

Poznań, dnia 10 maja 2010 r.

Prof. dr hab. inż. Wojciech Kabaciński
Katedra Sieci Telekomunikacyjnych i Komputerowych
Politechnika Poznańska

RECENZJA

rozprawy doktorskiej
mgr. inż. **Macieja Piechowiaka**

**pt.: Badania algorytmów heurystycznych dla połączeń rozgałęźnych
w sieciach pakietowych**

Tematyka i cel rozprawy

Recenzowana rozprawa doktorska mgr. inż. Macieja Piechowiaka dotyczy metod oceny algorytmów stosowanych do zestawiania połączeń rozgałęźnych w sieciach telekomunikacyjnych. Połączenia rozgałęźne, czyli od jednego nadawcy do wielu odbiorców, stanowią coraz większy procent wszystkich połączeń zestawianych w sieci. Przesyłanie informacji od nadawcy do wielu odbiorców może być realizowane przez przesyłanie przez sieć tej samej informacji wielokrotnie, osobno dla każdego odbiorcy. Jednakże prowadzi to do nieefektywnego wykorzystania zasobów sieci, a czasami wręcz do ich przeciążenia i niemożliwości realizacji innych połączeń. Dlatego zamiast przysyłać wielokrotnie tę same dane, stosuje się połączenia rozgałęźne, w których powielenie danych do większej liczby odbiorców następuje jak najpóźniej na drodze połączenia. Zestawienie takiego połączenia wymaga wyznaczenia w sieci odpowiedniej drogi połączeniowej, która nazywa się drzewem dystrybucyjnym. Zadaniem algorytmu sterowania połączeniami w sieci jest utworzenie takiego drzewa, przy czym powinno ono być jak najefektywniejsze, przy czym miarą efektywności może być koszt połączenia (cena zajmowanych zasobów), opóźnienie przesyłanej informacji, bądź inne parametry określające jakość połączenia i wykorzystania zasobów. Wyznaczenie takiego połączenia w sieci nie jest problemem trywialnym a problem wyznaczenia dokładnego takiego drzewa jest zadaniem trudnym (NP-zupełnym). Dlatego w praktyce stosuje się różne algorytmy heurystyczne, które z większą lub mniejszą dokładnością wyznaczają drogę bliską optymalnej z większą lub mniejszą dokładnością w ograniczonym czasie. Poważny problem stanowi ocena i porównanie między sobą różnych algorytmów heurystycznych. Z literatury dotychczasowej wiadomo, że model sieci przyjęty do oceny algorytmów ma istotny wpływ na wyniki algorytmów, to znaczy algorytmy, które osiągają dobre charakterystyki dla jednego typu sieci, mogą być znacząco gorsze w przypadku zastosowania ich w sieci o innej topologii. Dlatego sposób oceny algorytmów i ich porównywanie stanowi istotny problem nie tylko naukowy i praktyczny. Ma to szczególnie ważne znaczenie we współczesnych sieciach, w których rozwój usług multimedialnych, telewizji internetowej, czy różnego rodzaju połączeń konferencyjnych, wymaga stosowania tego typu połączeń. Uważam zatem, że cel jaki postawił sobie doktorant, czyli „... *zapropozowanie nowych, efektywnych algorytmów heurystycznych dla realizacji połączeń rozgałęźnych w sieciach pakietowych oraz opracowanie skutecznej metodologii badań porównawczych algorytmów routingu rozgałęźnego.*” jest ważny naukowo i aktualny.

Omówienie treści rozprawy

Praca składa się z wprowadzenia, pięciu rozdziałów i podsumowania. We wprowadzeniu Doktorant przytacza tematykę pracy oraz wyjaśnia znaczenie realizacji połączeń rozgałęźnych we współczesnych sieciach pakietowych. Formuluje także przytoczony wcześniej cel rozprawy oraz przedstawia zakres pracy.

Rozdział drugi stanowi wprowadzenie do problemu optymalizacji połączeń rozgałęźnych. Na początku zawiera podstawowe pojęcia teorii grafów, znajdowania najkrótszej ścieżki w grafie (bez ograniczeń i z ograniczeniami na opóźnienie) oraz wyznaczania minimalnego drzewa rozpinającego. Następnie omawia problem znajdowania minimalnego drzewa transmisji grupowej, zwanego też problemem Steinera. Dalej w rozdziale podano metodę dokładnego wyznaczania takiego drzewa MST (ang. *Minimum Steiner Tree*), która ze względu na złożoność problemu może być stosowana praktycznie tylko w sieciach o niewielkiej liczbie węzłów, oraz przykładowe rodzaje algorytmów heurystycznych.

