(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

PCT

(43) Internationales Veröffentlichungsdatum
3. August 2006 (03.08.2006)

(10) Internationale Veröffentlichungsnummer
WO 2006/079574 A1

(51) Internationale Patentklassifikation:
H04L 29/14 (2006.01) *H04L 29/12* (2006.01)
H04L 29/06 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2006/050028

(22) Internationales Anmeldedatum:
4. Januar 2006 (04.01.2006)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2005 004 151.5 28. Januar 2005 (28.01.2005) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **SIEMENS AKTIENGESELLSCHAFT** [DE/DE]; Wittelsbacherplatz 2, 80333 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: **LÖBIG, Norbert** [DE/DE]; Im Erlich 79, 64291 Darmstadt (DE). **TEGELER, Jürgen** [DE/DE]; Auf Der Leiten 2, 82377 Penzberg (DE).

(74) Gemeinsamer Vertreter: **SIEMENS AKTIENGESELLSCHAFT**; Postfach 22 16 34, 80506 München (DE).

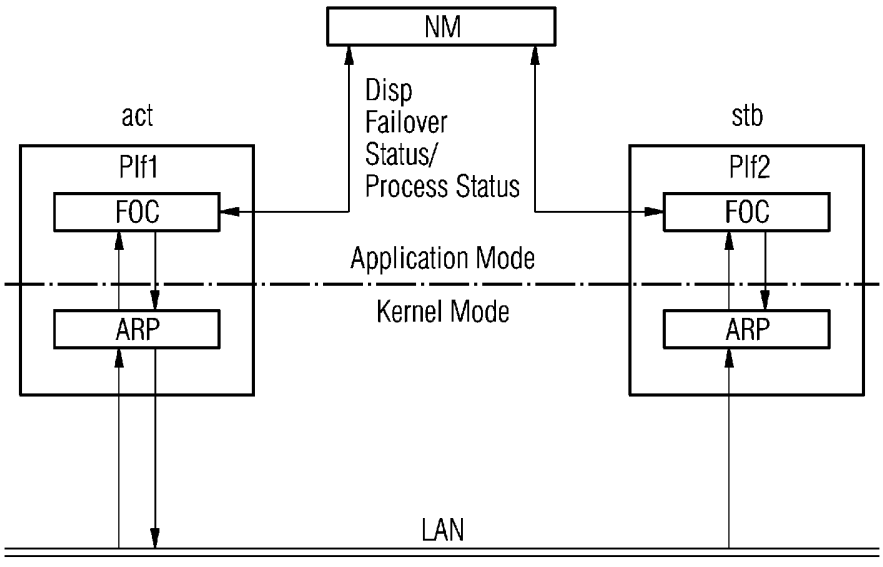
(81) Bestimmungsstaaten *(soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten *(soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,

[Fortsetzung auf der nächsten Seite]

(54) **Title:** METHOD AND APPARATUS FOR ASSIGNING PACKET ADDRESSES TO A PLURALITY OF DEVICES

(54) **Bezeichnung:** VERFAHREN UND VORRICHTUNG ZUR ZUORDNUNG VON PAKETADRESSEN EINER MEHRZAHL VON EINRICHTUNGEN



(57) **Abstract:** The aim of the invention is to unconditionally prevent a split brain scenario during high availability requests. In prior art, said aim is achieved by providing a separate hardware unit which monitors the individual devices on the IP level and makes sure that an IP failover occurs only when the device that has been identified as being down no longer responds to ARP requests. Said process has the disadvantage of requiring a lot of effort while being comparatively slow. The invention overcomes said disadvantage by providing a communication link between a switching control device (FOC) operating on the application level and the module (ARP) operating on the protocol level. Quitting of the correct physical address (ARP request) is shifted to the application level via said communication link.

[Fortsetzung auf der nächsten Seite]

WO 2006/079574 A1

ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

Veröffentlicht:

— *mit internationalem Recherchenbericht*

(57) Zusammenfassung: Ein „split brain" Scenario ist bei Hochverfügbarkeitsforderungen unbedingt zu vermeiden. Beim Stand der Technik ist zur Behandlung dieser Problematik eine separate Hardware Einheit vorgesehen, die die einzelnen Einrichtungen auf IP-Ebeneüberwacht. Diese Hardware Einheit sorgt dafür, dass nur dann ein IP-Failover stattfindet, wenn die als ausgefallen erkannte Einrichtung auch keine ARP requests mehr beantwortet. Problematisch an dieser Lösung ist, dass das entsprechende Verfahren aufwendig und vergleichsweise langsam ist. Die Erfindung löst dieses Problem, indem eine Kommunikationsbeziehung zwischen einer auf Applikationsebene arbeitenden Umschaltesteuervorrichtung (FOC) und dem auf Protokollebene arbeitenden Modul (ARP) bereitgestellt wird, über die eine Verlagerung des Quittierens der korrekten physikalischen Adresse (ARP Request) auf die Applikationsebene vorgenommen wird.

