



synrc research center s.r.o.
RONÁČOVA 141/18, PRAHA 3 13000, CZECH REPUBLIC

**Формальні методи верифікованих середовищ
послідовних обчислень узгоджених та розподілених
паралельних систем**

Максим Сохацький

Київський Політехнічний Інститут
Листопад 2015

1 Вступ

1.1 Що означає назва

«Формальні методи верифікованих середовищ послідовних обчислень розподілених узгоджених паралельних систем»

Формальні методи тут розглядаються у широкому сенсі, особливо нас цікавитиме та частина математичного забезпечення яку можна формалізувати для доведення машинним способом, що унеможливує виникнення цілого класу помилок. Зокрема розглядатимуться усі теорії які застосовують до подібних систем з підвищеними запитами якості.

Розглядаючи такі системи, які піддаються формальній верифікації, кажуть про сертифіковане або верифіковане програмне забезпечення. А у випадку коли такі системи розповсюджуються на всі прошарки моделі OSI, то кажуть про замкнені верифіковані середовища. Саме розробка такого середовища є основним завданням даної роботи.

Чому саме послідовні обчислення? У цій частині мова йде про природу обчислень, чи природу послідовності висловлювань у формальній логіці. Саме лінійність кінцевого процесу виконання певного структурованого або циклічного алгоритму відіграє ключову роль у моделюванні віртуальних обчислювальних середовищ. Такі лінеаризовані системи уже показали свою ефективність в певному класі обчислень, таких як ланцюгова реплікація, реактивні системи, та інші моделі напівгруп навколо певних типів – протоколів взаємодії між процесами. Крім того такі послідовності подій піддаються статистичній обробці для визначення первних кластеризацій та інших кореляцій у просторі та часі, тому можна говорити про певні зручні, нормалізовані системи типів для такого роду маніпуляцій.

Побудова розподілених та паралельних, тобто здатних виконуватися на багатьох машинах одночасно, та узгоджених, тобто не блокуючих, а значить лінеаризованих систем управління процесами є кінцевим результатом який очікується від цієї роботи.

1.2 Контекст та сфера дослідження

За багато років кількість теорій, які використовуються для побудови програмного забезпечення значно розширилися: починаючи з теорії компіляції сучасних функціональних мов, та систем програмування на основі теорії типів, включаючи сучасні моделі обчислень, які побудовані на основі лямбда числення та числення процесів, закінчуючи віртуальними машинами які працюють у семантиці захищених, простих та структурою процесів, час яких розподіляється у прозорий та ефективний спосіб.

Сучасний розвиток техніки та теоретична межа швидкості обробки процесорів вивів на передній план алгоритми та структури даних які ефективно використовують розподілені у просторі та часі ресурси, як то об'єми пам'яті та обчислювальні потужності. Принципи та підходи паралельного та узгодженого програмування дають змогу масштабувати системи та обчислення, однак анонсують нові теорії для забезпечення коректності в умовах підвищеної складності алгоритмів у розподілених системах, таких як алгоритми забезпечення консистентності та транзакційності у розподілених системах PAXOS та CR.

З розвитком систем програмування, підвищилась якість інструментів для автоматичного доведення коректності, починаючи від інструментів побудованих на темпоральній логіці як TLA+ та NuPRL закінчуючи системами побудованими на індуктивних типах і які вже використовуються метематиками як Agda та Coq.

Багато методологій виникло і багато підходів були випробувані в польових умовах на різних, по критичності, рівнях якості. Починаючи від асемблерів, через процедурний та об'єктно-орієнтований підхід, до функціональних мов програмування з розвиненими системами типів, таких як Haskell та ML, а також мов, спеціалізованих для розподілених систем як Erlang.

Застосовуючи формальні методи доведення коректності велику частину має повнота за замкненість системи, адже недоведені або неперифіковані частини можуть вплинути на детермінованість а отже якість системи. Тому має велике значення забезпечення виконання системи якомога ближче до апаратного забезпечення. Таких напрямків побудови замкнених середовищ є всього три, на мовах Erlang, Haskell та OCaml.

