مجله پژوهشي دانشگاه اصفهان (علوم پايه)
جلد بيست و يكم- شماره ۱- سال ۱۳۸٤
صص ۱۳۸ - ۱۲۵

طراحی و پیاده‌سازی یک کارگزار ناظر وب با قابلیت سرند هوشمند

بهزاد مهرابي، ناصر موحدي‌نيا و ناصر قاسم آقايي

دانشكده فني و مهندسي دانشگاه اصفهان

چكيده

با گسترش اينترنت و متنوع‌شدن ترافيك آن، نياز به روشي جهت كنترل و سرند[1] محتواي ترافيك، براساس نيازها و سياستهاي مشخص، احساس گرديده است. روش مؤثر جهت نيل به اين هدف، ايجاد يك كارگزار واسط مابين مشتريان و كارگزاران شبكه در لايه‌ی كاربرد و بررسي محتواي ترافيك ردوبدل شده مي‌باشد. اين كارگزار واسط مي‌تواند يك كارگزار پروكسي[2] باشد.

كارگزار پروكسي وظايف متعددي دارد كه كنترل دسترسي به شبكه و سرند ترافيك عبوري از آن جمله است. در اين مقاله به چگونگي ايجاد يك كارگزار پروكسي با قابليت سرند هوشمند ترافيك پرداخته شده است. كارگزار طراحي شده پس از ايجاد، در يك شرايط واقعي تست شده و كارايي آن برآورد گرديده است. نتايج اين آزمون نشان مي‌دهد كه هوشمندي كارگزار پروكسي اين امكان فراهم مي‌سازد كه بتوان محتوا و نوع صفحات وب را با دقت و ضريب اطمينان قابل قبولي تشخيص داد.

---

۱. Filtering

۲. Proxy Server

واژه‌هاي كليدي: سرند ترافيك ، كارگزار پروكسي، هوشمندي.

Designing and Implemantation of a Proxy Server with Intelligent Filtering Capability

B. Mehrabi, N. Movahhedinia and N. Ghasem Aghaee

Department of Engineering, University of Isfahan

Abstract

Recently the number of the Internet users has increased rapidly, and various traffic types are emerged as well. As such, the need for a method for traffic filtering has turned out to be very crucial. An effective method for this purpose is to investigate the traffic content in application layer. To provide such filtering, a proxy server can be utilized. A proxy server performs numerous duties which include network access control and traffic filtering.

This paper considers a proxy server with rule based intelligent traffic filtering. To evaluate its performance, the proposed proxy server has been implemented and tested in real situation. The test results show that the server has acceptable performance toward recognizing the content and categorizing of browsed web pages.

Keywords: Traffic, Filtering, Proxy server, Intelligence

مقدمه

شبكه جهاني اينترنت در سالهاي اخير بسيار رشد و عموميت پيدا كرده است. اين عموميت موجب شده است كه بسياري از داده‌هاي ذخيره شده و قابل بازيابي از طريق اينترنت نامربوط،

نامناسب و بعضاً غير اخلاقي باشند. از اينرو سرند اطلاعات اهميت ويژه‌اي يافته‌است (۲-۱).

در اين رابطه، کارگزار پروکسي مي‌تواند مابين کارگزاران ديگر (WWW, FTP, SMTP و...) و مشتريان قرار گرفته و علاوه بر اعتبارسنجي کاربران، داده‌هاي درخواست شده توسط آنان را کنترل نمايد. البته يك کارگزار پروکسي وظايف ديگري نيز مي‌تواند بر عهده داشته باشد که از آن جمله کنترل دسترسي[۱] و اعتبارسنجي کاربران[۲]، ايجاد ديواره آتش جهت تامين امنيت، ايجاد دروازه[۳]، تهيه آدرسهاي IP مجازي[٤] و مديريت آنها، نقد کردن[٥] و ذخيره تراکنشها[٦] مي‌باشد (٦-۲).

در اين پژوهش يك کارگزار پروکسي با قابليت سرند هوشمند اطلاعات و توانايي‌هاي ديگري مانند ذخيره‌ی تراکنشها، مديريت سيستم از راه دور و امکان ارتباط با کارگزارهاي پروکسي ديگر طراحي شده است. اين کارگزار در محيط سيستم‌عامل ويندوز ۲۰۰۰ بعنوان بستر عملياتي و محيط کاري ايجاد گرديده است.

در بخش بعدي مقاله به فرايند اعتبارسنجي کاربر توسط کارگزار پروکسي که يکي از مهمترين بخشها و جزو لاينفك آن مي‌باشد پرداخته و پس از آن مدل طراحي کارگزار پروکسي هوشمند ارايه مي‌گردد. در پايان، نتايج تست کارگزاري که براساس طراحي مذکور ايجاد شده است، گزارش مي‌گردد.

