Call-Level Mechanismen betrachten zur Entscheidung für die Verteilung einzig die Anzahl der assoziierten Stationen pro Access Point. Packet-Level-Basierende Mechanismen protokollieren die Performance einzelner APs. Dies umfasst unterschiedliche Netzwerkparameter wie den Datenverkehr, die Bandbreite oder ähnliches. Zur Realisierung einer Load-Balancing Lösung ist ein Algorithmus zur Verteilung der Last und zum Management der Stationen notwendig. Dieser Load Distribution Mechanism (LDM) kann in drei Phasen untergliedert werden. In Phase 1 erfolgt die Akquisition von Informationen über die Last nach ausgewählter Metrik sowohl auf Client- als auch auf AP-Seite. Phase 2 wendet den Load Balancing Algorithmus (LBA) zur Erreichung einer ausgewogenen Verteilung der Last auf die zur Verfügung stehenden APs auf die in Phase 1 gewonnenen Daten an. Dabei kann eine Reassoziierung eines Clients erfolgen. Die als Load Distribution Control (LDC) bezeichnete Phase kann zentralisiert und verteilt angelegt sein, wobei das Inter-AP Protocol (IAPP)[IE03] zur Interoperabilität von Geräten unterschiedlicher Hersteller eingesetzt werden kann. In der Phase 3 wird das Ergebnis aus Phase 2 durch Implementierung unterschiedlicher Mechanismen zur forcierten Reassoziierung von Clients realisiert. Dabei ist z.B. ein anwendungsgesteuertes Senden eines REASSOCIATION-Request an einen ausgewählten AP möglich.

2 Active Load-Balancing über Forced Roaming

Das hier vorgestellte Load-Balancing System basiert auf einem zentralisierten LDC auf Basis eines dedizierten Servers innerhalb des Netzwerkes. Der Client wird dabei analog zum IEEE 802.11 Standard als STA bezeichnet. Da wichtige physikalische Parameter des Clients abhängig von seiner Position zum AP sind, werden Sie kontinuierlich gemessen und an das LDC periodisch bzw. nach kritischen Änderungen übertragen. Die Phase 1 des Load-Balancing Mechanismus erfolgt dabei auf Seite des Clients ohne das eine Modifikation der Firmware nötig ist. Ausgewählte Informationen sind z. B. die ESS (Extended Service Set)-Adresse des Clients, die Liste verfügbarer APs auf verfügbare Kanäle in Reichweite des Clients, die BSS (Basic Service Set)-Adresse des assoziierten APs, die physikalische Verbindungsqualität über Signal-, und Rauschpegel sowie den Signal-Rausch-Abstand, die aktuell verfügbare Bandbreite auf Grund der Entfernung zum AP sowie die verwendeten Services des Clients. Zur Übertragung derartiger Informationen an das LDC wird ein Teil eines selbst entworfenen Protokolls, des so genannten Intelligent Management of WLAN-Cells Access (IMCA)-Protokolls, verwendet. Dieses bietet zusätzlich Möglichkeiten zur Last-Verteilung, Funktionalitäten zur Unterstützung von Quality of Service und zum mobilen Management. IMCA setzt auf TCP bzw. UDP auf. Im Folgenden werden die einzelnen Bestandteile des Protokolls näher beschrieben.

2.1 Das IMCA-Protokoll

Das IMCA-Protokoll unterstützt zwei unterschiedliche Architekturen. Im Client-Server-Modell ist kein direkter Zugriff auf die Firmware des Access Points notwendig, da ein dedizierter IMCA-Server eingesetzt wird. Dessen Aufgabe übernehmen im Controller-Modell die einzelnen APs selbst. Nur letzteres Modell erlaubt die Unterstützung verteilter

LoadBalancing-Mechanismen. Das Protokoll definiert unterschiedliche Methoden:

- REGISTER-Operationen zur Registrierung des IMCA-Clients als Bestandteil des Clients beim Server sowie zur Anfrage nach einer Reassoziatiion, falls diese notwendig ist. Operationen dieser Art werden entweder periodisch oder nach kritischen Änderungen überwachter Parameter durchgeführt.
- CHANGE-Operationen dienen der Anfrage des Servers an den Client während einer REGISTER-Operation zum AP-Wechsel.

