(19) Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 11 2006 001 378 T5** 2008.04.17

(12) **Veröffentlichung**

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2006/131906**
in deutscher Übersetzung (Art. III § 8 Abs. 2 IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2006 001 378.5**
(86) PCT-Aktenzeichen: **PCT/IL2006/000600**
(86) PCT-Anmeldetag: **21.05.2006**
(87) PCT-Veröffentlichungstag: **14.12.2006**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **17.04.2008**

(51) Int Cl.[8]: ***G06F 17/30*** (2006.01)

(30) Unionspriorität:

60/688,486	**07.06.2005**	**US**
11/258,256	**25.10.2005**	**US**

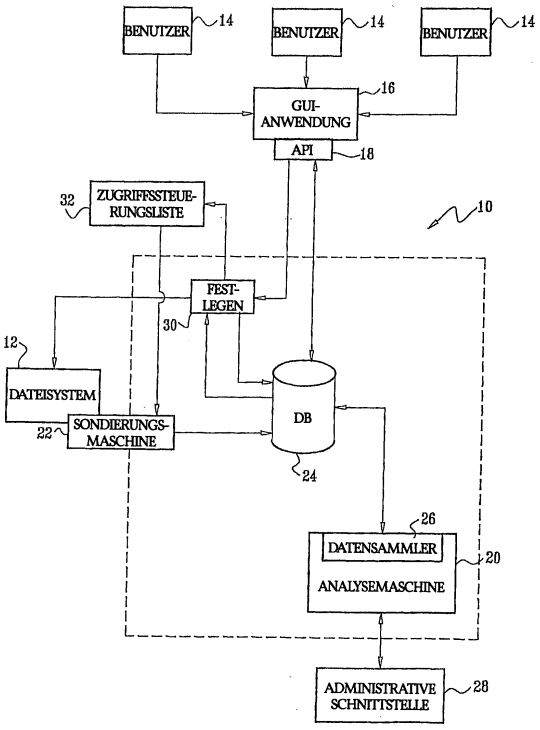
(71) Anmelder:
Varonis Inc., Saddle Brook, N.J., US

(74) Vertreter:
Kuhnen & Wacker Patent- und Rechtsanwaltsbüro, 85354 Freising

(72) Erfinder:
Faitelson, Yakov, Elkana, IL; Goldberger, Jacob, Tel-Aviv, IL; Korkus, Ohad, Tel-Aviv, IL

(54) Bezeichnung: **Automatische Verwaltung einer Speicherzugriffssteuerung**

(57) Hauptanspruch: Verfahren zum Steuern eines Datenspeicherzugriffs in einer Organisation mit Benutzern eines Dateisystems, wobei das Dateisystem Speicherelemente hat, mit folgenden Schritten:
Aufzeichnen von Zugriffen der Benutzer auf die Speicherelemente und Ableiten jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen;
Zweiclusterbildung aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird; und
Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer antwortend auf den Schritt der Zweiclusterbildung.



Beschreibung

HINTERGRUND DER ERFINDUNG

1. Gebiet der Erfindung

[0001] Diese Erfindung bezieht sich auf die Computersicherheit. Genauer gesagt bezieht sich diese Erfindung auf eine automatische Erzeugung und Verwaltung von Dateisicherheitsregelungen in Organisationen mit einer Vielfalt von Dateizugriffssteuerungsmodellen.

2. Beschreibung der verwandten Technik

[0002] Datensicherheitsregelungen bestimmen typischerweise, wer auf die gespeicherten Daten einer Organisation auf verschiedenen Computersystemen Zugriff hat. Diese Regelungen können nicht statisch sein. Benutzer von innerhalb der Organisation, z. B. Arbeitnehmer, Partner, Auftragsnehmer, können eine ebenso schwerwiegende Bedrohung wie Bedrohungen von außerhalb der Organisation darstellen. Daher sollte, da sich die Struktur und die Personalzusammensetzung der Organisation ändern, die Sicherheitsregelung von Zeit zu Zeit angepasst werden. Jedoch mangelt es Informationstechnologieabteilungen (IT-Abteilungen) an wirksamen Werkzeugen, um Benutzerzugriffsrechte zu verwalten und sicherzustellen, dass benötigte Informationen bequem verfügbar sind, während sie trotzdem die sensiblen Daten der Organisation schützen.

[0003] Aktuelle Verfahren, die dem IT-Personal zur Verfügung stehen, umfassen eine Überprüfung und Pflege von Zugriffssteuerungslisten, in Verbindung mit einer Administration von Benutzernamen, Passwörtern und der Erweiterung solcher Verfahren, um Biometrie, Verschlüsselung und eine Begrenzung des Zugriffs auf eine einzige Anmeldung zu umfassen. Solche Verfahren sind ineffizient, häufig ungenau und werden in dem Kontext großer, komplexer Organisationen, deren Struktur und Personal sich ständig ändern, unzweckmäßig.

[0004] Hilfsmittel für die Sicherheit sind für Unternehmen, die spezielle Betriebssysteme oder Umgebungen verwenden, verfügbar. Diese sind häufig auf einer rollenbasierten Zugriffssteuerung basiert, einem Verfahren, das in den letzten paar Jahren der Gegenstand eines beträchtlichen Interesses von Seiten von Regierungsorganisationen war und zuletzt in kommerziellen Unternehmen übernommen worden ist. Ein typischer Vorschlag für rollenbasierte Zugriffssteuerungen in einer Mehr-Benutzer-SQL-Datenbank ist in dem Dokument Secure Access Control in a Multi-user Geodatabase, Sahadeb De et al., das im Internet unter der URL http://www10.giscafe.com verfügbar ist, zu finden.

[0005] Trotzdem sind Zugriffssteuerungstechnologien in Unternehmen, die vielfältige Zugriffssteuerungsmodelle benutzen, nicht optimal implementiert worden. Der heutige Stand der Technik ist derart, dass es für Systemadministratoren keinen leichten Weg gibt, zu wissen, wer in solchen Umgebungen auf was zugreift. Als ein Resultat daraus haben in vielen Organisationen ein inakzeptabel hoher Anteil von Benutzern falsche Zugriffsvorrechte. Die verwandten Probleme redundanter Zugriffsvorrechte und verwaister Konten von Personal, das die Organisation verlassen hat, sind ebenfalls nicht vollständig gelöst worden. Daher besteht ein Bedarf an einem automatischen Verfahren zum Steuern von Benutzerdateiberechtigungen, um die Datensicherheit zu verbessern, Betrug zu verhindern und die Produktivität des Unternehmens zu verbessern.

ZUSAMMENFASSUNG DER ERFINDUNG

[0006] Gemäß offenbarten Ausführungsbeispielen der Erfindung werden Verfahren und Systeme zum automatischen Erzeugen und Verwalten einer Datensicherheitsregelung in vernetzten Organisationen mit vielfältigen Zugriffssteuerungsmodellen und Dateiserverprotokollen geschaffen. Der Zugriff auf Speicherelemente innerhalb des Organisationsnetzes wird kontinuierlich überwacht und analysiert, um gleichzeitige Datenzugriffsgruppierungen und Benutzergruppierungen zu definieren. Die tatsächliche Organisationsstruktur wird aus diesen Gruppierungen gelernt und wird die Basis einer dynamischen Datenzugriffssteuerungsregelung, die mit der Zeit ständig an Organisationsveränderungen angepasst wird. Eine Entscheidungshilfeschnittstelle für eine interaktive Verwaltung der Dateizugriffssteuerung wird geschaffen, und eine Einrichtung zum Erfassen und Verfolgen eines abnormalen Benutzerverhaltens wird geschaffen. So wird es Organisationen ermöglicht, den Zugriff auf ihre Daten und Anwendungen besser zu steuern.

[0007] In einigen Ausführungsbeispielen werden die Verfahren durch ein halb automatisches Verwalten der Dateizugriffssteuerung durch ein Koordinieren der Benutzer- und Datenzugriffsgruppierungen und herkömmlicher Zugriffssteuerungslisten ergänzt, um Modifikationen der Listen zu bewirken.

[0008] Zugriffssteuerungsregelungen, die durch ein Anwenden der Lehren der Erfindung entwickelt werden, haben begleitende Vorteile, z. B. ein Begrenzen der Ressourcenverwendung im Falle eines Dienstverweigerungsangriffs (engl.: denial-of-service attack).

[0009] Die Erfindung schafft ein Verfahren zum Steuern eines Datenspeicherzugriffs in einer Organisation, das durch ein Aufzeichnen von Zugriffen der Benutzer auf Speicherelemente und ein Ableiten jeweiliger Benutzerzugriffsprofile aus den aufgezeichneten Zugriffen ausgeführt wird. Dieses Verfahren wird ferner durch eine Zweiclusterbildung (engl.: biclustering) aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird, ausgeführt. Dieses Verfahren wird ferner, antwortend auf die Zweiclusterbildung, durch ein Definieren einer Steuerungsregelung für den Zugriff auf die Speicherelemente durch die Benutzer ausgeführt.

[0010] Gemäß einem Aspekt des Verfahrens erlaubt die Steuerungsregelung einen Zugriff auf Speicherelemente eines Datenclusters durch einen Benutzer lediglich, wenn auf mindestens eines der Speicherelemente in diesem Datencluster durch diesen Benutzer zugegriffen worden ist.

[0011] Gemäß einem zusätzlichen Aspekt des Verfahrens erlaubt die Steuerungsregelung einen Zugriff auf die Speicherelemente eines Datenclusters durch die Benutzer in einem Benutzercluster lediglich, wenn auf mindestens eines der Speicherelemente in diesem Datencluster durch mindestens einen der Benutzer dieses Benutzerclusters zugegriffen worden ist.

[0012] In einem weiteren Aspekt des Verfahrens wird die Struktur des Dateisystems des Speichersystems aus dem Zweiclusterbildungsverfahren abgeleitet.

[0013] Ein weiterer Aspekt des Verfahrens umfasst ein Ableiten von Benutzungsmustern des Dateisystems durch die Benutzer aus dem Zweiclusterbildungsverfahren.

[0014] Ein Aspekt des Verfahrens umfasst ein Erfassen abweichender Benutzungsmuster.

[0015] In noch einem weiteren Aspekt des Verfahrens wird die Zweiclusterbildung iterativ durchgeführt, wobei die Zugriffsprofile in jeder Iteration neu bestimmt werden, und die Steuerungsregelung folgend jeder Iteration aktualisiert wird.

[0016] In noch einem weiteren Aspekt des Verfahrens wird ein Definieren einer Steuerungsregelung durch ein Vorschlagen einer vorläufigen Version der Steuerungsregelung, ein Überwachen anschließender Zugriffe auf die Speicherelemente durch die Benutzer, ein Bestimmen, dass die anschließenden Zugriffe in Übereinstimmung mit der vorläufigen Version der Steuerungsregelung sind, und, antwortend auf die Bestimmung, ein Zulassen der vorläufigen Version als eine endgültige Version der Steuerungsregelung, ausgeführt.

[0017] Ein weiterer Aspekt des Verfahrens umfasst ein interaktives Modifizieren der Steuerungsregelung.

[0018] In einem weiteren Aspekt des Verfahrens wird das Definieren einer Steuerungsregelung automatisch und im Wesentlichen ohne einen menschlichen Eingriff durchgeführt.

[0019] Noch ein weiterer Aspekt des Verfahrens umfasst ein Bezugnehmen auf eine Zugriffssteuerungsliste, die mindestens einen Satz von Benutzern und mindestens einen Datensatz von Speicherelementen umfasst, wobei die Benutzer des Benutzersatzes in jeweiligen der Benutzercluster umfasst sind, und die Speicherelemente des Datensatzes in jeweiligen der Datencluster umfasst sind. Das Verfahren wird ferner durch ein Erfassen einer Abwesenheit von Zugriffen auf Mitglieder des jeweiligen Datenclusters durch Mitglieder des jeweiligen Benutzerclusters, und, antwortend auf den Mangel an Zugriffen, ein Entfernen mindestens eines Teils der Benutzer aus dem Benutzersatz und ein Entfernen mindestens eines Teils der Speicherelemente aus dem Datensatz ausgeführt.

[0020] Die Erfindung schafft ein Computersoftwareerzeugnis, umfassend ein computerlesbares Medium, in dem Computerprogrammanweisungen gespeichert sind, wobei diese Anweisungen, wenn sie von einem Computer gelesen werden, verursachen, dass der Computer ein Verfahren zum Steuern eines Datenspeicherzugriffs, welches durch ein Aufzeichnen von Zugriffen der Benutzer auf die Speicherelemente und ein Ableiten jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen ausgeführt wird, durchführt. Das Verfahren wird

ferner durch eine Zweiclusterbildung aus den Benutzern und den Speicherelementen ausgeführt, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird. Das Verfahren wird ferner, antwortend auf die Zweiclusterbildung, durch ein Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer ausgeführt.

