Czytaj lub słuchaj artykuł online

Zeskanuj kod QR, aby otworzyć artykuł na XO.PL

Źródło artykułu:

https://xo.pl/it-na-wesolo-net-traceroute

SIECI KOMPUTEROWE

Traceroute, czyli jak sprawdzić, przez które węzły przejechał pakiet

Na Autostradzie Pakietów trwa zwyczajny dzień. Samochody wożą dane, routery sprawdzają tablice, a kierownik **NAT** znowu szuka zeszytu z numerami rejestracyjnymi.

Do biura administratora przychodzi zgłoszenie.

— Internet nie działa.
— Cały?
— Nie wiem. Strona się nie otwiera.
— Jaka strona?
— Ta, co wczoraj działała.

Administrator wysyła pakiet. Cisza. Drugi pakiet. Nadal nic.

— Skoro samochód nie dojeżdża do celu, sprawdźmy, którędy jedzie.
— A jak? Przecież nie mamy kamer na każdym skrzyżowaniu!
— Kamer nie. Za to mamy **traceroute**.

Jak zrobić mapę drogi, nie mając mapy?

W poprzednim artykule o **TTL** ustaliliśmy, że pakiet dostaje ograniczoną liczbę przeskoków. Routery zmniejszają ten licznik, a kiedy się wyczerpie, mogą odesłać ICMP *Time Exceeded*.

Administrator postanawia wykorzystać ten mechanizm w sposób, którego zwykły kurier nie pochwaliłby nawet po trzech piwach.

— Pierwszemu samochodowi dajemy **TTL** = 1.
— Przecież on nie dojedzie!
— Właśnie o to chodzi.

Pierwszy router zatrzymuje samochód i może wysłać zawiadomienie. Administrator zapisuje adres. Drugi samochód dostaje **TTL** = 2, więc dociera do następnego routera. Potem **TTL** = 3, **TTL** = 4… I tak powstaje przybliżony obraz kolejnych węzłów.

Traceroute nie czyta cudzych tablic routingu. Odkrywa węzły, wysyłając serie prób z rosnącym limitem przeskoków i analizując otrzymane odpowiedzi.

Trzy samochody, trzy mandaty

Klasyczny traceroute często wysyła trzy próby dla każdego **TTL**. Jedna odpowiedź może nie wrócić, druga się spóźnić, a trzecia przejechać inną drogą.

Przykładowy wynik z adresami przeznaczonymi do dokumentacji:

```
traceroute to example.net (203.0.113.80)
 1  192.0.2.1       0.8 ms   0.7 ms   0.9 ms
 2  198.51.100.1    4.2 ms   4.1 ms   4.5 ms
 3  198.51.100.2    9.8 ms  10.1 ms  10.0 ms
 4  *               *        *
 5  203.0.113.80   18.3 ms  18.6 ms  18.4 ms
```

— Pierwszy węzeł jest, drugi jest, trzeci jest… A czwartego kto ukradł?!

Gwiazdki, czyli posterunek, który nie odbiera telefonu

Gwiazdka * oznacza, że w zadanym czasie nie przyszła odpowiedź na daną próbę. **Nie oznacza automatycznie, że router nie działa albo że pakiety nie mogą przejechać dalej.**

— Poproszę potwierdzenie przejazdu.
— Nie wydajemy.
— Ale samochody normalnie jeżdżą?
— Oczywiście.
— To dlaczego nie wydajecie?
— Bo nie wydajemy.

Odpowiedzi ICMP bywają filtrowane lub ograniczane. Jeśli czwarty węzeł milczy, a piąty odpowiada, nie ma sensu natychmiast ogłaszać katastrofy na czwartym skrzyżowaniu. Chociaż administratorzy lubią dramatyzować. 🤣

Windows i Linux: dwa wydziały jednego urzędu

Windowsowy `tracert` używa zwykle ICMP Echo. Klasyczny linuksowy `traceroute` często domyślnie wysyła **UDP** na wysokie porty. Istnieją też warianty korzystające z ICMP i **TCP**. Cel podobny, ale firewalle i routery nie muszą traktować tych prób tak samo.

— Pan ma **UDP**?
— Tak.
— A kolega?
— ICMP.
— A trzeci?
— **TCP**.
— Panowie przecież pytacie o tę samą drogę!
— Tak, ale każdy dostał inny formularz.

Milisekundy to nie czas między bramkami

Czasy przy kolejnych węzłach to zazwyczaj **RTT** — od wysłania próby do odebrania odpowiedzi. Nie jest to czas przejazdu między dwoma sąsiednimi routerami. Każdy pomiar obejmuje podróż do odpowiadającego węzła i z powrotem.

Router może przekazywać zwykły ruch szybko, ale obsługiwać odpowiedzi diagnostyczne wolniej. Jeśli pojedynczy hop odpowiada po 300 ms, a dalsze po 20 ms, nie oznacza to automatycznie problemu z przekazywaniem pakietów na tym pierwszym.

