

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS

**Formalių specifikacijų taikymas projektuojant
išskirstytas sistemas**

Applying Formal Specifications to Design Distributed Systems

Magistro darbo planas

Atliko: Matas Savickis (parašas)

Darbo vadovas: Karolis Petrauskas, Doc., Dr. (parašas)

Recenzentas: (parašas)

Vilnius – 2020

TURINYS

ĮVADAS	2
0.1. Temos aktualumas bei naujumas	3
0.2. Darbo tikslas	4
0.3. Uždaviniai	4
0.4. Laukiami rezultatai.....	4
LITERATŪRA	5

Įvadas

Šiais laikais, kai kurios, programų sistemos yra kuriamos išskirstytų sistemų principu[Smi18]. Šioms sistemoms plečiantis ir didėjant atsiranda problema, kaip efektyviai ir patikimai apdoroti gaunamus duomenis[BDD⁺18]. Vienas iš būdų tai padaryti yra naudotis išskirstytų srautinių duomenų platformomis. Viena iš tokių platformų yra Apache Kafka[20a].

Apache Kafka buvo pradėta kurti kompanijos LinkedIn[20a]. Programos tikslas buvo centralizuota įvykių valdymo platformą, skirta internetinių duomenų integravimo užduotims atlikti. 2012 metais Apache Kafka sistema buvo perduota į Apache Software Foundation tolesniam vystymui. Šiuo metu Apache Kafka platforma yra žinučių siuntimo sistema, kurios dizainas pasižymi lengvu plečiamumu, patvarumu, patikimumu ir greičiu. Duomenys Apache Kafka platformoje yra išsaugomi saugiu, trukdžiams atspariu būdu. Apache Kafka kūrėjų teigimu, šiuo metu platforma naudoja daugiau negu 80 procentų didžiausių Jungtinių Valstijų įmonių[20a]. Kafka platforma yra plačiai naudojama įvairiose srityse, tokiuose kaip žurnalistika, debesijos paslaugos, muzikos srauto paslaugos, telekomunikacijos, bankinės paslaugos ir daugelis kitų[20a].

Norint užtikrinti Apache Kafka platformos kokybę, kūrėjai yra įgyvendinę skirtingų testų[20b]. Testai padeda atskleisti programos klaidas arba pasakyti ar naujas kodas nepaveikė seniau parašyto funkcionalumo[Whi00]. Tačiau net ir laikantis gerųjų testavimo praktikų nepavyksta išvengti programos klaidų. Net ir paskyrus daugiau resursų testavimui, netrivialiuose sistemose, tokiose kaip Apache Kafka, pilnas sistemos testavimas yra neįmanomas[SYC⁺04]. Todėl norint atrasti subtilesnius sisteminius sutrikimus tenka naudoti kitus metodus. Vienas iš tokių metodų yra formalus verifikavimas.

Formalios specifikacijos yra matematinės technikos, skirtos apibūdinti sistemų elgseną ir padėti kuriant jos dizainą, naudojant griežtas ir veiksmingas priemones[HP95]. Turint formalią specifikaciją galima ja pasinaudoti darant formalų verifikavimą ir parodant, kad sistemos dizainas yra adekvatus pagal sukurtą specifikaciją. Tai leidžia objektyviai koreguoti sistemos dizainą dar prieš pradedant ją įgyvendinti. Formalios specifikacijos sudaromos pasinaudojant tam tikra kalbas arba įrankius. Viena iš tokių, formalaus specifikavimo kalbų, yra TLA⁺.

TLA⁺ yra formalios specifikacijos kalba sukurta Leslie Lamport[Lam02]. Leslie Lamport 1980 metais sukūrė laiko veiksmų logiką(angl. Temporal Logic of Actions)[Lam94] pasinaudodamas Amir Pnueli 1977 metais sukurta laiko logika(angl. Temporal Logic) [Pnu77]. 1999 metais Leslie Lamport, naudodamasis laiko veiksmų logika, sukūrė formalaus specifikavimo kalbą TLA⁺[Lam02].