Rozdział trzeci dotyczy algorytmów routingu rozgałęźnego bez ograniczeń. Doktorant dzieli w nim algorytmy na dwa rodzaje: algorytmy konstrukcji minimalnego drzewa Steinera oraz algorytmy konstrukcji drzewa najkrótszych ścieżek. W pierwszej kategorii przytacza algorytmy PPH (ang. *Pruned Prim Heuristic*), KMB (ang. *Kou, Markovskiy, Bermann*) oraz DDMC (ang. *Destination-Driven MultiCast*). W drugiej kategorii wymienia algorytmy PDH (ang. *Pruned Dijkstra Heuristic*) oraz SPT (ang. *Shortest Path Tree*). Na zakończenie tego rozdziału Doktorant proponuje własny algorytm STA (ang. *Switched Tree Algorithm*), który polega na wyznaczeniu drzewa algorytmem PPH i SPT oraz wybraniu z nich drzewa o mniejszym koszcie.

W rozdziale czwartym Doktorant omawia algorytmy routingu rozgałęźnego z ograniczeniami. Algorytmy te służą do wyznaczania minimalnego drzewa przy dodatkowym założeniu, że opóźnienie wzdłuż ścieżek nie może przekroczyć ustalonej wartości. Podobnie jak w przypadku algorytmów bez ograniczeń, rozważa algorytmy konstrukcji drzewa Steinera z ograniczeniem oraz wyznaczania drzewa ścieżek o minimalnym koszcie z ograniczeniami. Ponadto rozważa algorytmy wykorzystujące relaksację Lagrange'a oraz algorytmy wyznaczania K -najkrótszych ścieżek. W pierwszej grupie omawia algorytm KPP (ang. *Kompella, Pasquale, Polyzos*), CAO (ang. *Constrained Adaptive Ordering*) oraz BSMA (ang. *Bounded Shortest Multicast Algorithm*). W drugiej grupie prezentuje algorytmy CDKS (ang. *Constrained Dijkstra heuristics*) oraz DCSP (ang. *Delay Constrained Shortest Path Multicast Algorithm*). Dalej rozważa wykorzystanie relaksacji Lagrange'a do rozwiązywania problemu optymalizacji oraz przytacza algorytmy LARAC (ang. *Lagrange Relaxation based Aggregated Cost*), PLRA (ang. *Path's Lagrange Relaxation Algorithm*) oraz MLRA (ang. *Multicast Lagrange Relaxation Algorithm*), przy czym te dwa ostatnie są propozycjami doktoranta. Algorytm PLRA stanowi modyfikację algorytmu LARAC, a modyfikacja polega na zmianie parametru λ co prowadzi do zmniejszenia liczby iteracji. Algorytm MLRA polega na wyznaczaniu najkrótszych ścieżek przy pomocy algorytmu PLRA i dodawania ich kolejno do tworzonego drzewa dystrybucyjnego. Kolejna grupa algorytmów, w których wyznacza się K najkrótszych ścieżek w grafie, obejmuje proponowane przez Doktoranta algorytmy KSP (ang. *K-Shortest Path algorithm*) do wyznaczania najkrótszych ścieżek oraz KSPMA (ang. *K-Shortest Path Multicast Algorithm*) do wyznaczania drzewa najkrótszych ścieżek. Algorytm KSP stanowi modyfikację klasycznego algorytmu do wyznaczania K najkrótszych ścieżek a modyfikacja polega na uwzględnieniu opóźnienia wzdłuż każdej wyznaczonej ścieżki. Algorytm KSP jest wykorzystywany do wyznaczania najkrótszych ścieżek w algorytmie KSPMA. Ostatnim algorytmem rozważanym w tym rozdziale jest algorytm DCMA (ang. *Fast Delay-Constrained Multicast Routing Algorithm*), oraz dwie jego modyfikacje

zaproponowane przez Doktoranta: RDCMA (ang. *Reverse DCMA*) oraz OPTDCMA (ang. *Optimal DCMA*). W algorytmie DCMA poszukiwanie ścieżki rozpoczyna się w jednym z wierzchołków reprezentującego odbiorcę i jest prowadzone w kierunku wierzchołka reprezentującego nadawcę wzdłuż ścieżki o minimalnym opóźnieniu natomiast w algorytmie RDCMA poruszanie odbywa się wzdłuż ścieżki o minimalnym koszcie. W algorytmie OPTDCMA wybiera się drzewo o niższym koszcie spośród drzew wyznaczonych algorytmami DCMA i RDCMA.