1.3 Об'єкт дослідження

Об'єктом дослідження є усі можливі моделі обчислювальних середовищ в основному придатних до верифікованого аналізу та обробки формальними методами. Самі формальні методи теж є частиною об'єкта дослідження. Ми досліджуємо ті структури та алгоритми які дадуть максимально ефективний спосіб кодування та виконання, забезпечуючи при цьому семантику, яка використовується для машинного доведення коректності роботи алгоритмів та непротиворечивості структур даних.

Формальні мови програмування такі як Haskell та OCaml є надзвичайно потужними з точки зору теорії типів, що дало би змогу більш формально на ранніх етапах застосувати типізацію. З іншого боку Erlang пропонує більш довершені та прості механізми забезпечення віртуального середовища як на голому залізі так і під управлінням гіпервізора Xen. Так як нам усе одно доведеться конвертувати специфікації з розробленої нами формальної теорії Eхе, тому вибір мови з розвинутою системою типів не є основним критерієм. Для побудови замкненої системи була вибрана мова Erlang та її віртуальні машини LING та BEAM.

Усі моделі які побудовані в рамках запропонованої методології із застосуванням метамови та метамоделі операційного середовища Eхе виконуватимуться у контексті віртуальних машин на голому залізі без застосування операційної системи у ході виконання. Такі системи називаються системами з уніфікованим ядром яке пропонує свою модель виконання, у нашому випадку це віртуальні машини Erlang та комплекс програмного забезпечення розроблений для замкненого циклу виробництва.

1.4 Теоретичні сутності

У таксономії теоритичних сутностей умовно визначатимемо графічні топологічно-структурні, аналітичні, статистичні та логічні типи теорій, які ми використовуватимемо для опису комплексної теорії верифікованих середовищ послідовних обчислень розподілених узгоджених паралельних систем.

Теорія масового обслуговування

Теорія масового обслуговування застосовується для побудови статистичних моделей та запобігання відмов. Перші роботи у цій області належать шведському математику Агнеру Крурупу Ерлангу, який займався дослідженнями трафіка у телефонних мережах. Модель масового обслуговування достатньо адекватно описує роботу віртуальної машини Erlang, де клієнти – це процеси, які мають черги повідомлень, та здатні відправляти заявки на обслуговування у такі самі черги інших процесів. Ці заявки, чи повідомлення складають певний протокол взаємодії у системі таких процесів. Тому тут теорія масового обслуговування застосовується для визначення пропускну здатності системи.

Пі-числення

Теорія Пі-числення Роберта Мілнера є основним формалізмом обчислювальної теорії та її імплементації. З часів виникнення CSP числення розробленого Хоаром, Мілнеру вдалося значно розширити та адаптувати теорію до сучасних телекомунікаційних вимог, як наприклад хендовери в мобільних мережах. Основні теорми в моделі Пі-числення стосуються непротиворечивості та неблокованості у синхронному виконанні мобільних процесів. Так як сучасний Web можна розглядати як телекомунікаційну систему, тому у розробці додатків можна покладатися у тому числі і на такі моделі як Пі-числення.

Мережі Петрі

Мережі Петрі в даній роботі використовуються як прототип графічного лінгвістичного засобу структури категорій та системи типів. Оскільки такі графічні засоби як UML та різноманітні окремі технологічні стандарти як то BPMN більше допомагають ніж мішають, була розроблена також і графічна мова на базі графічної мови мереж Петрі, оскільки їх семантика ділить один простір з тематикою даної роботи. Ми візьмемо лінгвістичне забезпечення Мереж Петрі як прототип для нашої власної мови візуалізації структур обчислень.

Теорія категорій

Теорія категорій як робоча структурна теорія системи типів мови Eхе та категоріальна семантика числення процесів. Можна було би використовувати абстрактну алгебру загалом, та теорія напівгруп, а саме напівгруп активностей, проте ми будемо використовувати категоріальну семантику, що стало можливо завдяки роботам Лавіра та іншим.

