سلام رفقا!
امیدوارم که حالتون خوب باشه...
شنیدم که شما المپیاد کامپیوتری هستید! دنبال بهینه سازی و ایده زنی! ایده های خلاقانه!
ما به یک مسئله ای برخوردیم که دوست داشتیم با شما هم مطرحش کنیم ببینیم چه ایده هایی میتونید بزنید که کدتون بتونه تو زمان خوبی مسئله رو حل کنه!

قبلش نیازه یک سری توضیحات خدمتتون بدم...
میدونید اطلاعات داخل شبکه ی اینترنت (یا هر شبکه ی دیگه ای) چجوری جا به جا میشه؟
به صورت بسته بسته. به هر بسته از اطلاعات به انگلیسی میگن packet. هر بسته یا packet طبعا یک سری اطلاعات داخل خودش داره که قراره منتقل کنه که معمولا به اون اطلاعات میگن payload اون بسته. اما آیا هر بسته فقط payload داره؟ یعنی مثلا فرض کنید این بسته یه جایی رها شده. چجوری باید بفهمه به کی و کجا تعلق داره؟
به قول شاعر گفتنی، از کجا آمدم؟ و آمدنم بهر چه بود؟ به کجا می روم؟ و ...
طبعا هر بسته یه مبدا و مقصدی داره که نیازه مشخص بشه. نیازه بدونه از کجا اومده و داره به کجا میره.
خلاصه اینکه هر بسته باید بدونه مبدا و مقصدش کجاست و این اطلاعات رو داخل خودش ذخیره کنه.
حالا مبدا و مقصد رو چجوری نشون میدن؟
هر سیستمی یه آدرس IP داره که منحصر به فردش میکنه. (حالا ممکنه براتون سوال پیش بیاد که آیا به اندازه ی تمام سیستم های جهان IP address متفاوت وجود داره؟ بعید میدونم. جاش اینجا نیست حالا فعلا بماند. برید جستجو کنید. دوس داشتید بعدا میتونید راجع به IPv6 هم بیشتر بخونید.)
حالا خب آیا IP برای برقرار کردن ارتباط از یک سیستم کافیه؟
مثلا شما وقتی با سیستمتون کار میکنید مگه فقط در لحظه یک ارتباط با سیستمتون ایجاد میشه؟
طبعا نه! اگه هم که چند تا ارتباط ایجاد شه چجوری میشه این ارتباط ها رو از هم جدا کرد؟
این تصویر رو نگاه کنید. مثال خوبیه برای بیان این موضوع. چند تا کشتی هستن که هر کدوم یه سری بسته اوردن و یه سری بسته قراره ببرن. خب این کشتی ها که نمیتونن همشون همزمان داخل یک بندرگاه بسته هاشونو خالی کنن یا بسته هاشونو تحویل بگیرن! نیازه که هر کدوم تو یک بندرگاه جدا بسته هاشو خالی کنه و بسته های جدید رو تحویل بگیره. طبیعتا هر کدوم این کشتی ها هم میتونه بسته های مربوط به یک هدف یا مقصد متفاوت رو دنبال کنه. پس هر کدوم هم نیازه یک بندرگاه مشخص و متفاوت داشته باشن برای اینکه بارشون رو خالی کنن یا تحویل بگیرن. به عبارتی بیایم و بار کشتی ها رو بارگیری (Download) کنیم یا بیایم و داخلشون بارگذاری (Upload) کنیم!

حالا این بندرگاه یا همون port اینجا هم معنا یه همچین چیزیه. هر سیستم یه تعدادی port داره که هر کدوم با یک شماره ای مشخص میشه.
پس میشه گفت هر بسته دارای IP و port number برای مبدا و به طور مشابه IP و port number برای مقصده که تکلیف هر بسته رو مشخص میکنه.
در حال حاضر IP های رایج ۳۲ بیتی هستند و به صورت ۸ بیت ۸ بیت جدا میشن، سپس هر کدوم این ۸ بیت رو به صورت یک عدد در مبنای ۱۰ مینویسن. به طور مثال 11000000.10101000.00000001.00000001 نشون دهنده 192.168.1.1 است. همچنین شماره port ها نیز به طور رایج یک عدد ۱۶ بیتی هستند.

