طرح پیشنهادی سامانه پروکسی آنلاین برای Pinterest

Next Generation Proxy for Pinterest (Online)

تاریخ نگارش نخست: ۱۳۹۹/۰۵/۲۰

تاریخ نگارش جاری: ۱۳۹۹/۰۵/۲۱

طبقه‌بندی: محرمانه

ارائه شده به:

طرح پیشنهادی سامانه پروکسی آنلاین برای Pinterest

Next Generation Proxy for Pinterest (Online)

Published Jan, 2020

Trademarks

فهرست کلی

فهرست مطالب

۱ طرح پیشنهادی سامانه پراکسی سایت Pinterest

۱-۱ مقدمه

سامانه پیشنهادی یک پراکسی نسل جدید (Next Generation Proxy) NGP می باشد که قادر است تا یک سایت یا وب سرویس را با سایتی دیگر با نام و آدرس متفاوت جایگزین کند. از این رو کاربر می‌تواند از محتوای این سایت با نامی دیگر ولی با همان امکانات استفاده کند. با این روش قابلیت تغییر محتوای وب‌سایت و سفارشی‌سازی مورد نیاز برای نمایش به کاربران فراهم می‌گردد.

این سامانه پراکسی برای استفاده از pinterest به آدرسpinterest.com در مقیاس وسیع می باشد. این سرویس از نوع به اصطلاح آنلاین بوده، به این معنی که درخواست های کاربران از هر نقطه شبکه اینترنت (داخل کشور) به نام دامنه و آدرس جدید ارسال شده و سامانه با بازسازی درخواست کاربر در سیستم خود آن را بازارسال و پس از دریافت پاسخ و نتیجه از وب سرویس اصلی برای کاربر ارسال می نماید.

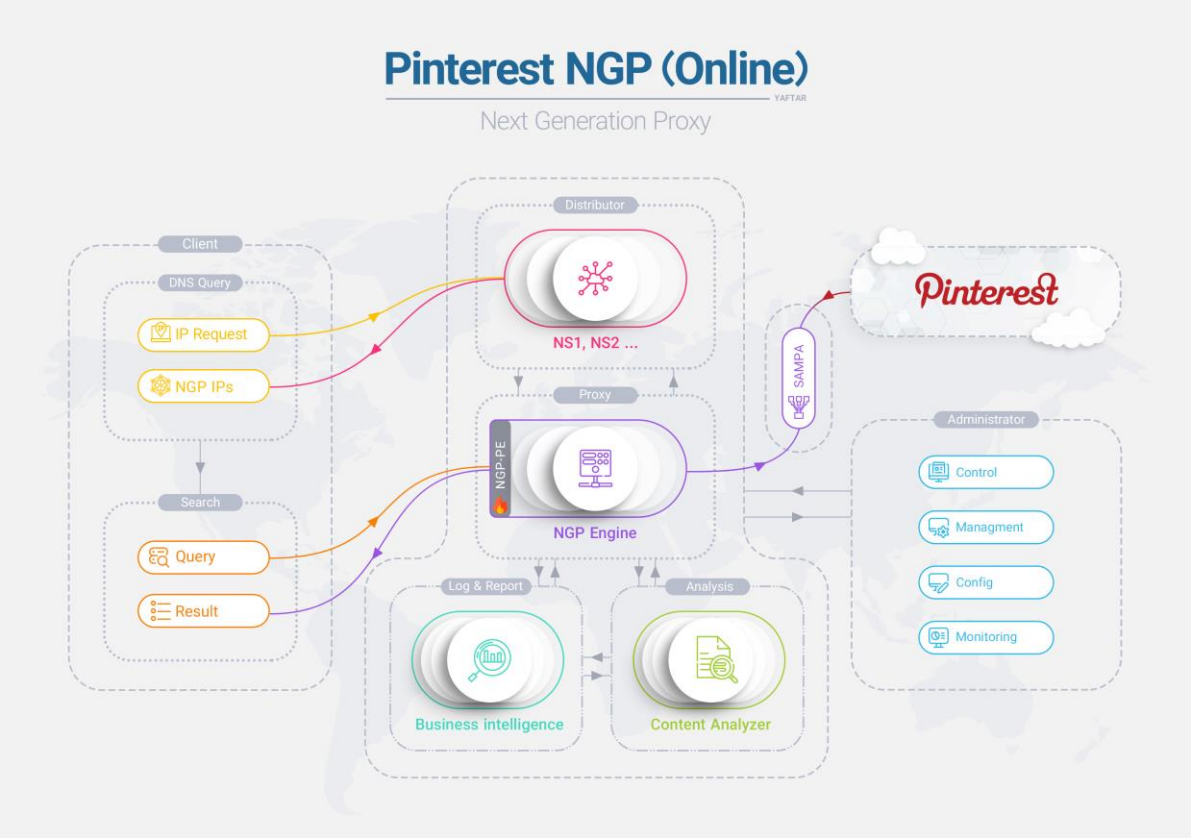
این سامانه از یک طرف امکان اعمال انواع سیاست های مدنظر برای دستیابی پاک تر و ایمن تر به اهداف مدنظر کارفرما را فراهم کرده و از طرف دیگر مناسب برای تحلیل رفتار کاربران استخراج آمارهایی با درصد خطای پایین برای تصمیم گیری های مدیریتی و فنی مبتنی بر تولیدات محصولات بومی نیز می باشد. با توجه به تکنولوژی به کار رفته در این سامانه امکان توسعه آن برای خدمات دهی به تعداد قابل توجه ای کاربر داخلی فراهم است.