Лямбда числення

Лямбда-числення як основна абстракція обчислювальної віртуальної машини. Будучи внутрішньою мовою декартово-замкненої категорії лямбда числення окрім змінних та констант у вигляді термів пропонує операції абстракції та аплікації, що визначає достатньо лаконічну та потужну структуру обчислень з функціями вищих порядків, та метатпізаціями, такими як System F, яка була запропонована вперше Робіном Мілнером в мові ML, та зараз присутня в більш складних, таких як System F ω системах.

Темпоральна логіка

Темпоральна логіка як індуктивна теорія верифікації розподілених алгоритмів застосовується до доведення коректності усіх нормалізованих підсистем. На основі теорії Леслі Лампорта [7], за яку він отримав премію Тюрінга.

Інтуїціоністська теорія типів Мартіна Льюфа

Системи з залежними типами як верифікаційні математичні формальні моделі для доведення коректності. Система Σ та Π типів, як кванторів існування та узагальнення.

1.5 Предмет дослідження

Предметом дослідження даної роботи є розробка формальних методів для побудови операційних середовищ та метамodelей для їх формальної специфікації. Розглянувши усі можливі математичні моделі опису формалізму процесів ми формуємо ряд вимог корисних для побудови ефективної моделі яка дозволить:

- Скоротити об'єм коду на порядок
- Нормалізувати дані для їх статистичної обробки
- Легко обчислювати показники системи масового обслуговування
- Легко доводити коректність
- Мінімізувати цикл розробки програмного забезпечення
- Підвищити ефективність виконання

Для цього ми визначаємо найбільш оптимальну модель потоку даних та функції для їх обробки, використовуючи категоріальний підхід для опису протоколів та їх категорій.

- Категоріальний погляд на протоколи процесів
- Визначення характеристик нормалізованих структур
- Побудова системи типів метамodelі
- Розробка системи додатків на основі метамodelі
- Впровадження та діагностика системи

Іншими словами словом ми розроблюємо теорію та імплементацію мінімалістичного сертифікованого верифікованого операційного середовища для компонентів замкненої системи: абстракції апаратного забезпечення, мова програмування, віртуальна машина, комунікаційні черги, бази даних, веб сервери, сервери додатків, та усе, що піддається верифікації та по можливості є коректним за побудовою. Починаючи з фундаментального формалізму лямбда та пі числення, через абстракції віртуальної машини до віртуальної апаратури, реальні сучасні додатки та протоколи, закінчуючи прикладами, засобами та документацією на створене обчислювальне середовище та його модель.

Алгебраїчний підхід

Ми будемо застосовувати алгебраїчні рекурсивні типи даних для побудови системи типів, тому зручно буде також застосовувати алгебраїчний підхід до генералізації теорії процесів. З алгебраїчної точки зору процеси, або кінечні автомати, представляють собою напівгрупи дій. З категоріальної точки зору процеси — це функтори які перетворюють категорії з декартовими добутками типів протоколу та стану процесу на категорію станів. Кожен процес визначається таким функтором, адже простір дій Σ для кожного процесу є унікальний.

$$\begin{array}{lcl}
 A & : & \Sigma \times X \rightarrow X \\
 X & : & \Sigma \times \Lambda_X \\
 & | & \perp \\
 \Sigma & : & \begin{array}{lcl} ok & \times & _ \\ error & \times & _ \\ io & \times & _ \\ \perp & & \end{array}
 \end{array}$$

У функціональних мовах програмування така категорія будем мати вигляд програми, де функтор буде представлений функцією перетворення станів процесу, а протокол буде типом-сумою усіх можливих запитів до процесу:

```

    action : ( protocol, state ) -> state
    state  : ( protocol, ----- )
           | []
protocol : ( ok,          ----- )
           | ( error,    ----- )
           | ( io,       ----- )
           | []

```

Протоколи повідомлень

Простір дій протоколу визначається сумою усіх можливих повідомлень.

$$\begin{array}{lcl}
 App & : & \begin{array}{lcl} init & \times & _ \\ client & \times & _ \\ replay & \times & _ \\ \perp & & \end{array} \\
 Store & : & \begin{array}{lcl} put & \times & _ \\ add & \times & _ \\ fold & \times & _ \\ \perp & & \end{array} \\
 Proc & : & \begin{array}{lcl} spawn & \times & _ \\ complete & \times & _ \\ run & \times & _ \\ \perp & & \end{array}
 \end{array}$$