Rozdział piąty pracy dotyczy metodologii badań algorytmów routingu. Autor rozważa w nim metody symulacyjne, które są w takich badaniach stosowane oraz parametry jakie mają wpływ na jakość symulacji. Pierwszym istotnym elementem jest dobór modelu sieci i sposób jej generowania. Autor zauważa, że najczęściej w literaturze stosowane są trzy modele sieci: regularne (pierścień, krata, gwiazda), rzeczywiste (ARPAnet, NSFnet, GEANT) i generowane losowo. Dalej zajmuje się metodami generowania grafów losowych i opisuje metody Waxmana i Barabasi-Alberta do generowania grafów losowych oraz prezentuje wybrane generatory takich grafów. W dalszej części tego rozdziału Doktorant omawia parametry opisujące topologię sieci, które umożliwiają ich porównanie. Przytacza tutaj takie parametry jak średni stopień grafu, średnica, liczba skoków (krawędzi) w najdłuższej ścieżce i jej długość euklidesowa, czy współczynnik gronowania. W celu oceny połączeń rozgałęźnych, Doktorant wprowadza jeszcze dwa dodatkowe parametry: gęstość grupy (zdefiniowanej jako stosunek liczby węzłów odbiorczych do liczby węzłów sieci) oraz współczynnik rozproszenia grupy (zdefiniowany jako stosunek obszaru ograniczonego okręgiem o średnicy d_m do wymiaru boku kwadratu, w którym mieści się badana sieć). Ważnym elementem procesu badawczego algorytmów wyboru drzewa dystrybucyjnego jest sposób rozmieszczenia członków grupy odbiorczej w sieci. W dalszej części rozdziału Autor opisuje zatem metody stosowane przy rozmieszczaniu węzłów odbiorczych grupy w sieci. W rozprawie zaproponował między innymi metody GroupRadius (członkowie grupy są wybierani z węzłów znajdujących się wewnątrz promienia o określonej długości) i GrupHighDegree (preferowane są węzły z większą liczbą łączy wychodzących). Następnie autor podaje parametry jakościowe stosowane do oceny porównawczej algorytmów. Są to koszt wybranego drzewa, średni koszt ścieżki, średnica drzewa, znormalizowany koszt drzewa oraz czas działania algorytmu. Na końcu autor opisuje metody oceny statystycznej wyników oraz zastosowane środowisko symulacyjne.

W rozdziale szóstym Doktorant prezentuje wyniki badań porównawczych algorytmów opisanych i zaproponowanych w rozdziałach trzecim i czwartym przy wykorzystaniu metodologii zaproponowanej w rozdziale piątym. Najpierw prezentuje wyniki badań dla algorytmów w małych sieciach, przy których daje się w rozsądnym czasie uzyskać także wyniki przy wykorzystaniu metody dokładnej. W ten sposób porównuje metody heurystyczne z metodą dokładną i algorytm najbliższy metodzie dokładnej stanowi następnie odniesienie dla badań w sieciach o większej liczbie węzłów. Badania przeprowadzono dla sieci o różnej liczbie węzłów przy określonej liczności grupy odbiorczej i średnim stopniu grafu, przy czym algorytmy były badane w grafach generowanych metodą Waxmana i Barabasi-Alberta a koszt przypisany dla łączy był generowany losowo lub wyznaczany jako odległość euklidesowa. Kolejne dwa podrozdziały zawierają wyniki badań dla algorytmów odpowiednio bez ograniczeń i z ograniczeniami. W każdym przypadku dla różnych algorytmów przeprowadzono analizę średniego kosztu drzewa, średniego kosztu ścieżki w drzewie, wpływ rozmieszczenia węzłów odbiorców na wyniki pracy algorytmów, czasu działania algorytmów, oraz podano statystyki opisowe wyników badań.

Pracę kończy podsumowanie, w którym Doktorant stwierdza, że postawiony na wstępie cel został osiągnięty i wymienia najważniejsze jego zdaniem osiągnięcia pracy.

Przedstawiona rozprawa ma charakter teoretyczny. Przeprowadzono w niej dogłębną analizę porównawczą algorytmów routingu rozgałęźnego, zaproponowano metodologię takich badań, oraz oceniono wybrane algorytmy znane z literatury oraz algorytmy zaproponowane przez doktoranta.