[0021] Die Erfindung schafft eine Vorrichtung zum Steuern eines Datenspeicherzugriffs in einer Organisation mit Benutzern eines Dateisystems, umfassend ein Computersystem, das betriebsfähig ist, um die Schritte eines Aufzeichnens jeweiliger Zugriffe der Benutzer auf die Speicherelemente und eines Ableitens jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen, und einer Zweiclusterbildung aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird, durchzuführen. Das Computersystem ist betriebsfähig zum Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer antwortend auf die Zweiclusterbildung.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0022] Für ein besseres Verständnis der vorliegenden Erfindung wird, unter Verwendung von Beispielen, auf die detaillierte Beschreibung der Erfindung Bezug genommen, die in Verbindung mit den folgenden Zeichnungen gelesen werden soll, in denen gleiche Elemente mit gleichen Bezugsziffern versehen sind und in denen:

[0023] Fig. 1 ein Blockdiagramm eines Datenverarbeitungssystems ist, in dem Datenzugriffssteuerungsregelungen gemäß einem offenbarten Ausführungsbeispiel der Erfindung automatisch definiert und verwaltet werden;

[0024] Fig. 2 ein Blockdiagramm ist, das eine Sondierungsmaschine in dem System, das in Fig. 1 gezeigt ist, gemäß einem offenbarten Ausführungsbeispiel der Erfindung darstellt;

[0025] **Fig.** 3 ein Blockdiagramm ist, das eine weitere Version einer Sondierungsmaschine in dem System, das in Fig. 1 gezeigt ist, gemäß einem offenbarten Ausführungsbeispiel der Erfindung darstellt;

[0026] Fig. 4 ein Flussdiagramm ist, das ein Verfahren einer Benutzerclusterbildung gemäß einem offenbarten Ausführungsbeispiel der Erfindung beschreibt;

[0027] Fig. 5 ein Flussdiagramm ist, das ein Verfahren für eine Speicherelementclusterbildung gemäß einem offenbarten Ausführungsbeispiel der Erfindung beschreibt; und

[0028] Fig. 6A und Fig. 6B, auf die hierin zusammengenommen als Fig. 6 Bezug genommen ist, ein Flussdiagramm sind, das ein Verfahren einer halb automatischen Dateizugriffssteuerung gemäß einem offenbarten Ausführungsbeispiel der Erfindung darstellt.

DETAILLIERTE BESCHREIBUNG DER ERFINDUNG

[0029] In der folgenden Beschreibung sind zahlreiche spezifische Details dargelegt, um ein gründliches Verständnis der vorliegenden Erfindung zu schaffen. Für einen Fachmann ist jedoch offensichtlich, dass die vorliegende Erfindung ohne diese spezifischen Details praktiziert werden kann. In anderen Fällen sind wohlbekannte Schaltungen, eine Steuerlogik und die Details von Computerprogrammanweisungen für herkömmliche Algorithmen und Verfahren nicht im Detail gezeigt, um die vorliegende Erfindung nicht unnötig undeutlich zu machen.

[0030] Softwareprogrammcode, der Aspekte der vorliegenden Erfindung ausführt, ist typischerweise in einem permanenten Speicher, wie einem computerlesbaren Medium, gehalten. In einer Client-Server-Umgebung kann ein solcher Softwareprogrammcode auf einem Client oder einem Server gespeichert sein. Der Softwareprogrammcode kann auf einem beliebigen aus einer Anzahl bekannter Medien für eine Verwendung mit einem Datenverarbeitungssystem ausgeführt sein. Dies umfasst, ist jedoch nicht begrenzt auf, magnetische und optische Speichervorrichtungen, wie Plattenlaufwerke, Magnetbänder, Kompaktplatten (engl.: compact discs; CDs), digitale Videoplatten (engl.: digital video discs; DVDs) und Computeranweisungssignale, die in einem Übertragungsmedium mit oder ohne eine Trägerwelle, auf der die Signale moduliert sind, ausgeführt sind. Zum

Beispiel kann das Übertragungsmedium ein Kommunikationsnetz, wie das Internet, umfassen. Zusätzlich können, obwohl die Erfindung in einer Computersoftware ausgeführt sein kann, die Funktionen, die notwendig sind, um die Erfindung zu implementieren, alternativ zum Teil oder im Ganzen unter Verwendung von Hardwarekomponenten, wie einer anwendungsspezifischen integrierten Schaltung oder anderer Hardware, oder einer Kombination von Hardwarekomponenten und Software, ausgeführt sein.

Systemüberblick

[0031] Nun wird, zuwendend zu den Zeichnungen, anfangs auf Fig. 1, die ein Blockdiagramm eines Datenverarbeitungssystems **10** ist, in dem gemäß einem offenbarten Ausführungsbeispiel der Erfindung Datenzugriffssteuerungsregelungen automatisch definiert und verwaltet werden, Bezug genommen. Das System **10** kann als ein Allzweckcomputer oder als eine Mehrzahl von Computern, die in einem Netz, zum Beispiel dem Internet, miteinander verbunden sind, implementiert sein.

[0032] Ein organisationsweiter Datenspeicher, auf den durch das System **10** zugegriffen werden kann, ist durch ein Organisationsdateisystem **12** dargestellt. Das Organisationsdateisystem **12** kann eine oder mehrere zusammengestellte Speichereinheiten aufweisen, oder es kann ein geografisch verteiltes Datenspeicherungssystem sein, wie es in der Technik bekannt ist. Es gibt kein Erfordernis, dass einzelne Speichereinheiten des Organisationsdateisystems **12** die gleichen Fähigkeiten haben.

[0033] Auf das Organisationsdateisystem **12** kann durch eine beliebige Zahl von Benutzern **14** unter Verwendung einer grafischen Benutzerschnittstellenanwendung **16** (engl.: graphical user interface; GUI), die über eine Anwendungsprogrammierungsschnittstelle **18** (engl.: application programming interface; API) zu anderen Elementen des Systems **10** in Beziehung steht, zugegriffen werden. Die Benutzer **14** sind typischerweise Mitglieder der Organisation, können jedoch ebenfalls Außenstehende, wie Kunden, umfassen. Die grafische Benutzerschnittstellenanwendung **16** ist die Schnittstelle des Verwaltungssystems, durch das die Benutzer **14** die Ergebnisse ihrer tatsächlichen Nutzungsanalyse, wie sie durch eine Analysemaschine **20** bestimmt worden sind, empfangen können. In einigen Ausführungsbeispielen können ausreichend qualifizierte Benutzer, z. B. administratives Personal, ihren aktuellen Status ansehen, und können Änderungen, die durch das System empfohlen werden, ansehen. Solche Benutzer können autorisiert sein, empfohlene Änderungen anzunehmen oder abzulehnen. Vor einem Auswählen irgendwelcher empfohlener Änderungen haben qualifizierte Benutzer die Möglichkeit, die Wirkung der empfohlenen Änderungen auf das System anzusehen. Systemadministratoren können dann den Berechtigungssatz, der sich als am geeignetsten erweist, auswählen oder bestätigen.

[0034] Eine Sondierungsmaschine **22** ist entworfen, um auf eine fortdauernde Art und Weise Zugriffsinformationen aus dem Organisationsdateisystem **12** zu sammeln, doppelte oder redundante Informationseinheiten herauszufiltern und den resultierenden Informationsstrom in einer Datenbank **24** zu speichern. Die Sondierungsmaschine **22** wird ebenfalls benutzt, um die aktuelle Dateisicherheitsregelung der Organisation, die aktuelle Struktur des Organisationsdateisystems **12** und Informationen über die Benutzer **14** zu sammeln. Die Sondierungsmaschine **22** kann in verschiedenen Umgebungen und Architekturen implementiert sein.

[0035] Die Analysemaschine **20** ist ein spezialisiertes Modul, das das Herzstück der Möglichkeit des Systems, den Speicherzugriff zu steuern, ist. Die Analysemaschine **20** schlägt automatisch die Sicherheitsregelung der Organisation vor und überarbeitet diese. Das Vorderende (engl.: front end) für die Analysemaschine **20** ist ein Datensammler **26**, der die Speicherzugriffsaktivitäten in der Datenbank **24** effizient aufzeichnet. Die Ausgabe der Analysemaschine **20** kann unter Verwendung einer interaktiven administrativen Schnittstelle **28**, die es Systemadministratoren ermöglicht, Abfragen an den gesammelten Daten durchzuführen, weiter manipuliert werden. Unter Verwendung der administrativen Schnittstelle **28** können die Administratoren die automatisch vorgeschlagene Sicherheitsregelung, falls notwendig, modifizieren und schließlich die neue oder überarbeitete Regelung aktivieren.

[0036] In Beziehung zu der Analysemaschine **20** steht ein Festlegemodul **30**, das eine vorgeschlagene Sicherheitsregelung unter Verwendung von Daten, die vor ihrer Implementierung gesammelt wurden, verifiziert. Das Festlegemodul **30** nimmt auf eine Zugriffssteuerungsliste **32** (engl.: access control list; ACL) Bezug. Aktivitäten des Festlegemoduls **30** sind in weiteren Details im Folgenden beschrieben.

Sondierungsmaschine

[0037] Sondierungsmaschinen sind auf spezielle Betriebssysteme und Umgebungen zugeschnitten. Die folgenden sind auf beispielhafte und nicht auf begrenzende Weise beschrieben.

Win-Sondierungsarchitektur

[0038] Nun wird auf Fig. 2, die ein Blockdiagramm ist, das ein Ausführungsbeispiel der Sondierungsmaschine **22** (Fig. 1) gemäß einem offenbarten Ausführungsbeispiel der Erfindung darstellt, Bezug genommen. Dieses Ausführungsbeispiel, hierin als das „Win-Sondierungsmodul" bezeichnet, verhält sich als eine Sonde für die Microsoft-Windows®-Plattform. Es ist verantwortlich für ein Überwachen von lokalen Dateisystemen, die Komponenten des Organisationsdateisystems **12** (Fig. 1) sind, auf einer Betriebssystemebene. Typischerweise gibt es ein Win-Sondierungsmodul, das alle Windows-Computer in der Organisation bedient. Das Win-Sondierungsmodul arbeitet parallel mit Sondierungsmaschinen, die an andere Betriebssysteme angepasst sind. Alternativ kann eine komplexe Organisation mehr als ein Win-Sondierungsmodul erfordern, um einen effizienten Betrieb sicherzustellen. Das Win-Sondierungsmodul hat einen Dateisystemfilter **34** (engl.: SIDFILE = SID-DATEI), der einen Kernel-Modus-Filtertreiber **36** zum Abfangen einer Aktivität eines lokalen Dateisystems **38** und zum Loggen (engl.: Jogging) desselben zusammen mit Sicherheitsinformationen, die die abgefangene Aktivität betreffen, einsetzt. Ein Dienst **40** (engl.: SIDFILE_SERVICE = SID-DATEI_DIENST) wechselwirkt mit dem Filtertreiber **36** und ruft neue Log-Einträge ab. Die Log-Einträge werden durch den Dienst **40** gefiltert. Der Dienst **40** ist verantwortlich für das Zusammentragen von Statistiken aus den gefilterten Log-Einträgen und das Weiterleiten sowohl der Roh-Log-Einträge als auch ihrer Statistiken zu der Datenbank **24** (Fig. 1) zu einem weiteren Verarbeiten. Der Filter **34** ist für das Betriebssystem transparent, und sein Mehraufwand ist auf das Extrahieren zugeordneter Sicherheitsattribute pro Eingabe/Ausgabe-Operation (engl.: input/outout; I/O) und das Loggen begrenzt. Die Kommunikation zwischen dem Filtertreiber **36** und dem Dienst **40** wird unter Verwendung von Betriebssystemeinrichtungen, wie einer Vorrichtungs-I/O-Steuerung, und vordefinierter Steuercodes, z. B. „sammle Statistiken", erreicht.