Лінійність обчислень

Списки є фундаментальними структурами та найпростішим рекурсивним алгебраїчним типом. Природа обчислень, як результат виконання процесів теж лінійна. Можна навіть визначити оператор похідної, як функтор, який для алгебраїчного типу процесу буде давати список, а для списку буде давати скаляр. Як було показано [14] мережі петрі, а значить і процеси є моноїдами які генерують лінійну послідовність подій.

$$\begin{array}{lcl}
 \Delta \Lambda^0 & : & \times^0 \\
 \Delta \Lambda^n & : & \Lambda^{n-1}
 \end{array}$$

Лямбда числення

Для розробки нашої теорії ми застосуємо інтуїціоністську теорію типів Мартіна Льюфа. Ця система типів буде застосовуватися для доведення робочих теорем. База мови складає лямбда числення як внутрішня мова декартово-замкненої категорії. Лямбда числення в своїй основі складається з конструктора типу яка визначає функції та елімінатора – аплікації функції до аргументу.

$$\frac{\Gamma x : A \vdash B}{\Gamma \vdash \text{fun } x : A \rightarrow B}$$

$$\frac{\Gamma f : A \rightarrow B \quad \Gamma a : A}{\Gamma \vdash f a : B}$$

Алгебраїчні типи даних

Далі йде внутрішня мова алгебраїчних типів даних яка складається з суми та добутку, за допомогою яких контруюються суми-протоколи та кортежі-пакети, а також **nil** типу-атому, яким зручно термінувати рекурсивні типи даних, такі як **cons**.

$$\frac{}{\Gamma \vdash \perp}$$

$$\frac{\Gamma \vdash a : A \quad \Gamma \vdash b : B}{\Gamma \vdash (a, b) : A \times B}$$

$$\frac{\Gamma x : A \times B}{\Gamma \vdash \text{fst } x : A; \Gamma \vdash \text{snd } x : B}$$

$$\frac{\Gamma \vdash a : A \quad \Gamma \vdash b : B}{\Gamma \vdash a \mid b : A \otimes B}$$

$$\frac{\Gamma x : A \otimes B}{\Gamma \vdash \text{inl } x : A; \Gamma \vdash \text{inr } x : B}$$

Процеси і протоколи

Також ми анонсуємо процес як фундаментальний тип даних, подібний до функції але який здатний тритати певний стан у вигляді типа коротежа та бути здатним реагувати на повідомлення суми.

$$\frac{\Gamma \vdash \Sigma, X \quad \Gamma \vdash a : \Sigma \times X \rightarrow X}{\Gamma \vdash \text{proc}_\Sigma a : \pi}$$

$$\frac{\Gamma \vdash \text{proc} : \pi \quad \Gamma \vdash m : \Sigma}{\Gamma \vdash \text{join}_\Sigma (m, \text{proc}) \rightarrow \Sigma; \Gamma \vdash \text{async}_\Sigma (m, \text{proc}) \rightarrow \Sigma}$$

Логіка та квантори

Далі йдуть квантори \forall та \exists які теж виражаються як конструкції типів:

$$\frac{\Gamma x : A \vdash B \quad \Gamma \vdash A}{\Gamma \vdash \Pi(x : A)B} \quad \frac{\Gamma \vdash a : A \quad \Gamma x : A \vdash B \quad \Gamma b : B(x = a)}{\Gamma \vdash (a, b) : \Pi(x : A)B}$$

$$\frac{\Gamma x : A \vdash B \quad \Gamma \vdash A}{\Gamma \vdash \Sigma(x : A)B} \quad \frac{\Gamma \vdash a : A \quad \Gamma x : A \vdash B \quad \Gamma b : B(x = a)}{\Gamma \vdash (a, b) : \Sigma(x : A)B}$$

$$\frac{\Gamma \vdash x : A \quad \Gamma \vdash x' : A}{\Gamma \vdash Id_A(x, x')}$$