اعتبار سنجي کاربر توسط کارگزار پروکسي

براي اعتبارسنجي کاربران، از کد شناسه کاربر و کلمه عبور او استفاده مي‌شود. اهم روشهايي که جهت اعتبارسنجي کاربر توسط کارگزار پروکسي بکارگرفته مي‌شود، عبارتند از Basic، NTLM[۷]، LSA[۸]، LDAP[۹] (۳، ٥، ۷).

روش NTLM و Basic روشهايي عمومي است که جهت شناسايي کاربران در شبکه براي

---

۱. Access Control
۲. Authentication
۳. Gateway
٤. Virtual IP Addresses
٥. Cashing
٦. Logging
۷. NT Lan Manager Method (NTLM)
۸. Local Security Authority (LSA)
۹. Lightweight Directory Access Protocol (LDAP)

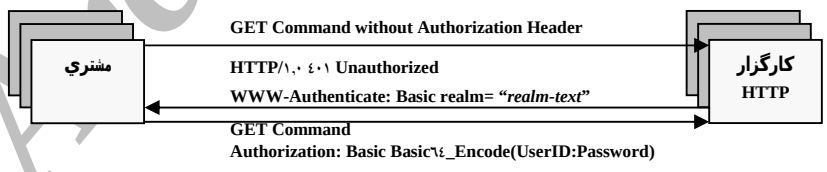
سيستم‌هاي عامل متداول ارايه شده است. روش LSA براي سيستم‌عامل ويندوز طراحي شده و توسط اين روش، كاربران ايستگاه‌هاي كاري[۱] در يك شبكه ويندوز NT يا ويندوز ۲۰۰۰ يا ويندوز XP شناسايي مي‌شوند. در روش LDAP از يك پايگاه داده مربوط به كاربران و يك يا چند سرويس دسترسي كه در طرف كارگزار اجرا شده‌اند، استفاده مي‌گردد. تمام درخواستهاي مشتريان از طريق اين سرويسها عبور كرده و پس از شناسايي كاربر، درخواستهاي مربوطه به كارگزار ارجاع داده مي‌شوند.

بدليل آنكه بيشتر كارگزاران روي شبكه، بخصوص در سيستم‌عامل ويندوز، براي شناسايي كاربران خود عموماً روشهاي Basic و NTLM را مورد استفاده قرار مي‌دهند، در قسمت بعدي اين روشها با تفصيل بيشتري شرح داده مي‌شوند.

الف- روش Basic

در اين روش كه در شكل يک نمايش داده شده است، با ردوبدل سه پيام بين كارگزار و مشتري، كارگزار، كاربر را شناسايي مي‌كند. ابتدا مشتري يك سند بدون معرفي خود درخواست مي‌كند. كارگزار با توجه به اين‌كه هويت كاربر مشخص نيست (درخواست حاوي سرآيند[۲] Proxy-Authorization نيست)، پيامي مبني بر لزوم ارسال مجدد درخواست با معرفي كاربر (با كد برگشتي ٤٠٧ با سرآيند Proxy-Authenticate: Basic) ارسال مي‌نمايد.

در اين لحظه، مشتري از كاربري كه با آن كار مي‌كند، كد كاربري[۳] و كلمه عبور[٤] را دريافت كرده، آنها را بصورت User ID : Password توسط كدينگ Basic_٦٤ كه يك كدينگ عمومي است، كد نموده و درخواست مجدد را با سرآيند مربوطه (Proxy-Authorization: Basic UP_Code) ارسال مي‌دارد.

