

Recenzja rozprawy doktorskiej mgr. Macieja Pacuta

Algorithmic aspects of contemporary networks

Rozprawa doktorska mgr. Macieja Pacuta dotyczy problemów algorytmicznych inspirowanych współczesnymi sieciami komputerowymi. Rozprawa składa się ze wstępu oraz dwóch niezależnych części.

Pierwsza część (rozdziały 2 i 3) poświęcona jest problemom związanym z lokalizacją oraz przydzielaniem zasobów maszynom wirtualnym w centrach obliczeniowych. Wyniki opisane w tej części zostały opublikowane na konferencjach ICNP, DISC oraz w Elsevier Journal of Theoretical Computer Science.

Druga część (rozdział 4) dotyczy problemów związanych z zarządzaniem pamięcią routera. Wyniki opisane w tej części zostały wcześniej opublikowane na konferencji SPAA.

Ocena merytoryczna

Rozdział pierwszy rozprawy zawiera wprowadzenie do tematyki pracy, krótki opis uzyskanych wyników oraz omówienie istniejących wyników związanych z tematyką pracy.

Rozdział drugi poświęcony jest statycznej wersji problemu lokalizacji maszyn wirtualnych. W szczególności rozważana jest tu sytuacja, w której fizyczna sieć ma topologię drzewa, a sieć wirtualna topologię kliki. Autor rozważa w definicji problemu pięć binarnych parametrów, co prowadzi do aż $32 (= 2^5)$ różnych wariantów. Autor dokonuje pełnej klasyfikacji tych wariantów ze względu na złożoność obliczeniową ich rozwiązania. Od strony algorytmicznej używane są techniki takie jak: przepływy, skojarzenia (w zasadzie szczególny przypadek przepływów, ale pozwalają na czytelniejszy opis), czy programowanie dynamiczne. Dowody NP-trudności są redukcjami z problemu 3D-MATCHING. Rozumowania przedstawione w tym rozdziale są dość standardowe. Różnorodność zastosowanych technik oraz skala przedsięwzięcia są jednak dość imponujące. Świetnie ilustruje to rysunek 2.2.

W rozdziale trzecim rozważana jest dynamiczna wersja problemu lokalizacji maszyn wirtualnych. W tej wersji maszyny mogą być relokowane w odpowiedzi na zmieniający się charakter komunikacji między nimi. Autor koncentruje się przy tym na algorytmach on-line i analizie kompetytywnej, oraz na bardzo prostej topologii sieci fizycznej – gwiazdzie. Głównym wynikiem tego rozdziału jest algorytm nazwany przez autora CREP, który jest $O((1 + \epsilon)k \log k)$ -kompetytywny dla k węzłów w sieci fizycznej, o ile pozwolimy mu alokować do każdego węzła o $2 + \epsilon$ razy więcej maszyn niż adwersarzowi (tzw. *resource augmentation*). To ostatnie założenie znacząco upraszcza sytuację, w szczególności pozwala uniknąć rozważania problemów rodzaju BIN-PACKING. To powiedziawszy, analiza algorytmu CREP nie jest ani prosta, ani całkiem standardowa. Punktem wyjścia jest klasyczna analiza RENT-OR-BUY, ale została ona tutaj połączona z rozważaniami dotyczącymi sformułowań liniowych wielościanu drzew rozpinających (ciekawe byłoby bardziej bezpośrednie potraktowanie tego wątku), a także z nieoczywistą amortyzacją.

W rozdziale trzecim rozpatrywany jest też szczególny przypadek, w którym każdy węzeł fizyczny może pomieścić dokładnie dwie maszyny wirtualne (nazwany tu ONLINE-REMATCHING).

Opisany algorytm jest dość prosty i naturalny, ale dobranie odpowiedniej funkcji potencjału w analizie zapewne nie było proste.

Rozdział trzeci zawiera także ograniczenia dolne, pokazane za pomocą dość standardowych redukcji z problemu ONLINE-CACHING.

Dość zaskakujące w kontekście tego rozdziału, a także rozdziału następnego, jest całkowite zignorowanie przez autora algorytmów randomizowanych. Wydaje się to naturalny kierunek badań, szczególnie biorąc pod uwagę, że autor rozważa problemy inspirowane zastosowaniami praktycznymi.

