Matematyczne podstawy poprawności, bezpieczeństwa i efektywności rozproszonych systemów informatycznych

Projekt 4 T11C 042 25

Czas: grudzień 2003 – listopad 2006.

Miejsce: Instytut Informatyki Uniwersytetu Warszawskiego.

Nakłady: 252, 000 zł.

dr hab. Damian Niwiński (kierownik)

dr hab. Jerzy Tyszkiewicz

dr Stefan Dziembowski

dr Sławomir Lasota

dr Mikołaj Bojańczyk

dr Marcin Peczarski

mgr Eryk Kopczyński

mgr Sławomir Leszczyński

mgr Filip Murlak

mgr Jacek Sroka

mgr Michał Strojnowski

Rezultaty

29 publikacji, w tym **20** w wydawnictwach z listy Filadelfijskiego Instytutu IN

+ 1 rozprawa doktorska (2004) + 3 manuskrypty złożone do publikacji.

Wyniki przedstawiane na konferencjach amerykańskich

- ACM Symposium on Theory of Computing (**STOC**),
- ACM Symposium on Principles of Database Systems (**PODS**),
- IEEE Symposium on Logic in Computer Science (**LICS**)(2),
- Annual International Cryptology Conference (**CRYPTO**)
- Theory of Cryptography Conference

i europejskich **ICALP** (4), **Eurocrypt**, **CONCUR**, **FOSSACS**, **CSL**, **MFCS**.

Czasopisma *Theoretical Computer Science* (3), *Information Processing Letters*.

Nagrody

EATCS: **Best ICALP Paper**

- **2004** *Tree-Walking Automata Cannot Be Determinized*,

  Mikołaj Bojańczyk i Thomas Colcombet,

- **2006** *The Wadge Hierarchy of Deterministic Tree Languages*,

  Filip Murlak.

ACM: **PODS Best Paper Award**

- **2006** *Two-Variable Logic on Data Trees and XML Reasoning*,

  Mikolaj Bojańczyk, A. Muscholl, Th. Schwentick, L. Segoufin, Claire David.

EACSL: **The Ackermann Award**

- **2005**, Mikołaj Bojańczyk.

Zadania badawcze

Analiza modeli systemów czasu rzeczywistego

Złożoność komunikacyjna w obecności błędów

Model hybrydowy

Metody projektowania protokołów odpornych na błędy

Efektywne protokoły wielopodmiotowe

Złożonościowe aspekty weryfikacji

Złożoność dostępu do baz danych w internecie

Algorytmy dla bisymulacji w systemach rozproszonych

Cele naukowe

Przedmiot badań — matematyczne modele, w których pojęcia poprawności, bezpieczeństwa i efektywności zyskują ścisły sens i mogą być obiektem twierdzeń matematycznych.

Cele badawcze — twierdzenia matematyczne uściślające kryteria oceny systemów i dostarczające algorytmy dla ich weryfikacji.

Cele aplikacyjne — wspomaganie projektantów wielkich systemów informatycznych, w szczególności systemów rozproszonych działających w sieciach.

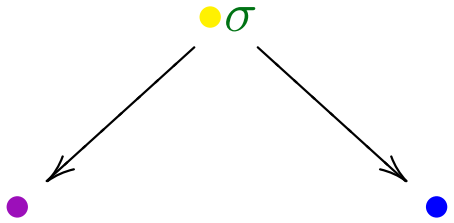
Poprawność

Taksonomia modeli ze względu na ich siłę obliczeniową.

Współczesne realizacje często reprezentują dane nie w formacie tablicowym, lecz drzewiastym (XML).