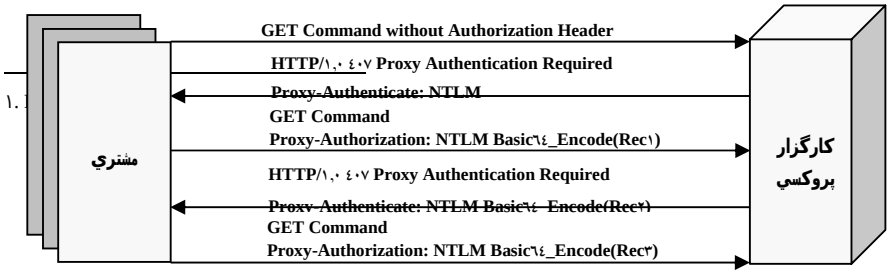


۱. Work Stations
۲. Header
۳. User ID
٤. Password

شکل۱: روش شناسايي Basic

ب- روش NTLM

روش Basic، امنيت چنداني ندارد و هر کس که پيام ارسالي از مشتري را دريافت کند، مي‌تواند کد کاربر و کلمه عبور او را پيدا کند. در روش NTLM که در ويندوز NT، ويندوز ۲۰۰۰ و ويندوزXP بکار رفته و در شکل دو نمايش داده شده است، يك کارگزار با ردوبدل پنج پيام، کاربر را شناسايي مي‌کند. ابتدا مشتري يك سند بدون معرفي خود درخواست مي‌کند. کارگزار با توجه به اين نکته که کاربر مشخص نيست (درخواست حاوي سرآيند Proxy-Authorization نيست)، پيامي مبني بر ارسال مجدد آن درخواست با معرفي کاربر به مشتري مي‌فرستد (با کد برگشتي ٤٠٧ با سرآيند Proxy-Authenticate: NTLM). در اين لحظه، مشتري از کاربري که با آن کار مي‌کند، کد و کلمه عبور و نام دامنه[1] را دريافت مي‌کند. سپس رکورد شماره يك روش NTLM را پر مي‌کند و آنرا با کدينگ Basic_٦٤، کد کرده (Rec۱_Code) و يك پيام درخواست با سرآيند Proxy-Authorization: NTLM Rec۱_Code ارسال مي‌کند. در اين لحظه، کارگزار رکورد شماره دوم NTLM را که حاوي يك کليد (براي رمز گشايي DES) است را پر کرده و آنرا با کدينگ Basic_٦٤ کد نموده (Rec۲_Code) و پيامي مبني بر تکرار درخواست (با کد برگشتي ٤٠٧) و يك سرآيند Proxy-Authenticate: NTLM Rec۲_Code به مشتري ارسال مي‌کند. مشتري کليد را از آن پيام استخراج کرده و توسط آن کليد، کلمه عبور کاربر را به دو روش LM (LM_Key) و NT (NT_Key) که روشهاي مختلف رمزنگاري هستند، رمزگذاري مي‌کند. سپس رکورد شماره سه NTLM که حاوي کد شناسه کاربر و LM_Key و NT_Key مي‌باشد را پر نموده و با کدينگ Basic_٦٤ کد مي‌کند (Rec۳_Code). سپس پيام درخواست مجدد را با سرآيند Proxy-Authorization: NTLM Rec۳_Code به کارگزار ارسال مي‌کند. کارگزار با توجه به کد شناسه کاربر و بوسيله کليد ارسالي خود، کلمه عبور آن کاربر را رمزگذاري کرده و با کلمه عبور رمزشده دريافتي چك مي‌کند و بدين ترتيب ادامه يا عدم ادامه کار را تعيين مي‌نمايد.



شكل۲: روش NTLM

چگونگي سرند اطلاعات در كارگزار پروكسي

هم اكنون روشهاي گوناگوني براي سرند اطلاعات در كارگزار پروكسي مورد استفاده قرار مي‌گيرد. اين روشها را به سه دسته مي‌توان تقسيم نمود كه در ادامه تشريح مي‌گردند:

۱. استفاده از پايگاه داده

در اين روش كارگزار پروكسي از يك پايگاه داده حاوي آدرس محل سايتهاي غير مجازي كه بايستي سرند شوند، استفاده مي‌كند. اين پايگاه داده همواره از طريق اينترنت به روز در آورده مي‌شود. همچنين از يك ليست قابل ويرايش استفاده مي‌شود كه مدير كارگزار مي‌تواند در آن سايتهاي مورد نظر خود را نيز جهت فيلتر كردن وارد نمايد. مثالهايي از اين شيوه، I-Net Filtering Server (۸) و يا Smart Filter (۹) مي‌باشند.

۲. استفاده از نرم‌افزاركمكي هوشمند

در اين شيوه كارگزار پروكسي از يك نرم‌افزار جانبي كمك مي‌گيرد. اين نرم‌افزار بطور اتوماتيك و براساس مؤلفه‌هايي كه از قبل براي آن تعريف مي‌شود، نوع سايتهاي معرفي شده در اينترنت را تشخيص داده و آنها را دسته‌بندي مي‌نمايد. پس از آنكه دسته‌هاي غيرمجاز توسط مدير شبكه مشخص شد، كارگزار پروكسي هنگام دريافت تقاضاي ارتباط با يك سايت، با پرسش از نرم‌افزار كمكي به مجاز يا غير مجاز بودن سايت مورد نظر پي مي‌برد (۱۰).

۳. استفاده از هوشمندي در پروكسي

دو روش فوق مبتني بر دسته‌بندي سايتها و سپس سرند داده‌هاي درخواست شده براساس

نوع سايت مي‌باشد. ليکن در بعضي از موارد، تمامي داده‌هاي ذخيره شده در يك سايت نامطلوب نيستند. بعبارت ديگر ممکن است يك سايت حاوي اطلاعات مطلوب در کنار اطلاعات نامطلوب باشد. در اين‌صورت، سرند سايت موجب محدود کردن دسترسي کاربران به اطلاعات مطلوب نيز مي‌گردد.