Sondierungsarchitektur eines an einem Netz angeschlossenen Speichers

[0039] Nun wird auf **Fig.** 3 Bezug genommen, die ein Blockdiagramm. ist, das ein weiteres Ausführungsbeispiel der Sondierungsmaschine **22** (Fig. 1), das an vernetzte Vorrichtungen angepasst ist, gemäß einem offenbarten Ausführungsbeispiel der vorliegenden Erfindung darstellt. Eine Sonde **42** eines an einem Netz angeschlossenen Speichers (engl.: network attached storage; NAS) ist verantwortlich für ein Sammeln von Zugriffsdaten von einer NAS-Speichervorrichtung **44**. Bei einigen Ausführungsbeispielen kann eine NAS-Sonde eine gesamte Organisation bedienen. Alternativ kann eine Mehrzahl von NAS-Sonden vorgesehen sein. Die Sonde **42** wechselwirkt unter Verwendung eines zweckgebundenen, typischerweise händlerspezifischen Protokolls mit der NAS-Vorrichtung **44**. Das Protokoll verursacht, dass die NAS-Vorrichtung **44** bei einer angefragten Dateizugriffsoperation, die von einem Benutzer **48** ausgeht, eine Benachrichtigung **46** zu der Sonde **42** sendet. Die Sonde **42** ermöglicht, gemäß einer aktuell herrschenden Regelung, entweder, dass die Anfragen durch die NAS-Vorrichtung **44** erfüllt werden, oder sie verweigert einen Zugriff auf die NAS-Vorrichtung **44**. Durch die Sonde **42** wird ein Log-Eintrag **50**, der eine ermöglichte Anfrage dokumentiert, angefertigt, und die Anfrage wird zu der NAS-Vorrichtung **44** zu einem herkömmlichen Verarbeiten gemäß ihrem eigenen Betriebssystem weitergegeben. Bei einigen Ausführungsbeispielen wird eine verweigerte Anfrage einfach verworfen. Alternativ können verweigerte Anfragen geloggt werden, um beim Verfolgen eines abnormalen Benutzerverhaltens zu helfen. In jedem Fall empfängt der Benutzer **48** eine Antwort **52** auf seine Anfrage, entweder in der Form einer Zugriffsverweigerung oder als eine Anzeige des Resultats der angefragten Dateioperation durch die NAS-Vorrichtung **44**. In beiden Fällen liegt eine minimale Leistungsbeeinträchtigung vor. Da die NAS-Vorrichtung **44** ihr eigenes proprietäres Betriebssystem hat, werden alle treiberbezogenen Fragen, z. B. ein Extrahieren von Systemidentifizierern (engl.: system identifiers; SIDs), Benutzeridentifizierern (engl.: user identifiers; UIDs) und der Typ des Dateizugriffs, der angefragt worden ist, auf der NAS-Vorrichtung **44** gehandhabt und durch die Sonde **42** einfach geloggt.

Analysemaschine

[0040] Wie im Vorhergehenden bemerkt, ist die Analysemaschine **20** (Fig. 1) das Herzstück des Systems **10**. Die Statistiken der tatsächlichen Zugriffe der Benutzer **14**, umfassend jedes Mitglied einer Organisation, auf jedes der Datenspeicherelemente in dem Organisationsdateisystem **12**, die durch die Sondierungsmaschine **22** gemeldet worden sind, werden verwendet, um eine gleichzeitige automatische Zweiclusterbildung aus den Benutzern und den Datenspeicherungselementen durchzuführen. Die Zweiclusterbildung wird auf eine solche Art und Weise durchgeführt, dass Benutzer, die Mitglieder des gleichen Benutzerclusters sind, ein ähnliches Datenzugriffsprofil gemeinsam verwenden, und auf Datenspeicherelemente (Dateien oder Verzeichnisse), die Mitglieder des gleichen Datenclusters sind, in erster Linie durch Benutzer mit ähnlichen Zugriffsprofilen zugegriffen wird. Die Cluster liefern ein globales Bild der Organisationsstruktur. Die Analysemaschine **20** kann aus den Resultaten der Clusterbildung ebenfalls ein lokales Maß einer Ähnlichkeit zwischen den Benutzern und

ein lokales Maß einer Ähnlichkeit zwischen den Datenelementen, die zu dem gleichen Cluster gehören, entwickeln. Außerdem sagt das Clusterbildungsverfahren einen zukünftigen Datenspeicherzugriff durch die Organisationsmitglieder zuverlässig voraus. Es darf mit einem hohen Grad an Vertrauen angenommen werden, dass, falls einer der Benutzer **14** auf eine bestimmte Datei oder ein bestimmtes Speicherelement nicht zugegriffen hat, und ähnliche Benutzer nicht auf ähnliche Dateien zugegriffen haben, dann dieser eine Benutzer in naher Zukunft keine Zugriffsrechte auf das entsprechende Speicherelement benötigen wird. Die Analysemaschine **20** liefert so IT-Administratoren ein klares globales Bild von Informationsnutzungsmustern und kann detaillierte Empfehlungen für eine Optimierung einer Sicherheitsregelung anbieten. Zur gleichen Zeit werden Administratoren auf ein anormales Benutzerverhalten aufmerksam gemacht. Die Analysemaschine **20** kann ebenfalls automatisch einen vollständigen forensischen Pfad aller verdächtigen Aktivitäten aufbauen. Das Resultat ist eine auf dramatische Weise größere Fähigkeit, ein Einhalten der Zugriffs- und Privatsphärenschutzregelungen sicherzustellen, und eine angemessene Informationsnutzung sicherzustellen, ohne dass dem IT-Personal zusätzliche administrative Belastungen auferlegt werden.

Zweiclusterbildungsalgorithmus

[0041] Die folgenden Clusterbildungsalgorithmen werden in dem aktuellen Ausführungsbeispiel verwendet. Die Erfindung ist jedoch nicht auf die speziellen Algorithmen, die im Folgenden beschrieben werden, begrenzt. Es ist für Fachleute offensichtlich, dass andere Clusterbildungsalgorithmen auf die Daten, die durch die Sondierungsmaschine **22** (Fig. 1) erhalten worden sind, angewendet werden können, um vergleichbare Resultate zu erhalten.

[0042] Angenommen, es liegt eine gemeinsame Verteilung von zwei diskreten Zufallsvariablen, X und Y, bezeichnet durch p(x, y) = p(X = x, Y = y), vor. In dem vorliegenden Fall steht X für den Satz von Benutzern in der Organisation, und Y ist der Satz von Dateiverzeichnissen, auf die durch die Mitglieder der Organisation zugegriffen wird. Der Wert p(x, y) ist die normalisierte Anzahl der Male, die Benutzer x das Datenspeicherelement y in einer Eintragungsphase angesprochen hat. Basierend auf den gesammelten Daten, die in einer Nachbarschaftstabelle der p(x, y) angeordnet sind, soll die wesentliche grundlegende Struktur der zwei Sätze und die gegenseitigen Beziehungen zwischen ihnen entdeckt werden. Genauer gesagt sollen aus den Zufallsvariablen X und Y Cluster aus disjunkten Sätzen von ähnlichen Elementen gebildet werden. Eine Clusterbildung der Zufallsvariable X ist ein Aufteilen der Elemente von X in disjunkte Cluster, die durch X' bezeichnet werden, und auf eine ähnliche Weise wird eine Aufteilung von Y durch Y' bezeichnet.

[0043] Angenommen, dass die Anzahl der Cluster vordefiniert ist (als ein Teil der Systemkonfigurationsparameter), sollen Clusterbildungen X' und Y' gefunden werden, derart, dass die gegenseitige Information I(X'; Y') zwischen den Benutzerclustern und den Datenclustern maximiert wird. Mit anderen Worten, das System benutzt das Kriterium der gegenseitigen Information als eine Kostenfunktion, um die Qualität von verschiedenen Clusterbildungsstrukturen zu bewerten.

[0044] Die gegenseitige Information ist auf die folgende Weise definiert:

$$\mathrm{I}(X;Y) = -\sum_{x,y} \mathrm{p}(X=x, Y=y)\ \log \mathrm{p}(X=x, Y=y) \qquad (1).$$

[0045] Die gegenseitige Information enthält das Ausmaß der Unsicherheit in einer der Zufallsvariablen, das zum Vorschein kommt, wenn die andere Zufallsvariable beobachtet wird. Es werden ebenfalls zwei verwandte Konzepte, die im Folgenden verwendet werden, definiert. Seien P = (P(1), ..., P(n)) und Q = Q(Q(1), ..., Q(n)) zwei diskrete Wahrscheinlichkeitsverteilungen. Die relative Entropie (Kullback-Leibler-Divergenz) zwischen den Verteilungen P, Q ist:

$$\mathrm{KL}(P \,||\, Q) = \sum_{i} P(i)\ \log\left(P(i)/Q(i)\right) \qquad (2).$$

[0046] Die Jensen-Shannon-Divergenz zwischen den Verteilungen P, Q gemäß einem Mischungskoeffizienten c ist:

$$\mathrm{JS}(P, Q) = c\mathrm{KL}(P||cP + (1-c)Q) + (1-c)\mathrm{KL}(Q||cP + (1-c)Q) \qquad (3).$$

[0047] Der nächste Schritt besteht darin, das Kriterium der gegenseitigen Information zu benutzen, um die optimale Zweiclusterbildung zu finden. Unterschiedliche Strategien werden für den Benutzersatz X und den Datensatz Y verwendet. Im Fall des Benutzersatzes X gibt es keine aktuelle Struktur, für die es notwendig ist,

dass sie beibehalten wird. In einigen Ausführungsbeispielen kann es jedoch wünschenswert sein, eine Organisationsbenutzerstruktur zu behalten. Im Gegensatz dazu basiert das Datendateisystem auf einer Baumstruktur, die beibehalten werden soll, da sie wahrscheinlich eine Betriebsähnlichkeit zwischen nahegelegenen Verzeichnissen in dem Baum widerspiegelt. Daher wird die Clusterbildung der Speicherelemente im Wesentlichen durch ein Beschneiden des Baums erreicht. Das Verfahren ist im Folgenden detaillierter beschrieben.

Benutzerclusterbildung

[0048] Nun wird auf Fig. 4 Bezug genommen, die ein Flussdiagramm ist, das ein Verfahren einer Benutzerclusterbildung gemäß einem offenbarten Ausführungsbeispiel der Erfindung beschreibt. Das Verfahren beginnt mit einer zufälligen Lösung und verbessert dann das Resultat aufeinanderfolgend auf eine monotone Art und Weise.

[0049] Bei einem Anfangsschritt **54** wird ein zufälliges Aufteilen der Benutzerliste in eine vorbestimmte Anzahl von Clustern als ein Startpunkt gewählt. Dieses Aufteilen wird in einem aktuellen Satz von Zyklen wie im Folgenden beschrieben verwendet. Für jeden Benutzer x steht die Wahrscheinlichkeitsverteilung p(y|x) für die normalisierte Datenzugriffsaktivität des Benutzers x, d. h. p(y|x) ist die Anzahl der Male, die der Benutzer x auf das Datenelement y zugegriffen hat, normalisiert durch die gesamte Anzahl der Datenaktivitäten, die von x in dem Eintragungszeitraum durchgeführt worden sind. Für jeden zufällig aufgebauten Cluster C ist p(y|C) als der Mittelwert der bedingten Wahrscheinlichkeitsverteilungen p(y|x), die sich auf die Benutzer, die Mitglieder des Clusters C sind, beziehen, definiert.

[0050] Als Nächstes wird bei einem Schritt **56** einer der Cluster, die in dem Anfangsschritt **54** aufgestellt worden sind, zufällig ausgewählt.

[0051] Als Nächstes wird bei einem Schritt **58** einer der Benutzer ausgewählt. Der Schritt **58** wird iterativ durchgeführt, und die Benutzer werden zyklisch ausgewertet. Die Reihenfolge der Auswertung in einem Zyklus ist jedoch nicht entscheidend.

[0052] Als Nächstes wird bei einem Schritt **60** der aktuelle Benutzer x vorläufig von seinem aktuellen Cluster zu dem Cluster, der in Schritt **56** ausgewählt worden ist, bewegt, um eine vorläufige neue Clusterbildung der Benutzer zu bilden.

[0053] Die Steuerung schreitet nun zu einem Entscheidungsschritt **62** fort, bei dem bestimmt wird, ob die globale gegenseitige Information I(X; Y) der neuen Clusterbildung größer als die der aktuellen Clusterbildung ist. Ein Abstand zwischen einem Benutzer x und einem Cluster C, der aus c Benutzern zusammengesetzt ist, ist auf die folgende Weise definiert:

$$d(x, C) = (c + 1)JS(p(y|x), p(y|C))$$
$$= KL(p(y||x)||(p(y|x) + cp(y|C))/(c + 1)) +$$
$$c * KL(p(y|c)||(p(y|x) + cp(y|C))/(c + 1))) \quad (4).$$

[0054] Jeder Benutzer x wird mit dem Cluster C, der den Abstand d(x, C) minimiert, verschmolzen. Die bedingte Zugriffswahrscheinlichkeit p(y|C) wird gemäß der Statistik des neuen Mitglieds x modifiziert. Es lässt sich verifizieren, dass ein Minimieren des Abstands d(x|C) äquivalent zu einem Maximieren der gegenseitigen Information zwischen den Clustern und den Datenaktivitäten ist.

[0055] Falls die Bestimmung bei dem Entscheidungsschritt **62** bejahend ist, dann schreitet die Steuerung zu einem Schritt **64** fort. Der aktuelle Benutzer x bleibt in dem Cluster, der in Schritt **56** ausgewählt worden ist, und die vorläufige neue Clusterbildung, die in Schritt **60** aufgestellt worden ist, wird bestätigt.

[0056] Falls die Bestimmung bei dem Entscheidungsschritt Schritt **62** negativ ist, dann schreitet die Steuerung zu einem Schritt **66** fort. Der aktuelle Benutzer x wird zu dem Cluster, aus dem er ausgewählt worden ist, zurückgegeben, und der vorläufige neue Cluster, der in Schritt **60** aufgestellt worden ist, wird abgelehnt.

[0057] In beiden Fällen schreitet die Steuerung nun zu einem Entscheidungsschritt **68** fort, bei dem bestimmt wird, ob in dem aktuellen Zyklus noch weitere Benutzer ausgewertet werden müssen. Wenn die Bestimmung in dem Entscheidungsschritt **68** bejahend ist, dann kehrt die Steuerung zu Schritt **58** zurück.