Durch Einsatz des IMCAProtokolls kann die grundsätzliche Funktionalität der Phasen 1 bis 3 realisiert werden. Eine Unterstützung für einen spezifischen LoadBalancing-Mechanismus fehlt jedoch. Der in diesem Artikel implementierte Algorithmus arbeitet nach dem Call-Level Prinzip und basiert auf dem IMCA-Protokoll. Er wird im Folgenden näher erläutert.

2.2 Active Call-Level Algorithmus

Im Unterschied zur Definition Call-Level-basierender Mechanismen wird ein Client in dem hier vorgestellten System als passiv oder als aktiv angesehen. Ein Client ist passiv, wenn er mit einem AP assoziiert ist, aber keine aktive Datenübertragung mit diesem Client durchgeführt. Er wird als aktiv bezeichnet, sobald eine Datenübertragung stattfindet, der Client also Bandbreite belegt. Befindet sich ein Client in Reichweite von mindestens zwei APs, wird er als shared bezeichnet. Ein Client kann jedoch nur bei einem AP zu einer Zeit als aktiv betrachtet werden. Aus Sicht des APs wird ein shared Client als covered bezeichnet, wenn er nicht mit ihm assoziiert ist. Daraus ergibt sich der Active Client Load (ACL) Wert eines APs aus der Anzahl momentan aktiver Clients. In Analogie dazu bezeichnet der Passive Client Load (PCL) Wert die passiven assoziierten Clients. Die Virtual Client Load (VCL) eines AP ergibt sich dann als Anzahl der als covered definierten Clients. Die Maximal Client Load (MCL) kann damit wie in Formel 1 definiert werden.

$$MCL = ACL + PCL + VLC \quad (1)$$

$$f = \frac{m_s}{m} \quad (2)$$

$$B = \frac{(\sum_{i=1}^n L_i)^2}{n \sum_{i=1}^n L_i^2} \quad (3)$$

Grundlage des vorgestellten Algorithmus ist die ausgewogene Verteilung des ACL-Wertes. Die Verteilung wird dabei auf Basis eines Balance-Index definiert [JDH84][CJ89] (siehe Formel 3). Dabei bezeichnet n die Anzahl der APs und L_i die ACL des APs i . Ein Balance Index von 1 gibt eine vollständig ausgewogene Verteilung an. Die Flexibilität der Last-Balancierung eines drahtlosen Netzwerkes ist prinzipiell durch die Anzahl der shared Clients bestimmt. Wäre diese 0, gäbe es keine Möglichkeit die Last zu verteilen. Eine Möglichkeit zur Einschätzung der Verteilung kann dabei mit dem in Formel 2 beschriebenen Faktor definiert werden. Dabei bezeichnet m die Anzahl aller Clients, m_s die Anzahl der gesamten shared Clients in einem Wireless LAN. Er ist analog zum Balance-Index quantitativ, jedoch nicht geeignet um präzise den Zustand der Clients zu beschreiben. Ein

Faktor $f = 1$ definiert hier, dass alle Clients verteilbar sind, was dem Best Case entspricht. Bei Einsatz anderer Load-Metriken, die z.B. auf der Bandbreite oder dem Paketverlust basieren stellt dies möglicherweise nicht das Optimum dar.

Im Fall einer Entscheidung des Load-Balancing Algorithmus zum Wechsel des assoziierten APs wird das so genannte Forced Roaming eingesetzt. Ein CHANGE-Request des Servers empfangen durch den Client löst die darin enthaltene MAC-Adresse des vom LDC ausgewählten neuen AP auf und leitet diese an das WLAN-Interface weiter. Drauf hin wird ein REASSOCIATION-Prozess gestartet. Nach Ende des Vorgangs informiert das WLAN-Interface den Client über den Erfolg oder den Misserfolg der Reassoziatio-

3 Messungen und Tests

Das in Abschnitt 4 erläuterte Verfahren wurde im Rahmen der ersten praktischen Untersuchungen in einem eigenständigen WLAN-Interface unter Linux implementiert. Die Implementierung setzt dabei auf das Open-Source Projekt HostAP mit Unterstützung für Chipsätze der Prism-Familie auf. Da der AP in den untersuchten Szenarien nicht verändert wird, ist ein Einsatz des Treibers nur auf den Clients im Managed Mode notwendig. Eine Veränderung des Source-Codes ist jedoch unnötig.