در روش سوم با استفاده از الگوريتمها و يا ساختارهاي هوشمند، نوع اطلاعات درخواست شده از يك سايت تشخيص داده شده و در صورت قرار گرفتن در انواع غير مجاز، اطلاعات درخواستي سرند مي‌گردد. به تعبير ديگر، ترکيب يك سيستم خبره با کارگزار پروکسي به کارگزار قابليت تصميم‌گيري هوشمندانه در مورد مجاز و يا غيرمجاز بودن اطلاعات را مي‌دهد. بدين جهت روش اخير کاراتر از دو روش قبلي مي‌باشد، ولي در عوض تأخير ناشي از فرايند سرند داده بيشتر خواهد بود (۱۲-۱۱).

روش ارايه شده در اين مقاله

در اين تحقيق روش سوم مورد مطالعه قرار گرفته و جهت هوشمند نمودن نظارت بر ترافيك، از يك سيستم مبتني بر قواعد[۱] و پايگاه دانش[۲] با قابليت فراگيري استفاده شده است (۱۲). با پرس‌وجو از اين پايگاه دانش، نوع و مشخصات يك ارتباط و شيوه‌ي مناسب رفتار با آن مشخص مي‌گردد. هر قاعده در اين سيستم داراي يك فاکتور عدم قطعيت مي‌باشد که در طي زمان با يادگيري از نمونه‌هاي متعدد تنظيم مي‌گردد. يك موتور استنتاج[۳] نيز در اين سيستم تهيه شده است تا براساس آموخته‌هاي آن و با توجه به قواعد مشخص از قبل تعيين شده، تصميم‌گيري مناسب انجام گيرد. با پرسش از اين سيستم خبره، نوع سند درخواستي و به تبع آن نحوه‌ی رفتار با آن سند مشخص مي‌گردد. حقايق (موضوعات) اوليه توسط مدير کارگزار وارد پايگاه دانش مي‌شود و پس از آن با يادگيري و تعميم آموخته‌ها، مجموعه‌ی حقايق تکميل مي‌گردد. قواعد فوق مي‌توانند بر روي کلمات و يا موضوع متن[٤] مورد تقاضا، آدرس محل[٥]، سرآيند و نيز عنوان[٦] سند درخواستي اعمال شوند.

---

۱. Rule Base Expert System
۲. Knowledge Base
۳. Inference Engine
٤. Content
٥. Uniform Resource Location (URL)
٦. Title

در اين سيستم، امكان ايجاد و تعريف سه نوع قاعده ممكن مي‌باشد :

۱- نوع اول كه خود شامل چهار گونه است:

۱-۱- اگر كلمه‌ی مشخصي در آدرس محل باشد:

مثال: Rule۱: If URL contain word “sport” then

this page is BANNED PAGE with CF=۰,۵

۱-۲- اگر كلمه‌ی مشخصي در متن صفحه وب باشد:

مثال: Rule۲: If CONTENT OF PAGE contain word “footbal” then

this page is ALLOWED PAGE with CF=۰,۹

۱-۳- اگر آدرس يا سايت ممنوع درخواست شده باشد:

مثال: Rule۳: If request come from site “WWW.SPORT.COM” then

requested page is BANNED PAGE with CF=۰,۱

۱-٤- اگر آدرس يا سايت مجاز در خواست شده باشد:

مثال: Rule٤: If request from site “WWW.SPORT.COM” then

requested page is ALLOWED PAGE with CF=۰,۳۳

۲- اگر صفحه وب درباره‌ی موضوعي بحث كند:

مثال: Rule۵: If subject of page is “sport” then

this page is BANNED PAGE with CF=۰,۹۵

۳- تركيب AND يا OR يا NOT ديگر قواعد:

مثال: Rule٦: If Rule۱ AND Rule۲ then

this page is BANNED PAGE with CF=۰,۵

مثال: Rule۷ : If NOT Rule۳ OR Rule ٤ then

this page is ALLOWED PAGE with CF=۰,۱

مثال: Rule۸ : If Rule۵ AND NOT Rule٦ then

this page is BANNED PAGE with CF=۰,٦۵

هر موضوع شامل يكسري كلمات با وزنهاي معين (عددي بين صفر و يك) مي‌باشد و اگر ميانگين وزني كلمات موضوعي در سند درخواستي از مقدار معيني بيشتر باشد، ميتوان گفت كه آن سند مربوط به موضوع مذكور مي‌باشد.

مثال:

Subject is “Sport” with MEAN OF CFs=۰,۸۵

Word۱: Sport , CF=۰,٦

Word۲: Football , CF=۰,۸۵

Word۳: Basketball , CF=۰,۷

Word٤: Tenis , CF=۰,۵

Word۵: Racket , CF=۰,۳

. . .

