



Internet measurement

Summary of the paper:

A Case Study of the Accuracy of SNMP Measurements

مینا فریدی

810100430

مقدمه و پیش زمینه

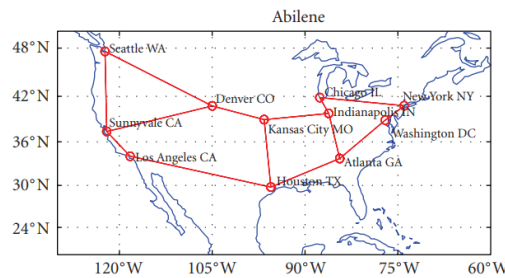
در این مقاله، یک مطالعه موردی را با تمرکز به دقت مجموعه‌ای از اندازه‌گیری‌های SNMP بارهای ترافیک لینک‌ها در شبکه Abilene ارائه شده است. SNMP یک پرتکل ساده است که کار کردن با آن بسیار آسان می‌باشد و از آن برای مدیریت شبکه تقریباً در همه جا استفاده می‌شود. به همین دلیل در نظر گرفتن دقت اندازه‌گیری‌های SNMP مهم است. مقاله دقت اندازه‌گیری‌های بار پیوند SNMP را به‌طور خاص ارزیابی می‌کند و نتایج آن نشان می‌دهد که خطاهای اندازه‌گیری لود لینک کوچک و جزئی هستند، اما توزیع آن‌ها heavy-tail است و تعداد خیلی کمی از خطاهای اندازه‌گیری می‌توانند به اندازه خود اندازه‌گیری‌ها بزرگ باشند. خطاها در الگوهای متمایز رخ می‌دهند و مقاله نیز علت آن‌ها را با استفاده از این ساختار توضیح می‌دهد.

داده‌های SNMP دارای محدودیت‌های زیادی هستند. چون SNMP از انتقال UDP غیرقابل اعتماد استفاده می‌کند داده‌ها ممکن است در حین انتقال از بین بروند، یا توسط NMS، برای مثال، اگر NMS خراب شود یا راه‌اندازی مجدد شود. داده‌ها ممکن است از طریق اجرای ضعیف عامل SNMP نادرست باشند. SNMP از شمارنده‌های 32 بیتی استفاده می‌کند و اینها می‌توانند به سرعت روی یک پیوند پرسرعت بیچند، به عنوان مثال، در کمتر از 4 ثانیه در یک پیوند 10 گیگابیت بر ثانیه، شمارنده بازنشانی می‌شود (مثلاً پس از راه‌اندازی مجدد روتر).

مشکل کالیبره کردن سیستم‌های SNMP در این است که منبع اصلی داده‌ها، داده‌های سطح جریان است که برای این کار نامناسب است، زیرا زمان‌بندی جریان‌ها تصادفی است و تنها منبع عملی داده‌های صحیح لیبل خورده، ردیابی بسته‌ها خواهد بود، و تعداد کمی از اپراتورها حاضر به پرداخت هزینه نصب و مدیریت دستگاه‌های لازم برای جمع‌آوری چنین داده‌هایی از لینک‌های پرسرعت هستند. جایگزین پیشنهادی در این مقاله استفاده از افزونگی موجود در بسیاری از مجموعه‌های داده SNMP برای کالیبره کردن داده‌ها است. به طور خاص، بسیاری از اپراتورها داده‌های SNMP را از واسط‌های دو طرف مجموعه قابل توجهی از پیوندها در شبکه خود جمع‌آوری می‌کنند. این روش از این افزونگی با انجام مقایسه بین اندازه‌گیری‌ها از دو طرف پیوند برای ارزیابی خطاها استفاده می‌کند.

داده‌ها

داده‌های استفاده شده در این مقاله، مناسب‌ترین و بزرگترین داده‌های قابل دسترسی تا آن زمان یعنی مجموعه داده Abilene است که اطلاعات اطلاعات چند ماه چندین لینک با ظرفیت بالا است که در سال 2006 گردآوری شده است که در شکل زیر نقشه این شبکه آورده شده است. همچنین از ابزار رایگان Round-Robin Database برای ذخیره و جمع‌آوری داده‌ها استفاده شده.



روش

تخمین خطاها

خطاها با مقایسه داده‌های هر انتهای یک لینک مطالعه شده است. کل ترافیک وارد شده به یک لینک در یک بازه زمانی باید تقریباً با ترافیک خروجی در همان بازه زمانی مساوی باشد. تفاوت در ترافیکی است که در ابتدا و انتهای بازه در جریان است یا به عبارتی داده‌هایی که در آن بازه زمانی روی خطوط انتقال هستند.

حداکثر ترافیک روی سیم برای یک پیوند را می‌توانیم محاسبه کنیم چون که این مقدار تنها حاصل از تاخیر پهنای باند است. در شبکه Abilene، این مقدار 109.8 مگابیت در پیوند هیوستون → لس آنجلس است. نسبت به ترافیک معمولی 5 دقیقه‌ای در این پیوند، این تنها حدود 1 قسمت در یک میلیون است. علاوه بر این، تفاوت ناشی از ترافیک روی سیم در شروع و پایان بازه زمانی (تقریباً در همه موارد) بسیار کوچک است.

تفاوت ترافیک اندازه‌گیری شده به صورت زیر است:

$$\hat{y}^{(1)}(t, t+s) - \hat{y}^{(2)}(t, t+s) = \epsilon^{(1)} - \epsilon^{(2)}$$

که در این رابطه:

$y(t, t+s)$: ترافیک در بازه $[t, t+s]$ است.

$\hat{y}(t, t+s)$: ترافیک مشاهده شده در بازه $[t, t+s]$ است.

منبع خطاها

دانستن منبع خطاها در اندازه‌گیری‌های SNMP بسیار مهم است. دو عامل اصلی که می‌توان برای این تاخیر برشمرد عبارت اند از: یک: برخی از انواع خطاهای ابزار دقیق، به عنوان مثال، عدم به‌روزرسانی صحیح شمارنده‌ها، و دو: وجود jitter در poll timing.



اگر بخواهیم یک سری اندازه گیری را بر روی $\hat{y}(t_1, t_2), \hat{y}(t_2, t_3), \dots, \hat{y}(t_n, t_{n+1})$ بازه های تجمیع کنیم تا را (t_i, t_{i+1}) داشته $\hat{y}(t_1, t_{n+1})$ باشیم، در این صورت خطا در مجموع نیز به همین ترتیب تجمیع می شود و بنابراین با فرض استقلال و ایستایی، واریانس خطا در مجموع به صورت زیر خواهد بود:

$$\text{Var}(\epsilon(t_1, t_{n+1})) = \sum_{i=1}^n \text{Var}(\epsilon(t_i, t_{i+1})) = n \text{Var}(\epsilon(t_1, t_2))$$