LANCOM Techpaper
Content Filter

Die Architektur

Content Filter können in verschiedenen Architekturen realisiert werden:

- Für private Zwecke eignen sich Client-Lösungen, bei denen eine Filter-Software direkt auf den Arbeitsplatzrechnern installiert wird. Diese Programme, die oft auch als „Kindersicherung" bezeichnet werden, fragen aufgrund der Konfiguration die aufgerufenen Webseiten bei einem zentralen Datenbankserver an und erlauben oder verhindern dementsprechend den Zugriff.
- In sehr großen Organisationen bietet sich der Betrieb von eigenen Datenbank-Servern an, die eine sehr spezielle und gezielte Anpassung der Filterregeln ermöglichen. Mit dem entsprechenden Aufwand für Backup-Server und die regelmäßige Pflege der Regeln erfordern diese Lösungen einen erheblichen finanziellen und organisatorischen Aufwand.
- Für mittlere und große Unternehmen bietet sich eine integrierte Lösung an, bei der ein Content Filter in das zentrale Gateway implementiert wird. Der LANCOM Content Filter stellt eine solche integrierte Lösung dar, bei dem der Filter die vorhandenen Funktionen der Firewall nutzt. Besonders vorteilhaft gegenüber den Client-Lösungen ist die zentrale Konfiguration des Content Filters: so werden alle Arbeitsstationen mit geringem Aufwand und jederzeit mit einer einheitlichen Richtlinie überwacht. Der LANCOM Content Filter besteht aus mehreren Komponenten:
  - Zeitsteuerung zur Auswahl der Richtlinien
  - Firewall zur Auswahl der Datenströme, die geprüft werden müssen
  - Content Filter, der als Proxy die zu prüfenden Daten übernimmt und die gewünschten Aktionen ausführt
  - Bewertungsserver (Rating-Server), der die besuchten Internetseiten auf ihre Zugehörigkeit zu bestimmten Kategorien hin prüft

Das Filter-Konzept

Der LANCOM Content Filter entfaltet seine Funktion in der Verbindung zwischen den konfigurierten Regeln und dem Bewertungsserver von IBM Tivoli Security Solutions. Auf der technischen Ebene untersucht der Content Filter jeden HTTP-Request nach den folgenden Kriterien:

- Wer versucht einen Zugriff auf das Internet? Die Antwort auf diese Frage liefern die Firewall-Regeln, in der bestimmte Benutzer oder Benutzergruppen über IP- oder MAC-Adressen identifiziert werden.
- Wann darf der Zugriff erfolgen? Hier greifen die Zeitrahmen, mit denen die zeitliche Gültigkeit der Filterregeln gesteuert wird.
- Welche URL soll geöffnet werden? Der Content Filter prüft die Zugehörigkeit der Webseite zu Black- oder Whitelists und zu einer oder mehreren Kategorien, die über benutzerdefinierte Profile für eine Benutzergruppe aktiviert werden.
- Welche Aktion soll der Content Filter ausführen? Die Regeln des Content Filters können den Zugriff auf bestimmte Webseiten erlauben, verbieten oder einen eingeschränkten Zugriff ermöglichen.

Die beiden ersten Punkte werden durch die Definition der entsprechenden Firewall-Regeln und Zeitfenster eingestellt. Für die Prüfung der aktuellen URL greift der Content Filter zunächst auf die konfigurierten Black- oder Whitelists zurück. Wenn die URL in der

LANCOM Systems

LANCOM Techpaper
Content Filter

Blacklist enthalten ist, wird sie ohne weitere Prüfung gesperrt. Wenn die URL in der Whitelist enthalten ist, wird sie ohne weitere Prüfung erlaubt, die Whitelist hat dabei Vorrang vor einem ggf. konkurrierenden Blacklist-Eintrag. Bei der Definition von Black- und Whitelists können innerhalb einer Domain gezielt verschiedene Pfade separat eingetragen werden, so dass bestimmte Bereiche einer URL unterschiedlich bewertet werden können.

