

**СПРАВКА ЗА НАУЧНИТЕ ПРИНОСИ
НА ТРУДОВЕТЕ ВКЛЮЧЕНИ В МОНОГРАФИЧНИЯ ТРУД
ПРЕДСТАВЕН ОТ 12 РАВНОСТОЙНИ ПУБЛИКАЦИИ**

ТЕМА:

**“МАСИВНА ПАРАЛЕЛИЗАЦИЯ И IN-SITU ВИЗУАЛИЗАЦИЯ
НА НАУЧНИ ПРИЛОЖЕНИЯ ЗА СУПЕРКОМПЮТРИ И
АНАЛИЗ НА ГОЛЕМИ МАСИВИ ДАННИ НА
ВИСОКОПРОИЗВОДИТЕЛНИ КОМПЮТЪРНИ КЛЪСТЕРИ”**

НА ГЛ. АС. Д-Р ДЕСИСЛАВА АНТОНОВА ИВАНОВА
КАТЕДРА „ИНФОРМАТИКА“,
ФАКУЛТЕТ ПРИЛОЖНА МАТЕМАТИКА И ИНФОРМАТИКА,
ТЕХНИЧЕСКИ УНИВЕРСИТЕТ - СОФИЯ
С КОИТО УЧАСТВА В КОНКУРСА ЗА ДОЦЕНТ
ПО ПРОФЕСИОНАЛНО НАПРАВЛЕНИЕ
4.6: „ИНФОРМАТИКА И КОМПЮТЪРНИ НАУКИ“

Приносите в монографичния труд представен от 12 равностойни научни публикации на гл. ас. д-р Десислава Антонова Иванова могат да бъдат групирани в две научни области, както следва: 1) *Масивна паралелизация и In-situ визуализация на научни приложения за суперкомпютри* и 2) *Анализ на големи масиви данни на високопроизводителни компютърни клъстери*.

4.1. Масивна паралелизация и In-situ визуализация на научни приложения за суперкомпютри.

Научно-приложни приноси, доказващи с нови средства на съществуващи проблеми при масивната паралелизация и In-situ визуализация на научни приложения за суперкомпютри:

- Разработени са паралелни алгоритми и модели на паралелни изчисления за обработка на мултиспектрални данни от дистанционно наблюдение на Земята за суперкомпютри с цел откриване и мониторинг на природни бедствия: пожари, обезлесяване, наводнения, замърсяване на водни басейни и почви, [I.1, I.2].
- Анализирани са особеностите на паралелни програмни модели за обработване на изображения за CPU и GPU (OpenMP и CUDA), за

филтрация и интерполация на изображения и уейвлетна декомпозиция. Проведени са експериментални изследвания и е оценена производителността (време на изпълнение, ускорение, ефективност) на OpenMP и CUDA моделите на базата на паралелни програмни имплементации, [I.3].

- Разработени са модели на паралелни изчисления за научни изследвания в областта на океанологията. Направени са оценка на паралелната производителност и анализ на скалирането на паралелното приложение за океанология на основата на масивно паралелни програмни имплементации, оптимирани за изпълнение на таргетни суперкомпютри и високопроизводителни хетерогенни компютърни клъстери с графични ускорители, [I.4, I.5, I.6, I.8, I.10].
- Разработени са модели на паралелни изчисления за научни изследвания в областта на космически изследвания и астрофизика. Направени са оценка на паралелната производителност и анализ на скалирането на паралелното приложение за астрофизика на основата на масивно паралелни програмни имплементации, оптимирани за таргетни суперкомпютри и високопроизводителни хетерогенни компютърни клъстери с графични ускорители, [I.7].
- Предложен е интегриран подход за интерактивна In-situ визуализация и пост-обработка на големи масиви данни от масивно-паралелни симулации, който се основава на софтуерна интеграция на паралелно изпълнявани софтуер за симулация и софтуер за визуализация и постобработка посредством SSH тунели и клиент-сървър комуникация, [I.11].
- Предложеният интегриран подход за интерактивна In-situ визуализация и пост-обработка на големи масиви данни от масивно-паралелни симулации е тестван за масивно-паралелни симулации в областта на океанологията и астрофизиката, [I.11].
- Направени са оценка и анализ на производителността на високопроизводителни компютърни архитектури от последно поколение, *NVIDIA GeForce GTX Titan, NVIDIA GeForce GTX 680 and AMD Radeon HD7970*, на основата на синтетични и реални еталонни програми, [I.9].
- Оценена е енергийната ефективност на високопроизводителни компютърни архитектури от последно поколение *NVIDIA GeForce GTX Titan, NVIDIA GeForce GTX 680 and AMD Radeon HD7970*, на базата на проведени тестове за потреблението на електроенергия при различни работни товари, [I.9].