در ادامه ابعاد فنی، تجهیزات و کیفیت سرویس مورد استفاده مورد بررسی قرار خواهد گرفت.

۱-۲ طرح پیشنهادی

سامانه سرویس دهی پراکسی آنلاین (NGP Online) با دریافت هر درخواست از سمت مصرف کننده، این درخواست را تبدیل به یک درخواست دیگر برای سایت pinterest کرده و برای آن ارسال می‌کند. سپس نتایج دریافتی و نتایج بدست آمده را در قالب همان صفحه ای کاملا مشابه سایت اصلی (اما با آدرس دامنه مختص به خود) در اختیار مصرف‌کننده قرار می‌دهد.

معماری سامانه به صورت درخواست – پاسخ است و درشکل ۲ نمایه معماری سامانه پراکسی Pinterest به صورت شماتیک به تصویر کشیده شده است.



شکل ۱ نمایه معماری

شکل ۲ نمایه معماری سامانه پراکسی Pinterest

همانگونه که در تصویر مشخص است، سامانه پراکسی NGP از ماژول‌های متعددی تشکیل شده است. در ادامه این بخش هر کدام از این ماژول‌ها توضیح داده خواهند شد.

۱-۲-۱ ماژول توزیع کننده[1]

وظیفه ماژول توزیع کننده، توزیع متناسب درخواست های کاربران بین ماشین های پراکسی می باشد. این مکانیزم با هدف توزیع بار کاربران، به نسبت ظرفیت پردازشی هر یک از ماشین های NGP-Engine از طریق مکانیزم های موجود در پروتکل DNS استفاده می شود. همچنین در صورتی که به هر دلیلی در عملکرد یکی از ماشین های پراکسی مشکلی پیش آید،

[1] Distribiutor

ماژول توزیع کننده به سرعت آدرس IP Valid داخلی آن ماشین را از لیست ارائه خود حذف می کند. بنابراین وظیفه دوم این ماژول برقراری Redundancy در عملکرد ماشین های NGP-Engine می باشد و از آنجا که خود این ماژول به صورت دو عدد NS ثبت شده است خود نیز دارای قابلیت Redundancy بصورت Active-Active می باشد.

بدیهی است در صورتی که هر دو ماشین های در نظر گرفته شده در این ماژول روی یک سخت افزار میزبانی شود، در زمان بروز اشکال نمی توان از Redundancy در نظر گرفته شده، استفاده نمود.

۲-۲-۱ ماژول موتور پراکسی [1]

این ماژول شامل تعدادی سامانه پراکسی متناسب با میزان بار تعیین شده در اوج ترافیک استفاده کاربر می باشد که در واقع سیستم و موتور اصلی پراکسی در این بخش قرار گرفته است. این ماژول می تواند با کمترین تاخیر درخواست کاربران را پردازش – ارسال– دریافت – باز ارسال کند.