— Panie kierowniku, samochody jadą bez opóźnień, tylko wydawanie zaświadczeń trwa trzysta milisekund!
— A samochody?
— Jadą.
— To proszę nie przeszkadzać urzędnikom.

Do pingów, **jittera** i strat pakietów wrócimy w osobnym odcinku.

Trzy próby i trzy różne drogi

Przy równoważeniu ruchu, np. przez ECMP, różne przepływy mogą korzystać z różnych tras. Traceroute wysyła kilka prób, więc wynik czasem łączy węzły z różnych ścieżek — zamiast pokazywać jedną podróż jednego samochodu. Odmiany takie jak Paris traceroute próbują ograniczać ten efekt.

— Jak on pojechał przez Warszawę, potem przez Poznań, a później wrócił do Warszawy?!
— To nie był ten sam samochód.
— Jak to?
— Wysłał pan trzy.

Także droga powrotna odpowiedzi ICMP może być zupełnie inna. Internet nie obiecuje tras symetrycznych: odpowiedź od routera dociera do nas swoją trasą. Traceroute nie rekonstruuje automatycznie całej drogi powrotnej.

Kiedy traceroute trafia do odległej galaktyki

Na koniec administrator sprawdza jeszcze jeden adres. Nie dlatego, że coś nie działa. Po prostu ktoś powiedział mu, że w Internecie istnieje trasa, która opowiada historię.

```
traceroute -6 -I -m 100 obiwan.qwt.ch
```

Pierwszych jedenaście węzłów wygląda zwyczajnie: **IPv6**, routery, milisekundy. A potem…

Fragment rzeczywistego pomiaru z produkcyjnego serwera

IPv6

Zwykły tekst

```
 1  2001:db8::1 (2001:db8::1)  0.798 ms * *
 2  2001:db8::2 (2001:db8::2)  0.174 ms  0.241 ms  0.244 ms
 3  2001:db8::3 (2001:db8::3)  0.650 ms  0.719 ms  0.831 ms
 4  2001:db8::4 (2001:db8::4)  0.132 ms  0.219 ms  0.221 ms
 5  * * *
 6  2001:db8::6 (2001:db8::6)  1.102 ms * *
 7  2001:db8::7 (2001:db8::7)  1.302 ms * *
 8  * * *
 9  2001:db8::9 (2001:db8::9)  22.028 ms * *
10  2001:db8::a (2001:db8::a)  20.438 ms * *
11  2001:db8::b (2001:db8::b)  20.561 ms * *
12  Episode.iv (2a03:4000:5a:cd:dead:beef:cafe:0)  21.518 ms * *
13  A.NEW.hope (2a03:4000:5a:cd:dead:beef:cafe:1)  23.054 ms * *
14  It.is.a.period.of.civil.war (2a03:4000:5a:cd:dead:beef:cafe:2)  23.005 ms * *
15  Rebel.spaceships (2a03:4000:5a:cd:dead:beef:cafe:3)  25.158 ms * *
16  striking.from.a.hidden.base (2a03:4000:5a:cd:dead:beef:cafe:4)  24.145 ms * *
17  have.won.their.first.victory (2a03:4000:5a:cd:dead:beef:cafe:5)  23.145 ms * *
18  against.the.evil.Galactic.empire (2a03:4000:5a:cd:dead:beef:cafe:6)  23.372 ms * *
19  During.the.battle (2a03:4000:5a:cd:dead:beef:cafe:7)  25.517 ms * *
20  Rebel.spies.managed (2a03:4000:5a:cd:dead:beef:cafe:8)  25.843 ms * *
21  to.steal.secret.plans (2a03:4000:5a:cd:dead:beef:cafe:9)  24.117 ms * *
22  to.the.Empires.ultimate.weapon (2a03:4000:5a:cd:dead:beef:cafe:a)  24.458 ms * *
23  the.DEATH.star (2a03:4000:5a:cd:dead:beef:cafe:b)  22.937 ms * *
24  an.armored.space.station (2a03:4000:5a:cd:dead:beef:cafe:c)  25.089 ms * *
25  with.enough.power.to (2a03:4000:5a:cd:dead:beef:cafe:d)  22.930 ms * *
26  destroy.an.entire.planet (2a03:4000:5a:cd:dead:beef:cafe:e)  23.784 ms * *
27  Pursued.by.the.empires (2a03:4000:5a:cd:dead:beef:cafe:f)  25.727 ms * *
28  sinister.agents (2a03:4000:5a:cd:dead:beef:cafe:10)  23.341 ms * *
29  Princess.Leia.races.home (2a03:4000:5a:cd:dead:beef:cafe:11)  25.539 ms * *
30  aboard.her.starship (2a03:4000:5a:cd:dead:beef:cafe:12)  23.941 ms * *
31  custodian.of.the.stolen.plans (2a03:4000:5a:cd:dead:beef:cafe:13)  24.389 ms * *
32  that.can.save.her (2a03:4000:5a:cd:dead:beef:cafe:14)  22.814 ms * *
33  people.and.restore (2a03:4000:5a:cd:dead:beef:cafe:15)  25.616 ms * *
34  freedom.to.the.galaxy (2a03:4000:5a:cd:dead:beef:cafe:16)  26.575 ms * *
35  0-----I-------I-----0.ll (2a03:4000:5a:cd:dead:beef:cafe:17)  22.838 ms * *
36  0------------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:18)  24.788 ms * *
37  0-----------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:19)  23.205 ms * *
38  0----------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1a)  22.882 ms * *
39  0---------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1b)  24.455 ms * *
40  0--------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1c)  23.004 ms * *
41  0-------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1d)  24.749 ms * *
42  0------------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1e)  22.795 ms * *
43  0-----------0.ll (2a03:4000:5a:cd:dead:beef:cafe:1f)  23.324 ms * *
44  0----------0.ll (2a03:4000:5a:cd:dead:beef:cafe:20)  25.684 ms * *
45  0---------0.ll (2a03:4000:5a:cd:dead:beef:cafe:21)  23.523 ms * *
46  0--------0.ll (2a03:4000:5a:cd:dead:beef:cafe:22)  22.794 ms * *
47  0-------0.ll (2a03:4000:5a:cd:dead:beef:cafe:23)  25.269 ms * *
48  0------0.ll (2a03:4000:5a:cd:dead:beef:cafe:24)  23.342 ms * *
49  0-----0.ll (2a03:4000:5a:cd:dead:beef:cafe:25)  22.880 ms * *
50  0----0.ll (2a03:4000:5a:cd:dead:beef:cafe:26)  24.875 ms * *
51  0---0.ll (2a03:4000:5a:cd:dead:beef:cafe:27)  24.452 ms * *
52  0--0.ll (2a03:4000:5a:cd:dead:beef:cafe:28)  23.143 ms * *
53  0-0.ll (2a03:4000:5a:cd:dead:beef:cafe:29)  24.602 ms * *
54  00.ll (2a03:4000:5a:cd:dead:beef:cafe:2a)  23.148 ms * *
55  0.ll (2a03:4000:5a:cd:dead:beef:cafe:2b)  22.966 ms * *
56  By.Felix (2a03:4000:5a:cd:dead:beef:cafe:2c)  23.204 ms * *
57  Read.more.at.blog.qwt.ch (2a03:4000:5a:cd:dead:beef:cafe:2d)  20.339 ms * *
```