рефлексивність	:	$Id_A(a, a)$
пiстановка	:	$Id_A(a, a') \rightarrow B(x = a) \rightarrow B(x = a')$
симетричність	:	$Id_A(a, b) \rightarrow Id_A(b, a)$
транзитивність	:	$Id_A(a, b) \rightarrow Id_A(b, c) \rightarrow Id_A(a, c)$
конгруентність	:	$(f : A \rightarrow B) \rightarrow Id_A(x, x') \rightarrow Id_B(f(x), f(x'))$

Контраваріантні процеси

Ко-процеси це моноїди з перевернутою стрілкою, які визначають зворотній шлях виконання події та обернену зміни стану в зворотньому порядку. З точки зору типів нічого не змінюється, однак оберненої функції-функтора процесу-моноїда може і не існувати. Процеси які є одночасно коваріантними та контраваріантними називаються інваріантними процесами, сигнатура яких алгебраїчно збігається з сигнатурою бінарного дерева.

$$\begin{array}{lcl} X & : & \times^i \\ \Lambda & : & \times^4 I X \Lambda_{next} \Lambda_{prev} \\ & | & \perp \end{array}$$

$st(p) = p(st - 1)$ — застосування функції до протокольної заявки та стану процесу
 $p - 1(st) = st - 1$ — обернена функція

Типи процесів

<i>action</i>	:	$proc_{Proc} (Proc \mid \Sigma) \times process \rightarrow process$
<i>event</i>	:	$proc_{App} (App \mid \Sigma) \times cx \rightarrow cx$
<i>operation</i>	:	$proc_{Store} (Store \mid \Sigma) \times container \rightarrow container$

Суми та ко-суми процесів

Віртуальне середовище виконує моноїдальні процеси як одну велику суму. Кожна така сума представляє собою віртуальну одиницю виконання – віртуальну машину. В нашій теорії нас будуть цікавити переважно категоріальні суми процесів. Ко-суми чи добутки процесів представляють собою виконання функцій активності в одному і тому самому процесі, перекриваючись всередині нього по протоколам. Стан такого процесу розширюється як добуток станів.

Віртуальна машина

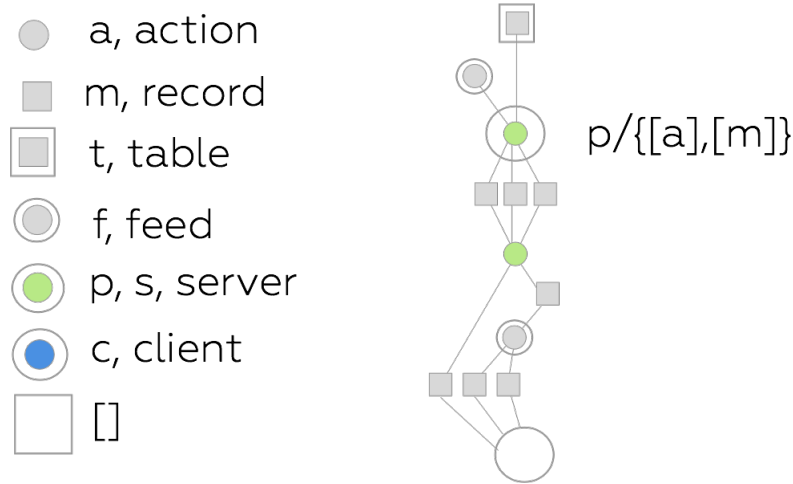


Рис. 1: Базові елементи віртуального середовища

action is the piece of code applied to message using pattern matching.

record is the tuple that defines a protocol.

client is the true customer of the system who is using service on daily basis.

server is the process nicely scheduled by the system.

feed is the server managed list of records.

table is the ixset or ets based queryable tables.