[0058] Falls die Bestimmung bei dem Entscheidungsschritt **68** negativ ist, dann schreitet die Steuerung zu

einem Entscheidungsschritt **70** fort, bei dem bestimmt wird, ob der letzte Zyklus eine Verbesserung in der gegenseitigen Information ergeben hat.

[0059] Falls die Bestimmung bei dem Entscheidungsschritt **70** bejahend ist, dann ist möglicherweise noch keine optimale Clusterbildung erreicht worden. Bei einem Schritt **72** wird die Benutzerliste zurückgesetzt, um einen weiteren Zyklus in dem aktuellen Satz von Zyklen zu beginnen. Die Steuerung kehrt zu Schritt **56** zurück, und der neue Zyklus beginnt durch Wählen eines neuen Clusters, unter Verwendung des gleichen zufälligen Aufteilens, das in dem Anfangsschritt **54** aufgestellt worden ist.

[0060] Falls die Bestimmung bei dem Entscheidungsschritt **70** negativ ist, dann schreitet die Steuerung zu einem Schritt **74** fort. Die beste Clusterbildung, die in dem aktuellen Satz von Zyklen erreicht worden ist, wird gespeichert.

[0061] Die Steuerung schreitet nun zu einem Entscheidungsschritt **76** fort, bei dem bestimmt wird, ob ein Abbruchkriterium erfüllt ist. Das Abbruchkriterium kann ein Abschluss einer vorbestimmten Anzahl von Iterationen des Anfangsschritts **54** sein. Alternativ kann ein Leistungsanzeiger als ein Abbruchkriterium verwendet sein.

[0062] Falls die Bestimmung bei dem Entscheidungsschritt **76** negativ ist, dann kehrt die Steuerung zu dem Anfangsschritt **54** zurück, und das Verfahren wird wiederholt, wobei ein neuer Startpunkt gewählt wird.

[0063] Falls die Bestimmung bei dem Entscheidungsschritt **76** bejahend ist, dann schreitet die Steuerung zu einem abschließenden Schritt **78** fort. Das beste Resultat, das in den Clusterbildungen, die in den Iterationen des Schritts **74** gespeichert worden sind, erhalten worden ist, wird als eine abschließende Clusterbildung, die die gegenseitige Information zwischen den Benutzerclustern und den Datenclustern maximiert, gemeldet.

Datenelementclusterbildung

[0064] Nun wird auf Fig. 5 Bezug genommen, die ein Flussdiagramm ist, das ein Verfahren für eine Speicherelementclusterbildung gemäß einem offenbarten Ausführungsbeispiel der Erfindung beschreibt. Dies ist ein agglomerierendes Verfahren, das auf einem Verschmelzen von Clustern, die durch Geschwisterelemente in dem Datendateibaum dargestellt sind, basiert. Es wird angenommen, dass eine Benutzerclusterbildung, wie im Vorhergehenden unter Bezugnahme auf Fig. 4 beschrieben, durchgeführt worden ist. In einer Anfangsphase findet ein Verschmelzen zwischen Geschwisterverzeichnissen oder Eltern-Nachkommen-Verzeichnissen, die bezüglich Benutzerzugangsereignissen nicht unterschieden werden können, statt. Diese Stufe resultiert in einem Verzeichnisbaum, der auf eine handhabbare Anzahl von Elementen beschnitten worden ist. In der nächsten Phase werden alle Blätter des aktuellen beschnittenen Baums besucht, und es findet ein weiteres Verschmelzen zwischen zwei Geschwister- oder Eltern-Nachkommen-Verzeichnissen statt, derart, dass eine minimale Reduzierung in der gegenseitigen Information zwischen den Benutzerclustern und den Datenclustern resultiert. Dieses Verfahren wird iteriert, bis ein Abbruchkriterium erfüllt ist, z. B., wenn eine vorbestimmte Anzahl von Clustern erhalten worden ist, oder wenn die aktuelle gegenseitige Information unter eine vorbestimmte Schwelle gesenkt worden ist. Das Verfahren wird nun detaillierter dargestellt.

[0065] Ein Anfangsschritt **80** beginnt einen Durchlauf der Verzeichnisse des Dateibaums. Bei einem Auswählen von Kandidaten für eine Clusterbildung werden Eltern-Nachkommen-Verzeichnisse und Geschwisterverzeichnisse sowie Cluster derselben betrachtet, und auf sie wird zusammengenommen als „Nachbarn" Bezug genommen. Die Durchlaufreihenfolge ist nicht entscheidend, solange alle Datenelemente besucht werden und alle gegenseitigen Nachbarn ausgewertet werden. Viele bekannte Algorithmen für einen Durchlauf eines Baums können eingesetzt sein. Zwei Nachbarn werden ausgewählt.

[0066] Die Steuerung schreitet nun zu einem Entscheidungsschritt **82** fort, bei dem bestimmt wird, ob die aktuellen Kandidaten bezüglich Benutzerzugriffsereignissen ununterscheidbar, oder nahezu ununterscheidbar gemäß vorbestimmten Ähnlichkeitskriterien, sind.

[0067] Falls die Entscheidung in Schritt **82** bejahend ist, dann schreitet die Steuerung zu einem Schritt **84** fort. Die Kandidaten werden miteinander verschmolzen, um einen neuen Datencluster zu bilden. Dieser Datencluster wird in anschließenden Iterationen des Anfangsschritts **80** wie ein einziges Speicherelement oder ein einziger Nachbar behandelt.

[0068] Nach einem Durchführen des Schritts **84**, oder falls die Bestimmung bei dem Entscheidungsschritt **82** negativ ist, schreitet die Steuerung zu einem Entscheidungsschritt **86** fort, bei dem bestimmt wird, ob der

Durchlauf des Datendateibaums abgeschlossen ist. Falls die Bestimmung bei dem Entscheidungsschritt **86** bejahend ist, dann kehrt die Steuerung zu dem Anfangsschritt **80** zurück, um eine weitere Iteration zu beginnen.

[0069] Falls die Bestimmung bei dem Entscheidungsschritt **86** negativ ist, dann ist eine Phase des Verfahrens abgeschlossen, was in einem beschnittenen Verzeichnisbaum resultiert. Im Allgemeinen bilden die Verzeichnisse und Cluster von Verzeichnissen in dem beschnittenen Baum eine handhabbare Anzahl von Elementen.

[0070] Die Steuerung schreitet nun zu einem Schritt **88** fort, der eine weitere Phase des Verfahrens beginnt, in der der beschnittene Baum erneut durchlaufen wird, mit einem zusätzlichen Verschmelzen von Kandidaten in einer Weise, die zu einer minimalen Reduzierung der gegenseitigen Information I(X; Y) führt. Die gegenseitige Information I(X; Y) zwischen den Benutzerclustern, die aus dem Verfahren, das unter Bezugnahme auf Fig. 4 beschrieben worden ist, resultieren, und den Datenclustern des aktuellen beschnittenen Baums wird gespeichert.

[0071] Als Nächstes werden bei einem Schritt **90** zwei Kandidaten ausgewählt. Wie im Vorhergehenden festgestellt, können diese Kandidaten Cluster, Verzeichnisse oder Kombinationen derselben sein, solange die Kandidaten eine Geschwister- oder Eltern-Nachkommen-Beziehung haben.

[0072] Als Nächstes werden bei einem Schritt **92** die aktuellen Kandidaten vorläufig verschmolzen, um eine neue Clusterbildung der Benutzer und der Datenelemente zu bilden. Die gegenseitige Information I'(X; Y) der vorläufigen Anordnung wird bestimmt.

[0073] Die Steuerung schreitet nun zu einem Entscheidungsschritt **94** fort, bei dem bestimmt wird, ob die Reduzierung der gegenseitigen Information I'(X; Y) – I(X; Y), die durch die vorläufige Clusterbildung verursacht worden ist, geringer als die Reduzierung der gegenseitigen Information ist, die durch die beste vorhergehende vorläufige Clusterbildung verursacht worden ist. Diese Bestimmung ist in der ersten Iteration des Entscheidungsschritts **94** immer bejahend.

[0074] Falls die Bestimmung bei dem Entscheidungsschritt **94** bejahend ist, dann schreitet die Steuerung zu einem Schritt **96** fort. Die aktuelle vorläufige Clusterbildung wird gespeichert und als eine Höchststandsmarke gesetzt. Sie ist die beste neue Clusterbildung, die bislang verfügbar ist.

[0075] Nach dem Durchführen des Schritts **96**, oder falls die Bestimmung bei dem Entscheidungsschritt **94** negativ ist, schreitet die Steuerung zu einem Entscheidungsschritt **98** fort, bei dem bestimmt wird, ob noch weitere Kandidaten in dem Baum ausgewertet werden müssen. Falls die Bestimmung bei dem Entscheidungsschritt **98** bejahend ist, dann kehrt die Steuerung zu dem Schritt **90** zurück.

[0076] Falls die Bestimmung bei dem Entscheidungsschritt **98** negativ ist, dann schreitet die Steuerung zu einem Entscheidungsschritt **100** fort, bei dem bestimmt wird, ob ein Abbruchkriterium erfüllt ist. Dieses Kriterium kann die Aufstellung einer vorbestimmten Anzahl neuer Cluster sein. Alternativ kann das Verfahren abbrechen, wenn die aktuelle beste Reduzierung in der gegenseitigen Information weniger als eine vorbestimmte Schwelle ist.

[0077] Falls die Bestimmung bei dem Entscheidungsschritt **100** negativ ist, dann wird das Verfahren unter Verwendung der gegenseitigen Information der aktuellen besten Clusterbildung als einen Startpunkt wiederholt. Die Steuerung kehrt zu dem Schritt **88** zurück, bei dem ein neuer Wert der gegenseitigen Information I(X; Y) gesetzt wird.

[0078] Falls die Bestimmung bei dem Entscheidungsschritt **100** bejahend ist, dann schreitet die Steuerung mit einem abschließenden Schritt **102** fort. Die Clusterbildung, die zuletzt bei dem Schritt **96** abgespeichert worden ist, wird als eine optimale Datenelementclusterbildung gemeldet.

[0079] Am Ende des Clusterbildungsalgorithmus sind sowohl die Benutzer als auch die Datenspeicherelemente in disjunkten Clustern angeordnet. Eine hierarchische Baumstruktur ist unter den Datenspeicherelementen beibehalten, während die Benutzer ohne eine hierarchische Anordnung auf einen Benutzerraum verteilt sind. Ein stabiles Ähnlichkeitsmaß zwischen den Benutzern in der Organisation kann dann extrahiert werden. Man sagt, dass sich Benutzer ähnlich verhalten, wenn sie zu dem gleichen Benutzercluster gehören, was anzeigt, dass die zwei Benutzer auf ähnliche Teile des Datenspeichersystems zugreifen. Zwei Verzeichnisse oder andere Speicherelemente werden als ähnlich betrachtet, wenn sie zu dem gleichen Datencluster gehö-

ren.

Speicherzugriffssteuerung

[0080] Die Clusterbildung, die unter Verwendung des im Vorhergehenden, unter Bezugnahme auf Fig. 5 beschriebenen Verfahrens erhalten worden ist, kann verwendet werden, um automatisch unnötige Zugriffsberechtigungen zu eliminieren. Zum Beispiel wird die Berechtigung für einen Benutzer x, auf ein Speicherelement y zuzugreifen, eliminiert, wenn der Benutzer während eines Eintragungszeitraums auf das Element y (oder Elemente, die zu y ähnlich sind) nicht zugegriffen hat. Es wird vorausgesagt, dass der Benutzer x in der nahen Zukunft nicht auf das Element y zugreifen muss. Die Voraussage basiert auf dem Zugriffsprofil ähnlicher Mitglieder der Organisation. Es darf angenommen werden, dass, falls keine Benutzer mit einem ähnlichen Zugriffsprofil auf das Element y, die daher in dem gleichen Cluster wie der Benutzer x sind, auf das Element y zugegriffen haben, noch auf Speicherelemente ähnlich zu dem Element y zugegriffen haben, der Benutzer x dann in der nahen Zukunft nicht auf das Element y zugreifen wird. Daher kann, um den Grad der Organisationsdatensicherheit zu erhöhen, die Zugriffsberechtigung für den Benutzer x hinsichtlich des Elements y aufgehoben werden. Eine Überprüfung der Benutzer wird bei vorbestimmten Intervallen iterativ durchgeführt, und die Zugriffsregelung wird dementsprechend aktualisiert.