Zastosowana metoda badawcza i ocena rezultatów rozprawy

Do oceny algorytmów Doktorant stosuje metodę symulacji cyfrowej. Jest to obecnie jedyna metoda stosowana w literaturze, umożliwiająca porównywanie pracy różnych algorytmów routingu, w tym rozgałęźnego. Ponieważ na wyniki działania algorytmów wpływ ma zarówno topologia badanej sieci oraz rozmieszczenie węzłów uczestniczących w połączeniach, Doktorant wykorzystuje w ocenie różne sieci, generowane różnymi metodami i o różnej liczbie węzłów i średnim stopniu grafu. Dla każdego punktu symulacji wykonano 5 serii pomiarów dla każdej z 1000 struktur sieci, przy czym w punkcie symulacji ustalone były parametry grupy odbiorczej (liczba węzłów odbiorczych, maksymalne opóźnienie, metoda wyboru węzłów odbiorczych w sieci) oraz sieci (liczba węzłów w sieci, metoda generowania topologii, średni stopień sieci). Wyznaczono także, zgodnie z zasadami, 95-procentowe przedziały ufności dla pomierzonych parametrów. Na uwagę zasługuje liczba badanych sieci a zwłaszcza uwzględnienie różnych sieci generowanych różnymi metodami i o różnym średnim stopniu grafu reprezentującego sieć. Według mojej wiedzy są to pierwsze badania tak szerokie algorytmów routingu. W dotychczasowych publikacjach najczęściej ograniczano się do porównywania algorytmów w kilku sieciach. Doświadczenia pokazują jednak, że zmiana topologii sieci może znacząco wpłynąć na otrzymywane wyniki. Potwierdzają to także wyniki prezentowane w rozdziale 6 pracy, w których widać różnice w wynikach otrzymanych dla sieci generowanych metodą Waxmana i Barabasi-Alberta.

Ocenę algorytmów Doktorant odnosi do algorytmu KMB. Wykazał on bowiem, że dla sieci o małej liczbie węzłów, dla której możliwe jest stosowanie metody dokładnej, koszt otrzymanych drzew jest tylko nieznacznie (w granicach 5%) większy od kosztu drzew otrzymanych metodą dokładną. Z wykresów przedstawionych w dalszej części pracy wynika, że otrzymane wyniki dla sieci z większą liczbą węzłów są także najlepsze przy stosowaniu tego algorytmu. Badania symulacyjne potwierdzają także, że zaproponowany przez Doktoranta algorytm STA do wyznaczania drzewa dystrybucyjnego bez ograniczeń jest nieznacznie gorszy od algorytmu KMB, a jest znacznie mniej złożony obliczeniowo (na przykład dla sieci o 100 węzłach i grupie odbiorczej zawierającej 20 członków czas realizacji algorytmu KMP w sieci generowanej metodą Waxmana wynosi 11,5 ms a dla algorytmu STA wynosi on 0,9 ms). STA konstruuje drzewa transmisji grupowej o mniejszych kosztach niż algorytm SPT, a jego efektywność jest szczególnie widoczna w sieciach o dużej gęstości grupowej.

W przypadku badań algorytmów z ograniczeniami, Doktorant przedstawia osobno wyniki porównania proponowanych przez niego algorytmów RDCMA i OPTDCMA oraz MLRA i KSPMA. Utrudnia to czytelnikowi ocenę tych algorytmów. Na przykład parametry badanych sieci na rysunkach 6.10 i 6.19 oraz 6.11 i 6.20 są takie same i przedstawione na nich wyniki dla algorytmów KPP i CSPT powtarzają się. Umieszczenie tych wyników na jednym wykresie ułatwiłoby ocenę i porównanie proponowanych algorytmów. Pod względem średniego kosztu drzewa, proponowane przez Doktoranta algorytmy są gorsze tylko od algorytmu KPP, natomiast pod względem średniego kosztu ścieżki, algorytmy MLRA i KSDMA są najlepsze, tzn. średni koszt ścieżki w wyznaczonych przy ich pomocy drzewach dystrybucyjnych jest najniższy. Pod względem złożoności czasowej algorytmy MLRA i RDCNA są znacznie szybsze niż algorytm KPP i porównywalnie szybkie z pozostałymi

algorytmami. Jedynie algorytm KSPMA cechuje się znacznie większą złożonością czasową. Brak jednak w porównaniu złożoności czasowej (tabela 6.12 i 4.1) algorytmu OPTDCMA.

Uważam jednak, że Doktorant zaproponował interesującą i dogłębną metodę analizy porównawczej algorytmów i skutecznie wykorzystał ją do porównania znanych z literatury algorytmów i algorytmów zaproponowanych w pracy. Algorytmy te prowadzą do wyznaczenia drzew o kosztach porównywalnych lub niższych niż koszty drzew wyznaczanych przez inne algorytmy, a jeżeli koszty te są większe, to proponowane algorytmy cechują się mniejszą złożonością czasową. Zatem otrzymane w rozprawie rezultaty są znaczące.