Приложни приноси, доказващи с нови средства на съществуващи проблеми при развитието на масивно-паралелни приложения за суперкомпютри:

- Разработен е софтуер BlueVision за обработка на мултиспектрални данни от дистанционно наблюдение на Земята за суперкомпютри и хетерогенни многопроцесорни компютърни клъстери с цел откриване и мониторинг на природни бедствия, съставен от пет модула: PARFIRE (откриване на пожари), PARDEFOR (за определяне на обезлесяване), PARSOIL (за откриване на области с осолени почви), PARWATER (за търсене на замърсени водни площи), PARFLOOD (за определяне на наводнени райони), [I.1, I.2].
- Верификация и валидиране на софтуерните модули на софтуера BlueVision са направени с данни от MODIS (Moderate-resolution Imaging Spectroradiometer) - сателитен сензорен инструмент пуснат в земната орбита от НАСА през 1999 г. на борда на сателит Terra и през 2002 г. – на борда на сателит Aqua, [I.1, I.2].
- Оценка на производителността на софтуерните модули е извършена на базата на паралелни програмни имплементации на таргетни високопроизводителни архитектури (българския суперкомпютър IBM BlueGene/P и хетерогенен компютърен клъстер локализиран в ТУ-София), [I.1, I.2].

4.2. Анализ на големи масиви данни на високопроизводителни компютърни клъстери.

Научно-приложни приноси, доказващи с нови средства на съществуващи проблеми при анализа на големи масиви данни на високопроизводителни компютърни клъстери:

- Предложен е концептуален модел за анализ на големи масиви данни за океанология в екосистемата на Интернет на нещата (ИIoT) с използване на средствата за машинно обучение, [I.12].
- Верификация и валидиране на предложения концептуален модел са направени с данни предоставени от американската агенция по атмосферата и океаните (NOAA), [I.12].
- Направена е оценка на прецизността на предложения концептуален модел на базата на паралелни програмни имплементации в среда на Apache Spark, [I.12].

С П Р А В К А

ЗА НАУЧНИТЕ ПРИНОСИ НА ТРУДОВЕТЕ ИЗВЪН МОНОГРАФИЯТА

НА ГЛ. АС. Д-Р ИНЖ. ДЕСИСЛАВА АНТОНОВА ИВАНОВА,

Катедра “Информатика”
Факултет Приложна Математика и Информатика”
Технически университет – София

с които участва в конкурс за доцент по
НАУЧНО НАПРАВЛЕНИЕ 4.6 “ИНФОРМАТИКА И КОМПЮТЪРНИ НАУКИ”

Научните приноси на представените трудове извън монографичния труд (представен от 12 равностойни публикации) могат да бъдат групирани в следните области: 1) *Техническа информатика*: Архитектура на системни мрежи за суперкомпютри; 2) *Приложна информатика*: Анализ на големи масиви данни за персонализирана медицина; 3) *Индустриална информатика*: Интелигентно компонентно-базирано управление на индустриални процеси и системи.