Halb automatische Clusterbildung

[0081] In dem vorhergehenden Abschnitt wurde eine Beschreibung geliefert, wie der Benutzer-Daten-Clusterbildungsansatz benutzt werden kann, um eine Zugriffssteuerungsregelung zu definieren, die die tatsächliche Struktur der Organisation widerspiegelt. Aufgezeichnete Datenaktivitäten sind lediglich eine der Informationsquellen, die extrahiert werden können, um die optimale Datenzugriffssteuerungsregelung zu definieren. Um eine neue oder aktualisierte Datenzugriffsregelung vorzuschlagen, sollten die aktuelle Benutzer-Daten-Gruppenstruktur und die aktuelle Datensicherheitsregelung ebenfalls in Betracht gezogen werden. Eine weitere wesentliche Quelle von Wissen über die Organisation ist die aktuelle (manuell gesetzte) Zugriffssteuerungsliste **32** (Fig. 1). Die ACL kann als ein Satz von Paaren angesehen werden, wobei jedes Paar aus einer Gruppe von Benutzern und einer Gruppe von Datenelementen, auf die durch die Benutzergruppe zugegriffen werden kann, besteht. Obwohl die aktuelle ACL Fehler enthalten kann, ist es vernünftig, anzunehmen, dass sie dennoch mit der gewünschten Steuerungsregelung hoch korreliert ist. Das Verfahren, das im Folgenden dargestellt wird, kann das unüberwachte Clusterbildungsverfahren, das im Vorhergehenden erläutert worden ist, verwenden, um die aktuelle ACL zu modifizieren und dadurch eine verbesserte Regelung zu erhalten. Die Organisationsstruktur, die aus den aufgezeichneten Benutzerzugriffsdaten gelernt worden ist, wird dann verwendet, um unnötige Datenzugriffsberechtigungen zu eliminieren. Der Algorithmus basiert auf der aktuellen ACL und funktioniert getrennt für jede Benutzer-Daten-Gruppe auf die folgende Weise: Es wird für jeden Benutzer geprüft, ob ein Zugriff auf eines der Datenelemente, die in dem Paar definiert sind, aufgezeichnet worden ist. Falls nicht, wird geprüft, ob ein ähnlicher Benutzer während des Eintragungszeitraums auf das Datenelement zugegriffen hat. Hier hat eine Ähnlichkeit die gleiche Bedeutung wie im Vorhergehenden angegeben. Falls kein solcher Benutzer gefunden worden ist, kann geschlossen werden, dass der bestimmte Benutzer in der nahen Zukunft nicht auf das Datenelement zugreifen muss. Wenn dies für die Datenelemente, die in der Datengruppe erscheinen, ebenfalls der Fall ist, wird der Benutzer aus dem Zugriffssteuerungspaar eliminiert. Eine zweite Phase des Verfahrens wird, wie im Folgenden erklärt, angewandt, um Datenelemente aus dem Zugriffssteuerungspaar zu eliminieren.

[0082] Nun wird auf Fig. 6 Bezug genommen, die ein Flussdiagramm ist, das ein Verfahren für eine teilweise überwachte Dateizugriffssteuerung gemäß einem offenbarten Ausführungsbeispiel der vorliegenden Erfindung darstellt. Die Schritte des Verfahrens sind in Fig. 6 für eine Klarheit der Darstellung in einer beispielhaften Folge gezeigt. Es ist jedoch für Fachleute offensichtlich, dass viele von ihnen parallel, asynchron oder in unterschiedlichen Reihenfolgen durchgeführt werden können.

[0083] Das Verfahren beginnt bei einem Anfangsschritt **104**. Die Zweiclusterbildungsverfahren, die im Vorhergehenden unter Bezugnahme auf Fig. 4 und Fig. 5 beschrieben worden sind, werden durchgeführt und angewendet.

[0084] Als Nächstes wird bei einem Schritt **106** eine Zugriffssteuerungseinheit aus der ACL ausgewählt. Diese Einheit ist ein Paar, das aus einer Benutzergruppe und einer Verzeichnisgruppe zusammengesetzt ist.

[0085] Als Nächstes wird bei einem Schritt **108** ein Benutzer aus den Benutzern der aktuellen Zugriffssteuerungseinheit gewählt.

[0086] Als Nächstes wird bei einem Schritt **110** ein Datenelement aus der aktuellen Zugriffssteuerungseinheit gewählt.

[0087] Die Steuerung schreitet nun zu einem Entscheidungsschritt **112** fort, bei dem bestimmt wird, ob der aktuelle Benutzer auf das aktuelle Datenelement zugegriffen hat.

[0088] Falls die Bestimmung bei dem Entscheidungsschritt **112** bejahend ist, dann muss keine Modifikation der ACL hinsichtlich des aktuellen Benutzers vorgenommen werden. Die Steuerung schreitet zu einem Schritt **114** fort, der im Folgenden beschrieben wird.

[0089] Falls die Bestimmung bei dem Entscheidungsschritt **112** negativ ist, dann werden Benutzer, die (in dem Clusterbildungsverfahren, das in dem Anfangsschritt **104** durchgeführt worden ist) als ähnlich zu dem aktuellen Benutzer bestimmt worden sind, ausgewertet. Die Steuerung schreitet zu einem Schritt **116** fort. Ein ähnlicher Benutzer wird ausgewählt.

[0090] Die Steuerung schreitet nun zu einem Entscheidungsschritt **118** fort, bei dem bestimmt wird, ob der aktuelle ähnliche Benutzer auf das aktuelle Datenelement zugegriffen hat.

[0091] Falls die Bestimmung bei dem Entscheidungsschritt **118** bejahend ist, dann muss, basierend auf einer Ähnlichkeit der Zugriffsbedürfnisse zwischen dem aktuellen Benutzer und dem aktuellen ähnlichen Benutzer, keine Modifikation der ACL hinsichtlich des aktuellen Benutzers vorgenommen werden. Die Steuerung schreitet zu dem Schritt **114** fort.

[0092] Falls die Bestimmung bei dem Entscheidungsschritt **118** negativ ist, dann wird bei einem Entscheidungsschritt **120** bestimmt, ob weitere ähnliche Benutzer vorhanden sind, die betrachtet werden müssen.

[0093] Falls die Bestimmung bei dem Entscheidungsschritt **120** bejahend ist, dann kehrt die Steuerung zu dem Schritt **116** zurück.

[0094] Falls die Bestimmung bei dem Schritt **120** negativ ist, dann wird bei einem Schritt **122** der aktuelle Benutzer aus der aktuellen Zugriffssteuerungseinheit entfernt.

[0095] Als Nächstes wird bei einem Entscheidungsschritt **124** bestimmt, ob noch weitere Benutzer in der aktuellen Zugriffssteuerungseinheit ausgewertet werden müssen. Falls die Bestimmung bei dem Entscheidungsschritt **124** bejahend ist, dann kehrt die Steuerung zu dem Schritt **108** zurück.

[0096] Falls die Bestimmung bei dem Entscheidungsschritt **124** negativ ist, dann wird bei einem Entscheidungsschritt **126** bestimmt, ob noch weitere Zugriffssteuerungseinheiten ausgewertet werden müssen. Falls die Bestimmung bei dem Entscheidungsschritt **126** bejahend ist, dann kehrt die Steuerung zu dem Schritt **106** zurück, um eine neue Iteration zu beginnen.

[0097] Falls die Bestimmung bei dem Entscheidungsschritt **126** negativ ist, dann schreitet die Steuerung zu einem abschließenden Schritt **128** fort. Die Speicherzugriffssteuerung kann nun die ACL-Liste so, wie sie modifiziert worden ist, aufnehmen.

[0098] Der Schritt **114**, auf den im Vorhergehenden Bezug genommen worden ist, beginnt eine neue Phase des Algorithmus, die den Status des aktuellen Datenelements in der aktuellen Zugriffssteuerungseinheit betrifft. Diese Phase wird lediglich durchgeführt, wenn weder der aktuelle Benutzer noch ein ähnlicher Benutzer auf das aktuelle Datenelement zugegriffen hat. Der Zweck der folgenden Schritte besteht darin, zu untersuchen, ob durch einen der Benutzer in der aktuellen Zugriffssteuerungseinheit auf Datenelemente, die (gemäß dem Clusterbildungsverfahren, das in dem Anfangsschritt **104** durchgeführt worden ist) als ähnlich zu dem aktuellen Datenelement betrachtet werden, zugegriffen worden ist. Wenn nicht, dann wird das aktuelle Datenelement aus der aktuellen Zugriffssteuerungseinheit entfernt. Ist diese Handlung einmal erfolgt, kann danach kein Mitglied der aktuellen Benutzergruppe auf das aktuelle Datenelement zugreifen. Ein ähnliches Datenelement wird aus der Clusterbildung, die in dem Anfangsschritt **104** durchgeführt worden ist, ausgewählt.

[0099] Als Nächstes wird in einem Schritt **130** erneut ein Benutzer aus der aktuellen Zugriffssteuerungseinheit ausgewählt. Es ist beabsichtigt, dass alle Benutzer in der aktuellen Zugriffssteuerungseinheit in Iterationen des Schritts **130** einer Auswertung unterzogen werden.

[0100] Die Steuerung fährt nun mit einem Entscheidungsschritt **132** fort, bei dem bestimmt wird, ob der aktuelle Benutzer auf das aktuelle ähnliche Datenelement zugegriffen hat. Falls die Bestimmung bei dem Entscheidungsschritt **132** bejahend ist, dann muss das aktuelle Datenelement nicht aus seiner Zugriffssteuerungseinheit entfernt werden. Die Steuerung schreitet zu dem Entscheidungsschritt **124** fort, der im Vorhergehenden beschrieben worden ist.

[0101] Falls die Bestimmung bei dem Entscheidungsschritt **132** negativ ist, dann wird bei einem Entscheidungsschritt **134** bestimmt, ob sich weitere Benutzer in der aktuellen Zugriffssteuerungseinheit befinden. Falls die Bestimmung bei dem Schritt **134** bejahend ist, dann kehrt die Steuerung zu dem Schritt **130** zurück.

[0102] Falls die Bestimmung bei dem Schritt **134** negativ ist, dann wird bei einem Entscheidungsschritt **136** bestimmt, ob weitere ähnliche Datenelemente vorhanden sind, die gegen die Benutzer in der aktuellen Zugriffssteuerungseinheit getestet werden müssen.

[0103] Falls die Bestimmung bei dem Entscheidungsschritt **136** bejahend ist, dann kehrt die Steuerung zu dem Schritt **114** zurück.

[0104] Falls die Bestimmung bei dem Entscheidungsschritt **136** negativ ist, dann sind alle Benutzer der aktuellen Zugriffssteuerungseinheit gegen alle Datenelemente, die ähnlich zu dem aktuellen Datenelement (das in der letzten Iteration des Schritts **110** gewählt worden ist) sind, auf einen Zugriff getestet worden. Kein Zugriff ist gefunden worden. Bei einem Schritt **137** wird das aktuelle Datenelement nun aus der aktuellen Zugriffssteuerungseinheit eliminiert.

[0105] Die Steuerung schreitet nun zu einem Entscheidungsschritt **138** fort, in dem bestimmt wird, ob sich weitere Datenelemente in der aktuellen Zugriffssteuerungseinheit befinden. Falls die Bestimmung bei dem Entscheidungsschritt **138** bejahend ist, dann kehrt die Steuerung zu dem Schritt **110** zurück, um eine neue Iteration, unter Verwendung eines unterschiedlichen Datenelements aus der aktuellen Zugriffssteuerungseinheit, zu beginnen.

[0106] Falls die Bestimmung bei dem Entscheidungsschritt **138** negativ ist, dann schreitet die Steuerung zu dem Entscheidungsschritt **124** fort, der im Vorhergehenden beschrieben worden ist.

Virtuelles Festlegen zum Verifizieren einer vorgeschlagenen Regelung

[0107] Erneut Bezug nehmend auf **Fig. 1**, werden die Clusterbildungsverfahren, die im Vorhergehenden beschrieben worden sind, auf die Speicherzugriffsaktivitäten, die während eines Eintragungs- oder Trainingszeitraums für das System gesammelt worden sind, angewandt. Diese Verfahren können von Zeit zu Zeit wiederholt werden, zum Beispiel nach Zusammenschlüssen und Erwerbungen in der zugrunde liegenden Organisation. Es ist wünschenswert, sicherzustellen, dass eine vorgeschlagene oder vorläufige neue oder aktualisierte Zugriffssteuerungsregelung bezüglich einer Benutzeraktivität, die nach dem Eintragungszeitraum auftritt, bestätigt ist. Daten, die nach dem Eintragungszeitraum gesammelt worden sind, werden verwendet, um die Bestätigung der vorläufigen Regelung vor ihrer Einrichtung zu verifizieren. Diese Funktion wird durch das Festlegemodul **30** ausgeführt, das Benutzerzugriffsaktivitäten aufzeichnet und Verletzungen der vorläufigen Regelung erfasst. Falls die Benutzeraktivitäten die vorläufige Regelung nicht verletzen, dann wird sie als eine endgültige Speicherzugriffssteuerungsregelung zugelassen. Ansonsten wird sie abgelehnt oder zu einer weiteren Auswertung oder Überarbeitung zurückgegeben. Das Festlegemodul **30** liefert so einen Querbestätigungsmechanismus, um die Qualität einer vorgeschlagenen Speicherzugriffssteuerungsregel vor ihrer tatsächlichen Implementierung zu überprüfen.