از آنجایی که وب سایت ها مشهور نسبت به حفظ داده های خود بسیار حساس هستند، (چرا که داده ها دارایی مهم آنها محسوب می‌شوند) جهت محافظت از این داده ها، مکانیزم های مختلفی توسط آنها به کار گرفته می شود. این مکانیزم ها مبتنی بر محدود کردن تعداد درخواست برای یک کاربران است. روش های مختلفی برای تشخیص و تمایز بین کاربران و کلاینت های مختلف ایجاد شده است که تحت عنوان Fingerprinting شناخته می شوند. جهت اجرا و مدیریت پراکسی ها در بلندمدت و در مقیاس وسیع، نیاز است که این روش ها شناسایی شده و روش های مقابله با آن جهت استخراج داده وسیع استفاده شوند. این روش ها شامل استفاده مناسب از IP های مختلف، نشست های وب و موارد مهم دیگر است.

در این NGP با عدم تغییر در کوکی کاربر و استفاده از تعداد زیادی IP مختلف و ایجاد رفتار ترافیکی عادی از بلاک شدن، پیشگیری خواهد شد.

ماژول پراکسی بایستی تعداد زیادی آدرس IP Valid در اختیار داشته باشد تا بتواند بصورت Random درخواست های کاربران به وب سرور اصلی را باز ارسال نماید. هر چه تنوع و پراکندی آدرس های در اختیار بیشتر باشد، امکان شناسایی و مسدودسازی سامانه ها برای به وب سرور Pinterest کمتر خواهد بود. برای این منظور و اجرای عملیات توزیع بار، از ماجول سمپا در این خصوص استفاده شده است.

۳-۲-۱ ماژول سمپا[2]

در مواقعی که تعداد درخواست از سمت یک IP، از یک حد تعیین شده در بازه زمانی مشخص تجاوز کند، وب سایت ها مانند pinterest به آن حساسیت بیشتری پیدا خواهند داد. در چنین مواقعی نیاز است که IP مورد استفاده تغییر کند. همچنین

[1] NGP Engine

[2] SAMPA

جهت مقیاس پذیر کردن تعداد درخواست ها، نیازمند استفاده از چندین IP همزمان و توزیع بار مناسب درخواست ها به این IP خواهیم بود. سامانه سمپا جهت حل این چالش ها ایجاد شده است.

سمپا از IP های چرخشی به عنوان پراکسی استفاده می‌کند. و رابط سیستمی ارایه می دهد که استفاده کننده از این سرویس بتواند آن را مدیریت کند. تا نهایتا بتوان رفتار مودبانه ای متناسب را محدوده های تعیین شده داشته باشد.

لازم به ذکر است برای بهره مندی از IP متنوع و تعداد بالا بایستی از مکانسیم هایی مانند CG-NAT اپراتورها استفاده شود که آدرس های Valid غیرثابت و متغیر مشابه آدرس IPهای کاربران موبایل تخصیص می دهند.

۴-۲-۱ ماژول اعمال سیاست پراکسی[1]

برای اعمال سیاست مد نظر کارفرما دو روش بصورت توامان در این مرحله قابل پیاده سازی است.

- بصورت مسدودسازی برخی کلیدواژه ها که از طریق یک لیست سیاه کلید واژه اعلامی اعمال می شود و می تواند بصورت کلیدواژه های ترکیبی با روش Regex باشد، صورت می گیرد.
- روش به این صورت است که هر پاسخ درخواست کاربر (عکس ها و طرح ها) که از طریق موتور پراکسی دریافت می شود برای بار اول به سامانه تحلیل محتوا ارسال شده و در آنجا پردازش های لازم صورت می گیرد و در صورت مغایرت با سیاست های از قبل تعیین شده باشد، لینک URL آن در سایت آدرس جدید NGP برای مسدودسازی به این ماژول ارسال شده و اعمال سیاست صورت برای درخواست های بعدی صورت می گیرد.

۵-۲-۱ ماژول تحلیل محتوا

شاید بتوان گفت مهم ترین بخش این سامانه به لحاظ فنی و کار تحلیلی سنگین این ماژول می باشد. این ماژول بصورت کنار خط[2] عمل می کند و با استفاده از تصاویر دانلود شده کاربران (برای بار اول) و متن توضیحات (یا Tagهای هر تصویر) همچنین لاگ های بدست آمده از ماژول رویدادنگاری اقدام به پردازش تصاویر و متن ها نموده و تصاویر نامناسب (مطابق با سیاست اعلام شده) را با درصد خطای مشخصی شناسایی و برای ایجاد قواعد اعمال سیاست به ماژول Administrator ارسال می شود. در نهایت نتایج حاصل برای ماژول اعمال سیاست NGP-PE ارسال شده و در آنجا URL تصویر نامناسب برای کاربران پروکسی مسدود خواهد شد.

