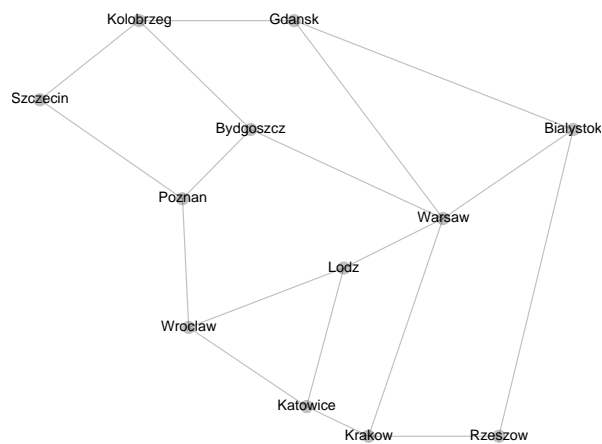


Od początku powstania Internetu obserwujemy ciągle wzrost wolumenu ruchu w sieci. Szybkość tego wzrostu nie słabnie. Wręcz przeciwnie – wymuszone pandemią COVID-19 przeniesienie wielu aktywności do sieci wręcz przyspieszyło jego skalę. Aby sprostać rosnącemu zapotrzebowaniu operatorzy muszą rozbudowywać swoje sieci, co często wiąże się z kosztownymi inwestycjami infrastrukturalnymi (np. położenie nowego kabla między miastami). Innym rozwiązaniem, pozwalającym na obsłużenie gwałtownie wzrastającego wolumenu ruchu jest pogorszenie jakości oferowanych usług. Zostało ono zastosowane na przykład podczas pierwszego lockdownu, kiedy Komisja Europejska wymusiła na dużych dostawcach VoD (YouTube, Netflix) pogorszenie domyślnie oferowanej rozdzielczości wideo.

Istnieją jednak rozwiązania alternatywne, nie wiążące się z kosztownymi inwestycjami ani pogorszeniem jakości oferowanych usług. Polegają one na bardziej inteligentnym zarządzaniu ruchem w istniejącej infrastrukturze. Obecnie standardem w sieciach jest używanie wielu tras do przesyłania pakietów pomiędzy dwoma określonymi węzłami w sieci, ale tylko wtedy kiedy koszty (długości) tych tras są takie same (tzw. *ECMP*). Oznacza to, że np. do przesyłania danych pomiędzy Krakowem i Łodzią zostaną użyte dwie trasy: Kraków-Warszawa-Łódź i Kraków-Katowice-Łódź, ale pomiędzy Krakowem i Warszawą zostanie użyta tylko trasa bezpośrednia (Kraków-Warszawa), ponieważ nie istnieją w sieci trasy alternatywne o takiej samej długości. Gdy trasa ta zostanie przeciążona ruchem, alternatywne trasy okrężne (np. Kraków-Rzeszów-Białystok-Warszawa) nie będą używane, mimo że może być na nich dostępny duży zapas przepustowości.



Rozwiązaniem tego problemu jest zastosowanie routingu adaptacyjnego i wielodrogowego opartego na przepływach. Polega on na dynamicznym dobieraniu tras dla poszczególnych przepływów (połączeń) w sieci z uwzględnieniem aktualnego czy też przewidywanego rozkładu ruchu w sieci. Jako, że za większość wolumenu ruchu odpowiedzialna jest stosunkowo niewielka liczba przepływów (tzw. *elephant flows*), bardziej korzystny rozkład ruchu można osiągnąć sterując indywidualnie tylko tą niewielką grupą.

Od kilku lat jest to najbardziej obiecujący obszar badań w inżynierii ruchu sieciowego (*traffic engineering, TE*). Rocznie publikowanych jest kilkadziesiąt artykułów naukowych prezentujących kompletne systemy TE, które, według ich autorów w najlepszy możliwy sposób rozwiązują ww. problemy. Kluczową zmienną wpływającą na wyniki prezentowane w tych badaniach jest jednak użyty model ruchu, a w szczególności rozkłady takich parametrów jak długość i wielkości przepływów. Odpowiednie i wiarygodne modele nie są dostępne w literaturze. W rezultacie, w badaniach prowadzonych przez różne zespoły używane są różne modele ruchu, co uniemożliwia porównywanie ich wyników, wyłonienie najlepszych rozwiązań, a w efekcie postęp naukowy i technologiczny. Co więcej, modele te często są zbyt uproszczone lub celowo dobierane tak, aby wykazać wyższość proponowanego przez autorów systemu nad innymi.

Mając na uwadze powyższe, celem projektu będzie stworzenie precyzyjnych i reprodukowalnych modeli statystycznych przepływów w Internecie. Stworzone modele będą mogły być używane zarówno w obliczeniach analitycznych, jak i w celu generowania realistycznego ruchu w symulatorach sieciowych. Stworzone zostaną modele ogólnego ruchu, jak też i modele podzbiorów ruchu (TCP, UDP, z wykluczeniem DNS, itp.). Rozpatrywane będą różne definicje przepływów, zarówno w warstwie sieciowej (host do hosta) jak i warstwie transportowej (przepływy definiowane przez *5-tuple*), jak też różne wartości timeout, zaczynając od milisekund (tzw. *flowlets*), aż do 60 sekund. Gotowe do użycia modele zostaną upublicznione w formule open source, gdzie użytkownik po wybraniu źródła ruchu i definicji przepływu będzie mógł ściągnąć odpowiedni model. Kończącym zadaniem w projekcie będzie próba stworzenia ujednoliconego sparametryzowanego modelu obejmującego przepływy z różnymi wartościami timeout.