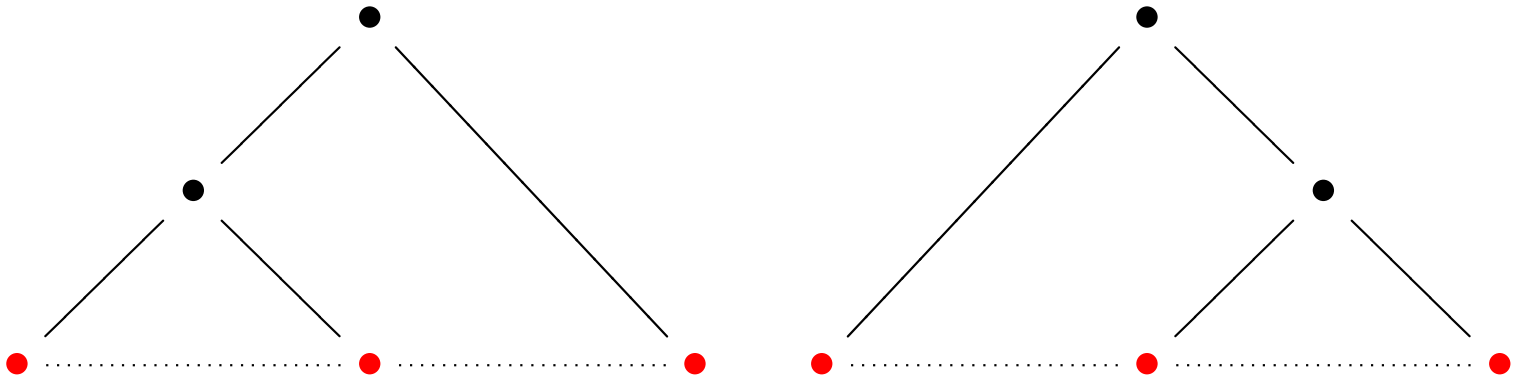
Automaty na drzewach stanowią użyteczny model dla mechanizmów obliczania zapytań (ang. *queries*) i znajdowania wzorca.

Automaty ścieżkowe przetwarzają drzewo w sposób sekwencyjny, w odróżnieniu od tradycyjnych automatów alternujących, przetwarzających je w sposób równoległy.



Twierdzenie Następujące modele nie są równoważne.

- deterministyczne automaty ścieżkowe
- niedeterministyczne automaty ścieżkowe
- alternujące automaty ścieżkowe .



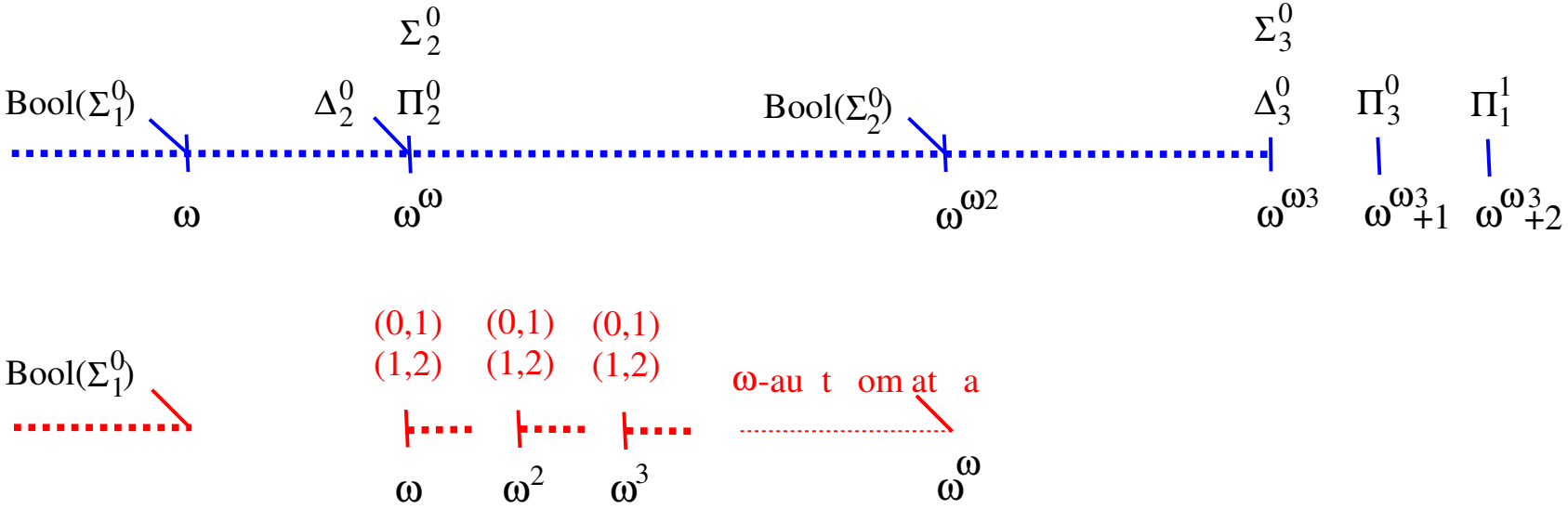
M.Bojańczyk, Th.Colcombet, Tree-Walking Automata Cannot Be Determinized, **ICALP 2004**.