---

[1] NGP - PE

[2] NearLine

۱-۲-۶ ماژول رویدادنگاری و هوشمندسازی کسب و کار[1]

در این بخش عبارت‌هایی که کاربر برای تصاویر درخواست جست‌وجو داده است از طریق موتور پراکسی به صورت Log برای سامانه رویدادنگاری ارسال می‌شوند. در سامانه رویدادنگاری، Logها ابتدا تحلیل سپس توسط آن گزارش های پیشرفته سفارشی سازی شده از جنبه های کسب و کار و مدیریتی مبتنی بر مفاهیم BI تولید نمود. در مولفه رویدادنگاری تحلیل وزن دار کلمات درخواستی با استفاده از پاسخ ها برای مراحل بعدی جهت اعمال سیاست و مسدودسازی صورت می گیرد همچنین از طریق همین ماژول می توان به یک لیست کلمات پیشنهادی جهت مسدودسازی بصورت یک چرخه آفلاین دستیافت.

۱-۲-۷ ماژول رابط مدیریتی

جهت پایش و همچنین کنترل سامانه پراکسی ، ماژول رابط مدیریتی اجرا خواهد شد. این ماژول امکان گزارشگیری های مدیریت از منابع مصرفی، نرخ درخواست ها، وضعیت عملکرد سامانه، گزارش تفکیکی مربوط به عملکرد ماژول‌های مختلف و ارایه گزارش های تجمیعی و آماری را فراهم خواهد کرد. پارامترهای بالا یا پائین بودن سرویس پروکسی، تعداد بسته های دریافتی و ارسالی و حجم آنها نیز ارائه می شوند.

همچنین تغییرات در پیکربندی، به روزرسانی سایر ماژول ها، قوانین اعمال سیاست و سایر مولفه های مدیریتی از طریق این بخش صورت می گیرد.

۱-۳ ابعاد طرح

طرح فعلی سامانه پروکسی Pinterest ، جهت پوشش 1GPS ترافیک کاربر و حداکثر ۱۰۰ کاربر همزمان ارایه شده است. در صورت تهیه منابع، سامانه قادر است مقیاس پذیر باشد.

۱-۳-۱ محدودیت ها

نرخ پاسخ دهی موثر به ازای هر IP آماده و قابل استفاده برای درخواست به وب سایت اصلی ، ۱ درخواست بر ثانیه است و این عدد به صورت شبه‌خطی قابل مقیاس‌پذیری است. یعنی با افزایش IP، ظرفیت سامانه افزایش پیدا خواهد کرد بنابراین رسیدن به نرخ پاسخ دهی ۱ درخواست در ثانیه، با اجرای موازی درخواست ها روی IPهای متنوع قابل استفاده، ممکن می شود. البته باید متذکر شد که زمان دریافت پاسخ یک درخواست مشخص، بسته به کیفیت اینترنت کاربر و IP مربوطه پروکسی، ممکن است متغیر باشد.

[1] NGP-BI

ذکر این نکته نیز ضروری است که به دلایل مختلفی مانند طول کشیدن بیش از حد پاسخ، اختلالات اینترنتی و مواردی از این دست، بخشی از درخواست ها ممکن است در زمان تعیین شده به نتیجه نرسند. این بخش حداکثر ۵٪ خواهد بود و به تعبیر دیگر، نرخ موفقیت درخواستها بیش از ۹۵٪ خواهد بود.

۱-۴ تجهیزات و منابع

جهت پوشش سقف مورد نیاز در این طرح، تجهیزات ذیل با در نظر گرفتن اینکه بایستی همه پارامترهای مدنظر اعمال سیاست که کمترین اثر منفی در سرعت کاربران را ایجاد کند، ارائه شده است. در برآورد ذیل امکان High Availability بصورت Failover به شکل N+1 برای سرورهای پراکسی در سطح ماشین مجازی در نظر گرفته شده است.