موتور استنتاج ايجاد شده، متشكل از يك تابع بازگشتي است. اين تابع، نتيجه‌ی آخرين قاعده‌اي كه صفحه وب درخواستي در آن صدق مي‌كند، را باز مي‌گرداند. اين موتور استنتاج، قواعد مذكور را به روش زير براي آن صفحه وب درخواستي بررسي مي‌كند:

- ✓ اگر قاعده از نوع اول باشد، كلمه‌ی مورد نظر را در صفحه‌ی درخواستي، آدرس محل، نام سايت ورودي يا نام سايت خروجي جستجو كرده، نتيجه‌ی آن قاعده برگشت داده مي‌شود.
- ✓ اگر قاعده از نوع دوم باشد، ميانگين وزن كلمات مربوط به موضوع را در صفحه‌ی درخواستي پيدا مي‌شود. اگر ميانگين مذكور از مقدار تعريف شده براي آن موضوع بيشتر باشد، صفحه درخواستي در آن قاعده صدق مي‌كند.
- ✓ اگر قاعده تركيبي(نوع سوم) باشد، قواعد مورد استفاده در آن بررسي شده و در نهايت نتيجه‌ی آنها با هم تركيب مي‌گردد.

هرگاه يك قاعده از نوع دوم در مورد يك صفحه وب درخواستي صدق كند، روش زير براي يادگيري پايگاه دانش مورد استفاده قرار مي‌گيرد:

۱. وزن تمام كلمات آن موضوع كه در صفحه درخواستي وجود دارند، به نسبت تكرارشان، اضافه مي‌شود.

۲. وزن تمام كلمات آن موضوع كه در صفحه درخواستي وجود ندارند، به يك ميزان معين كم مي‌شود. اگر وزن اينگونه كلمات به صفر رسيد، اين كلمات از پايگاه دانش حذف مي‌شوند.

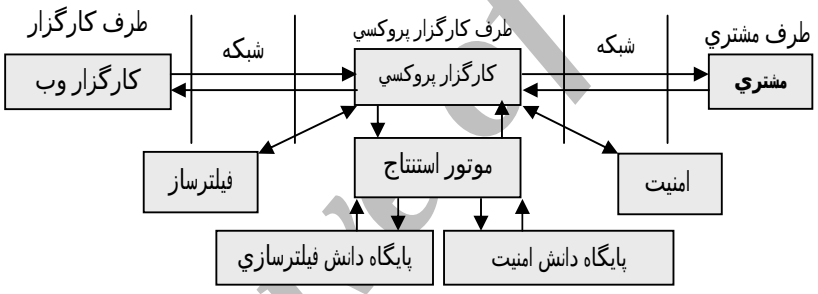
۳. كلمه‌اي كه بيشترين تكرار را در صفحه درخواستي دارد و جزء كلمات آن موضوع نمي‌باشد، با يك وزن پايين به كلمات آن موضوع (پايگاه دانش) اضافه مي‌شود.

مدل كارگزار پروكسي هوشمند

در زير ابتدا مدل مربوط به سيستم هوشمند سرند اطلاعات، و سپس مدل كلي كارگزار پروكسي هوشمند ارايه خواهد شد:

۱. مدل مربوط به سرند هوشمند

اين مدل كه در شكل سه نشان داده شده است، چگونگي انجام سرند هوشمند اطلاعات و نيز چگونگي استفاده از پايگاه دانش با قابليت يادگيري را نشان مي دهد.