حالا مسئله چیه؟
اینترنت شرکت داره سر یه سری موضوعاتی صرف میشه که نمیدونیم چیه و اصلا لازمه یا نه! و تو یه شرکت به این بزرگی این حجم اینترنت هزینه ی محسوسی داره و باید به فکر صرفه جویی و مصرف مدیریت شده و بهینه ازون باشیم.
برای اینکار، تصمیم گرفتیم ترافیک عبوری از شبکه ی شرکت رو بررسی کنیم و ببینیم هر بسته ای که از شبکه رد میشه چقدر سود برای ما داره و چقدر استفاده ی مفید داره ازش میشه.
برای این کار یک سری شرط مشخص کردیم که به کمک اونها میزان سود هر بسته رو محاسبه کنیم. سود هر بسته اندازه ی تعداد شرط هاییه که اون بسته در آنها صدق میکنه.
حالا ما اومدیم ترافیک عبوری سیستم های شبکه مون رو capture کردیم. به عبارت خودمونی تر ما یک فایل pcap ذخیره شده داریم که حاوی یک سری بسته (packet) است. حالا ما یک سری شرط داریم که می خوایم بدونیم هر کدوم از این بسته ها توی چند تا از این شرط ها صدق می کنه. به عبارتی میزان سودی که هر بسته داره رو محاسبه کنیم.
ازونجایی که میزان ترافیک شرکت خیلی زیاد و سنگینه، نیازه که از روش هایی استفاده کنیم که بتونه مجموعا بسته هارو تو کمترین زمان ممکن پردازش کنه و میزان سود هر کدوم رو بدست بیاره.
برای اینکار میتونید هر ایده ای که دلتون میخواد بزنید هر بهینه سازی و به قول خودمون تفی که دلتون میخواد بزنید تا کدتون بتونه سریع ترین نحوی که بنظرتون ممکنه عمل کنه.

ورودی مسئله چیه و چجوریه؟
- یک فایل rules.conf داریم که داخلش شرط هایی که گفتیم رو نوشتیم. این فایل 10.000 خط داره که هر خط یک شرط رو نشون میده. هر شرط به ترتیب از چپ شامل موارد زیر است.
“source IP range” “source port num” “destination IP range” “destination port num”

IP range یا بازه IP به شکل a.b.c.d/e نمایش داده میشه که e تعداد بیت های ثابت سمت چپشه. یعنی مثلا e تا بیت های سمت چپ این رشته ۳۲ بیتی مشخص و ثابته اونارو مینویسیم و e – ۳۲ بیت سمت راست رو هم صفر میذاریم. حالا رشته ی دودویی رو بعد ۴ تکه شدن در مبنای ۱۰ مینویسم میشه a.b.c.d که نشان دهنده ی شروع بازه ای از IP هاست که این e – ۳۲ بیت متغیر سمت راست میسازند.
port number هم که به صورت یک عدد ۱۶ بیتی در مبنای ۱۰ عه. اگر صفر بود یعنی port number محدودیت خاصی نداره و نیاز به بررسی کردنش نیست اگر صفر نبود یعنی باید port number مورد نظر مساوی اون باشه.
تطبیق با یک شرط در صورتی رخ می دهد که IP مبدا و مقصد در بازه ی مورد نظر باشد و شماره port مبدا و مقصد با آن صدق کند.
- یک فایل با فرمت pcap نیز در اختیار شما قرار خواهد گرفت که حاوی n بسته ی ذخیره شده است.

یک ابزار کاربردی تحلیل فایل های pcap نرم افزار wireshark است که می توانید از پیوند زیر آنرا دریافت نمایید. https://soft98.ir/internet/network/13907-wireshark.html

خروجی چیه و چجوری؟
یک فایل txt با n خط. که هر خط نشون دهنده ی اینه که بسته ی متناظر با اون خط تو چند تا شرط صدق میکنه یا به عبارتی میزان سودش چقدره.

نحوه داوری هم اینجوریه که یک script (اسکریپت یه ورد جادوییه که توش یه سری دستورات رو مینویسن که به جایی که این دستور هارو تک به تک تو ترمینال اجرا کنن، یک راست اسکریپت رو اجرا کنن که خودش خط به خط اون دستورا رو تو ترمینال اجرا کنه) داریم که میاد کد شمارو اجرا می کنه و زمان اجرا و بدست اوردن خروجی رو براش محاسبه میکنه. اول از همه میاد مقایسه میکنه ببینه خروجیش خروجی درستی هست یا نه بعدش نشون میده که زمان اجرای کد شما چقدر بوده.
حالا میخوایم ببینیم کدوم کد بهترین زمان اجرا رو داره.
این اسکریپت رو برای آزمایش کردن نمونه هایی که در اختیارتون قرار دادیم بهتون میدیم.