Beschreibung

Verfahren und Vorrichtung zur Zuordnung von Paketadressen einer Mehrzahl von Einrichtungen

In zuverlässigen Rechnersystemen, wie sie in Kommunikationssystemen eingesetzt sind, werden beim Stand der Technik redundante Ressourcen bereitgehalten. Diese können beispielsweise als Rechnerkapazität, Speicher- oder Ein-/Ausgabe-Kapazitäten ausgebildet sein, wobei die Ressource von einer Plattform oder Einrichtung bereitgestellt wird. Beim Ausfall einer Einrichtung übernimmt eine redundante Einrichtung unmittelbar die Aufgaben der ausgefallenen Einrichtung, so dass von außen gesehen Funktion und Verfügbarkeit nur minimal beeinträchtigt sind.

Weist eine hinsichtlich ihrer Funktion zu ersetzende ausgefallene Einrichtung paketbasierte Schnittstellen auf, über die sie in Kontakt mit anderen Einrichtungen steht, so gibt es prinzipiell zwei Möglichkeiten, die Operabilität wiederherzustellen:

Im ersten Fall hat die ersatzgeschaltete Einrichtung eine eigene Paketadresse, die von derjenigen der ausgefallenen Einrichtung verschieden ist. In diesem Fall müssen alle Kommunikationspartner explizit auf diese neue Paketadresse umgeschaltet werden.

Im zweiten Fall hat die die Funktion der ausgefallenen Einrichtung übernehmende Einrichtung die Fähigkeit, die Paketadresse der ausgefallen Einrichtung bereitzustellen. Dies hat den Vorteil, dass keine im Hinblick auf Hochverfügbarkeitsszenarien vermeidbaren Zeitverzüge und Nichterreichbarkeiten entstehen. Außerdem hat das den Vorteil geringerer Anforderungen an die Kommunikationspartner, die in diesem zweiten Falle in Ausfallszenarien nicht auf eine andere Paketadresse umschalten müssen bzw. umgeschaltet werden müssen. Voraus-

setzung hierfür ist aber, dass die über eine Mehrzahl von Einrichtungen ersatzschaltbare logische Paketadresse eindeutig und zuverlässig auf die ebenfalls paketbasierte physikalische Adressierung des vorgelagerten Netzes aufgelöst d. h. zugeordnet wird. Dies ist notwendig, da das Netz jede der Einrichtungen mit ihrer eigenen physikalischen HW-Adresse ansprechen muss.

Im Hinblick auf das Beispiel eines IP-basierten Paketnetzes ist die logische Paketadresse als IP-Adresse und die physikalische Adresse als weltweit eindeutige MAC-Adresse der beteiligten Ethernet-Schnittstelle ausgebildet. Damit die Beeinträchtigung der Funktion durch den Ausfall der aktiven Einrichtung aus Sicht des IP-Netzes und der fernen Kommunikationspartner besonders gering gehalten werden kann, wird eine sogenannte „IP-Failover" - Funktion ausgeführt. Hierbei übernimmt eine zu der ausgefallenen Einrichtung redundante Einrichtung mit deren Funktionen auch deren IP-Adresse. Dies bedeutet, dass die IP-Adresse der ausgefallenen Einrichtung auf eine funktionsfähige Einrichtung mit derselben IP-Adresse übertragen wird. Daneben kann die Einrichtung weitere IP-Adressen haben, die keine IP-Failover Adressen sind.

Die IP-Failover Funktion kann so realisiert sein, dass die übernehmende Einrichtung zu den IP-Adressen, die sie übernimmt, jeweils eine Nachricht zur Adressauflösung (gratuitous ARP) auf ihr lokales Netz (LAN) absetzt und damit de facto eine Broadcast-Nachricht in das sie umgebende Netz sendet. Alle empfangenden Einrichtungen aktualisieren daraufhin ihre Adresslisten (ARP caches), indem sie in ihre Caches befindlichen IP-Adressen hinsichtlich der empfangenen Adressauflösungen (logische IP-Adresse zu physikalischer MAC Adresse am Ethernet) überprüfen, gegenebenfalls die MAC-Adresse der übernehmenden Einrichtung eintragen und dabei die MAC-Adresse der ausgefallenen Einrichtung überschreiben. Dies ist die Voraussetzung dafür, dass die über IP (Internet Protokoll) mit dem Rechnersystem kommunizierenden, entfernten Partner-

systeme zumindest auf Applikationsebene den Ausfall innerhalb des Rechnersystems nicht bemerken.

Für die Zeit zwischen dem Ausfall einer Einrichtung bis zur Ersatzschaltung durch die redundante Einrichtung werden insbesondere Nachrichten, die von externen Kommunikationspartnern stammen, durch den der ausgefallenen Einrichtung vorgelagerten Router nicht erfolgreich zugestellt werden können. Gleiches gilt für weitere Kommunikationspartner der Einrichtung am gleichen Netz (LAN). Spätestens mit dem Ablauf der Gültigkeitszeit der Adressauflösung wird ein Kommunikationspartner, z. B. der Router, die ausgefallene IP-Adresse erneut auflösen wollen. Hierzu generiert er eine Adressauflösungsanfrage an sein Netz (ARP request). Dies hat die Wirkung, dass alle am Netz befindlichen Einrichtungen die Anfrage erhalten, ob sie für eine vorgegebene logische Paketadresse zuständig sind und dies gegebenenfalls in der Quittung unter Angabe ihrer physikalischen MAC-Adresse am Netz offenbaren. Erfolgt dies nun in der kurzen Zeit zwischen Ausfall und Ersatzschaltung, so wird die betroffene IP-Adresse sogar aus der Adressliste des Routers herausfallen, da keine gültige Adressauflösung auf eine MAC-Adresse gefunden werden kann. Für hochzuverlässige Systeme ist es daher wichtig, den Zeitraum zwischen Fehlerbeginn und Ersatzschaltung auf ein Minimum zu begrenzen.