Rozdział czwarty dotyczy zarządzania pamięcią routera, a w szczególności radzenia sobie z sytuacją, gdy cała tablica routingu nie mieści się w tej pamięci. Autor proponuje rozwiązanie polegające na przechowywaniu części informacji poza routerem. Dostęp do tych zewnętrznych danych jest wolniejszy, przez co staje się konieczne rozwiązywanie dynamicznie problemu wyboru danych, które są przechowywane w pamięci routera. Autor abstrahuje kluczowe elementy opisanej sytuacji w definicji problemu ONLINE-TREE-CACHING, który jest uogólnieniem klasycznego problemu ONLINE-CACHING. Algorytm TC, zaproponowany przez autora, osiąga współczynnik kompetytywny analogiczny do optymalnego współczynnika dla problemu ONLINE-CACHING, modulo dodatkowy czynnik odpowiadający głębokości drzewa routingu. Sam algorytm jest, jak to z reguły bywa w przypadku algorytmów on-line, dość naturalny. Zdecydowanie najciekawszym rozumowaniem w tym rozdziale jest oszacowanie dolne na koszt optimum off-line. Rozumowanie to opiera się na sprytnym modyfikowaniu sekwencji wejściowej tak, że koszt optimum off-line nie wzrasta, ale jednocześnie sekwencja jest dużo bardziej *regularna* i przez to łatwiejsza do analizowania.

Redakcja

Praca językowo i redakcyjnie stoi na bardzo wysokim poziomie. Jeśli chodzi o usterki, to w całej pracy znalazłem ich zaledwie kilka (zob. Dodatek). Praca jest bardzo dobrze skonstruowana. Rysunki umieszczone w pracy sprawiają wrażenie dobrze przemyślanych i znacząco ułatwiają zrozumienie tekstu. Bibliografia została dobrze przygotowana i użyta. Podsumowując, praca została przygotowana bardzo dobrze.

Konkluzja

Poziom naukowy rozprawy oceniam wysoko. Rozważane problemy są dość naturalne i dobrze umotywowane. Autor korzysta z bardzo bogatego arsenału technik, niektóre z rozumowań wymagały sporo pomysłowości i bardzo dobrego zrozumienia rozważanych problemów. Praca jest bardzo dobrze zredagowana i napisana. Opisywane w pracy wyniki zostały opublikowane na prestiżowych konferencjach.

Uważam, że przedstawiona rozprawa spełnia wymogi ustawowe i wnoszę o dopuszczenie kandydata do dalszych etapów przewodu doktorskiego.

Prof. dr hab. inż. Andrzej Kozłowski

Dodatek: uwagi szczegółowe

- str. 12, akapit 1 „where all components are of size at most n/k ” - powinno być n/ℓ .
- str. 12, akapit 2 „Our dynamic model...” - „online caching” pojawia się dwukrotnie, za drugim razem powinno być „paging”.
- str. 20-21 W definicji problemu nie jest całkiem jasne, czy przepustowość ma być gwarantowana dla każdej pary z osobna, czy wspólnie. Równanie na stronie 21 wyjaśnia sprawę, ale pojawia się dość późno.
- str. 24, analiza W jaki sposób i jak szybko zaokrągla się przepływ?
- str. 46, akapit 2 Jedno z $F(C)$ powinno być $\mathcal{F}(C)$.
- str. 46, akapit 2 „The union (...) cover all possible node-time pairs.” - za wyjątkiem ogonów.
- str. 52, nad równaniem (3.3) Greedy dokonuje zamiany w trzeciej fazie, ale w drugiej nie płaci, jeśli zamienia. To dziwne.
- str. 53, dowód Lematu 3.13 Krawędzie po zamianie to (x, y) i (x', y') .
- str. 66 i dalej Tu pojawia się notacja kwadratowej sumy zbiorów, czy ta notacja jest gdzieś wcześniej zdefiniowana?
- str. 65 i 66 Podział na Lemat 4.1 i 4.2 wydaje się dość sztuczny.