Сигнатури сервісів

app is the web clients monoidal state machine.

proc is the oriented graph monoidal executional engine.

store is the 2-category key-value database.

$$\begin{aligned}
 store^3 &: \times^3 \Lambda_{id_{seq}} \Lambda_{table} \Lambda_{feed} \\
 roster^2 &: \times^2 \Lambda_{feed} \Lambda_{feed}^2 \\
 app^2 &: \times^2 \Lambda_{session} \Lambda_{proc} \\
 depot^1 &: \times^1 \Lambda_{account} \\
 act^1 &: \times^1 \Lambda_{proc}
 \end{aligned}$$

1.6 Персистентність послідовностей

Для забезпечення персистентного зберігання послідовностей використовується очевидний спосіб зберігання елементів послідовності у key-value сховищі разом з референсами на наступні, та у випадку контраваріантних процесів, також і на попередні елементи у послідовності. Таким чином ми можемо перелічити усі елементи послідовності. Акумулятор розділеної у просторі та часі операції накопичує усі необхідні агрегації після кожної операції з послідовністю. Таким чином похідна від алгебраїчної структури списку буде скаляр-акумулятор після завершення перелічення.

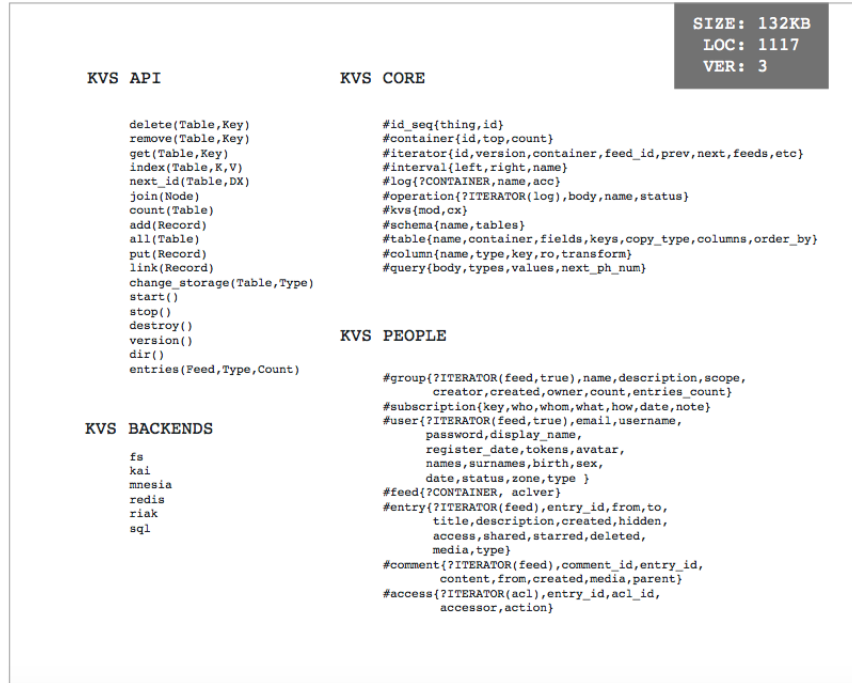


Рис. 2: Типи, функції та модулі

В даній роботі ми розглянемо формальні визначення сховища, його референсів, послідовностей через рекурсивні алгебраїчні типи даних та процеси, черги яких забезпечують впорядкованість послідовності запитів до послідовностей, що веде до лінійності операції у часі.