Leider können auch mit dieser Vorgehensweise nicht alle möglichen Ersatzschaltescenarien behandelt werden. So ist von kommerziellen Rechnerplattformen bekannt, dass diese dauerhaft auf Applikationsebene hängen bleiben können, d. h. einzelne Applikationsprozesse oder Applikationsprozeßanteile (Threads) oder gegebenenfalls sogar alle Applikationsprozesse können nicht mehr zum Ablauf gebracht werden. Dies bedeutet aber nicht, dass auf IP-Ebene die Kommunikation mit vorgelagerten Routern oder anderen Kommunikationspartnern am Netz ebenfalls unterbrochen ist, diese kann unter Umständen weiter aufrechterhalten bleiben. Die Überwachung der Rechnerplatt-

form erkennt nun den Ausfall und schaltet inklusive IP-Failover Adressen auf eine redundant vorgehaltene Einrichtung um. Damit befinden sich aber plötzlich zwei Einrichtungen am Netz, die beide die betroffenen logischen IP-Failover Paketadressen auf Anforderung auf ihre eigene physikalische MAC-Adresse auflösen.

Damit ist es dem Zufall überlassen, ob ein vorgelagerter Router zum Zeitpunkt des Zustellens einer Nachricht gerade die ausgefallene Einrichtung oder die operable, ersatzgeschaltete Einrichtung adressiert. Wenn also der ARP Modul (Teil des IP Stack) der ausgefallenen Einrichtung noch ARP requests beantworten kann und aufgrund anderer Gegebenheiten eine Ersatzschaltung inklusive IP-Failover stattgefunden hat, kann es zu einer gravierenden andauernden Kommunikationsstörung und damit zum Funktionsverlust der Einrichtung kommen. Wenn in dieser Situation beide Einrichtungen einen ARP request für eine IP-Failover Adresse empfangen, für die eine Ersatzschaltung stattgefunden hat, beantworten ihn beide. Wenn nun die ausgefallene Einrichtung die ARP response zufällig später schickt als die operable ersatzgeschaltete Einrichtung, speichern alle Kommunikationspartner und Router des Netzes die MAC-Adresse der ausgefallenen Einrichtung als die zugehörige MAC-Adresse zur IP-Failover Adresse, wenn sie diese IP fail-over Adresse bereits aktuell auf eine MAC Adresse aufgelöst haben, also zu der genannten Einrichtung bereits Kommunikation besteht. Die IP-Nachrichten werden also weiterhin an die ausgefallene Einrichtung geschickt statt an die ersatzgeschaltete Einrichtung, die die Funktion der ausgefallenen übernehmen sollte. In einem solchen Fall ist es also nicht kontrollierbar, welche der beiden Einrichtungen die IP-Nachrichten empfängt und inwieweit sie noch bearbeitet werden. Die Gefahr, dass Fehlfunktionen oder Funktionsverlust eintreten oder eventuell sogar inkonsistente Daten entstehen, ist sehr groß.

Ein solches Szenario wird in der Fachliteratur auch als „split brain" Scenario bezeichnet und ist bei Hochverfügbarkeitsforderungen unbedingt zu vermeiden. Beim Stand der Technik ist zur Behandlung dieser Problematik eine separate Hardware Einheit vorgesehen, die die einzelnen Plattformen einer Redundanz Einheit (i.a. eines Redundanz Paares) auf IP-Ebene überwacht. Diese Hardware Einheit sorgt dafür, dass nur dann ein IP-Failover stattfindet, wenn die als ausgefallen erkannte Einrichtung oder Plattform auch keine ARP requests mehr sendet/ beantwortet. Problematisch an dieser Lösung ist, dass das entsprechende Verfahren aufwendig und vergleichsweise langsam ist (im ungünstigen Fällen im Minutenbereich).

Der Erfindung liegt die Aufgabe zugrunde, einen Weg und eine Vorrichtung aufzuzeigen, mit deren Hilfe das Auftreten und die Auswirkungen von „split brain" Scenarien minimiert werden können.

Die Erfindung wird ausgehend von den in den Oberbegriffen der Patentansprüche 1 und 10 angegebenen Merkmale durch die kennzeichnenden Merkmale gelöst.

Der Vorteil an der Erfindung ist im Adresshandling auf Applikationsebene zu sehen. Die Beurteilung, ob eine Plattform funktionsfähig ist oder durch eine IP-Failover Funktion ersatzgeschaltet werden muss, wird auf Applikationsebene von der Funktion der Fail-over Control aufgrund einer Mehrzahl von Umgebungskriterien (Umgebungsparametern) bestimmt. Ein Umgebungskriterium oder Umgebungsparameter kann die Verfügbarkeit/ Nichtverfügbarkeit der Kommunikation zur redundanten Einrichtung sein, die Möglichkeit der Kommunikation zu weiteren Einrichtungen, die die Ersatzschaltung steuern, die Zugreifbarkeit zentraler Ressourcen, die für die ordnungsgemäße Bearbeitung der Aufgaben der Einrichtung notwendig sind. Die Fail-over Control sorgt dafür, dass zu jeder Zeit die zwei redundanten Einrichtungen aus einer Mehrzahl von redundanten Einrichtungen zugeordneten