Tree-Walking Automata Do Not Recognize All Regular Languages, **ACM STOC 2005**.

M.Bojańczyk, M.Samuelides, T.Schwentick, L.Segoufin, Expressive Power of Pebble Automata, **ICALP 2006**.

Taksonomia modeli cd.

Efektywna hierarchia topologiczna deterministycznych automatów na drzewach nieskończonych



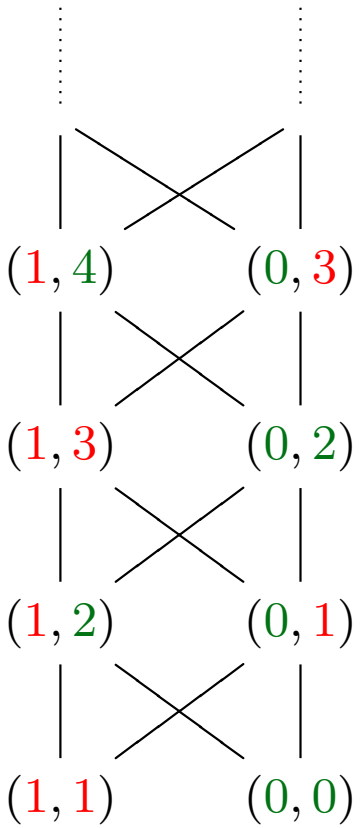
Wysokość hierarchii jest $\omega^{\omega \cdot 3} + 2$ (*vs* ω^ω dla języków słów).

F.Murlak, The Wadge Hierarchy of Deterministic Tree Languages, **ICALP 2004**.

Taksonomia modeli cd.

Efektywna hierarchia indeksów Mostowskiego

dla deterministycznych automatów na drzewach



D.Niwiński, I. Walukiewicz, Deciding Nondeterministic Hierarchy of Deterministic Tree Automata, **WOLLIC 2005**.

Wybór właściwego modelu

Aby model automatowy jak najwierniej odpowiadał rzeczywistości, rozszerza się go m.in. o

— wymiar czasu rzeczywistego,

— mechanizm stosu wyższego rzędu,

— dostęp do danych.

W ogólności prowadzi to do nieskończonej liczby stanów i **nierozstrzygalności** większości problemów weryfikacyjnych.

Jednak przy odpowiednim doborze struktury i parametrów, można nadal efektywnie rozstrzygać wiele pytań !

- Równoważność automatów czasowych jest nierozstrzygalna, ale...

  równoważność alternujących automatów czasowych z jednym zegarem jest rozstrzygalna.

  S.Lasota, I.Walukiewicz, Alternating Timed Automata, **FOSSACS 2005**.

- **Core-XPath** jest nierozstrzygalny, ale...

  fragment $(FO^2, \sim, +1)$ jest rozstrzygalny.