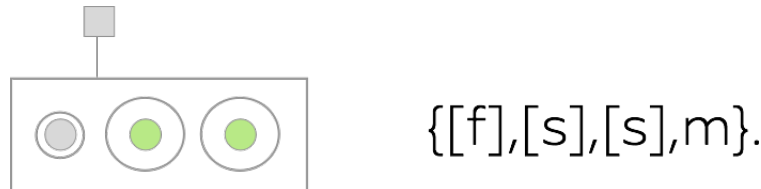


Рис. 3: Послідовності ключів, таблиць, індексів та інтерфейс

1.7 Мультипротокольний чат

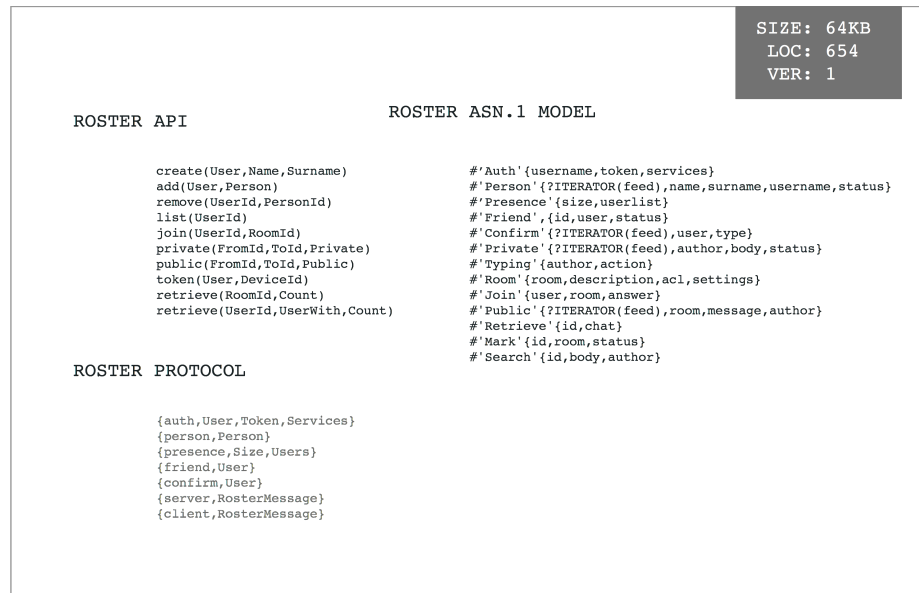


Рис. 4: Типи, функції та протокол

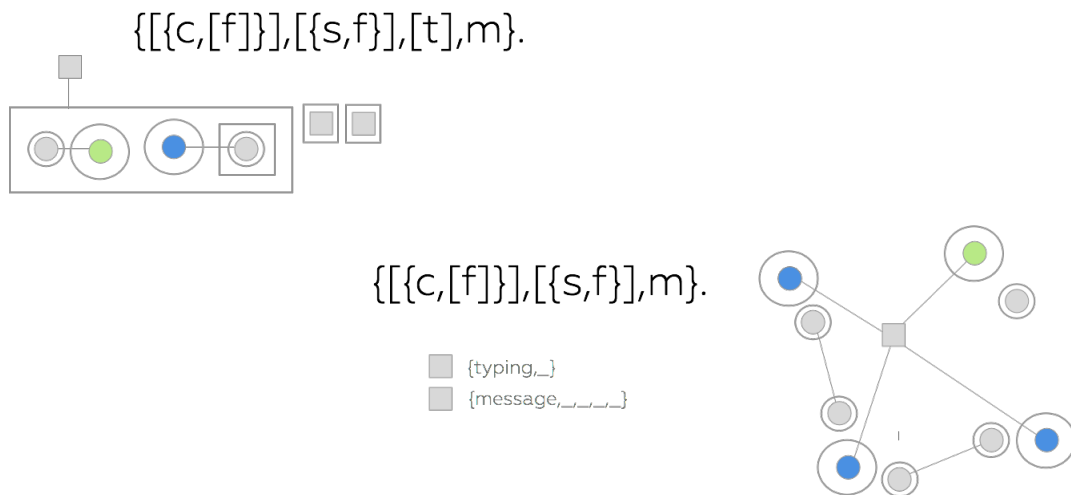
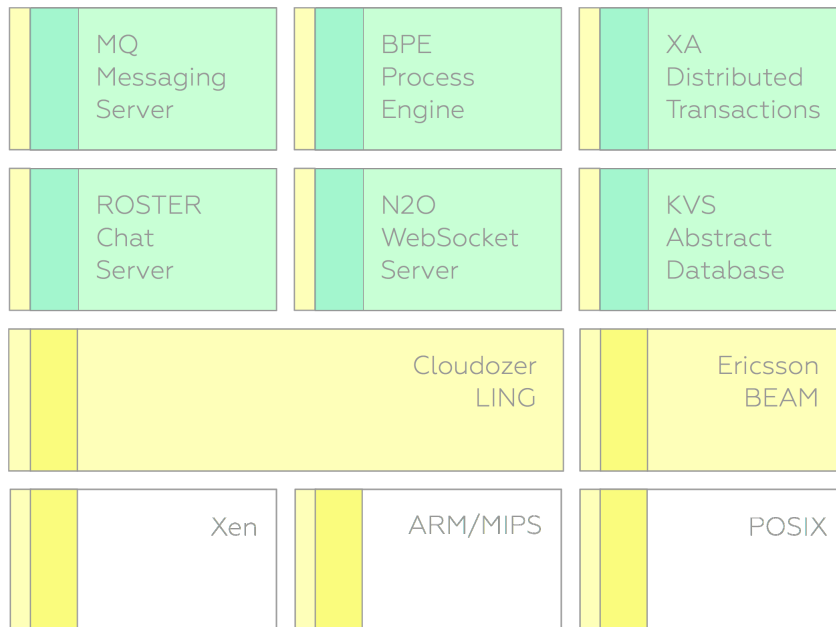


