

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS
INFORMATIKOS INŽINERIJOS STUDIJŲ PROGRAMOS
DOKTORANTAS ALBERTAS JURGELEVIČIUS

Doktorantūros metinė ataskaita

2016 M. SPALIO MĖN. 1 D. – 2017 M. RUGSĖJO
MĖN. 30 D.

Tema

Disertacijos tema:

- **„VIEŠŪJŲ PASKIRSTYTŲ SKAIČIAVIMŲ PLATFORMA DIDELIŲ DUOMENŲ TYRYBAI“**

Darbo vadovas:

- **PROF. HABIL. DR. LEONIDAS SAKALAUSKAS**

- **Įstojimo į doktorantūrą metai: 2016 m. spalio mėn. 1 d.**
- **Doktorantūros baigimo metai: 2020 m. rugsėjo mėn. 30 d.**

Tyrimas

Tyrimo objektas:

- Viešieji paskirstyti skaičiavimai
- Didelių duomenų tyrybos metodai

Tyrimo tikslas:

- Sukurti viešųjų paskirstytų skaičiavimų platformą didelių duomenų tyrybai

Tyrimo uždaviniai:

- Analitiškai apžvelgti viešųjų paskirstytų skaičiavimų modelio taikymo galimybes didelių duomenų tyrybai
- Identifikuoti mokslines problemas, kylančias taikant šį modelį
- Sukurti viešųjų paskirstytų skaičiavimų platformą didelių duomenų tyrybai

Planuojami rezultatai:

- Viešųjų paskirstytų skaičiavimų platforma didelių duomenų tyrybai
- Platformos taikymo praktiniams uždaviniams spręsti rezultatų apibendrinimas

Ataskaitinių metų darbo planas

Išlaikyti egzaminus:

- Stochastinis programavimas
- Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika

Atlikti mokslinių tyrimų disertacijos tema apžvalgą ir analizę:

- Atlikti didelės apimties duomenų gavybos metodų bei technologijų analitinę apžvalgą
- Identifikuoti mokslines problemas, kylančias uždaviniuose, susijusiuose su viešųjų paskirstytų skaičiavimų platformos panaudojimu didelės apimties duomenų tyrybai

Dalyvauti mokslinėje konferencijoje

Parengti 1 publikaciją

Atlikti darbai

Išlaikyti egzaminai:

- **Stochastinis programavimas.** Gautas įvertinimas 10 (puikiai). Egzamino komisijos sudėtis: prof. habil. dr. Leonidas Sakalauskas (pirmininkas), dr. Virginijus Marcinkevičius, doc. dr. Saulius Minkevičius. Pasiūlytas šiuo kursu paremtas matematinis modelis, skirtas viešųjų paskirstytų skaičiavimų platformos kaštų bei efektyvumo vertinimui didelių duomenų tyryboje. Pagal studijų planą, šis modelis toliau bus nagrinėjamas 2017 m. lapkričio mėn. – 2018 m. rugsėjo mėn.
- **Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika.** Gautas įvertinimas 6 (šeši). Egzamino komisijos sudėtis: prof. dr. Albertas Čaplinskas (pirmininkas), prof. dr. Saulius Gudas, doc. dr. Audronė Lupeikienė.

Atlikti darbai

Siekiant susipažinti bei išmokti dirbti su viena iš viešųjų paskirstytų skaičiavimų platformų, sukurtos BOINC karkaso pagrindu paremtos programos šiems uždaviniams spęst:

- atsitiktinių skaičių sumavimo uždavinio sprendimas, bei sukurtos programos pritaikymas veikti įvairiose kompiuterių platformose;
- atsitiktinių skaičių sumos minimizacijos uždavinio sprendimas, bei sukurtos programos pritaikymas veikti įvairiose kompiuterių platformose;
- duomenų transformavimo bei perkėlimo į kitą duomenų bazę (ETL) uždavinio sprendimas, bei sukurtos programos pritaikymas veikti įvairiose kompiuterių platformose. Šis bei kiti didelių duomenų apdorojimo uždaviniai pagal studijų planą bus toliau tiriami 2017 m. lapkričio mėn. – 2018 m. rugsėjo mėn.

Atlikti darbai

Siekiant susipažinti bei išmokti dirbti su viena iš viešųjų paskirstytų skaičiavimų platformų, ruošiamas BOINC karkaso pagrindu veikiantis [MII@home](#) projektas, skirtas sujungti institute esantiems nenaudojamiems skaičiavimų ištekliams, bei juos panaudoti tolimesniems tyrimams. Šiuo metu yra gauti reikiami kompiuteriniai ištekliai – prieiga prie serverio (**boinc.mii.vu.lt**).