Приносите на научните трудове в първата област (Техническа информатика: Архитектура на системни мрежи за суперкомпютри) се състоят в следното:

В научния труд II.1 е направен сравнителен анализ на архитектурните аспекти на най-мощните суперкомпютри от класация топ500, включващ системните комуникационни мрежи и топологичните особености на суперкомпютрите.

В труд II.3 е направена оценка на комуникационната производителност на системна мрежа с топология *TOFU* за високопроизводителни компютърни системи на базата на паралелни компютърни симулации в среда на мрежов симулатор на дискретни събития *OMNeT++*. Изследвано е влиянието на разпределението на трафика и размера на пакетите върху латентността на системната мрежа.

В труд II.4 е предложен архитектурен проект на високоскоростен комутатор и системна мрежа с топология Омега с рециркулация за високопроизводителни компютърни клъстери. Предложените архитектури на високоскоростен комутатор и системна мрежа с топология Омега с рециркулация са верифицирани на основата на паралелни компютърни симулации в среда на мрежов симулатор на дискретни събития *OMNeT++*. На основата на симулационни експерименти са оценени и анализирани параметрите на комуникационната производителност на системната мрежа. Изследвано е влиянието на разпределението на трафика и размера на пакетите върху латентността и пропускателната способност на системата мрежа.

В труд II.7 е направена оценка на комуникационната производителност на системна мрежа с топология „BBN Butterfly“ на базата на паралелни симулации в среда на мрежов симулатор на дискретни събития OMNET++ (MPI), които са изпълнени на IBM HS22 Blade Center, локализиран в лабораторията по „Високопроизводителни компютърни системи и ГРИД технологии“ на Техническия Университет - София. Изследвано е влиянието на разпределението на трафика и размера на пакетите върху параметрите на комуникационната производителност на системата мрежа. Изследвано е ускорението от паралелното изпълнение на симулациите за системната мрежа при мащабиране на входните данни.

В труд II.8 е предложен обобщен симулационен модел на комутатор с входно буфериране и скалируем радикс. Работоспособността на предложени обобщен модел е доказана за случая на системни мрежи с топологии от типа „k-ary n-cube“, изградени с комутатори с входно буфериране. Симулационните модели са верифицирани на основата на паралелни компютърни симулации в среда на мрежов симулатор на дискретни събития OMNeT++. На основата на симулационни експерименти са оценени и анализирани параметрите на комуникационната производителност на системните мрежи.

Приносите на научните трудове във втората област (Приложна информатика: Анализ на големи масиви данни за персонализирана медицина) се състоят в следното:

В научния труд II.2 е направен анализ на съществуващите методи и средства за машинно обучение и приложението на ИТ за ранно откриване на рак на щитовидна жлеза в екосистемата на Интернет на нещата в медицинската образна диагностика (IoMIT).

В труд II.23 е предложен концептуален модел за ранно откриване на рак на гърдата, базиран на машинно обучение за анализ и извличане на знания за персонализирано лечение. Работоспособността на предложени концептуален модел е доказана на основата на програмна имплементация на PYTHON за класификатор на Бейс (Naive Bayes). Програмните имплементации са проведени в среда на *Apache Spark*. За верификацията и валидирането са използвани бази данни на *Wisconsin* за рак на гърдата.

Изследването в труд II.24 е в областта на екосистемата на Интернет на нещата в медицинската образна диагностика (IoMIT). В научния труд е предложен концептуален модел за ранна диагностика на рак на щитовидната жлеза. Най-съществената част от предложени концептуален модел е сегментирането на изображенията от ултразвук и детектирането на доброкачествен и злокачествен тумор. Откриването на аномалии за рак на щитовидната жлеза се базира на използването на средствата за машинно обучение. Работоспособността на предложени концептуален модел е доказана на основата на програмна имплементация на PYTHON за конволюционни невронни мрежи (CNNs) и машини с поддържащи вектори