Das drahtlose Test-WLAN besteht aus APs von Enterasys mit einer 802.11b Karte, dem IMCA-Server auf einem dedizierten Rechner und dem Client. Da der HostAP-Treiber nur Prism-Chipsätze unterstützt, kam im Client eine MA401-Netzwerkkarte zum Einsatz. In zwei Szenarien wurden entweder mit aktivem oder passivem Client Messungen durchgeführt. Im aktiven Szenario wurde während des Roamings eine FTP-Übertragung durchgeführt. Um Erkenntnisse im Einsatz des verwendeten Roaming-Prozesses zu ermitteln wurde in den ersten Tests die Zeit des forcierten Roamings (Forced-Roaming-Time) mit der IMCA-Client-Software ermittelt. Die FRT wird als die Zeit zwischen Versenden des REASSOCIATION-Requests bis zur software-seitigen Erkennung der neuen Assoziation definiert.

In der Auswertung der Messungen zeigt sich, dass während der Untersuchung trotz Variation des SNR-Wertes sowie automatischer Bandbreitenanpassung der APs nur in zwei Fällen eine geringere Bandbreite auftritt. Deutlich wird, dass nun keine Abhängigkeit zwischen SNR und Roaming vorliegt. Insgesamt überraschend ist, dass die Roaming-Zeit selbst mit im Mittel unter 50 ms vielfach geringer ist, als beim automatischen Roaming im regulären Betrieb. Zeiten von über 2 s unter Einsatz des Host-AP-Treibers im Hard-Roaming-Modus [A103] können damit drastisch reduziert werden. Im Minimum wurden 14 ms, im Maximum 90 ms für die Roaming-Zeit gemessen. Mit Werten unter 50 ms ist sogar eine Unterstützung echtzeitbasierter Anwendungen möglich. Ein Grund für die geringeren Zeiten ist der verwendete Forced Roaming Mechanismus, der im Gegensatz zum Hard Roaming keine Suche nach neuen Access Points zur Assoziation benötigt, was viel Zeit kostet in der keine Übertragung möglich ist.

4 Ausblick

Die Untersuchungen zeigten, das Forced Roaming neben guten Roaming-Zeiten komfortable Ansätze ohne direkte Hardware-Veränderung oder Firmware-Eingriffe erlaubt. Eine optimale Lastverteilung in WLANs ist jedoch nicht garantiert, da alle bekannten Verfahren einzig die Neuordnung von shared Clients auf mögliche APs unterstützen. Hier gilt es

neue Strategien zu finden, um die anfallende Last optimal zu verteilen.

Das vorgestellte Protokoll wird ausgehend von den beschriebenen Messungen an Bedürfnisse echtzeitbasierter Anwendungen angepasst werden. Unserer Erkenntnis nach bieten die entwickelten Algorithmen viele Potentiale zur Steigerung der Effizienz von Load-Balancing-Algorithmen sowie der Verringerung der Roaming-Zeiten. Dies ist ein wesentlicher Schritt hin zur ubiquitären Nutzung von WLAN auch in echtzeitbasierten Systemen.

References

- [AI03] Aleo, V. Load distribution in IEEE 802.11 cells. March 2003.
- [CJ89] Chiu, D. und Jain, R.: Analysis of the increase/decrease algorithms for congestion avoidance in computer networks. In: *Journal of Computer Networks and ISDN, Vol. 17, No. 1*. June 1989.
- [FZ02] Fang, Y. und Zhang, Y.: Call admission control schemes and performance analysis in wireless mobile networks. In: *IEEE Transactions on Vehicular Technology vol. 51, No. 2*. March 2002.
- [IE98] IEEE. IEEE standard 802.11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. 1998.
- [IE03] IEEE. IEEE trial-use recommended practice for multi-vendor access point interoperability via an inter-access point across distribution system supporting IEEE 802.11 operations. 2003.
- [JDH84] Jain, R., D. Chiu, und Hawe, W.: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. In: *DEC REsearch Report TR-301*. September 1984.
- [Ju04] Judge, P.: New standard could give wi-fi a voice. In: *ComputerWeekly.com*. January 2004.