برای اجرای اسکریپت و حالا احتمالا library هایی که در ادامه نیاز دارید و نوشتن cmake احتمالا خوبه که linux داشته باشید. اگه ندارید پیشنهاد میکنم از wsl استفاده کنید. wsl چیه و چجوری نصبش کنیم؟ برید بخونید دیگه. مثلا کامپیوتری اید. از قدیم گفتن کامپیوتریه و سرچش.
سعی کنید از IDE مناسب برای نوشتن پروژه استفاده کنید. پیشنهاد من Clion یا نهایتا vscode عه.

شاید یکی از نکاتی که توجهتونو جلب کرده باشه این باشه که خب چجوری پکت هارو بخونیم؟ چجوری پردازششون کنیم؟ چجوری باهشون کار کنیم؟
اینجاست که باز اون روز کامپیوتری خودتون رو باید نشون بدید و برید تو اینترنت جستجو کنید.
حالا بخاطر محدودیت زمان ما هم یکم راهنمایی و سرنخ بهتون میدیم که کمکتون کنه.
دو تا کتابخونه یا library خوب برای کار با پکت، libpcap و PcapPlusPlus هستن که من به شخصه دومی رو تجربه کردم تو پروژه ی C++ ای، و تجربه ی خوبی داشتم. رو این حساب همونو پیشنهاد میکنم بهتون. میتونید برای جزئیات بیشتر به https://pcapplusplus.github.io/ مراجعه نمایید. خوشبختانه مستندات خیلی خوب و روونی هم داره برای کار باهاش.
اول کار که باید فایل های باینری کامپایل شده ی این کتابخونه رو با توجه به سیستم عاملتون از گیتهابش دانلود کنید. https://github.com/seladb/PcapPlusPlus/releases/tag/v23.09
خب اول از همه باید اینو راه بندازیمش. آموزشش تو همون مستنداتی که گفتم هست. داخل لینک اولیه GET STARTED رو بزنید میتونید جزئیاتشو بر اساس سیستم عاملتون مشاهده کنید.
" همونطور که قبل تر گفتم، پیشنهاد ما به شما استفاده از لینوکسه. اگر اوبونتو رو سیستم تون داشته باشید که خیلی خوبه اگر نه هم پیشنهاد میکنیم با wsl راه بندازیدش. "
حالا این وسط اوایل کار احتمالا برخوردید به یه چیزی به اسم Cmake و واستون سوال شده که این چیه؟ در وهله اول خوبه خودتون سعی کنید جوابشو پیدا کنید. اینترنت!
حالا منم یکم تقلب میرسونم بهتون این دفعه فقط خیلی توش عمیق نشید فعلا نیازی نیستش. گاماس گاماس. صرفا الان چند خط بخونید شهود بگیرید چیه. https://melec.ir/cmake-tutorial/
حالا ما هم برای اینکه یه پروژه بزنیم که توش از مثلا کتابخونه ی++pcap هم بخوایم استفاده کنیم، نیازه پروژه مون یه CMakeLists.txt داشته باشه. اگر که با Clion دارید کار میکنید خودش یه نسخه ی اولیه براش میسازه که پروژه عادیتون رو اجرا کنه و شما صرفا باید بابت استفاده از کتابخونه اتون یه چیزای کوچیکی بهش اضافه کنید. بهتون پیشنهاد میکنم الان توش خیلی عمیق نشید تو مثالایی که واسه ++pcap داخل مستنداتش هست cmakelist های ساده هم کنارشون گذاشته که میتونید همونارو کپی کنید و ازشون استفاده کنید.
حالا برگردیم به ++pcap، داخل همون مستنداتش بخش توتوریال یه سری بخش آموزش داره که اینم پیوندشه: https://pcapplusplus.github.io/docs/tutorials اینجا میتونید بخشای مختلف آموزش کار با ++pcap رو بررسی کنید. بخش دوم و بخش چهارم دو تا بخش اصلی ایه که تو این پروژه شما بهش نیاز دارید.

ما هم برای کمک کنارتون هستیم.

موفق باشید!