Umgebungskriterien nicht zugleich auf aktiven Betrieb, d. h. Bereitstellung der IP fail-over Adressen eingestellt sein können. Erfindungsgemäß besitzt die Fail-over Control Funktionen in jeder der redundanten Einrichtungen mit IP fail-over Adressen und läuft auf Applikationsebene als Applikationsprozeß oder Teil davon, der bei jedem vom ARP Modul empfangenen ARP request befragt wird, ob eine ARP response zu schicken ist. Es erfolgt also grundsätzlich kein unbedingtes ARP response auf einer unter der Applikationsebene befindlichen Ebene. Der ARP request wird erfindungsgemäss davon abhängig gemacht, ob insbesondere ein vorgegebener Nachrichtenaustausch (Handshake) zwischen ARP Modul und Failover Control auf Applikationsebene durchgeführt werden kann.

Durch das Handshake Verfahren wird auch im Falle von nicht mehr ausgeführten Applikationen einfach, sicher und zweifelsfrei erkannt, dass keine ARP requests beantwortet werden dürfen. Damit kann eine redundante Einrichtung gemäß Steuerung der Fail-over Control die IP fail-over Adressen auch in diesem Falle übernehmen, ohne dass die ausgefallene, bzgl. der Applikationsebene nicht mehr funktionierende Einrichtung die Adressumschaltung auf einem Level unterhalb des Applikationslevels durch ungewollte ARP responses stört. Es wird also eine Abhängigkeit eingeführt zwischen einer Fail over Control (FOC) in der Anwenderschicht und dem ARP Modul auf der niedrigeren Schicht des IP-Stack. Diese stellt sicher, dass bei beliebigen Ausfallszenarien der ARP Modul logisch synchron zum Zustand der Fail over Control arbeitet und er in der Lage ist, jeweils die richtige Information (Nachrichten zur IP-Failover Adresse an diese Plattform schicken/ nicht schicken) ins Netz kommuniziert. Eine externe Überwachung der Plattform, die über die Fail over Control auf Applikationsebene hinausgeht, wie dies beim Stand der Technik der Fall ist, ist damit nicht erforderlich.

Vorteilhafte Weiterbildungen der Erfindung sind in den Unteransprüchen angegeben.

So ist weiterhin vorgesehen, die Zeitfenster für unzulässigen Cache Inhalt durch freilaufende Adressauflösungsnachrichten, die zusätzlich nach ARP Request oder als Folge empfangener inkonsistenter Adressauflösungen (ARP Request /Response, Gratuitous ARP) aus der Applikationsebene ausgebracht werden, zu verkürzen.

Darüber hinaus können HW-Mechanismen zum Rücksetzen der Einrichtung in der Weise genutzt werden, dass als erste Aktion des automatischen Wiederanlaufs die paketbasierten Schnittstellen deaktiviert werden, um das Zeitfenster für mögliche Adresskonflikte nach SW-Problemen weiter zu reduzieren.

Schließlich ist ein weiterer Vorteil der Erfindung darin zu sehen, dass ein Lauschen am Netz zum Zweck der Bedieneralarmierung durchgeführt wird im Falle dauerhafter Inkonsistenz zwischen dem durch die Fail-over Control vorgegebenen Zustand einer der redundanten Einrichtungen und den im Netz ausgetauschten Adressauflösungen.

Durch Kombination aller voranstehend aufgezeigten Mechanismen wird die höchste funktionale Verfügbarkeit erreicht.

Die Erfindung wird im Folgenden anhand eines figürlich dargestellten Ausführungsbeispiels näher erläutert. Demgemäß wird die Erfindung anhand zweier Plattformen (m = 2) beschrieben, die an ein LAN angeschlossen sind.

Dabei sind unter Umgebungsparametern Zustandsgrößen des eigenen Systems oder fremder Systeme, Kommunikations- -und Protokollbeziehungen, Zugriffe auf abgesetzte Einheiten wie Datenbasen zu verstehen. Bei vorliegendem Ausführungsbeispiel soll m = 2 gelten, was aber keinerlei Einschränkung des Verfahrens hinsichtlich einer beliebigen Anzahl von Einrichtungen sein

soll. Die Einrichtungen sind als (Rechner) Plattformen ausgebildet. Ferner kann im Netz ein Network Management NM vorgesehen sein.

Im fehlerfreien Betrieb sind alle m = 2 Plattformen aktiviert und alle Prozesse verfügbar. Jede der Plattformen Plf1, Plf2 hat mindestens eine paketbasierte Schnittstelle und ist in der Lage, eine Mehrzahl von Failover Adressen mit einer definierten Funktionalität zu bedienen. Die beiden Plattformen bilden folglich eine Menge redundanter Einrichtungen im Hinblick auf die genannten Failover Adressen und die für sie zu erbringende Funktionalität.