Tyrimai

Atliktas 28 dienas trukęs kompiuterių elektros suvartojimo tyrimas. Tyrimo tikslas – nustatyti BOINC karkaso pagrindu sukurtos programinės įrangos naudojimo sukeltus papildomus kaštus.

- Tyrimas atliktas naudojant du atsitiktinai parinktus kompiuterius dviejose įstaigose (A ir B). Kompiuteriuose buvo įdiegta BOINC programa, dvi savaites vykdanči SETI@home projekto užduotis. Skaiciavimų metu buvo matuojami elektros energijos vartojimas, bei procesoriaus užimtumo laikas naudojant “Performance monitor” (programa prieinama Microsoft Windows sistemoje) bei elektroninį elektros energijos vartojimo matuoklį. Po to procesas buvo pakartotas nevykdant BOINC projektų. Tyrimo metu abu kompiuteriai buvo naudojami įstaigų darbuotojų darbo reikmėms. Rezultatai bus naudojami tolimesniuose tyrimuose vertinant tokios programinės įrangos naudojimo galimybes įmonėse bei organizacijose.

Tyrimo rezultatai



Tyrimo rezultatai

Organizacija	BOINC projektas	Procesoriaus laikas nevykdant skaičiavimų	Elektros energijos suvartojimas
A	-	98,77%	16,61 kWh
A	SETI@home	65,23%	22,03 kWh
B	-	83,49%	1 kWh
B	SETI@home	26,86%	2,09 kWh

Skaityti pranešimai

Pagal studijų planą atlikta mokslinių tyrimų disertacijos tema apžvalga ir analizė, identifikuotos mokslinės problemos. Rezultatai 2017-08-30 pristatyti tarptautinėje konferencijoje „BOINC: Fundamental and Applied Science and Technology (BOINC:FAST 2017)“ (Petrozavodskas, Rusija). Remiantis atlikta apžvalga, bei 2016 m. spalio mėn. – 2017 m. sausio mėn. atliktais elektros suvartojimo tyrimais paruošti straipsniai:

„**BOINC from the view point of Cloud computing**“. Straipsnis bus skelbiamas „CEUR Workshop Proceedings“ (CEUR-WS.org, ISSN 1613-0073);

„**Big Data mining using public distributed computing**“. Straipsnį planuojama skelbti tarptautiniame leidinyje „Information Technology and Control“. Šiuo metu straipsnis redaguojamas pagal gautas pastabas.

Skaityti pranešimai

2016-11-09 LOTD ir VU MII seminare "Operacijų tyrimas ir taikymai" skaitytas pranešimas. Pranešimo tema: "Viešųjų paskirstytų skaičiavimų platforma didelių duomenų gavybai".

2017-09-22 dieną buvo skaitomas pranešimas XVIII tarptautinėje mokslinėje kompiuterininkų konferencijoje („Kompiuterininkų dienos – 2017“, Kaunas). Pranešimo metu apžvelgtos viešųjų paskirstytų skaičiavimų taikymo galimybės didelių duomenų tyrybai.

Privalumai

- Nemokami kompiuterinių skaičiavimų išteklių;
- Dalinai sprendžiama konfidencialių duomenų apsaugos problema.

Trūkumai

- **Konfidencialių duomenų apsaugos trūkumas**
- Paslaugos kokybės neužtikrintumas
- Sistemų saugumas
- Skaičiavimų kainos vertinimo neapibrėžtumas
- **Skaičiavimų pajėgumo neapibrėžtumas**
- Vykdomųjų programų pritaikymas skirtingoms platformoms

Didžiųjų duomenų tyryba naudojant BOINC

- BOINC-MR
- distributedDataMining.org
- V-BOINC
- SCADAMAR

Išvados

Nustatytos mokslinės problemos:

- viešųjų paskirstytų skaičiavimų platformos kaštų bei efektyvumo vertinimas didelių duomenų tyryboje;
- duomenų saugumo bei konfidencialumo užtikrinimo metodai viešųjų paskirstytš skaičiavimų platformoje.

Kitų metų darbo planas

Išlaikyti egzaminus:

- Daugiamačių duomenų vizualizavimo metodai
- Lygiagretieji ir paskirstytieji skaičiavimai

Atlikti mokslinį tyrimą:

- Tyrimo metodikos sudarymas
- Teorinis tyrimas

Dalyvauti mokslinėje konferencijoje

Parengti 1 publikaciją