Verfolgen eines abnormalen Verhaltens

[0108] Ein weiterer wesentlicher Aspekt der Datenanalyse, die auf den aufgezeichneten Daten durchgeführt wird, ist eine Erfassung und ein Verfolgen eines abnormalen Verhaltens. Das Festlegemodul **30** ist angepasst, um diese Funktion folgend der Implementierung einer Speicherzugriffssteuerung durchzuführen. Ein abnormales Verhalten kann identifiziert werden, wenn sich ein Benutzer nicht übereinstimmend mit anderen Benutzern, die zu dem gleichen Benutzercluster gehören, verhält.

[0109] Es ist für Fachleute offensichtlich, dass die vorliegende Erfindung nicht auf das, was im Vorhergehenden im Einzelnen gezeigt und beschrieben worden ist, begrenzt ist. Stattdessen umfasst der Schutzbereich der vorliegenden Erfindung sowohl Kombinationen als auch Unterkombinationen der verschiedenen Merkma-

le, die im Vorhergehenden beschrieben worden sind, ebenso wie Variationen und Modifikationen derselben, die nicht in dem Stand der Technik enthalten sind, und die Fachleuten nach einem Lesen der vorangegangenen Beschreibung einfallen würden.

Zusammenfassung

Automatische Verwaltung einer Speicherzugriffssteuerung

[0110] Verfahren und Systeme zum Definieren und Erzeugen einer automatischen Dateisicherheitsregelung und eines halb automatischen Verfahrens zum Verwalten einer Dateizugriffssteuerung in Organisationen mit mehreren vielfältigen Zugriffssteuerungsmodellen und mit mehreren vielfältigen Dateiserverprotokollen werden geschaffen. Das System überwacht einen Zugriff auf Speicherelemente innerhalb des Netzes. Der aufgezeichnete Datenverkehr wird analysiert, um gleichzeitige Datenzugriffsgruppierungen und Benutzergruppierungen, die die tatsächliche Organisationsstruktur widerspiegeln, zu bewerten. Die gelernte Struktur wird dann in eine dynamische Dateisicherheitsregelung, die mit der Zeit ständig an Organisationsänderungen angepasst wird, umgewandelt. Das System schafft eine Entscheidungshilfeschnittstelle für eine interaktive Verwaltung der Dateizugriffssteuerung und für ein Verfolgen eines abnormalen Benutzerverhaltens.

Patentansprüche

1. Verfahren zum Steuern eines Datenspeicherzugriffs in einer Organisation mit Benutzern eines Dateisystems, wobei das Dateisystem Speicherelemente hat, mit folgenden Schritten:
Aufzeichnen von Zugriffen der Benutzer auf die Speicherelemente und Ableiten jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen;
Zweiclusterbildung aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird; und
Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer antwortend auf den Schritt der Zweiclusterbildung.

2. Verfahren nach Anspruch 1, bei dem die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch einen der Benutzer lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch den einen Benutzer zugegriffen worden ist.

3. Verfahren nach Anspruch 1, bei dem die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch die Benutzer eines der Benutzercluster lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch mindestens einen der Benutzer des einen Benutzerclusters zugegriffen worden ist.

4. Verfahren nach Anspruch 1, das ferner den Schritt eines Ableitens einer Struktur des Dateisystems antwortend auf den Schritt der Zweiclusterbildung aufweist.

5. Verfahren nach Anspruch 1, das ferner den Schritt eines Ableitens von Nutzungsmustern des Dateisystems durch die Benutzer antwortend auf den Schritt der Zweiclusterbildung aufweist.

6. Verfahren nach Anspruch 5, das ferner den Schritt eines Erfassens abweichender der Nutzungsmuster aufweist.

7. Verfahren nach Anspruch 1, bei dem der Schritt der Zweiclusterbildung iterativ durchgeführt wird, wobei die Zugriffsprofile bei jeder Iteration davon neu bestimmt werden, und die Steuerungsregelung folgend jeder der Iterationen aktualisiert wird.

8. Verfahren nach Anspruch 1, bei dem der Schritt des Definierens einer Steuerungsregelung die folgenden Schritte aufweist:
Vorschlagen einer vorläufigen Version der Steuerungsregelung;
Überwachen anschließender Zugriffe auf die Speicherelemente durch die Benutzer;
Bestimmen, dass die anschließenden Zugriffe in Übereinstimmung mit der vorläufigen Version der Steuerungsregelung sind; und
Zulassen der vorläufigen Version als eine endgültige Version der Steuerungsregelung antwortend auf den

Schritt des Bestimmens.

9. Verfahren nach Anspruch 1, das ferner den Schritt eines interaktiven Modifizierens der Steuerungsregelung aufweist.

10. Verfahren nach einem der Ansprüche 1-10, bei dem der Schritt des Definierens einer Steuerungsregelung automatisch und im Wesentlichen ohne menschlichen Eingriff durchgeführt wird.

11. Verfahren nach einem der Ansprüche 1-10, das ferner die folgenden Schritte aufweist:
Bezug nehmen auf eine Zugriffssteuerungsliste mit mindestens einem Benutzersatz der Benutzer und mindestens einem Datensatz der Speicherelemente, wobei die Benutzer des Benutzersatzes in jeweiligen der Benutzercluster umfasst sind, und die Speicherelemente des Datensatzes in jeweiligen der Datencluster umfasst sind;
Erfassen einer Abwesenheit von Zugriffen durch Mitglieder des jeweiligen Benutzerclusters auf Mitglieder des jeweiligen Datenclusters; und
Entfernen mindestens eines Teils der Benutzer aus dem Benutzersatz und Entfernen mindestens eines Teils der Speicherelemente aus dem Datensatz antwortend auf den Schritt des Erfassens.

12. Computersoftwareerzeugnis, das ein computerlesbares Medium, in dem Computerprogrammanweisungen abgespeichert sind, umfasst, wobei die Anweisungen, wenn sie von einem Computer gelesen werden, verursachen, dass der Computer ein Verfahren zum Steuern eines Datenspeicherzugriffs in einer Organisation mit Benutzern eines Dateisystems, wobei das Dateisystem Speicherelemente aufweist, durchführt, mit folgenden Schritten:
Aufzeichnen von Zugriffen der Benutzer auf die Speicherelemente und Ableiten jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen;
Zweiclusterbildung aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird; und
Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer antwortend auf den Schritt der Zweiclusterbildung.

13. Computersoftwareerzeugnis nach Anspruch 12, bei dem die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch einen der Benutzer lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch den einen Benutzer zugegriffen worden ist.

14. Computersoftwareerzeugnis nach Anspruch 12, bei dem die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch die Benutzer eines der Benutzercluster lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch mindestens einen der Benutzer des einen Benutzerclusters zugegriffen worden ist.

15. Computersoftwareerzeugnis nach Anspruch 12, bei dem der Schritt der Zweiclusterbildung iterativ durchgeführt wird, wobei die Zugriffsprofile bei jeder Iteration desselben neu bestimmt werden, und die Steuerungsregelung folgend jeder der Iterationen aktualisiert wird.

16. Computersoftwareerzeugnis nach Anspruch 12, bei dem der Schritt des Definierens einer Steuerungsregelung die folgenden Schritte aufweist:
Vorschlagen einer vorläufigen Version der Steuerungsregelung;
Überwachen anschließender Zugriffe auf die Speicherelemente durch die Benutzer;
Bestimmen, dass die anschließenden Zugriffe in Übereinstimmung mit der vorläufigen Version der Steuerungsregelung sind; und
Zulassen der vorläufigen Version als eine endgültige Version der Steuerungsregelung antwortend auf den Schritt des Bestimmens.

17. Computersoftwareerzeugnis nach einem der Ansprüche 12-16, das ferner die folgenden Schritte aufweist:
Bezug nehmen auf eine Zugriffssteuerungsliste mit mindestens einem Benutzersatz der Benutzer und mindestens einem Datensatz der Speicherelemente, wobei die Benutzer des Benutzersatzes in jeweiligen der Benutzercluster umfasst sind, und die Speicherelemente des Datensatzes in jeweiligen der Datencluster umfasst sind;

Erfassen einer Abwesenheit von Zugriffen durch Mitglieder des jeweiligen Benutzerclusters auf Mitglieder des jeweiligen Datenclusters; und
Entfernen mindestens eines Teils der Benutzer aus dem Benutzersatz und Entfernen mindestens eines Teils der Speicherelemente aus dem Datensatz antwortend auf den Schritt des Erfassens.

18. Vorrichtung zum Steuern eines Datenspeicherzugriffs in einer Organisation mit Benutzern eines Dateisystems, wobei das Dateisystem Speicherelemente aufweist, mit einem Computersystem, das betriebsfähig ist, um die folgenden Schritte durchzuführen:
Aufzeichnen von Zugriffen der Benutzer auf die Speicherelemente und Ableiten jeweiliger Zugriffsprofile aus den aufgezeichneten Zugriffen;
Zweiclusterbildung aus den Benutzern und den Speicherelementen, um jeweils Benutzercluster und Datencluster zu definieren, wobei die Zugriffsprofile der Benutzer in den Benutzerclustern gegenseitig ähnlich sind, und auf die Speicherelemente in den Datenclustern lediglich durch die Benutzer mit den gegenseitig ähnlichen Zugriffsprofilen zugegriffen wird; und
Definieren einer Steuerungsregelung für einen Zugriff auf die Speicherelemente durch die Benutzer antwortend auf den Schritt der Zweiclusterbildung.

19. Vorrichtung nach Anspruch 18, bei der die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch einen der Benutzer lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch den einen Benutzer zugegriffen worden ist.

20. Vorrichtung nach Anspruch 18, bei der die Steuerungsregelung einen Zugriff auf die Speicherelemente eines der Datencluster durch die Benutzer eines der Benutzercluster lediglich erlaubt, wenn auf mindestens eines der Speicherelemente in dem einen Datencluster durch mindestens einen der Benutzer des einen Benutzerclusters zugegriffen worden ist.

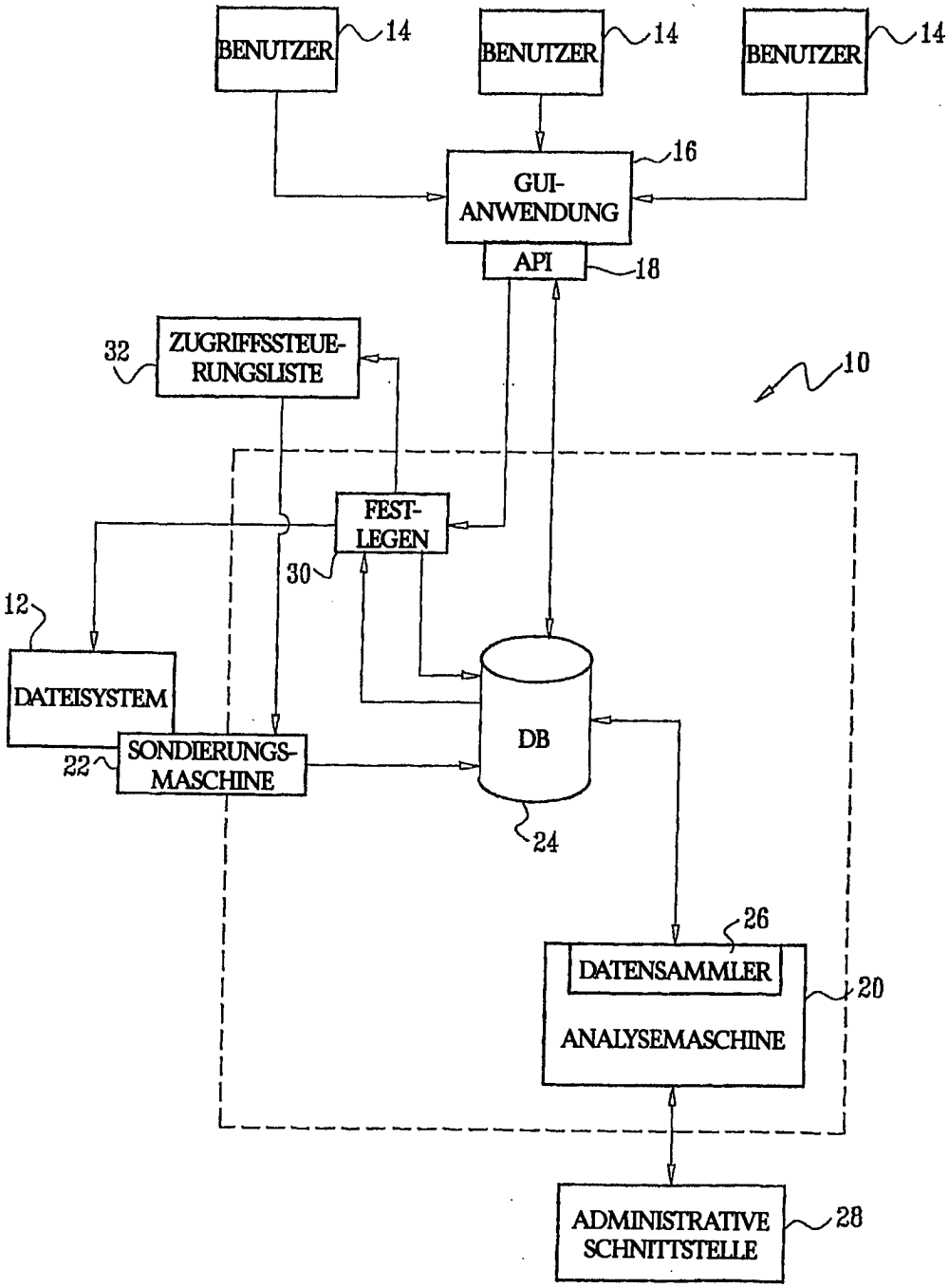
21. Vorrichtung nach Anspruch 18, bei der der Schritt des Definierens einer Steuerungsregelung die folgenden Schritte aufweist:
Vorschlagen einer vorläufigen Version der Steuerungsregelung;
Überwachen anschließender Zugriffe auf das Speicherelement durch die Benutzer;
Bestimmen, dass die anschließenden Zugriffe in Übereinstimmung mit der vorläufigen Version der Steuerungsregelung sind; und
Zulassen der vorläufigen Version als eine endgültige Version der Steuerungsregelung antwortend auf den Schritt des Bestimmens.