Wenn die Prüfung der URL aufgrund der Black- und Whitelists kein Ergebnis bringt, leitet der Content Filter die Anfrage an einen Bewertungsserver weiter. Dabei werden die Daten anonymisiert verarbeitet, so dass kein Rückschluss auf den Anwender möglich ist. Der Bewertungsserver prüft die aufgerufene URL gegen seine Datenbank und liefert das Ergebnis über den LANCOM Content Filter an den Browser des Clients zurück.

Die Einträge in der Datenbank enthalten im Wesentlichen die gespeicherte URL und die Kategorie, der die URL zugeordnet wurde. Dabei kann eine URL auch mehreren Kategorien zugeordnet sein. Außerdem können die einzelnen Unterseiten unterschiedlich kategorisiert sein, so kann z. B. für jede der drei folgenden URLs eine andere Kategorisierung gelten:

- www.mycompany.de
- www.mycompany.de/aktuell
- www.mycompany.de/downloads

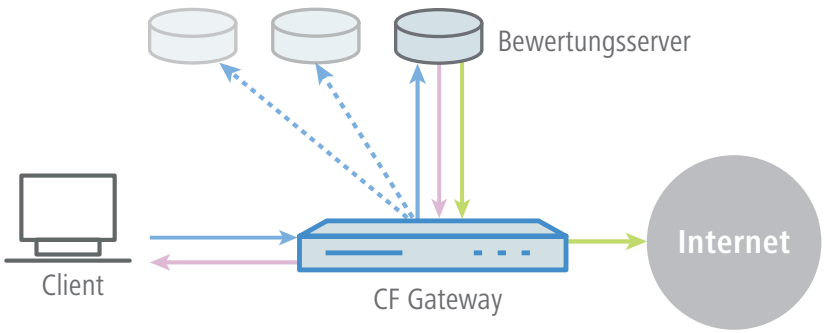
Analog können auch einzelne Hosts innerhalb der Domain unterschiedlich kategorisiert sein:

- www.mycompany.de
- downloads.mycompany.de

Die Einträge in der Datenbank sind wie in folgendem Beispiel aufgebaut:

- Domains: mycompany.de
- Hosts: www.mycompany.de, bilder. mycompany.de
- Verzeichnisse: www.mycompany.de/pics/
- HTML-Seiten: www.mycompany.de/pics/index.html
- Bild-URL: www.mycompany.de/pics/001.jpg
- IP-Adressen: http://123.123.123.123
- Protokolle: http:// vs. ftp://
- Ports: http://www.mycompany.de:80 und http://www.mycompany.de:81

Die Prüfung der URL liefert bei Erfolg die Zugehörigkeit der Webseite zu einer oder mehreren von 58 Kategorien zurück. Welche Webseiten bzw. Kategorien von Webseiten zu welcher Aktion führen, kann der Administrator individuell für das Unternehmen anpassen. Die 58 Kategorien sind für eine komfortablere Konfiguration zu 14 Gruppen thematisch zusammengefasst, z. B. “Pornografie/Nacktheit”, “Einkaufen” oder “Kriminelle Aktivitäten”. Für jede dieser Gruppen lassen sich die enthaltenen Kategorien aktivieren oder deaktivieren. Die Unterkategorien für “Kriminelle Aktivitäten” sind z. B. “Illegale Aktivitäten”, “Computerkriminalität”, „Politischer Extremismus/Hass/Diskriminierung“.



Aktionen des Content Filters

Mit den Aktionen des Content Filters werden gezielt die erforderlichen Maßnahmen eingeleitet, die aufgrund der Prüfung nach Benutzergruppe, Zeit und

LANCOM Systems

LANCOM Techpaper
Content Filter

Zugriffsziel ausgeführt werden sollen. Neben den üblichen Aktionen „erlauben“ und „blockieren“ bietet der LANCOM Content Filter eine weitere Möglichkeit, mit deren Hilfe der Zugriff auf die Webseiten flexibel gehandhabt werden kann.