TLA⁺ kalba yra skirta kurti konkurencinių ir išskirstytų sistemų formalios specifikacijoms ir šias specifikacijas verifikuoti. Naudojant TLA⁺ galima specifikuoti šias išskirstytų sistemų savybes:

1. Gyvumas - turint laisvai pasirinktą sistemos būseną A, sistema turi pasiekti būseną B.
2. Saugumas - turint laisvai pasirinktą sistemos būseną A, sistema niekada nepateks į būseną B.

Kadangi TLA⁺ specifikacija yra rašoma formalia kalba tai leidžia patikrinti sukurtos specifikacijos saugumo ir gyvumo savybes.

Šias savybes mes galime patikrinti naudodamiesi TLC Model Checker(modelio tikrintojas). TLC yra nurodytos būsenos(explicit-state) modelio tikrintojas, kurio paskirtis yra pažingsniui pasiekti visas galimas sistemos būsenas pagal nurodytą formalią specifikaciją. Tačiau, kartais, pagal sukurta formalią specifikaciją, susidaro labai daug sistemos būsenų ir jas visas pasiekti naudojant TLC tampa nepraktiška. Tokiu atveju galime naudotis TLA⁺ specifikacijos įrodymo sistema TLAPS. TLAPS yra įrodymų sistema skirta patikrinti TLA⁺ įrodymus. Šios sistemos paskirtis yra patikrinti pateiktus teoremų įrodymus. Įrodžius teoremą laikoma, kad TLA⁺ specifikacija yra adekvati.

Viena iš formalių specifikacijų ir TLA⁺ panaudojimo industrijoje sėkmės istorijų yra Amazon Web Service(AWS) komandos 2014 metais išleistas straipsnis[NRZ⁺14]. Straipsnyje rašoma, kad AWS komanda naudojo TLA⁺ sudarant formalias specifikacijas dešimtyje projektų. Tuo metu AWS turėjo 7 komandas, kurios naudojos TLA⁺ kurdamos naujas programų sistemas. AWS sistemos specifikavimo metu buvo surasti 10 iki šiol neatrastų sisteminių klaidų, kurių atradimas ir pasiūlyti ištaisymai atskleidė tolimesnes sistemines klaidas, kurios taip pat buvo ištaisytos. Straipsnyje įvardinta ir kita, netiesioginė, nauda gauta formaliai specifikuojant sistemą: pagerėjęs bendras sistemos suvokimas, padidėjęs produktyvumas ir inovacijos.

Dar viena sėkmės istorija yra 2018 metais Kafka Summit konferencijoje pristatyta Apache Kafka TLA⁺ formali specifikacija sukurta Jason Gustafson[Gus04]. Pristatyme buvo parodyta, kad pritaikius TLA⁺ specifikuojant Apache Kafka duomenų replikavimo algoritmą buvo surastos ir pataisytos 3 retais atsitinkančios programos klaidos. Betaisant rastos ir pataisytos dar kelios klaidos.

0.1. Temos aktualumas bei naujumas

Iki šiol, kiek mums žinoma, Apache Kafka platforma buvo specifikuota tik vieną kartą[Gus04] neakademiniame kontekste ir sukurta specifikacija atnešė naudos padedant surasti sisteminės klaidas. Panašią mokslininkų sėkmę matome ir Amazon Web Service formalios specifikacijos sudarymo tyrimuose[NRZ⁺14]. Dėl papildomų Apache Kafka formalių specifikacijų stokos ir praeityje pasisėkusio formalios specifikavimo išskirstytuose sistemose manome, kad papildomi tyrimai Apache Kafka platformoje atneštu naudos surandant sisteminių klaidų arba užtikrinant, kad specifikuotoje sistemos dalyje jų nėra. Šiuo metu Apache Kafka sisteminių klaidų registre[20c] yra išspręstų ir neišspręstų klaidų kurių verifikavimas padėtų atskleisti naujas klaidas arba įrodyti kad klaidos ištaisytos adekvačiai. Apache Kafka platforma turi daug naudotojų[20a], todėl tolimesnis kokybės užtikrinimas Apache Kafka platformoje atneštų naudą. Sėkmingai identifikavus išskirstytų sistemų architektūrinius šablonus[hohpe2004enterprise] būtų galima įrodyti adekvatumą daug didesnei išskirstytų sistemų aibei ir šio darbo rezultatais būtų galima vadovautis kuriant atitinkamas išskirstytas sistemas.