Рис. 5: Списки кімнат та користувачів і їхні чати, дві таблиці та протокол

1.8 Завдання дослідження

Тактична мета даної роботи – це реалізація усіх верхніх компонентів OSI на базі метамоделі Eke на мові Erlang. Побудова зручної сучасної гнучкої верифікованої теорії з компактною системою типів є основним завданням даного дослідження.



1.9 Мета дослідження

Основною метою дослідження є забезпечення виконання критеріїв відповідності показників ефективності, детермінованості та якості шляхом впровадження результату даної теорії, комплексу програмного забезпечення у виробничий процес.

За рахунок компактності передбачається значна економія складності та простота у експлуатації, що веде до зниження собівартості впровадження. Коректність передбачає заміну автоматизованому тестуванню, а закритість середовища дозволяє гнучко використовувати середовище на різних виробничих платформах.

1.10 Результати

Одна з останніх систем, яка була впроваджена для заміни існуючої системи, мала показники у співвідношенні розміру 10, швидкості 20 та виконання 3. Інші системи у історичній послідовності мали ці показники такими 10, 2, 8, та 19, 5, 5. В середньому показники ефективності відрізняються більше ніж на пів порядку в кращу сторону після впровадження продуктів на базі розробленої моделі. Існують продукти також і на інших мовах програмування.

Список литературы

- [1] Per Martin-Löf *Intuitionistic Type Theory* 1984
- [2] J.Armstrong. *Making reliable distributed systems in the presence of software errors* 2003.
- [3] Robin Milner. *A Calculus of Communicating Systems*. 1986.
- [4] William Lawvere. *Conceptual Mathematics*. 1997.
- [5] Benjamin Pierce. *Basic category theory for computer scientist*. 2004.
- [6] Сандерс Мак Лейн. *Категории для работающего математика*. 2004.
- [7] Leslie Lamport. *Specifying Systems* 2004.
- [8] Michael Barr and Charles Wells. *Toposes, Triples and Theories*. 2000.
- [9] Michael Barr and Charles Wells. *Category Theory for Computing Science*. 1995.
- [10] И.Бакур, А.Деляну. *Введение в теорию категори и функторов* 1972.
- [11] Robin Milner. *Communicating and Mobile Systems: The π -calculus*. 1999.
- [12] Robin Milner. *The Polyadic π -Calculus: A Tutorial*. 1993.
- [13] Коваленко. *Теория Массового Обслуживания*. 1965.
- [14] Meseguer, Montanari. *Petri Nets Are Monoids* 1990.
- [15] Philip Wadler *Call-by-Value is Dual to Call-by-Name* 1000
- [16] D.Mostrous, V.Vasconcelos *Session Typing for a Featherweight Erlang* 1990.
- [17] S.Marlow, P.Wadler *A practical subtyping system for Erlang* 1997
- [18] A.Lindgren *A Prototype of a Soft Type System for Erlang* 1996

З повагою, команда Synrc.

Андрій Задорожній, генеральний директор
Сохацький Максим, технічний директор
Кирилов Володимир, співзасновник