ماشین های مجازی				
Row	Item	HA	Machine Description	Qty
1	NGP-Engine	N+1	CPU Real Core 4(Reserved), Logical Core 8(Reserved), RAM 16GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 300GB-10K(RAID Type5), NetAdapter=4*10Gbps,	4
2	NGP-Distributor	N+1	CPU Real Core 5(Reserved), Logical Core 20(Reserved), RAM 16GB(Reserved), Storage Partition1(OS): 100GB (RAID Type1)& Partition2(/var): 150GB(RAID Type5), NetAdapter=4*10Gbps,	2
3	NGP-LogServer	N+1	CPU Real Core 16(Reserved), Logical Core 32(Reserved), RAM 128GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 50TB (RAID Type5), NetAdapter=4*10Gbps,	2
4	NGP-BI	N+1	CPU Real Core 24(Reserved), Logical Core 48(Reserved), RAM 256GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 16TB (RAID Type5), NetAdapter=4*10Gbps,	2
5	NGP-MUM	N+1	CPU Real Core 16(Reserved), Logic Core 32(Reserved), RAM 256GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 2TB-10K(RAID Type5), NetAdapter=4*10Gbps,	2
6	Content-Analyzer	N+1	CPU Real Core 16(Reserved), Logical Core 32(Reserved), RAM 96GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 3TB (RAID Type5) **GPU:4* RTX 2080 ti** , NetAdapter=4*10Gbps,	2

7	NGP-Sensor	N	CPU Real Core 16(Reserved), Logical Core 32(Reserved), RAM32GB(Reserved), Storage Partition1(OS): 100GB-10K(RAID Type1) & Partition2(/var): 600GB-10K(RAID Type5),NetAdapter=4*10Gbps	2

لازم به ذکر است که از میان ماژول‌های ارائه شده در طرح پیشنهادی تنها ماژول سنسور احتیاجی به افزونگی نداشته است. همچنین نیازمندی ها شبکه و امنیت شبکه فرض بر این گرفته شده است که در شبکه کارفرما بطور کامل تامین خواهد شد.

۱-۴-۱ جزییات عملکرد

1 کاربر ابتدا در خواست DNS برای دامنه (جدید داخلی) مثلا pintterest.ir مربوط به NGP را به DNS Server خود ارسال می کند.

2 DNS Server در خواست را با لحاظ نمودن توزیع بار، با IP یکی از سامانه های NGP-Engine برمی گرداند.

3 کاربر درخواست خود را وب سایت .pintterest.ir بر روی پروتکل https به سمت سامانه ارسال می کند.

4 سامانه در خواست را به pinterest.com تغییر داده و با IP های خود به سمت وب سرویس اصلی pinterest ارسال می کند.

5 سامانه جواب را از pinterest دریافت می کند.

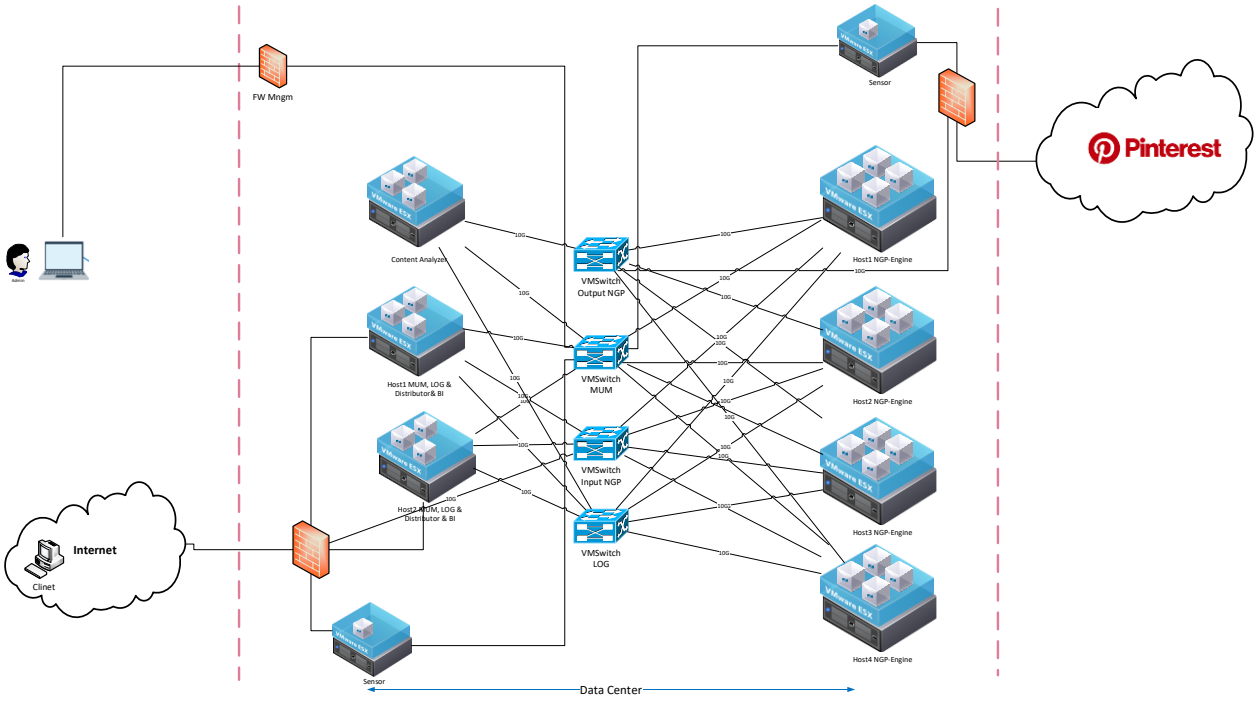
6 سامانه جواب دریافتی را به pintterest.ir تبدیل کرده و به کاربر ارسال می کند.