(SVM). Програмните имплементации са проведени в среда на Apache Spark с използване на библиотеката за машинно обучение MLlib. Верификацията и валидирането са направени с използването на набор за обучение и тестване съставени съответно от 282 и 250 изображения. Наборът от данни съдържа 96 изображения със злокачествен и 186 изображения с доброкачествен рак на щитовидната жлеза. Направен е сравнителен анализ и оценка на точността на получените резултати при използването на конволюционни невронни мрежи (CNNs) и машини с поддържащи вектори (SVM) за ранна диагностика на рак на щитовидната жлеза.

В труд II.25 е разработено мобилно приложение за разпознаване на ръкописни обекти (букви и цифри) с използване на изкуствени невронни мрежи (ANN). Мобилното приложение е имплементирано на операционна система iOS и е написано на програмния език SWIFT. Верификацията и валидирането на приложението е направено с набора от данни MNIST, който съдържа 60 000 примера ръкописни обекти за обучение и 10 000 примера за валидиране. Направена е предварителна обработка на данните с библиотека Tensor-Flow. Проведени са тестове и е оценена точността на разпознаване на приложението.

Приносите на научните трудове в третата област (Индустриална информатика: Интелигентно компонентно-базирано управление на индустриални процеси и системи) се състоят в следното:

В труд II.5 е разработена компютърно базирана система за безконтактно управление на процеса на коагулация на млякото използваща дискретни уейвлитни трансформации. Проведени са тестове в реална среда. Анализът на получените резултатите показват, че предложената система предлага поредица от предимствата в сравнение с конвенционалните подходи, а именно: 1) предложената система дава възможност за наблюдение на динамиката на процеса на трансформация на млякото; 2) предложената система дава възможност за безконтактно управление на процеса в реално време, което не нарушава структурата на продукта; 3) системата покрива хигиенните и санитарните изисквания на хранително вкусовата промишленост.

В труд II.6 е предложен модел на паралелни изчисления за разпознаване на материали с използване на алгоритъма на Малат. За целта е избрана парадигмата "главен/подчинен". Разработен е класификатор за разпознаването на материалите, който използва спектрален анализ на входните сигнали чрез дискретни уейвлитни трансформации (DWT). Обучението на класификатора се извършва чрез алгоритъма *kMeans* и бърза уейвлитна трансформация на входни данни и извличане на коефициентите. Работоспособността на класификатора е доказана на базата на хибридни паралелни програмни имплементации (OpenMP+MPI), проведени на компактен хетерогенен компютърен клъстер.

В труд II.9 е предложен подход за управление на компонентно-базирани системи основан на три етапа: 1) Разработка на функционални софтуерни компоненти (базово управление) с цел подобряване на гъвкавостта. Това се постига чрез използване на общи функционални многократно използваеми компоненти, които могат да бъдат използвани многократно при управлението на различни съоръжения и процеси с подобна функционалност. 2) Предложен е модел, базиран на сигнално интерпретиращи мрежи на Петри (Signal Interpreted Petri Nets - SIPN) за описание на процедурите, съгласно стандарта за проектиране на периодични процеси *ISA S88*. По този начин може да бъде извършен формален анализ на модела. 3) Предложен е модел на съответствието за трансформация от модел, базиран на сигнално интерпретиращи мрежи на Петри в мрежа от функционални блокове посредством софтуерната архитектура S^3 (Scheduler-Selector-Synchronizer). Моделът е интерпретиран като оперативен график (Schedule of operation (SOP)), който се управлява от S^3 компоненти. Работоспособността на предложения подход е доказана на базата на компютърни симулации за инсталация на Festo, представяща процесите на преработка на целулоза (Festo Mini Pulp Process - MPP). Предложеният подход повишава гъвкавостта при проектиране и проправя пътя към ефективно компонентно базирано управление на оперативния график.