Ein Umgebungsparameter wird zu jeder Zeit für genau eine der beiden Plattformen gesetzt ist und für die restliche Plattform(en) nicht gesetzt. Gemäß vorliegendem Ausführungsbeispiel soll der Umgebungsparameter die Betriebszustände act/ standby sein. Ist der Umgebungsparameter für eine der beiden Plattformen gesetzt, so soll diese die Kommunikation über die Failover Adressen und die gewünschte Verarbeitungsfunktion bereitstellen. Ist das Umgebungskriterium für eine Plattform nicht gesetzt, soll diese die Failover Adressen nicht bereitstellen. Durch externe Steuerung werden die Umgebungsparameter in der beschriebenen Art zur Verfügung gestellt. Aus Gründen der Einfachheit wird im folgenden eine Plattform mit gesetztem Umgebungsparameter als aktiv (act) bzgl. der Fail over Adressen, eine ohne gesetzten Umgebungsparameter als standby (stb) bzgl. der Failover Adressen bezeichnet. Es ist hierbei insbesondere ausgeschlossen, dass ein Umgebungsparameter dauerhaft für mehr als eine Plattform auf aktiv gesetzt ist. Gemäß der Figur gibt es auf jeder der beiden Plattformen Plf1, Plf2 einen Applikationsprozess „Fail over Control" FOC, der eine Schnittstelle zu einem Modul ARP der Plattform aufweist und die Information besitzt, ob sich die Einrichtung bezüglich der Failover Adressen im Zustand „act" oder „stb" befindet.

Der Zustand "act" bedeutet dabei, dass die Plattform die Fail over Adressen bedient. Ein vorgelagerter Router muss Nachrichten mit diesen logischen Paketadressen zu dieser Plattform senden. Daher beantwortet der ARP Modul dieser Einrichtung die zugehörigen Adressauflösungsanforderungen (ARP requests). Der Zustand "Stb" bedeutet hingegen, dass die Plattform die Failover Adresse nicht bedient. Ein vorgelagerter Router darf Nachrichten mit diesen logischen Paketadressen nicht zu dieser Einrichtung senden. Der ARP Modul dieser Einrichtung beantwortet die zugehörigen ARP requests daher nicht.

Der ARP Modul teilt der Fail Over Control FOC mit, dass ein ARP request empfangen wurde. Ein ARP request wird nur quittiert, falls die Fail Over Control FOC einer Einrichtung dies explizit erlaubt. Damit werden ARP Requests de facto nicht über eine low level Funktion des IOP stacks oder des Betriebssystems autark beantwortet, sondern aus der Applikationsebene heraus. Ein Ausfall oder Hängenbleiben der Applikationsebene der Plattform führt folglich dazu, dass die Fail Over Control FOC dieser Einrichtung auch nicht mehr verfügbar ist, womit nachfolgend eintreffende ARP requests nicht mehr beantwortet werden. ARP requests werden dann nicht beantwortet, wenn entweder die Fail over Control FOC dies für richtig hält oder wenn die Kommunikation des ARP Moduls mit der Fail over Control gestört ist. Die Fail over Control FOC beantwortet ARP Requests insbesondere dann nicht, wenn die Einrichtung im Zustand stb ist. Ein Hängenbleiben auf Applikationsebene führt zur Abschaltung des ARP Handlings der betroffenen Einrichtung.

Durch diese innerhalb der Plattform erfolgte Verlagerung des Quittierens von ARP Requests auf die Applikationsebene, d. h. aus einem Anwenderprozess der Plattform heraus, wird vermieden, dass die ARP-Funktion noch aktiv ist, obwohl der Applikationslevel nicht mehr funktionstüchtig ist. (Letzteres

könnte zum Ersatzschalten der Einrichtung dienen, also der Veränderung des Umgebungsparameters der ungestörten redundanten Einrichtung auf aktiv). Sollte es dennoch zum Split Brain Effekt kommen, so wird die Auswirkung optional durch folgende Maßnahme eingeschränkt:

Erhält eine bezüglich der Failover Adresse aktive Plattform einen ARP Request, so beantwortet sie diesen, geht aber davon aus, dass weitere Plattformen unberechtigterweise ebenfalls die IP- Failover Adresse beanspruchen könnten. Also sendet sie vorsorglich nach gewissem zeitlichem Abstand der Quittung auf den ARP Request mindestens ein Gratuitous ARP zur Auflösung der Absenderadresse sicherheitshalber hinterher. Dies erfolgt auf Initiative der Fail over Control aus der Applikationsebene der Einrichtung heraus. Damit kann ein fälschlich bestehender ARP Cache Inhalt eines Kommunikationspartners am Netz oder eines vorgelagerten Routers schnell korrigiert werden, so dass die Zeitfenster, in denen wegen des Split brain Effektes unrichtige Cache Inhalte vorhanden sein können, extrem verkürzt werden.

Bleiben ARP Requests aus, so wird gratuitious ARP vorsorglich zyklisch in gewissem Mindestabstand generiert. Dies erfolgt ebenfalls auf Initiative der Fail over Control aus der Applkationsebene der Einrichtung heraus. Auch für diesen Fall ist eine Schnittstelle zwischen Fail Over Control FOC auf Applikationslevel und dem ARP Modul erforderlich. Jeder ARP Request wird notifiziert und führt zusätzlich oder alternativ zum aus dem Applikationslevel angereiztem eventuell mehrfach ausgebrachten Gratuitous ARP.