  M.Bojańczyk, A. Muscholl, Th. Schwentick, L. Segoufin, C.David, Two-Variable Logic on Data Trees and XML Reasoning, **ACM PODS 2006**.

- Teorie monadyczne drzew obliczeń schematów rekurencyjnych drugiego rzędu są rozstrzygalne.

  T.Knapik, D.Niwiński, P.Urzyczyn, I.Walukiewicz, Unsafe Grammars and Panic Automata, **ICALP 2005**.

Wybór właściwego modelu cd.

Klasyczne automaty Büchiego i Rabina, prowadzące nieskończone obliczenia, nie uwzględniają własności ilościowych w rodzaju

$$\ldots 0 \underbrace{111\ldots11}_{\text{dowolnie dużo}} 0\ldots$$

Zaproponowano model uwzględniający takie warunki i wykazano jego rozstrzygalne własności.

M.Bojańczyk, A Bounding Quantifier, **CSL 2004**.

M.Bojańczyk, Th.Colcombet, Bounds in ω-Regularity **IEEE LICS 2006**.

Otwiera to nowe ścieżki w weryfikacji modelowej.

Wybór właściwego języka

Zasadniczym problemem weryfikacji jest znalezienie właściwego *trade–off*

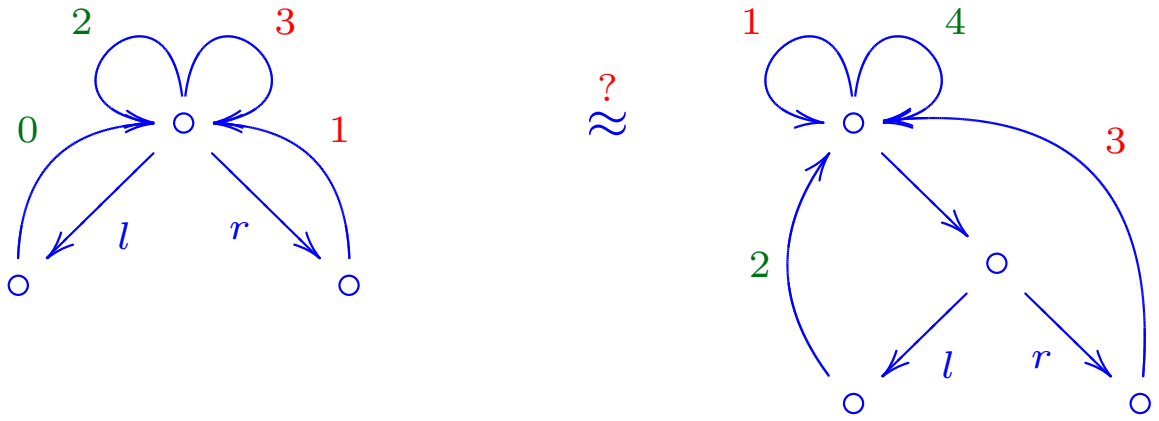
siła wyrazu języka *vs* złożoność obliczeniowa weryfikacji

Pożądane są algorytmy **upraszczania** specyfikacji tam, gdzie to możliwe.

Podano algorytmy efektywnego upraszczania formuł logiki **MSO** nad drzewami do logiki temporalnej **EF + EX**.

M.Bojańczyk, I.Walukiewicz, Characterizing EF and EX Tree Logics, **CONCUR 2004**.

Rozstrzyganie bisymulacyjnej równoważności procesów



$$L(G) \quad \overset{?}{=} \quad L(G')$$

$$(a.\mathbf{0} + b.\mathbf{0})||a.\mathbf{0} + a.\mathbf{0}||a.\mathbf{0} \quad \overset{?}{\Longleftrightarrow} \quad (a.\mathbf{0} + b.\mathbf{0})||a.\mathbf{0}.$$

Rozstrzyganie bisymulacyjnej równoważności procesów cd.