Der „Override“ bietet dem Benutzer die Gelegenheit, eine eigentlich gesperrte Webseite dennoch zu besuchen. Mit dieser Option kann der Administrator kritischen Situationen ausweichen, in denen der Zugriff auf bestimmte, normalerweise gesperrte Webseiten für die geschäftliche Nutzung unbedingt erforderlich ist. Manchmal ist in diesen Fällen keine Zeit oder Gelegenheit, den Zugriff zu prüfen und die Regeln ggf. anzupassen. Wenn der Benutzer dem Override zustimmt, kann er je nach Konfiguration für einige Minuten bis zu einem Tag auf die entsperrte Webseite zugreifen. Der Administrator kann sich automatisch über diese Vorgänge informieren lassen, um ggf. nachträglich den Zugang zu häufiger entsperrten URLs zu aktivieren. Mit dem Override-Typ kann die Auswirkung der temporären Freischaltung genauer definiert werden. So kann mit dem Override die Kategorie freigeschaltet werden, zu der die gerade aufgerufene Domain gehört. Damit werden dann automatisch auch alle anderen Webseiten erreichbar, die zu dieser Kategorie gehören. Alternativ kann die Domain vorübergehend entsperrt werden, unabhängig von den Kategorien, denen die verschiedenen Unterseiten zugeordnet sind. Besonders restriktiv ist die Kombination von Kategorie und Domain für den Override: Damit werden nur die URLs erlaubt, die sowohl zu dieser Domain als auch zu den Kategorien der aktuellen Webseite gehören.

Der Bewertungsserver

Der Bewertungsserver benutzt verschiedene Web Crawler um die Inhalte auf den Webseiten zu bewerten. Web Crawler durchsuchen automatisch und schnell das Internet und analysieren Schlüsselwörter, Text und Bilder auf Webseiten, klassifizieren sie und speichern sie in einer Datenbank. Die Kombination von verschiedenen intelligenten Algorithmen für Text-Klassifizierungen und Bild-Erkennung, mit z. B. Symbol-, Logo-, Haut- und Gesichtserkennung und OCR (Optical Character Recognition) sichert eine extrem zuverlässige Bewertung. Die Datenbank beinhaltet eine Liste von über 100 Millionen URLs, die etwa 10 Milliarden Web Sites abdecken, und wird durch das dynamische und automatische Verfahren der Web Crawler täglich um fast 150.000 Einträge aktualisiert und erweitert.

Web Crawler

Die Algorithmen der Web Crawler sind dabei gezielt auf die Anforderungen der Bewertungsserver ausgelegt. Auf der technischen Ebene müssen die Web Crawler bei Performance-Schwankungen, nicht erreichbaren Servern, Spam-Domains, geparkten Domains, mehrsprachigen und suchmaschinen-optimierten Webseiten die jeweils passende Strategie anwenden. Auf der inhaltlichen Ebene sind kulturelle, moralische und ethische Aspekte wichtig für die geografisch korrekte Bewertung von Webseiten.

Um alle relevanten URLs untersuchen zu können, verfolgen die Web Crawler die auf den Webseiten gefundenen Hyperlinks – nicht innerhalb einer Domain, sondern gezielt auch über deren Grenzen hinweg in andere Domains. Um auch die nicht verlinkten Webseiten zu finden, greifen die Web Crawler zusätzlich auf Domain Registration Informationen und andere externe Quellen über neue Domains zurück.

Textklassifizierung

Die Textklassifizierung in den Bewertungsservern geht weit über die einfache Untersuchung von Schlüsselwörten hinaus. Auch wenn die Suche anhand dieser Merkmale leicht zu konfigurieren ist, hat die einfache Bewertung anhand des Vorkommens bestimmter Wörter einen entscheidenden Nachteil: Manche Wörter haben mehrere Bedeutungen, die nur mit Betrachtung

LANCOM Techpaper
Content Filter

des Kontextes aufgelöst werden können. Das Wort „Sex“ kann z. B. sowohl auf pornografische wie auch medizinische Inhalte hindeuten.

Intelligente Suchfunktionen nutzen daher auch die Häufigkeit von Wörtern und die Kombination der Wörter untereinander. Mit diesen heuristischen Verfahren ist eine nahezu fehlerfreie Klassifizierung der Texte möglich, sofern die Textmenge für eine valide Bewertung ausreicht.