В труд II.10 е предложен подход, който комбинира стандарта за проектиране на периодични процеси *ISA S88* със стандарта за имплементация на разпределено управление *IEC 61499* на ниво фазови модели и тяхната имплементация. За да отговори на йерархичната структура на моделите, съгласно стандарта *ISA S88* е предложено за целите на формализацията да бъдат използвани йерархични модели, базирани на сигнално интерпретиращи мрежи на Петри. За да се постигне онлайн реконфигурация на системата за управление се използва софтуерна архитектура S^3 (Scheduler-Selector-Synchronizer). Работоспособността на предложения подход е доказана на базата на компютърни симулации за инсталация на Festo, представяща процесите на преработка на целулоза (Festo Mini Pulp Process - MPP). Използването на формални модели, представени чрез йерархични сигнално интерпретиращи мрежи на Петри, описващи графика за изпълнение на управлението по формален начин, повишат надеждността на системите за управление.

В трудовете II.11, II.12 е предложен нов подход, който комбинира стандарта *ANSI/ISA-S88* и сигнално интерпретиращите мрежи на Петри (SIPN) за управление на последователността на супервайзорно ниво на периодични процеси. Целта на предложения подход е да повиши надеждността и многократната използваемост и да се позволи бързата реконфигурация на инсталации за периодични процеси. Основните предимства от използването на предложения подход е, че не само поддържа формалната верификация на разработения модел за последователно управление, а също така предлага и добре структурирани механизми за имплементация на контролери.

Йерархичните сигнално интерпретиращи мрежи на Петри и модулността на прилаганите техники за проверка на модела (NuSMV) позволява имплементация на методологията, заложена в стандарта ANSI/ISA-S88. Работоспособността на предложения подход е доказана на базата на компютърни симулации с примери на инсталация на въздушен компресор и инсталация за периодични процеси за дестилация и извличане на терпентини. Тази комбинация увеличава надеждността на компонентно-базираните системи за управление и съкращава жизненият цикъл на разработка.

В труд II.13 е предложен подход за управление на периодични процеси включващ: 1) Декомпозиране на управлението съгласно стандарта за управление на периодични процеси *ISA S88*, на секционни процедури, чиято управляваща последователност зависи от рецептата за управление, дефинираща графика на изпълнение. 2) Функционалните компоненти са имплементирани като функционални блокове съгласно стандарта IEC-61499. 3) Оперативният график на компонентите за управление е имплементиран, съгласно стандарта IEC-61499, на базата на софтуерна архитектура S^3 (Scheduler-Selector-Synchronizer) и 4) Модели, базирани на сигнално интерпретиращи мрежи на Петри, описващи последователността на изпълнение на управлението. Работоспособността на предложения подход е доказана на базата на компютърни симулации за реална инсталация за дестилация на сулфатен терпентин, намираща се в завод „Кристал” във Велинград.

В труд II.14 е предложен подход за разработка на софтуерна библиотека от многократно използвани и верифицирани компоненти за изграждане на следващото поколение системи за управление на периодични процеси и производства, базирани на стандарта *IEC 61499*. Представени са някои от основните компоненти на библиотеката и е илюстрирана тяхната многократно използваемост за две различни рецепти за управление на системи за създаване на вакуум: 1) В система за създаване на вакуум, която е част от инсталация за дестилация на сулфатен терпентин, където компонентната «клапан» е използвана 4 пъти и 2) В система за създаване на вакуум, която е част от инсталация за електронно-лъчево топене на метали, същата компонента е използвана 5 пъти. Предложеният подход е изключително полезен за етапите на разработка на следващото поколение системи за управление на периодични процеси и производства, поради факта, че неговото използване води до значително съкращаване на етапите на разработка на системите за управление, което от своя страна води до икономия на време и финансови средства.