- Algorytm dla dziedzicznej równoważności zachowującej historię procesów w BPP w czasie $\tilde{\mathcal{O}}(n^2)$.

  S.Froeschle, S.Lasota, Decomposition and Complexity of Hereditary History Preserving Bisimulation on BPP, **CONCUR 2005**.

- Algorytmiczna rozstrzygalność równoważności efektywnościowej w BPP, wielomianowa w p BPP.

  S.Lasota, Decidability of Performance Equivalence for Basic Parallel Processes, **TCS**.

- Algorytm $\tilde{\mathcal{O}}(n^8)$ dla równoważności procesów w n BPA (poprzedni $\mathcal{O}(n^{13})$).
  Algorytm $\tilde{\mathcal{O}}(n^6)$ dla równości języków gramatyk prostych (poprzedni $\mathcal{O}(n^7)$).

  S.Lasota, W.Rytter, Faster Algorithm for Bisimulation Equivalence of Normed Context-Free Processes, **MFCS 2006**.

Modele oparte na grach

Problem weryfikacji modelowej systemów otwartych można sprowadzić do konstrukcji strategii w grze

system $\leftarrow$ kontra $\rightarrow$ środowisko.

Efektywność tej metody zależy od jakości strategii.

Znaleziono szeroką klasę warunków zapewniających systemowi strategię pozycyjną.

E.Kopczyński, Half-Positional Determinacy of Infinite Games, **ICALP 2006**.

Bezpieczeństwo

Model hybrydowy

Odkryto fundamentalny problem modelu hybrydowego, który łączy metody kryptografii asymetrycznej z założeniami modelu teorio-informacyjnego z ograniczoną pamięcią.

Protokoły oparte na modelu hybrydowym nie zawsze są bezpieczne **!**

S.Dziembowski, U.Maurer, On Generating the Initial Key in the Bounded-Storage Model, **EUROCRYPT 2004**.

Bezpieczeństwo cd.

Protokoły wielopodmiotowe

Opracowano nową metodę uzyskiwania protokołów odpornych na wtargnięcia intruza. Klucz zostaje sztucznie powiększony tak, by podszycie się pod ofiarę było możliwe dopiero po skradzeniu znaczącej części klucza.

Siłę tej metody wykazano na przykładzie protokołów uzgadniania klucza sesyjnego.

S.Dziembowski, Intrusion-Resilience Via the Bounded-Storage Model, **Theory of Cryptography Conference 2006**.

Pokazano sposób wykorzystania powyższej metody do zwiększenia bezpieczeństwa zaszyfrowanych danych przechowywanych w komputerze. Wykazano związki z teorią kompresowalności instancji NP.

S.Dziembowski, On Forward-Secure Storage, **CRYPTO 2006**.

Efektywność

Odporność na błędy

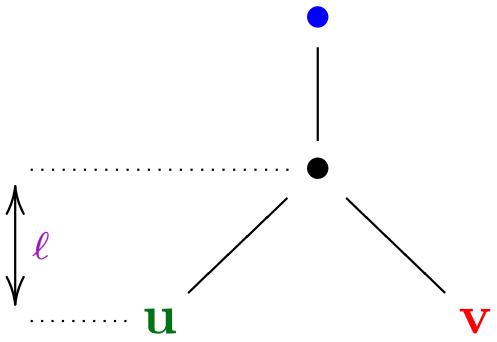
Metoda konstrukcji protokołów obliczania funkcji n-argumentowej przez n procesorów dzielących argumenty, odpornych na błędy transmisji.

Strategia w grze Ulama z kłamstwami $\longmapsto$ protokół.

M.Peczarski, Strategy in Ulam's Game and Tree Code Give Error-Resistant Protocols, 2004, złożone do FI.

Odporność na błędy cd.