Eine derartige Vorgehensweise ist immer dann vorteilhaft, wenn die Umgebungsparameter durch unterschiedliche Laufzeiten in den beteiligten Plattformen kurzzeitig inkonsistent sein können. Ist also z. B. im Rahmen einer routinemäßigen oder manuellen Umschaltung die aktivierte Plattform bereits aktiv, wohingegen die zu deaktivierende Plattform noch nicht deakti-

viert wurde, und kommt es in diesem Zeitfenster zu einem ARP Request eines vorgeschalteten Routers, so antworten beide. Bei symmetrischer Implementierung werden beide dann auch ein zeitverzögertes Gratuitous ARP nachsetzen wollen. Durch den Zeitversatz wird die Änderung des Umgebungskriteriums die abzuschaltende Plattform mit hoher Wahrscheinlichkeit vor dem Ausbringen des zusätzlichen Gratuitous ARPs erreichen, so dass dies seitens der abzuschaltenden Plattform unterbleibt, und die u. U. inkorrekte Adressauflösung von der aktivierten Plattform korrigiert wird.

Optional kann zusätzlich zu dieser Vorgehensweise eine HW-Überwachung der Plattform auf der Applikationsebene erfolgen. Dies hat zur Folge, dass ein Hängenbleiben der Plattform oder der Fail-over Control dieser Plattform wegen SW-Defekt oder Überlastung schnell erkannt werden kann und zu dem durch HW-Mechanismen bewirkten Wiederanlauf der Plattform oder von Teilen der Plattform führt, im Rahmen dessen ggf. auch die paketbasierten Schnittstellen als erste Aktion deaktiviert bzw. zurückgesetzt werden. Mit dem Anlauf einer ehemals für eine Paketadresse aktiven Plattform werden sich die Umgebungsparameter derart verändern, dass eine der weiteren vorgesehenen Plattformen deren Funktionen übernimmt und so lückenlose Verfügbarkeit der von der Plattform gebotenen Funktionalität gewährleistet. Die HW-Überwachung setzt das Vorhandensein entsprechender HW-Funktionen voraus.

In einer weiteren Ausgestaltung der Erfindung ist vorgesehen, dass eine Plattform ihre IP-Failover Adressen deaktiviert, falls sie den dauerhaften Verlust der Zugreifbarkeit der Umgebungsparameter feststellt. Sie geht in diesem Falle davon aus, dass hiermit die Umgebungsparameter einer weiteren Plattformen derart eingestellt werden, dass diese fortan die genannten IP fail-over Adressen bedienen wird und damit keine eigene Interaktion bzgl. dieser Failover Adressen zu ungewollten Störungen und Fehlfunktionen führen würde. Die Ressourcenverfügbarkeit (z. B. interne Ressourcen für Verar-

beitung oder die ungestörte Kommunikation mit weiteren externen Verarbeitungsinstanzen) gehen in den Umgebungsparameter mit ein.

Befinden sich Plattformen zur alternativen Bedienung einer Failover Adresse am gleichen Netz, so kann jede Plattform durch Lauschen am Netz die Quittungen auf ARP-Requests und die Gratuitous ARPs weiterer Plattformen vom ARP Modul in die Fail over Control FOC übertragen bekommen. Die Fail over Control instruiert hierzu den ARP Modul, alle ARP Responses, die ja nicht als Broadcast Nachrichten gesendet werden, und alle ARP Requests und gratuitous ARPs die im Broadcast Mode verschickt werden und zur Adressauflösung herangezogen werden der Fail Over Control dann zuzustellen, wenn die von der Fail Over Control bedienten IP-Failover Adressen hierbei eine andere als die Auflösung auf die eigene MAC-Adresse erfahren.

Hiermit kann die Fail over Control optional die Konsistenz der Inhalte mit dem eigenen Umgebungsparameter überprüfen. Dies bedeutet, dass auf diesem Wege festgestellt werden kann, welche weitere Plattform sich zu einer Zeit für die Failover Adresse zuständig fühlt. Die bezüglich der Adresse aktive Plattform kann die physikalische Adresse einer unberechtigt zugreifenden weiteren Plattform damit zum Alarmierungszwecke ausgeben und Gratuitous ARPs zur Korrektur ausbringen. Eine bezüglich der Adresse inaktive Plattform oder eine beliebige weitere Plattform am Netz kann damit die an einer split brain Situation beteiligten Plattformen erkennen und den Bediener alarmieren. Eine bezüglich der Adresse inaktive Plattform kann den eigenen Umgebungsparameter gegen die Adressauflösung im Netz auf Konsistenz prüfen und ebenfalls den Bediener alarmieren. Durch die Alarmierung und den als letztes Mittel anzusehenden Bedienereingriff wird die Zeitdauer eines möglichen Adresskonflikts weiter verkürzt und die Verfügbarkeit der zur Verfügung gestellten Funktion weiter erhöht.

Patentansprüche

1. Verfahren zur Zuordnung von logischen Adressen zu physikalischen Adressen einer Mehrzahl von Einrichtungen, die miteinander in Wirkverbindung stehen und die jeweils einen auf Protokollebene arbeitenden Modul (ARP) aufweisen, mit dessen Hilfe eine Zuordnung der logischen zu den physikalischen Adressen von Kommunikationspartnern vorgenommen wird, indem vom Modul eine Anfragenachricht (ARP Request) eines Kommunikationspartners empfangen und von einer der Einrichtungen mit ihrer physikalischen Adresse quittiert (ARP Response) wird, dadurch gekennzeichnet,
dass eine Kommunikationsbeziehung zwischen einer auf Applikationsebene arbeitenden Umschaltesteuerfunktion (FOC) und dem auf Protokollebene arbeitenden Modul (ARP) bereitgestellt wird, über die eine Verlagerung des Quittierens einer Anfragenachricht (ARP request) mit der physikalischen Adresse auf Applikationsebene vorgenommen wird.