۱-۴-۲ معماری استقرار و توپولوژی

همانطور که در در شکل ۳ معماری استقرار و توپولوژی سامانه پروکسی نشان داده شده است برای اطمینان از عدم تاثیر سایر ماشین های مجازی دیتاسنتر بر کارائی سامانه و همچنین رعایت Redundancy در سطح سخت افزار میزبان هر چهار ماشین NGP-Engine ، روی یک سرور فیزیکی جداگانه در نظر گرفته شده است. این کار علاوه بر اطمینان از منابع پردازشی در اختیار، پهنای باند شبکه داخلی مورد نیاز مابین ماژول های مختلف سامانه را نیز فراهم می کند.

همچنین وجود چهار VM-Switch مجزا نیز برای پردازش بهتر پهنای باند داخلی ماشین های مجازی و جداسازی ترافیک Log، مانیتورینگ و مدیریت از ترافیک کاربران را محیا می کند.

نحوه استقرار ماژول سنسور باید به گونه ای باشد که کمترین تاثیر و عوارض را از سخت افزارهای ماژول های اصلی سیستم بگیرد تا بتواند به بهترین نحو وضعیت کارکردی را بررسی و اشکال یابی کند. همانطور که در شکل هم مشخص است کارت شبکه یکی از سنسورها مانند کاربران قبل از ورود به شبکه و در Zone ورودی فایروال Input قرار گرفته و ارتباط دیگر آن نیز با VM-Switch MUMبرقرار خواهد شد. یک کارت شبکه از سنسور دوم نیز به Zone خارجی فایروال Output در اینترنت و کارت شبکه دوم آن نیز مانند سنسور اول با VM-Switch MUM بر قرار می شود.



شکل ۳ معماری استقرار و توپولوژی سامانه پروکسی Pinterest

۳-۴-۱ ملزومات

۱. فراهم نمودن ماشین های مجازی مطابق طرح معماری استقرار و توپولوژی
۲. فراهم نمودن پهنای باند مورد نیاز
۳. تهیه دامنه و گواهی مطلوب جهت ارائه سرویس
۴. تهیه حداقل ۱۰۰ آدرس IP Valid متنوع برای ایجاد پراکندگی آدرس های IP

۴-۴-۱ قابلیت‌های نظارتی

۱. ارسال گزارش از وضعیت پردازنده، رم، دیسک با استفاده از snmp v3
۲. ارسال گزارش از میزان پهنای باند ورودی و خروجی به تفکیک ماشین ها

۳. قابلیت ارسال گزارش از میزان ترافیک پردازش شده

۴. ارسال پارامترهای بالا یا پائین بودن سرویس پروکسی، تعداد بسته های دریافتی و ارسالی و حجم آنها

۲ زمان‌بندی

زمان‌بندی تخمینی اجرای طرح **۳ ماه** می‌باشد.

۳ خدمات پشتیبانی و بروزرسانی

با توجه به ماهیت ابزارهای نرم‌افزاری و سخت افزاری استفاده شده در این طرح، ارائه خدمات مستمر و بروزرسانی جزء مهمی از این طرح می‌باشد. این خدمات شامل موارد زیر می‌باشد:

۱. خدمات به روزرسانی موتور پروکسی به صورت دوره ای
۲. خدمات بروزرسانی و رفع مشکلات تجهیزات نرم افزاری مورد استفاده
۳. ارائه خدمات سفارشی‌سازی محصول بر اساس نیازمندی‌های جدید
۴. پشتیبانی 24*7