Elementem powyższej metody jest efektywna konstrukcja **kodu drzewiastego Schulmana** (dotychczas znany był tylko dowód jego istnienia).



$$H(\mathbf{u},\mathbf{v}) \leq \tfrac{1}{2} \cdot \ell.$$

Zredukowano przy tym liczbę potrzebnych etykiet

($16d^2 \searrow d^{2+\lceil \log_d 4 \rceil}$, np. $64 \searrow 16$).

M.Peczarski, An improvement of the tree code construction, **IPL**, 2006.

Odporność na błędy cd.

Zaproponowano nowe protokoły dla wymiany informacji w zbiorze procesorów, wśród których niektóre są uszkodzone. Najlepszy algorytm rozwiązuje problem w czasie $\mathcal{O}(\log n)$ używając $\mathcal{O}(n \log n + t^2)$ wiadomości.

Wykazano dolne ograniczenia ze względu na czas działania i złożoność komunikacyjną.

D.Kowalski, M.Strojnowski, Gossiping in the presence of omission failures, 2006, złożone w IPL.

Efektywność cd.

Przepływy danych w Internecie

Wiele zastosowań, jak np. eksperymenty *in silico* w biologii obliczeniowej, wymagają pobierania ogromej ilości danych z licznych baz dostępnych w Internecie. *European Bioinformatics Institute* rozwinął w tym celu system **Taverna**.

Zaproponowano rozszerzenie funkcjonalności Taverny przez wzajemne tłumaczenie z językiem XQuery.

Taverna $\Longleftrightarrow$ **XQuery** .

J.Sroka, G.Kaczor, J.Tyszkiewicz, A.M.Kierzek, XQTav: an XQuery processor for Taverna environment, **Bioinformatics**, 2006.

Przepływy danych w Internecie cd.

Zdefiniowano język DFL do formalnego definiowania przepływów danych, łączący sieci Petriego i rachunek zagnieżdżonych relacji NRC. Udowodniono podstawowe twierdzenie o efektywności przepływów zdefiniowanych w tym systemie.

J.Hidders, N.Kwasnikowska, J.Sroka, J.Tyszkiewicz, J.Van den Bussche, DFL: A dataflow language based on Petri nets and nested relational calculus, 2006, złożone.

Artykuły w czasopismach

1. M.Bojańczyk, T. Colcombet, Tree-Walking Automata Cannot Be Determinized, **Theoretical Computer Science**, 350(2-3), **2006**, 164-173.
2. T.Colcombet, D.Niwiński, On the positional determinacy of edge-labeled games, **Theoretical Computer Science**, 352 (1-3), **2006**, 190-196.
3. S.Lasota, I.Walukiewicz, Alternating Timed Automata, **ACM Transactions on Computational Logic**, przyjęte.
4. S.Lasota, Decidability of Performance Equivalence for Basic Parallel Processes, **Theoretical Computer Science**, 360, **2006**, 172-192.
5. M.Peczarski, An improvement of the tree code construction, **Information Processing Letters**, 99, **2006**, 92-95.
6. J.Sroka, G.Kaczor, J.Tyszkiewicz, A.M.Kierzek, XQTav: an XQuery processor for Taverna environment, **Bioinformatics**, 22(10), **2006**, 1280-1281.

Sprawozdania z konferencji – *Springer Lecture Notes*

1. M.Bojańczyk, A Bounding Quantifier, **CSL 2004**, LNCS 3210, 41-55.
2. M.Bojańczyk, T. Colcombet, Tree-Walking Automata Cannot Be Determinized, **ICALP 2004**, LNCS 3142, 246-256.
3. M.Bojańczyk, M.Samuelides, T.Schwentick, L.Segoufin, Expressive Power of Pebble Automata, **ICALP 2006**, LNCS 4052, 7-16.
4. M.Bojańczyk, I.Walukiewicz, Characterizing EF and EX Tree Logics, **CONCUR 2004**, LNCS 3170, 131–145.
5. S.Dziembowski, U.Maurer, On Generating the Initial Key in the Bounded-Storage Model, **EUROCRYPT 2004**, LNCS 3027, 126-137.
6. S.Dziembowski, Intrusion-Resilience Via the Bounded-Storage Model, **Theory of Cryptography Conference 2006**, LNCS 3876, 207-224.