В труд II.15 е доказана работоспособността на предложения подход за разработка на софтуерна библиотека от многократно използвани и верифицирани компоненти за изграждане на следващото поколение системи за управление на периодични процеси и производства, базирани на стандарта *IEC 61499* (II.14) за управление на вакуума в инсталация за Електронно лъчевото топене и рафиниране (ЕЛТР) на Института по електроника към БАН. Едно от

основните предимства от използването на този подход е възможността за многократна употреба на софтуерните компоненти от библиотеката, базирани на стандарта IEC61499 и създадени в инженеринговата среда FBDK (Function Block Development Kit).

В труд II.16 е предложен подход за формална верификация на софтуерни компоненти, разработени съгласно стандарта IEC61499. Подходът обхваща следните стъпки: 1) Спецификация на формалния модел на алгоритъма за управление, с използване на сигнално-интерпретиращи мрежи на Петри (SIPN), описващи планираното поведение на системата по формален начин; 2) Верификация базирана на методите за проверка на модела (model checking). 3) Предложена е процедура за верификация на разработените компонентни модели, базирани на стандарта IEC61499 и представляващи част от текущата софтуерна библиотека за управление на периодични процеси. Работоспособността на предложения подход е доказана за случая на управление на вакуума в инсталация за електронно лъчевото топене и рафиниране с използване на софтуерния инструмент *NuSMV*.

В труд II.17 е предложен подход за спецификация и верификация на основните функционални блокове, съгласно стандарта IEC 61499 и на базата на сигнално-интерпретиращите мрежи на Петри (SIPN) и формални методи за верификация с използване на крайни автомати в средата на *NuSMV*. Работоспособността на предложения подход е доказана на базата на компютърни симулации с пример за управление на компресорна система.

В труд II.19 е предложена архитектура ориентирана към услуги и модел на уеб услуги за периодични процеси. Предложени са модели на *back-end* услуги за управление на периодични процеси, съобразени със стандарта *ANSI/ISA SP88*: услуги за данни, услуги за управление на достъпа, информационни услуги и услуги за визуализация и мониторинг.

В труд II.20 е предложена архитектура ориентирана към услуги и модел на уеб услуги за мониторинг и визуализация на периодични процеси. Предложените архитектура и модели на уеб услуги са приложени за инсталация за дестилация на сулфатен терпентин.

В научния труд II.21 са анализирани и обобщени подходите и методите в областта на разработката на разпределени системи за управление и свързаните с тяхното прилагане задачи и проблеми. Предложени са два примерни подхода: 1) Първият подход е базиран на комбинираното използване на стандарта IEC61499 за управление на процеси и инструментариума на сигнално интерпретиращите мрежи на Петри (SIPN), предоставящи възможност както за формална верификация на моделите за управление, така също и подпомагаща валидацията на затворената система за управление. Работоспособността на предложения подход е доказана на базата на компютърни симулации за управление на инсталация за ректификация на сулфатен терпентин. 2) Вторият подход се основава на унифицирания език за моделиране *UML*, осигуряващ реализацията на имплементиращ обектно-ориентиран подход, покриващ пълния жизнен цикъл на разработка на

разпределени приложения за управление и позволяващ ранно откриване на грешки. Вторият подход е илюстриран с пример за управление на пилотна разпределителна станция на *FESTO*.

В научния труд II.22 са обобщени: 1) Разработените модели съгласно стандарта ISA S88 за управление на инсталация за дестилация на сулфатен терпентин на различни нива на декомпозиция; 2) За всяка една от процедурите за управление е представено формалното описание чрез сигнално-интерпретиращи мрежи на Петри; 3) Имплементация на базата на предварително дефинирани IEC-61499 функционални блокове или чрез йерархична агрегация на блокове; 4) Представен е и графика за изпълнение на функционалните блокове посредством използването на софтуерната архитектура S³ (Scheduler-Selector-Synchronizer), която гарантира реконфигуриране чрез детайлно планиране на задачите на основните функции в стандарта IEC-61499.