Redakcja rozprawy

Praca jest zredagowana bardzo starannie, logicznie ułożona i napisana poprawnym językiem. Autor nie ustrzegł się wprawdzie drobnych błędów wymienionych w uwagach szczegółowych, ale nie są one znaczące do ogólnej oceny pracy.

Literatura przytoczona w rozprawie obejmuje 128 pozycji. Są to moim zdaniem pozycje w całości związane z tematyką rozprawy. W zamieszczonej w rozprawie bibliografii w dwudziestu czterech pozycjach Doktorant jest współautorem. Wśród nich 8 to artykuły w czasopismach (w tym jedno czasopismo PAN i jedno zagraniczne) i 6 na konferencjach międzynarodowych.

Uwagi szczegółowe

Algorytmy rozważane w pracy są przedstawione za pomocą pseudokodu. Proponowałbym jednak w przyszłości nie stosować powołania w tekście na pseudokod o jakimś numerze, a raczej na algorytm o określonym numerze. Zwłaszcza że na początku pracy podany jest spis algorytmów a nie pseudokodów. Numeracja powinna być podobna jak do numeracji rysunków czy tabel, czyli dwustopniowa a nie trzystopniowa. Poza tym opis w pseudokodzie jest niejednorodny, na przykład w kodzie 4.2.3 na stronie 39 opis jest w języku polskim, w 4.2.5 na stronie 43 – tylko w j. angielskim, a w 4.2.4 na stronie 42 jest stosowana mieszanka języka polskiego i angielskiego

Opisy na rysunkach są mało czytelne, są obcięcia liter i cyfr co powoduje, że czytający musi domyślać się, co jest napisane. Jest to prawdopodobnie wina źle dobranej czcionki lub rozmiaru opisu. Dotyczy to także wykresów prezentowanych w rozdziale 6.

Str. 21, l. 11 – powtórzenie definicji grafu jest tutaj niepotrzebne, jest ono już podane na stronie 18.

Tabela 4.1 – w złożoności obliczeniowej algorytmów występuje $|M|$ i oraz m , przy czym z definicji stosowanych oznaczeń wynika, że $m = |M|$. Jaki sens ma zatem stosowanie obu tych symboli w tym samym wzorze?

Str. 64, l. 23 – czy chodzi tutaj o uzależnienie, czy raczej o uniezależnienie wyników od topologii sieci?

Podpis pod rysunkiem 6.1 – rysunek przedstawia zależność średniego kosztu drzewa transmisji grupowej od liczby węzłów sieci dla trzech algorytmów w tym algorytmu dokładnego jakim jest MST a nie tylko dla algorytmu dokładnego.

Rysunek 6.3 – w legendzie podane są oznaczenia dla algorytmu PPH w sieciach Waxmana i Barabasi-Alberta, a na wykresie brak jest tych linii.

Tabela 6.5 – średni czas wykonywania algorytmu STA zarówno w sieciach Waxmana jak i w sieciach Barabasi-Alberta maleje wraz ze wzrostem liczności grupy dystrybucyjnej, podczas gdy w pozostałych algorytmach ten czas rośnie. Jaka jest przyczyna tego?

Str. 87, l. 2 pod tabelą 6.7 – zamiast c_Y i c_X powinno być Y i X , gdyż to oznacza algorytmy a nie koszty drzew uzyskanych tymi algorytmami. (Podobnie linia 4 na stronie 114).

Najważniejsze osiągnięcia rozprawy

Za najważniejsze osiągnięcia rozprawy uważam:

- zaproponowanie algorytmu STA do wyznaczania drzewa dystrybucyjnego bez kryterium ograniczającego,
- zaproponowanie algorytmów wykorzystujących relaksację Lagrange'a do wyznaczania drzewa dystrybucyjnego z uwzględnieniem kryterium ograniczającego,
- zaproponowanie modyfikacji algorytmu DCMA, które są przydatne do wyznaczania drzewa dystrybucyjnego w sieciach o dużej średniej wartości stopnia wężła i przy małych wartościach opóźnienia maksymalnego,
- zdefiniowanie parametrów
- oraz zaproponowanie metodologii umożliwiającej porównanie algorytmów do wyznaczania drzewa dystrybucyjnego i przeprowadzenie badań porównawczych różnych algorytmów znanych z literatury oraz zaproponowanych w pracy

Podsumowanie

Podsumowując stwierdzam, że przedłożona do recenzji rozprawa spełnia wymagania stawiane pracom doktorskim i może być dopuszczona do publicznej obrony.