Pierwsze jedenaście węzłów zanonimizowano, zachowując ich numery i czasy odpowiedzi. Poniżej pełne 57 hopów, bez skracania dalszej trasy.

Administrator przeciera oczy.

— Zaraz. Co to jest?!
— Gwiezdne Wojny.
— Ale ja uruchomiłem traceroute!
— Zgadza się.
— A dlaczego router opowiada historię o statkach kosmicznych?!
— Bo może. 🤣

To rzeczywista demonstracja **IPv6**, Hop Limit i odwrotnego **DNS**, czyli rekordów PTR. Autor przygotował adresy i odpowiedzi tak, aby nazwy kolejnych hopów tworzyły opowieść. Nie znaczy to, że pakiet odwiedził 57 oddzielnych fizycznych routerów — część trasy jest specjalnie zaaranżowana. O PTR pisaliśmy też przy **DNS**.

Możesz spróbować sam, jeśli masz działające **IPv6**. Dodaj później `-n`: bez rozwiązywania nazw zamiast opowieści zobaczysz same adresy. I właśnie dlatego informatyka potrafi być świetną zabawą.

— To wracamy do diagnozowania awarii?
— Za moment. Najpierw sprawdzę, czy ktoś zrobił Władcę Pierścieni.

Ostatni posterunek

Administrator wraca do zgłoszenia. Na ekranie kilkanaście adresów, trochę gwiazdek i mnóstwo milisekund.

— No i co? Wiemy, dlaczego strona się nie otwiera?
— Jeszcze nie.
— To po co wysyłaliśmy te samochody?!
— Żeby sprawdzić, gdzie je widziano.
— Czyli awaria na szóstym?
— Niekoniecznie.
— To gdzie?!
— Właśnie dlatego mamy pracę.

Dzwoni telefon.

— Halo, administracja? Strona już działa.
— Co się zmieniło?
— Nic. Samo się naprawiło.

Administrator zamyka terminal i zapisuje w zgłoszeniu: **„Problem ustąpił przed zakończeniem diagnostyki”.**

Gdzieś na Autostradzie Pakietów sześciu policjantów nadal zastanawia się, dlaczego od rana zatrzymywali samochody z karnetami na jeden, dwa i trzy przejazdy. Jeden podobno wystąpił o dodatek za pracę z administratorami. 🤣