Bildklassifizierung

Für die korrekte Klassifizierung der Bilder auf den Webseiten werden mehrere Technologien verwendet. Pornografische Bilder werden z. B. anhand der Bildanteile mit Hauttönen ermittelt. Damit Porträtaufnahmen nicht fälschlicherweise als Pornografie bewertet werden, sind die Bewertungsserver mit einer speziellen Gesichtserkennung ausgestattet, die Gesichter und deren Anteil an einem Bild zuverlässig erkennen können. Aber auch Symbole mit politischen Hintergründen o. ä. werden von der Bildklassifizierung erkannt, um verbotene Inhalte zu identifizieren. Da die Bilder in vielen Fällen auch Texte enthalten, wird zusätzlich eine Optical Character Recognition (OCR) verwendet, um auffällige Botschaften in den Bilder für die Klassifizierung nutzen zu können.

Kombination der Ergebnisse

Die Ergebnisse der einzelnen Bewertungs-Engines führen in manchen Fällen nicht zu eindeutigen Ergebnissen. So können auf einer Webseite z. B. viele Bilder mit großen Hautanteilen gezeigt werden, die auf pornografische Bilder schließen lassen. Der zugehörige Text signalisiert aber evtl, einen medizinischen Zusammenhang. Aus diesem Grund werden die Ergebnisse mit einer gewichteten Bewertung zusammengefasst, die eine zuverlässige Klassifizierung ermöglicht.

Automatische Auswahl des Bewertungsservers

Der LANCOM Content Filter nutzt einen zentralen Bewertungsserver in einem der größten Rechenzentren der Welt. Neben der ständig aktuellen Datenbank durch das hochprofessionelle Management der Kategorien und der geprüften Webseiten bietet diese Lösung einen weiteren entscheidenden Vorteil: sollte der Bewertungsserver gestört sein, greift der LANCOM Content Filter automatisch auf einen Backup-Server zu. Mehrere redundante Datenbanken sind weltweit verteilt verfügbar, so dass jederzeit der Zugriff auf einen aktuellen Bewertungsstand möglich ist. Der LANCOM Content Filter wählt dabei immer den Server mit der schnellsten Antwortzeit, um die Leistungsfähigkeit des Filters zu optimieren.

Benutzerdefinierte Kategorie-Profile

Mit den benutzerdefinierten Kategorie-Profilen kann der LANCOM Content Filter ganz gezielt auf die Anforderungen in einer Organisation angepasst werden. Blacklists und Whitelists gelten in vielen Fällen für alle Anwender gleichermaßen – aber die Sperrung oder Freigabe der Webseiten von bestimmten Kategorien muss nicht für alle Unternehmensbereiche einheitlich geregelt werden. Denn die Pressesprecher in der PR-Abteilung müssen zur Ausführung ihrer Aufgaben vielleicht auch auf URLs zugreifen können, die für Mitarbeiter der Produktion nicht zugänglich sein müssen. Um diese speziellen Regelungen abzubilden werden gezielt Profile eingerichtet, die über eine eigene Konfiguration der Kategorien verfügen. Über die Firewall wird ein solches benutzerdefiniertes Profil einer bestimmten Benutzergruppe zugeordnet, die damit über ihre eigenen Filtereinstellungen verfügt.

LANCOM Techpaper
Content Filter

Protokollierungs- und Alarmierungsfunktionen

Den Protokollierungs- und Alarmierungsfunktionen des Content Filters kommt vor allem vor dem Hintergrund der kontinuierlichen Weiterentwicklung der Filterregeln eine hohe Bedeutung zu. Je nach Konfiguration kann der Content Filter den Administrator über Lizenzüberschreitung und Lizenzablauf, Fehler oder den Override von gesperrten Webseiten informieren. Als Informationskanäle kann der Administrator zwischen E-Mail, SNMP oder SYSLOG wählen. Darüber hinaus kann der Content Filter in regelmäßigen Abständen einen Snapshot erstellen, der eine ausführliche Statistik über den Zeitraum auch für einen späteren Vergleich mit anderen Zeiträumen bereitstellt. Für einen gezielten Einblick in den Zustand des Content Filters zur Laufzeit bietet der LANmonitor komfortable Funktionen an.