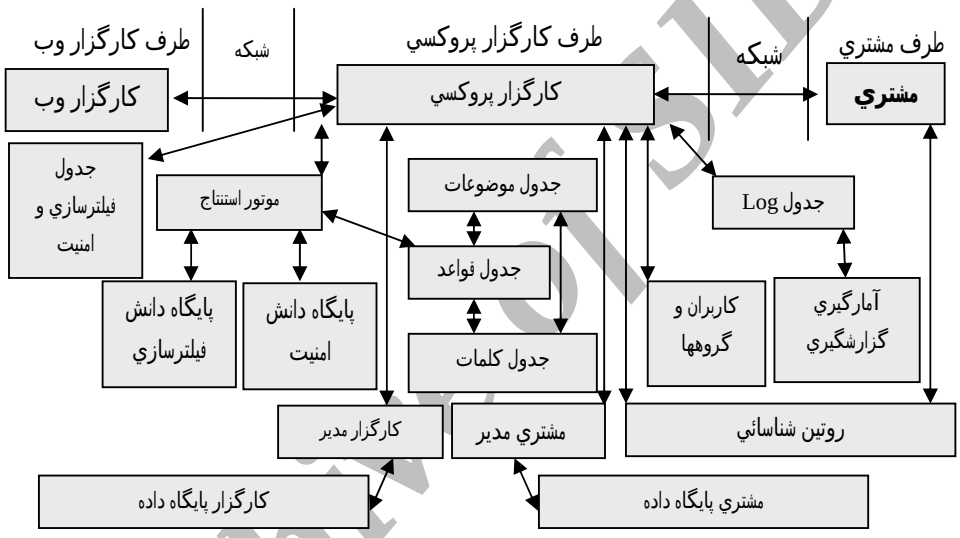
شكل۳: مدل مربوط به فيلترسازي هوشمند

همانطور كه در شكل فوق ملاحظه مي‌شود، پس از ارسال درخواست بوسيله مشتري، اين درخواست توسط موتور استنتاج در يك پايگاه دانش مورد بررسي قرار گرفته و در صورت سازگاري با قواعد و حقايق موجود در پايگاه دانش به كارگزار وب ارسال مي‌شود. با دريافت پاسخ از كارگزار وب، دوباره پاسخ توسط موتور استنتاج در پايگاههاي دانش مورد استنتاج قرار گرفته و در صورت سازگاري با قواعد و حقايق موجود در پايگاه دانش، اين پاسخ به مشتري ارسال مي‌شود. پايگاه دانش ايجاد شده با تنظيم "فاكتور عدم قطعيت[۱]" مربوط به قواعد و همچنين اضافه و يا حذف يك كلمه جديد از مجموعه حقايق به مرور زمان دانش خود را تكميل مي‌كند.

۱. Uncertainty Factor (CF)

۲. مدل کلي کارگزار پروکسي

مدل کلي کارگزار پروکسي از ترکيب زير مدلهايي با وظايف مشخص بصورت ترسيم شده در شکل چهار ايجاد گرديد. همانگونه که در شکل مشاهده مي‌شود، تمام عمليات مربوط به کارگزار پروکسي ايجاد شده، مانند سرند هوشمند و غير هوشمند، ذخيره‌ي تراکنشها، مديريت از راه دور کارگزار پروکسي، مديريت کاربران و شناسايي آنان، گزارشات و آمارگيريها در اين مدل در نظر گرفته شده است.



شکل٤: مدل کلي کارگزار پروکسي

بررسي کارايي پروکسی هوشمند

جهت بررسي کارايي الگوريتم هوشمند ارايه شده، کارگزار پروکسي پيشنهادي با زبان Visual C++ بر روي يک کامپيوتر Pentium II با کلاک ۳۳۳ مگاهرتز و سيستم عامل ويندوز ۲۰۰۰ پياده‌سازي گرديد. اين کارگزار در يک شبکه آزمايشي به مدت يك ماه تحت آزمون قرار گرفت. هدف از آزمون، شناسايي سايتهاي اينترنتي مرتبط با موضوع "COMPUTER" بود. انتظار مي‌رفت که کارگزار هوشمند پس از يک مرحله يادگيري بتواند با

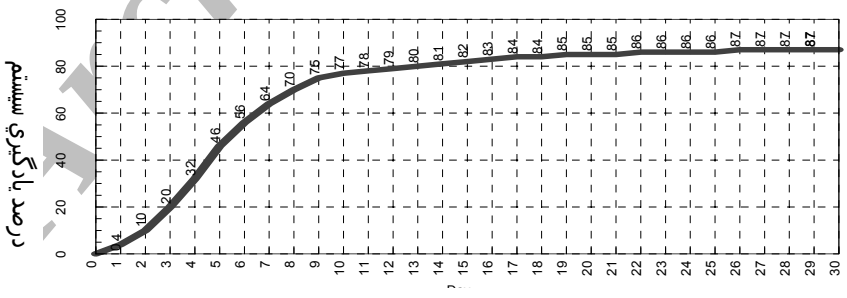
خطايي قابل قبول، اطلاعات مرتبط با موضوع مذكور را مشخص و سرند نمايد. در انتهـاي هـر روز، پايگاه دانش مربوطه بررسي گرديد. كلمات آموخته شده توسط پايگاه دانش و ضرايب مربوط به هر يک از كلمات در انتهاي سه روز اول آزمون در جدول يک، و در انتهـاي روز سـي‌ام در جـدول دو آورده شده است. همچنين، نمودارهاي مربوط بـه درصـد يـادگيري و درصـد خطـا در تـشخيص موضوعات صفحات وب درشكلهاي پنج و شش ارايه شده‌است.

جدول۱: محتوي پايگاه دانش در انتهاي سه روز اول آزمون پروكسي هوشمند براي شناسايي سايتهاي مرتبط با موضوع "COMPUTER"