0.2. Darbo tikslas

Pagerinti Apache Kafka platformos kokybę surandant sisteminės klaidas arba įrodant, kad sukurtoje specifikacijoje klaidų nėra. Įrodyti didesnės išskirstytų sistemų aibės algoritmų adekvatumą specifikuojant architektūrinius šablonus.

0.3. Uždaviniai

1. Išnagrinėti literatūrą susijusią su formaliais metodais, TLA⁺ specifikavimo kalba bei Apache Kafka platforma.
2. Specifikuoti išskirtas Apache Kafka platformos dalis naudojant TLA⁺ specifikavimo kalbą.
3. Įvertinti, ar pagal sukurta specifikaciją Apache Kafka platforma veikia adekvačiai.
4. Esant poreikiui įrodyti specifikacijos teoremas.
5. Patikrinti ar Apache Kafka įgyvendinimas atitinka specifikaciją verifikuojant naudojant įvykių žurnalus.
6. Apache Kafka platformoje surasti architektūrinius šablonus, kurie yra taikomi ir kituose išskirstytuose sistemose, bei juos specifikuoti.

0.4. Laukiami rezultatai

1. Pasirinktų Apache Kafka modulių specifikacija.
2. Įrodymas apie specifikacijos adekvatumą.
3. Apache Kafka specifikacijos ir įgyvendinimo sutapimo įvertinimas.
4. Išskirti ir specifikuoti išskirstytų sistemų šablonai taikomi kitose platformose.

Literatūra

- [20a] Apache kafka, 2020. URL: <https://kafka.apache.org/>.
- [20b] Apache kafka mirror tests, 2020. URL: <https://github.com/confluentinc/kafka/tree/master/tests>.
- [20c] Kafka - - asf jira, 2020. URL: <https://issues.apache.org/jira/projects/KAFKA/issues/KAFKA-10635?filter=allopenissues>.
- [BDD⁺18] Antonio Bucchiarone, Nicola Dragoni, Schahram Dustdar, Stephan T Larsen ir Manuel Mazzara. From monolithic to microservices: an experience report from the banking domain. *Ieee Software*, 35(3):50–55, 2018.
- [Gus04] Jason Gustafson. Hardening kafka replication. <https://kafka-summit.org/sessions/hardening-kafka-replication/>; <https://github.com/hachikuji/kafka-specification>, 2004. Pristatymas konferencijoje.
- [HP95] Gerard J Holzmann ir Doron Peled. An improvement in formal verification. *Formal Description Techniques VII*, p. 197–211. Springer, 1995.
- [Lam02] Leslie Lamport. *Specifying systems*, tom. 388. Addison-Wesley Boston, 2002.
- [Lam94] Leslie Lamport. The temporal logic of actions. *ACM Trans. Program. Lang. Syst.*, 16(3):872–923, 1994-05. ISSN: 0164-0925. DOI: 10.1145/177492.177726. URL: <https://doi.org/10.1145/177492.177726>.
- [NRZ⁺14] Chris Newcombe, Tim Rath, Fan Zhang, Bogdan Munteanu, Marc Brooker ir Michael Deardeuff. Use of formal methods at amazon web services. See <http://research.microsoft.com/en-us/um/people/lamport/tla/formal-methods-amazon.pdf>, 2014.
- [Pnu77] A. Pnueli. The temporal logic of programs. *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, p. 46–57, 1977. DOI: 10.1109/SFCS.1977.32.
- [SYC⁺04] Kevin Sullivan, Jinlin Yang, David Coppit, Sarfraz Khurshid ir Daniel Jackson. Software assurance by bounded exhaustive testing. *Proceedings of the 2004 ACM SIGSOFT international symposium on Software testing and analysis*, p. 133–142, 2004.
- [Smi18] Tom Smith. New research shows 63% of enterprises are adopting microservices architectures, 2018. URL: <https://www.globenewswire.com/news-release/2018/09/20/1573625/0/en/New-Research-Shows-63-Percent-of-Enterprises-Are-Adopting-Microservices-Architectures-Yet-50-Percent-Are-Unaware-of-the-Impact-on-Revenue-Generating-Business-Processes.html>.
- [Whi00] J. A. Whittaker. What is software testing? and why is it so hard? *IEEE Software*, 17(1):70–79, 2000. DOI: 10.1109/52.819971.