Die Statistiken des Content Filters sind grundsätzlich anonym und lassen keine Rückschlüsse auf den Anwender zu. Lediglich die Meldungen über die Nutzung der Override-Funktion können je nach Einstellung die IP- und MAC-Adresse des zugehörigen Rechners und die aufgerufene URL enthalten.

Benutzerdefinierte Blockseiten

Zur Anzeige der Aktionen „Blockieren“ und „Override“ bietet der LANCOM Content Filter vordefinierte Seiten an, die bei Auslösen der Aktionen im Browser angezeigt werden. Für die vollständige Integration des Content Filters in das Unternehmen können auch eigene Block- und Override-Seiten entwickelt werden. Dabei stehen alle Funktionen für moderne Webseiten bis hin zu Javascript o.ä. zur Verfügung. Zum Einblenden der aktionsbezogenen Texte und Schaltflächen können einfache HTML-Tags genutzt werden, die zur Laufzeit in den Quelltext eingesetzt werden.

LANCOM Content Filter setzen Unternehmensrichtlinien durch

Mit einem Content Filter können die Administratoren gezielt Inhalte im Netzwerk filtern und dadurch den Zugriff auf z. B. illegale, anstößige oder gefährliche Internetseiten verhindern. Weiterhin wird so das private Surfen während der Arbeitszeit unterbunden. LANCOM Content Filter erlauben die zentrale Konfiguration und Verwaltung des Internetzugriffs für die gesamte Organisation und damit eine konsistente und effektive Umsetzung der Internet-Richtlinien. Das steigert die Produktivität der Mitarbeiter und die Sicherheit des Netzwerks, reduziert die rechtlichen Risiken in vielen Bereichen und reserviert außerdem die volle Bandbreite ausschließlich für Geschäftsprozesse.

Der LANCOM Content Filter erlaubt eine flexible Anpassung an die unternehmensspezifischen Anforderungen mit einer ganzen Reihe von Funktionen – von der gezielten Einschränkung nach Gruppen, IP-Adressen oder bestimmten Stationen, zeitgesteuerten Regeln, Warnungen oder der vollständigen Sperrung von bestimmten Webseiten. Die Basis für die Filterung stellt eine zentrale Datenbank dar, in der aktuell mehr als 100 Millionen URLs erfasst sind. Die gespeicherten URLs sind in 58 Kategorien gruppiert, die eine kundenspezifische Konfiguration unterstützen.

Die kontinuierliche Pflege der Datenbank geht weit über die Konzepte vieler anderer Lösungen hinaus. Denn der Ansatz einer manuellen Pflege der URL-Listen führt nur bedingt zum Ziel, da die resultierenden Fehler zu Lücken in der Netzwerksicherheit führen können. Der LANCOM Content Filter nutzt eine Datenbank mit automatisierter Prüfung von Webseiten, was zu einer sehr schnellen Integration neuer und geänderter Web-Inhalte führt. Web Crawler aktualisieren und ergänzen die zentrale Datenbank täglich um fast 150.000 Web Sites. Dabei werden komplexe Analysemethoden zur

LANCOM Techpaper
Content Filter

automatischen Kategorisierung angewendet. Der Administrator kann beim LANCOM Content Filter gezielt und zugleich komfortabel auf Basis von Kategorien entscheiden, welche Internet-Inhalte im Netzwerk erlaubt sind und welche nicht.

Zusammenfassung

Der Zugang zum Internet ist für viele Unternehmen ein geschäftsnotwendiger Vorgang. Die Durchsetzung von klaren und einheitlichen Richtlinien für den Umgang mit dem Internet ist unentbehrlich, um die Gefahren durch Malware etc. abzuwehren, die Produktivität der Mitarbeiter zu sichern und gesetzliche Vorgaben zu erfüllen. Der LANCOM Content Filter bietet die Möglichkeit einer zentralen, leistungsfähigen und einfach zu konfigurierenden Filterlösung für Unternehmensnetzwerke. Mit stets aktuellen Datenbanken und gezielt an die jeweiligen Anforderungen angepassten Regeln erlaubt der LANCOM Content Filter die Umsetzung der Unternehmensrichtlinien und steigert so die Sicherheit des Netzwerks.