2. Verfahren nach Anspruch 1,
dadurch gekennzeichnet,
dass nach einem vorgegebenen zeitlichen Abstand, nachdem die korrekte physikalische Adresse (ARP Response) über die Applikationsebene quittiert wurde, erneut mindestens eine auf Applikationsebene durch die Umschaltesteuerfunktion (FOC) initiierte Nachricht zur Zuordnung der physikalischen Adresse zu den logischen Adressen (Gratuitous ARP) freilaufend über den Modul (ARP) an alle Kommunikationspartner gesendet wird.

3. Verfahren nach Anspruch 1 oder 2,
dadurch gekennzeichnet,
dass wahlweise zusätzlich eine HW-Überwachung der Verfügbarkeit der Applikationsebene der Einrichtungen oder der Verfügbarkeit der Umschaltesteuerfunktion der Einrichtungen vorgesehen ist, deren Ansprechen zu Maßnahmen des Wiederanlaufs der nichtverfügbaren Teile der Einrichtung führt.

4. Verfahren nach Anspruch 1 bis 3,
dadurch gekennzeichnet,
dass die Umschaltesteuerfunktion (FOC) einer Einrichtung zur Entscheidung, ob eine Anfragenachricht (ARP request) beantwortet wird, einen oder eine Mehrzahl von Umgebungsparameter heranzieht.

5. Verfahren nach Anspruch 1 bis 4,
dadurch gekennzeichnet,
dass die logischen Adressen wenigstens einer der Einrichtungen deaktiviert werden, falls ein dauerhafter Verlust der Zugreifbarkeit der Umgebungsparameter festgestellt wird.

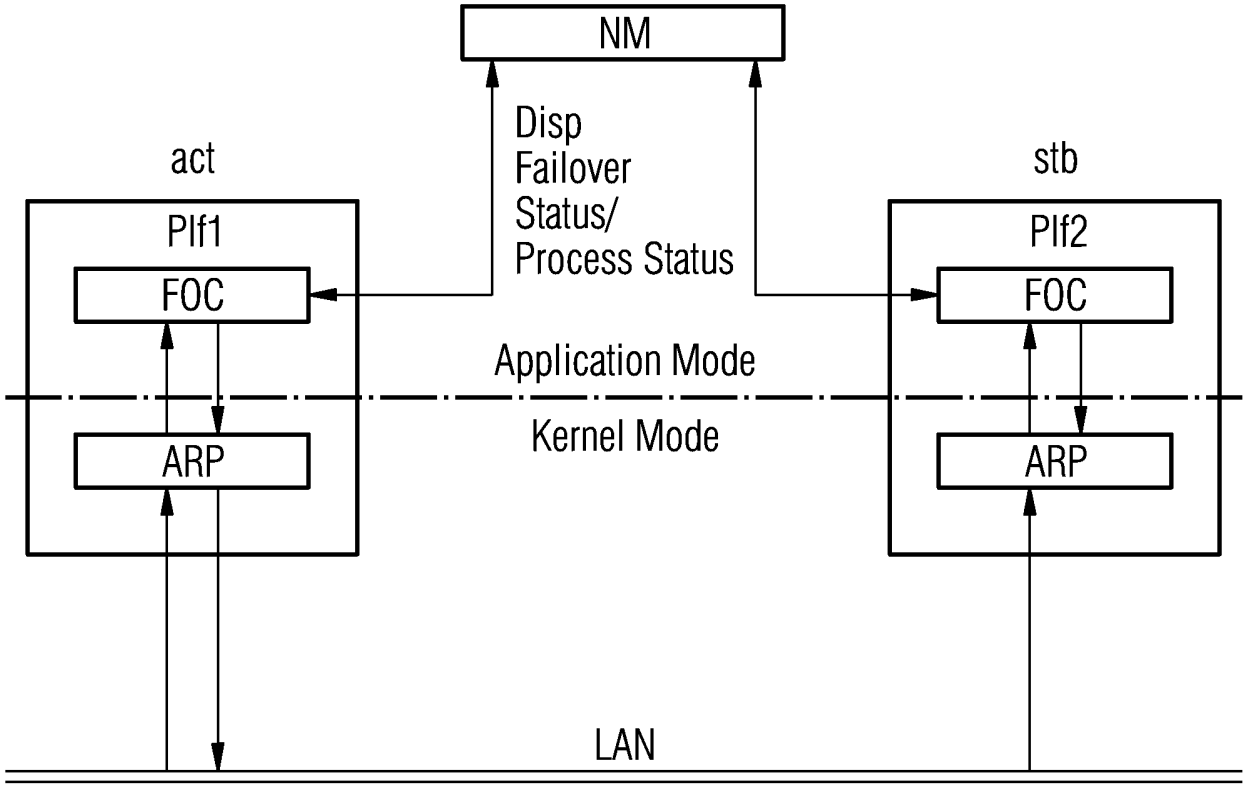
6. Verfahren nach einem der vorstehenden Ansprüche,
dadurch gekennzeichnet,
dass eine der Einrichtungen zur alternativen Bedienung einer logischen Adresse durch Lauschen am Netz die Nachrichten zur Zuordnung von logischen und physikalischen Adressen (ARP response, gratuitous ARP) der weiteren Einrichtungen in die Umschaltesteuervorrichtung (FOC) übertragen bekommen.