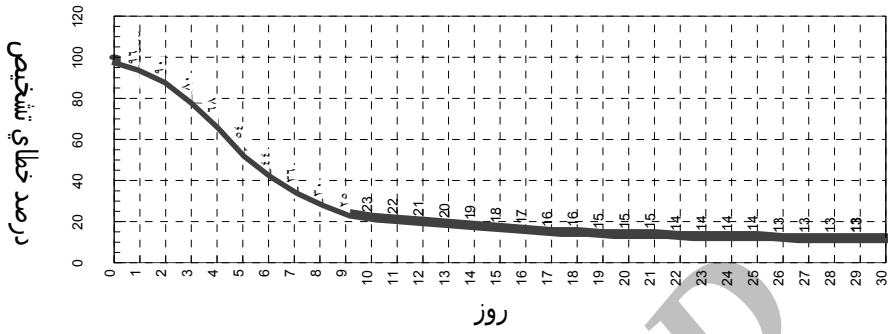
محتواي پايگاه دانش در انتهاي روز اول		محتواي پايگاه دانش در انتهاي روز دوم		محتواي پايگاه دانش در انتهاي روز سوم	
ARTICLE	CF=۰,٦۳	ARTICLE	CF=۰,٥۳	ARTICLE	CF=۰,٥۸
BOOK	CF=۰,٤٧	BOOK	CF=۰,٥۲	AUTOMATE	CF=۰,۰۹
COMPUTER	CF=۰,۹٤	COMPUTER	CF=۰,۹۱	BOOK	CF=۰,٦۲
ENGINEERING	CF=۰,۲۱	CPU	CF=۰,۲۳	CDROM	CF=۰,۰۷
GAME	CF=۰,۳۱	ENGINEERING	CF=۰,۲٦	COMPUTER	CF=۰,۹۳
GLASS	CF=۰,۰۲	FACULTY	CF=۰,۰۹	CPU	CF=۰,۳٦
KEY	CF=۰,٥۹	GAME	CF=۰,۳۸	DRIVE	CF=۰,۱۷
LIGHT	CF=۰,۰۷	HARDWARE	CF=۰,۳٤	ENGINEERING	CF=۰,۲۱
MOBILE	CF=۰,٤۲	KEY	CF=۰,٤۷	FACULTY	CF=۰,۰٥
MONITORING	CF=۰,٥۲	MOBILE	CF=۰,٥۲	FLOPPY	CF=۰,۰۹
NOTE	CF=۰,۲٦	MONITORING	CF=۰,٦۳	GAME	CF=۰,٤۲
PAD	CF=۰,۰٤	MOUSE	CF=۰,۱۸	HARDWARE	CF=۰,٤۳
SCIENCE	CF=۰,٤۱	NOTE	CF=۰,۲۳	INTERNET	CF=۰,۲۷
UNIVERSITY	CF=۰,٦٦	SCIENCE	CF=۰,۳٥	KEY	CF=۰,٥۳
		SOFTWARE	CF=۰,۲۷	MOBILE	CF=۰,٥۹
		UNIVERSITY	CF=۰,٥٦	MONITORING	CF=۰,٥٤
				NOTE	CF=۰,۱۹
				RAM	CF=۰,۱٥
				SCIENCE	CF=۰,۳۲
				SOFTWARE	CF=۰,٤۹
				SOUND	CF=۰,۰۸
				UNIVERSITY	CF=۰,۳٥

جدول۲: محتوي پايگاه دانش در انتهاي روزسي‌ام آزمون پروکسي هوشمند براي شناسايي سايتهاي مرتبط با موضوع "COMPUTER"

محتواي پايگاه دانش در انتهاي روز سي‌ام					
ARTICLE	CF=۰,۵۱	FLOPPY	CF=۰,۱۷	PACKET	CF=۰,۲۷
AUTOMATE	CF=۰,۲۳	GAME	CF=۰,۳۷	PUSH	CF=۰,۱۱
BOOK	CF=۰,٤۲	HARDWARE	CF=۰,۵۳	RAM	CF=۰,۱۵
BOUND	CF=۰,۱۲	INTERNET	CF=۰,٦٤	SATELLITE	CF=۰,۱۱
CLASS	CF=۰,۰۹	KEY	CF=۰,٤۳	SCIENCE	CF=۰,۲۲
CLIENT	CF=۰,۲۹	LAN	CF=۰,۳۷	SERVER	CF=۰,۳۷
CDROM	CF=۰,۲۰	MESSAGE	CF=۰,۲۹	SERVICE	CF=۰,٤۱
COMPUTER	CF=۰,۹۱	METHOD	CF=۰,۲۱	SOFTWARE	CF=۰,٦۹
CPU	CF=۰,٦۱	MOBILE	CF=۰,۵۵	UNIVERSITY	CF=۰,۳٤
DATA	CF=۰,٤۹	MONITORING	CF=۰,٤٤	WEB	CF=۰,٤۳
DRIVE	CF=۰,۱۷	NOTE	CF=۰,۱۹	WINDOWS	CF=۰,٦٤
EVENT	CF=۰,۱٦	OBJECT	CF=۰,۲	WWW	CF=۰,٦۹



شكل۵: درصد يادگيري سيستم در مدت يك ماه براي موضوع "COMPUTER"



شكل٦: درصد خطاي تشخيص در مدت يك ماه براي موضوع "COMPUTER"

با توجه به نتايج تست الگوريتم ارايه شده، سرعت مناسب رشد پايگاه دانش و فراگيري سريع سيستم از مزاياي الگوريتم فوق مي‌باشد. به عبارت ديگر، لختي سيستم چندان بزرگ نبوده و در صورت نياز به تغيير معيارهاي سرند، قواعد مربوط به اين تغييرات پس از روزهاي معدودي آموخته شده و در پايگاه دانش ذخيره مي‌گردد. همچنين استفاده از فاكتور عدم قطعيت، باعث افزايش ضريب اطمينان سيستم شده و دقت بالاتري را در تشخيص موضوع صفحات وب بدست مي‌دهد. با استفاده از سيستم طراحي شده، تشخيص و درك هوشمندانه موضوع صفحات وب و سرند اطلاعات ناخواسته تا ۸۵% امكان‌پذير مي‌گردد.