22. Vorrichtung nach einem der Ansprüche 18-21, bei der der Schritt des Definierens einer Steuerungsregelung die folgenden Schritte aufweist:
Bezug nehmen auf eine Zugriffssteuerungsliste mit mindestens einem Benutzersatz der Benutzer und mindestens einem Datensatz der Speicherelemente, wobei die Benutzer des Benutzersatzes in jeweiligen der Benutzercluster umfasst sind, und die Speicherelemente des Datensatzes in jeweiligen der Datencluster umfasst sind;
Erfassen einer Abwesenheit von Zugriffen durch Mitglieder des jeweiligen Benutzerclusters auf Mitglieder des jeweiligen Datenclusters; und
Entfernen mindestens eines Teils der Benutzer aus dem Benutzersatz und Entfernen mindestens eines Teils der Speicherelemente aus dem Datensatz antwortend auf den Schritt des Erfassens.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

FIG. 1



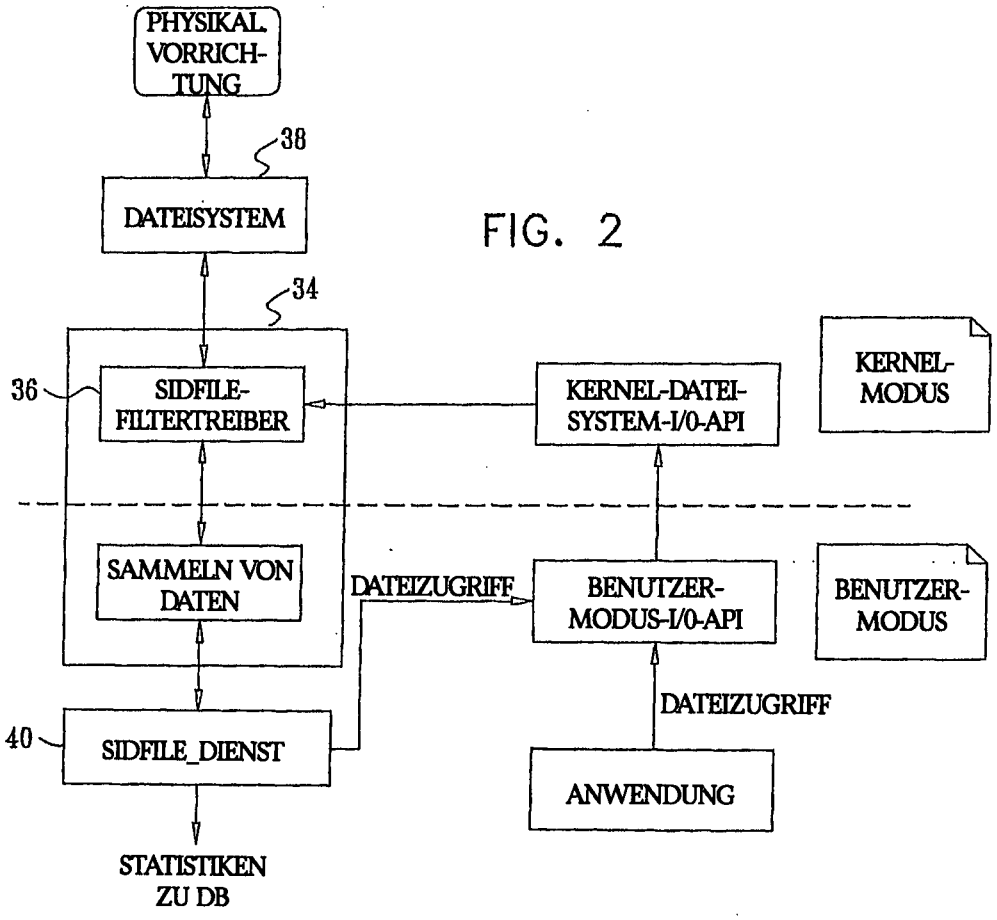
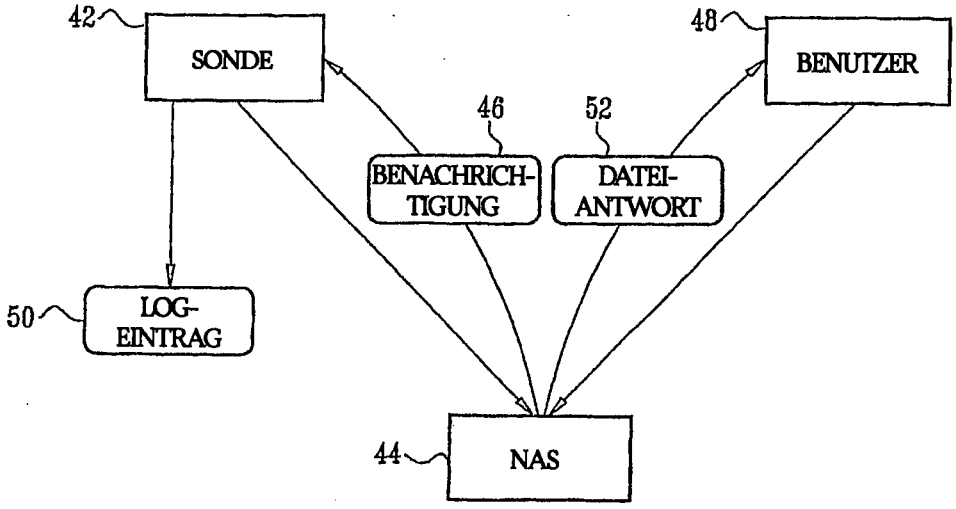

PHYSIKAL. VORRICH-TUNG
38
DATEISYSTEM
34
36
SIDFILE-FILTERTREIBER
KERNEL-DATEI-SYSTEM-I/0-API
KERNEL-MODUS
SAMMELN VON DATEN
DATEIZUGRIFF
BENUTZER-MODUS-I/0-API
BENUTZER-MODUS
DATEIZUGRIFF
40
SIDFILE_DIENST
ANWENDUNG
STATISTIKEN ZU DB

FIG. 2

FIG. 3


42
SONDE
48
BENUTZER
46
52
BENACHRICH-TIGUNG
DATEI-ANTWORT
50
LOG-EINTRAG
44
NAS

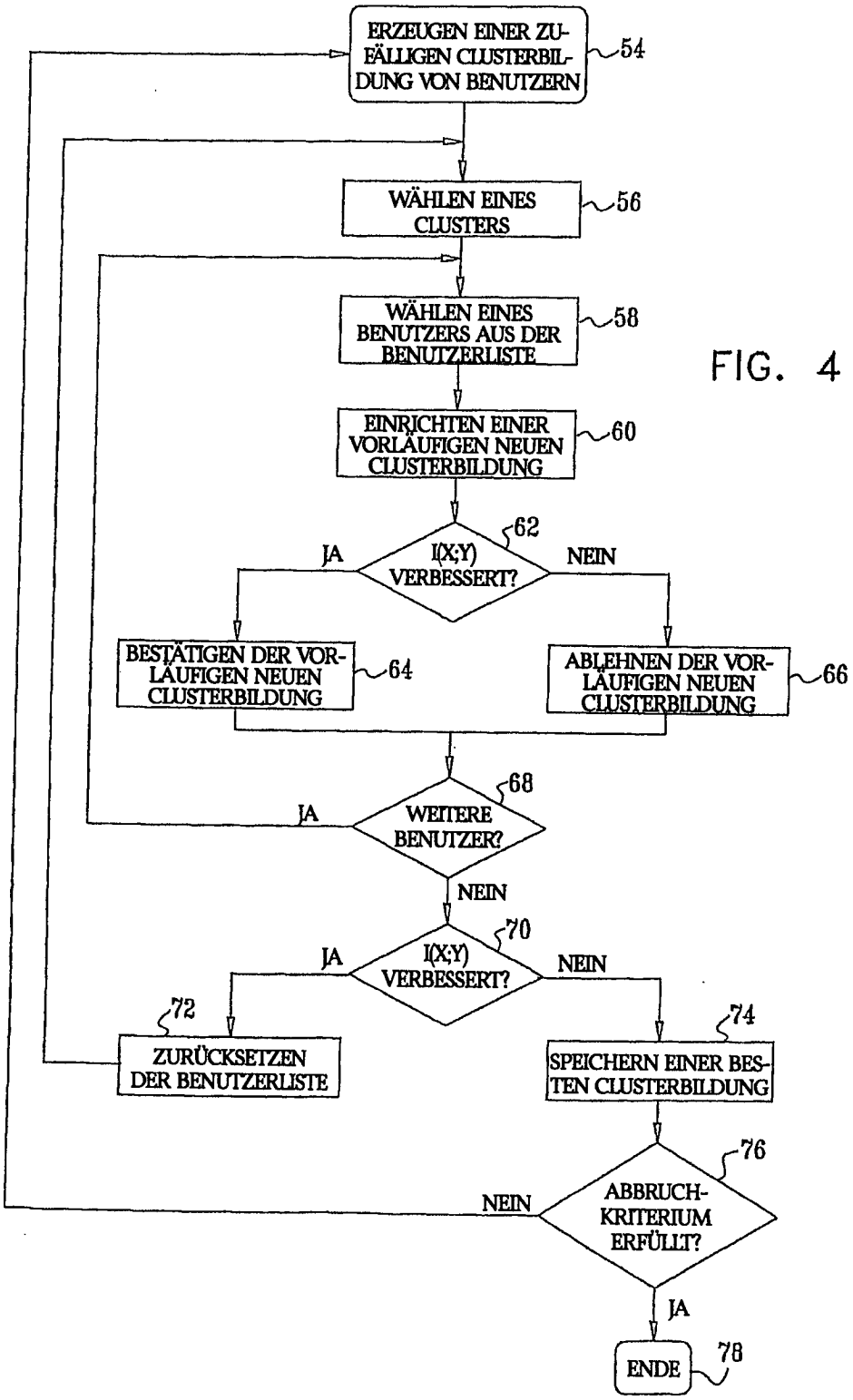

ERZEUGEN EINER ZUFÄLLIGEN CLUSTERBILDUNG VON BENUTZERN
54
WÄHLEN EINES CLUSTERS
56
WÄHLEN EINES BENUTZERS AUS DER BENUTZERLISTE
58
EINRICHTEN EINER VORLÄUFIGEN NEUEN CLUSTERBILDUNG
60
62
I(X;Y) VERBESSERT?
JA
NEIN
BESTÄTIGEN DER VORLÄUFIGEN NEUEN CLUSTERBILDUNG
64
ABLEHNEN DER VORLÄUFIGEN NEUEN CLUSTERBILDUNG
66
68
WEITERE BENUTZER?
JA
NEIN
70
I(X;Y) VERBESSERT?
JA
NEIN
72
ZURÜCKSETZEN DER BENUTZERLISTE
74
SPEICHERN EINER BESTEN CLUSTERBILDUNG
76
ABBRUCHKRITERIUM ERFÜLLT?
NEIN
JA
ENDE
78