7. Verfahren nach Anspruch 5,
dadurch gekennzeichnet,
dass die empfangenen Adresszuordnungen gegen die Umgebungsparameter plausibilisiert werden und eine Alarmierung des Bedieners über eine Network Management Funktion unter Ausgabe von Fehlerindizien erfolgt.

8. Verfahren nach einem der vorstehenden Ansprüche,
dadurch gekennzeichnet,
dass die logische Adresse als IP-Failover Adresse ausgebildet ist.

9. Verfahren nach einem der vorstehenden Ansprüche,
dadurch gekennzeichnet,
dass die physikalische Adresse als MAC-Adresse ausgebildet ist.

10. Vorrichtung zum Umschalten zwischen wenigstens zwei redundanten Einrichtungen,
dadurch gekennzeichnet,
dass eine Kommunikationsbeziehung zwischen einer auf Applikationsebene arbeitenden Umschaltesteuervorrichtung (FOC) und dem auf Protokollebene arbeitenden Modul (ARP) bereitgestellt ist, über die die Verfügbarkeit von logischen Adressen unter einer physikalischen Adresse einer Einrichtung davon abhängig gemacht wird, ob ein Nachrichtenaustausch (Handshake) zwischen der auf Applikationsebene arbeitenden Umschaltesteuervorrichtung (FOC) und dem auf Protokollebene arbeitenden Modul (ARP) durchgeführt werden kann.


NM
Disp
Failover
Status/
Process Status
act
stb
Plf1
Plf2
FOC
FOC
Application Mode
Kernel Mode
ARP
ARP
LAN

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2006/050028

A. CLASSIFICATION OF SUBJECT MATTER
H04L29/14 H04L29/06 H04L29/12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WENSONG ZHANG: "Linux Virtual Server for Scalable Network Services" PAPER, [Online] 2000, pages 1-10, XP002372188 Ottawa Linux Symposium 2000 Retrieved from the Internet: URL:http://www.linuxvirtualserver.org/ols/lvs.pdf> [retrieved on 2006-03-15] page 2 - page 5 page 4, paragraph 3.3 page 7, paragraph 5 - page 8 ----- -/--	1-10

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 March 2006	28/03/2006

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Riegler, J

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2006/050028

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STEVENS W R ED - STEVENS ET AL: "TCP/IP Illustrated, ARP: ADDRESS RESOLUTION PROTOCOL" TCP/IP ILLUSTRATED. VOL. 1 : THE PROTOCOLS, PROFESSIONAL COMPUTING SERIES, READING, MA : ADDISON WESLEY, US, vol. VOL. 1, 1994, pages 53-68, XP002247360 ISBN: 0-201-63346-9 page 53 - page 56	1-10
A	FAN C C ET AL: "The raincore distributed session service for networking elements" PARALLEL AND DISTRIBUTED PROCESSING SYMPOSIUM., PROCEEDINGS 15TH INTERNATIONAL SAN FRANCISCO, CA, USA 23-27 APRIL 2001, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 23 April 2001 (2001-04-23), pages 1673-1681, XP010544581 ISBN: 0-7695-0990-8 page 1 - page 3, left-hand column page 7	1-10

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/EP2006/050028

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
H04L29/14 H04L29/06 H04L29/12

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
H04L

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	WENSONG ZHANG: "Linux Virtual Server for Scalable Network Services" PAPER, [Online] 2000, Seiten 1-10, XP002372188 Ottawa Linux Symposium 2000 Gefunden im Internet: URL:http://www.linuxvirtualserver.org/ols/lvs.pdf> [gefunden am 2006-03-15] Seite 2 - Seite 5 Seite 4, Absatz 3.3 Seite 7, Absatz 5 - Seite 8 ----- -/--	1-10

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen [ ] Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :
"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
15. März 2006	28/03/2006

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Riegler, J

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/EP2006/050028

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	STEVENS W R ED - STEVENS ET AL: "TCP/IP Illustrated, ARP: ADDRESS RESOLUTION PROTOCOL" TCP/IP ILLUSTRATED. VOL. 1 : THE PROTOCOLS, PROFESSIONAL COMPUTING SERIES, READING, MA : ADDISON WESLEY, US, Bd. VOL. 1, 1994, Seiten 53-68, XP002247360 ISBN: 0-201-63346-9 Seite 53 - Seite 56	1-10
A	FAN C C ET AL: "The raincore distributed session service for networking elements" PARALLEL AND DISTRIBUTED PROCESSING SYMPOSIUM., PROCEEDINGS 15TH INTERNATIONAL SAN FRANCISCO, CA, USA 23-27 APRIL 2001, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 23. April 2001 (2001-04-23), Seiten 1673-1681, XP010544581 ISBN: 0-7695-0990-8 Seite 1 - Seite 3, linke Spalte Seite 7	1-10