Sprawozdania z konferencji – *Springer Lecture Notes* cd.

7. S.Dziembowski, On Forward-Secure Storage, **Advances in Cryptology CRYPTO 2006**, LNCS 4117, 251-270.
8. S.Froeschle, S.Lasota, Decomposition and Complexity of Hereditary History Preserving Bisimulation on BPP, **CONCUR 2005**, LNCS 3653, 263–277.
9. T.Knapik, D.Niwiński, P.Urzyczyn, I.Walukiewicz, Unsafe Grammars and Panic Automata, **ICALP 2005**, LNCS 3580, 1450-1461.
10. S.Lasota, I.Walukiewicz, Alternating Timed Automata, **FOSSACS 2005**, LNCS 3441, 250–265.
11. S.Lasota, W.Rytter, Faster Algorithm for Bisimulation Equivalence of Normed Context-Free Processes, **MFCS 2006**, LNCS 4162, 646-657.
12. E.Kopczyński, Half-Positional Determinacy of Infinite Games, **ICALP 2006**, LNCS 4052, 336–347.
13. F.Murlak, On deciding topological classes of deterministic tree languages, **CSL 2005**, LNCS 3634, 428-442.

Sprawozdania z konferencji – *Springer Lecture Notes* cd.

14. F.Murlak, The Wadge Hierarchy of Deterministic Tree Languages, **ICALP 2006**, LNCS 4052, 408-419.

15. A.Szałas, J.Tyszkiewicz, On the Fixpoint Theory of Equality and its Applications, **Relations and Kleene Algebra in Computer Science 2006**, LNCS 4136, 388-401.

Sprawozdania z konferencji – inne

1. M.Bojańczyk, T. Colcombet, Tree-Walking Automata Do Not Recognize All Regular Languages, **ACM STOC 2005**, 234-243.

2. M.Bojańczyk, T. Colcombet, Bounds in ω-Regularity, **IEEE LICS 2006**, 285-296.

3. M.Bojańczyk, A. Muscholl, Th. Schwentick, L. Segoufin, C.David, Two-Variable Logic on Data Trees and XML Reasoning, **ACM PODS 2006**, 10–19.

4. M.Bojańczyk, C.David, A. Muscholl, Th. Schwentick, L. Segoufin, Two-Variable Logic on Words with Data, **IEEE LICS 2006**, 7-16.

Sprawozdania z konferencji – inne cd.

5. S.Froeschle, S.Lasota, Normed Processes, Unique Decomposition, and Complexity of Bisimulation Equivalences, **INFINITY 2006**, ENTCS, Elsevier.
6. S.Lasota, D.Nowak, Yu Zhang, On completeness of logical relations for monadic types, **ASIAN 2006**.
7. S.Leszczyński, When Small Number of Alternations is Not Enough, **SofSem 2004**, Charles University's Publishing House.
8. D.Niwiński, I.Walukiewicz, Deciding Nondeterministic Hierarchy of Deterministic Tree Automata, **WOLLIC 2004**, ENTCS, Elsevier.

Rozprawa doktorska

M.Bojańczyk, Decidable Properties of Tree Languages, Uniwersytet Warszawski, **2004**.

Manuskrypty złożone do publikacji

1. J.Hidders, N.Kwasnikowska, J.Sroka, J.Tyszkiewicz, J.Van den Bussche,

   DFL: A dataflow language based on Petri nets and nested relational calculus, 2006.

2. M.Peczarski, Strategy in Ulam's Game and Tree Code Give Error-Resistant Protocols, 2004.

3. D.Kowalski, M.Strojnowski, Gossiping in the presence of omission failures, 2006.