بايد توجه داشت كه به هنگام بكارگيري عملي كارگزار ايجاد شده، تنها قسمت سرند كه شامل باز بيني كليد واژه‌ها و محاسبه وزن آنها است، در مسير انتقال داده قرار مي‌گيرد. از اينرو تأخير ناشي از اين كارگزار با كارگزارهاي غيرهوشمند كه از پايگاه داده استفاده مي‌نمايند تقريباً يكسان است. به عبارت ديگر، فرايند يادگيري و تازه نمودن پايگاه دانش مي‌تواند بطور موازي و يا بطور ناهمزمان و يا حتي توسط پردازنده‌اي ديگر انجام پذيرد، لذا تاخيري كه سربار استفاده از هوشمندي ايجاد مي‌نمايد آزار دهنده نمي‌باشد.

نتيجه‌گيري

روشهاي مختلفي جهت كنترل و سرند ترافيك در شبكه اينترنت مورد استفاده قرار گرفته است. يكي از مؤثرين اين روشها استفاده از يك كارگزار پروكسي هوشمند مي‌باشد. در اين تحقيق، كارگزار پروكسي و متدهاي تحليل ترافيك عبوري از آن و تركيب آن با متدهاي هوشمند مطرح گرديد. پس از آن طراحي يك كارگزار پروكسي با توانايي سرند اطلاعات با بهره‌گيري از يك پايگاه دانش

با قابليت يادگيري هوشمند ارايه گرديد. نتايج حاصل از تست سيستم طراحي شده نشان مي‌دهد که سرعت فراگيري سيستم رضايت‌بخش بوده، و خطا موجود در تشخيص موضوع صفحات وب به ميزان قابل توجهي کاهش مي‌يابد.

منابع

۱. Palme J., “Rating and filtering of scientific, technical and other network documents,” Research group for CMC (Computer Mediated Communication), Dep. of Computer and Systems Science, Stockholm University and KTH, ۱۹۹۸.

۲. Belkin. N. J., Croft, W. B., “Information filtering and information retrieval: two sides of the same coin?,” Communications of the ACM, ۳۵ (۱۲), pp. ۲٦-۳۸, ۱۹۹۲.

۳.Weber, W., “Firewall Basics,” ٤th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Yugoslavia, pp. ۳۰۰-۳۰٥, Oct ۱۹۹۹.

٤. Wong, K. Y., Yeung, K. H., “Site-based Approach in HTTP Proxy Design,” in Proceedings of the ۱۹۹۹ ICPP Workshops, Wakamatsu, Japan, pp. ۲۲۸-۲۳۳, ۱۹۹۹.

٥. Cooper & Lybrond, L. L. P., “Microsoft Proxy Server Security Evaluation,” Price-waterhouse-Coopers L. L. P. Information Technology Security Services, April ۱۹۹۷.

٦. Law, K. L. E., Nady, B. and Chapman, A., “A Scalable and Distributed WWW Proxy System,” Proceedings of IEEE International Conference on Multimedia Computing and Systems, Ottawa, Canada, pp. ٥٦٥-٥۷۱, ۱۹۹۷.

۷.Tschalar, R., “NTLM Authentication Scheme for HTTP,” available at

http://www.innovation.ch /java/ntlm.html, visited Feb. ۲۰۰۵.

۸. "iNet Filter technical specification", Rhazes Inc., available at http://www.rhazes.com/download/Technical_desc/INetFilter_Tech_desc .html , visited Feb. ۲۰۰۵.

۹. "SmartFilter ver۳,۱ product overview", Secure Computing, available at http://www.securecomputing.com/pdf/SF۳۱ProductOverview.pdf, visited Feb. ۲۰۰۵.

۱۰. Luotonen, A., Altis, K., "World-Wide Web Proxies," Computer Networks and ISDN Systems, Vol. ۲۷, No. ۲, pp. ۱٤۷-۱۵٤, ۱۹۹٤.

۱۱. Hussain, S., McLeod, R. D., "Intelligent Prefetching at a Proxy Server," Canadian Conference on Electrical and Computer Engineering, Halifax, Canada, pp ۲۰۹-۲۲۱, March ۲۰۰۰.

۱۲. Durkin, J., *Expert Systems Design & Development*, University of AKRON, Prentice Hall International Inc., ISBN ۰-۰۲-۳۳۰۹۷۰-۹, New York, ۱۹۹٤.