FIG. 4

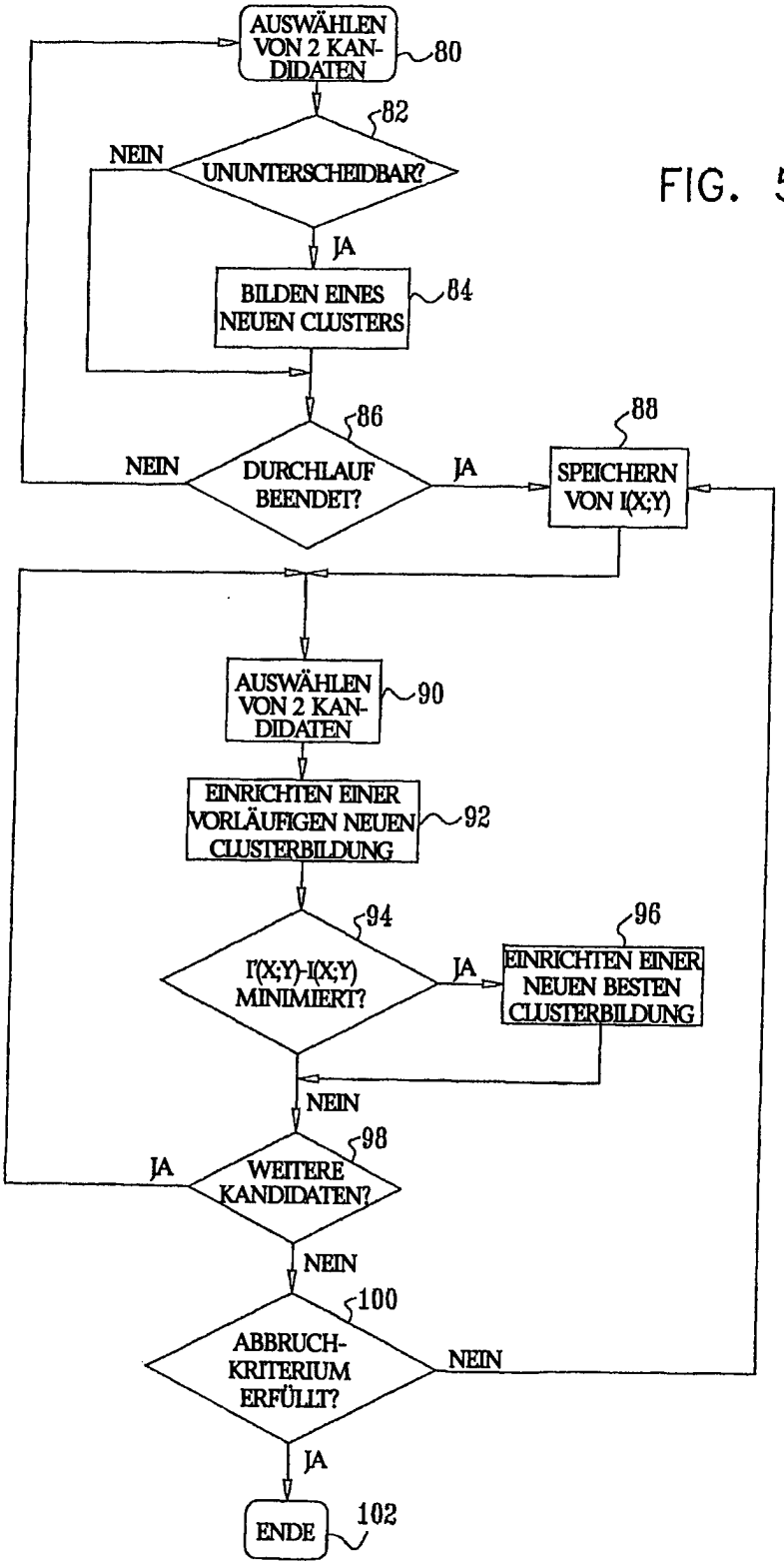

AUSWÄHLEN VON 2 KANDIDATEN
80
82
NEIN
UNUNTERSCHEIDBAR?
JA
BILDEN EINES NEUEN CLUSTERS
84
86
NEIN
DURCHLAUF BEENDET?
JA
88
SPEICHERN VON I(X;Y)
AUSWÄHLEN VON 2 KANDIDATEN
90
EINRICHTEN EINER VORLÄUFIGEN NEUEN CLUSTERBILDUNG
92
94
I'(X;Y)-I(X;Y) MINIMIERT?
JA
96
EINRICHTEN EINER NEUEN BESTEN CLUSTERBILDUNG
NEIN
98
JA
WEITERE KANDIDATEN?
NEIN
100
ABBRUCH-KRITERIUM ERFÜLLT?
NEIN
JA
ENDE
102

FIG. 5

FIG. 6

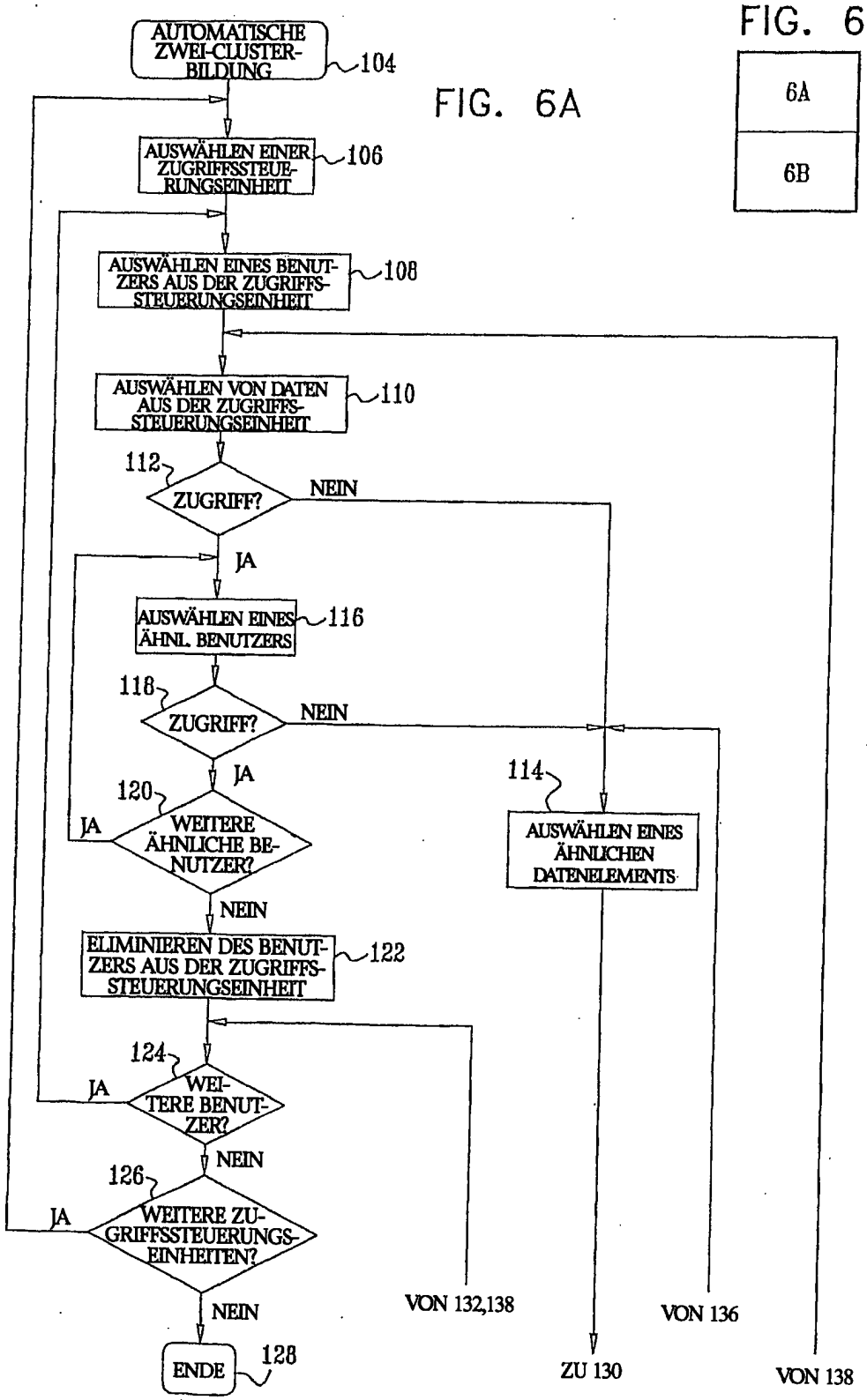
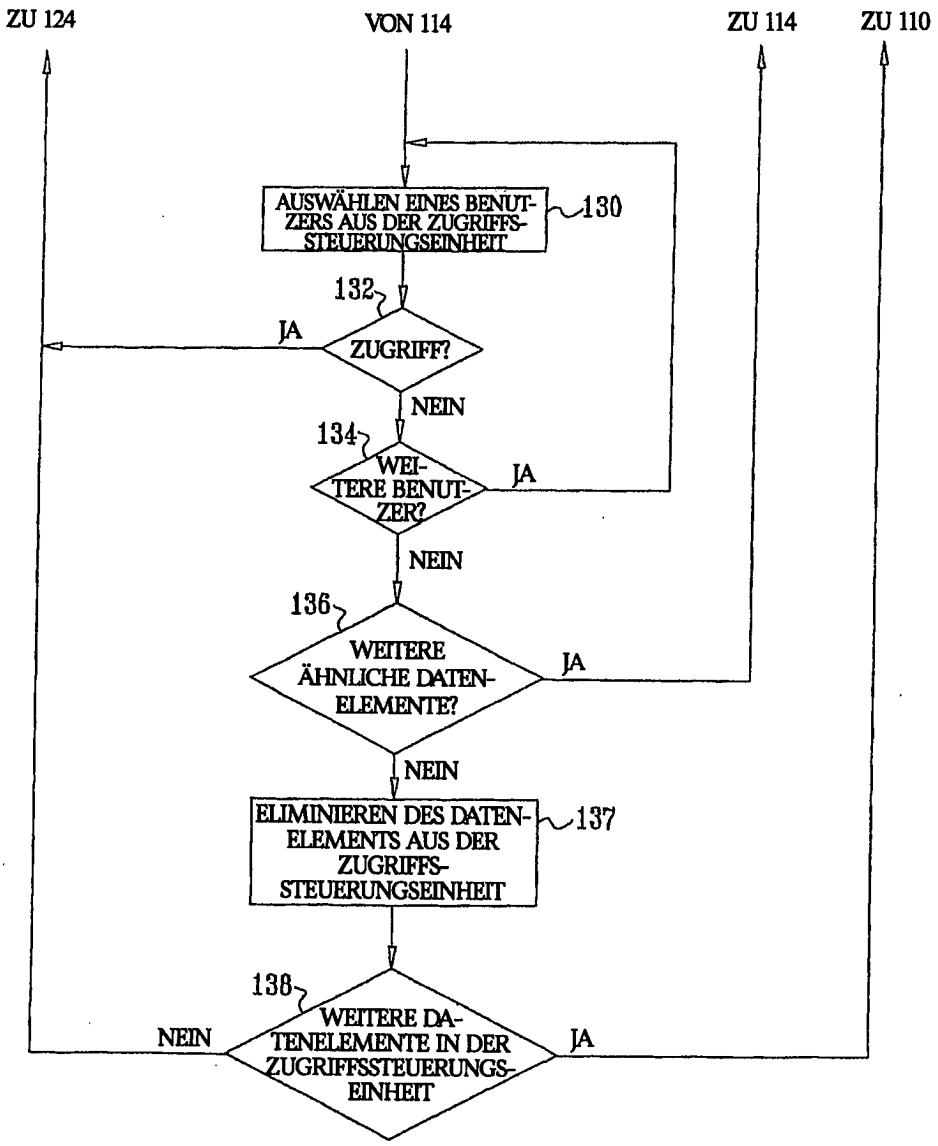

AUTOMATISCHE ZWEI-CLUSTER-BILDUNG
104
AUSWÄHLEN EINER ZUGRIFFSSTEUE-RUNGSEINHEIT
106
AUSWÄHLEN EINES BENUT-ZERS AUS DER ZUGRIFFS-STEUERUNGSEINHEIT
108
AUSWÄHLEN VON DATEN AUS DER ZUGRIFFS-STEUERUNGSEINHEIT
110
112
ZUGRIFF?
NEIN
JA
AUSWÄHLEN EINES ÄHNL. BENUTZERS
116
118
ZUGRIFF?
NEIN
JA
120
JA
WEITERE ÄHNLICHE BE-NUTZER?
NEIN
114
AUSWÄHLEN EINES ÄHNLICHEN DATENELEMENTS
ELIMINIEREN DES BENUT-ZERS AUS DER ZUGRIFFS-STEUERUNGSEINHEIT
122
124
JA
WEI-TERE BENUT-ZER?
NEIN
126
JA
WEITERE ZU-GRIFFSSTEUERUNGS-EINHEITEN?
NEIN
ENDE
128
VON 132,138
ZU 130
VON 136
VON 138

FIG. 6A

6A
6B

FIG. 6B


ZU 124
VON 114
ZU 114
ZU 110
AUSWÄHLEN EINES BENUTZERS AUS DER ZUGRIFFSSTEUERUNGSEINHEIT
130
132
ZUGRIFF?
JA
NEIN
134
WEITERE BENUTZER?
JA
NEIN
136
WEITERE ÄHNLICHE DATENELEMENTE?
JA
NEIN
ELIMINIEREN DES DATENELEMENTS AUS DER ZUGRIFFSSTEUERUNGSEINHEIT
137
138
WEITERE DATENELEMENTE IN DER ZUGRIFFSSTEUERUNGSEINHEIT
NEIN
JA