КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 004.8, 004.4'2, 004.738.5 DOI:10.34229/2707-451X.25.3.8

А.О. ДОВЖЕНКО, В.С. ЯРЕМЕНКО

ПІДХІД ДО ПОБУДОВИ СИСТЕМИ ДОВІРИ В МУЛЬТИАГЕНТНІЙ СИСТЕМІ НА ОСНОВІ ТЕХНОЛОГІЇ БЛОКЧЕЙН

У статті представлено систему довіри TrustLedger для мультиагентних розподілених середовищ, розроблену на основі технології блокчейн. Вона поєднує смартконтракти, токени стандарту ERC-20 та голосування, зважене токенами, для забезпечення прозорої, безпечної та децентралізованої взаємодії між агентами. TrustLedger дозволяє виявляти та ізолювати недобросовісних агентів, мінімізуючи їх вплив на колективні рішення. Проведене моделювання показало ефективність системи в умовах часткової недовіри, втрат зв'язку та зловмисної поведінки. Запропонований підхід демонструє практичну реалізацію механізму "довіри за дизайном" без потреби у централізованому контролі.

Ключові слова: *мультиагентні системи, розподілені системи, TrustLedger, блокчейн, система довіри, смартконтракти, ERC-20, децентралізоване управління, репутація агентів, голосування.*

Вступ. Сучасні розподілені системи, такі як розумні енергетичні мережі, мережі Інтернету речей (IoT) та системи спільної робототехніки, характеризуються зростаючою складністю та автономністю їхніх компонентів – агентів. У таких середовищах, де агенти взаємодіють автономно та можуть не мати вродженої довіри один до одного, забезпечення надійних механізмів довіри стає невід'ємним, але водночас складним завданням.

Ця робота представляє систему довіри, розроблену на основі технології блокчейн, яка пропонує децентралізовану, прозору та захищену основу для надійної взаємодії між агентами, навіть в умовах часткової недовіри та відсутності централізованого контролю – TrustLedger.

Мета даної роботи – дослідження полягала у розробці та емпіричному вивченні цієї системи для виявлення та обмеження впливу недобросовісних агентів, а також для підтримки справедливої та надійної взаємодії між учасниками.

TrustLedger використовує смарт-контракти, токени стандарту ERC-20, як описано у [1], та механізм голосування, зваженого токенами, який досліджено у [2], для забезпечення своєї функціональності. Успішна симуляція системи в сценарії управління розподілом електроенергії продемонструвала її ефективність у ізоляції недобросовісних агентів, підтверджуючи життєздатність запропонованого підходу.

Проблема полягає у тому, що традиційні централізовані підходи до управління довірою, які покладаються на єдиний контролюючий орган, мають значні недоліки. Вони вразливі до зловживань та мають обмежену масштабованість, що робить їх непрактичними для динамічних, великомасштабних мультиагентних середовищ. Ці традиційні моделі страждають від "єдиної точки відмови", де компрометація центрального органу може зруйнувати всю систему довіри. Крім того, їхня здатність до масштабування обмежена, оскільки зростання кількості агентів та обсягу взаємодій призводить до експоненційного зростання навантаження

на центральний сервер, що призводить до зниження продуктивності та надійності. Це фундаментальні архітектурні недоліки, які перешкоджають досягненню бажаного рівня стійкості та зростання в сучасних розподілених системах. Таким чином, існує нагальна потреба у розробці нових парадигм управління довірою, які б долали ці обмеження, забезпечуючи децентралізовану, стійку та масштабовану основу для взаємодії агентів.

Роль технології блокчейн. Як зазначено у [3], децентралізація, прозорість, незмінність та криптографічна безпека безпосередньо вирішують недоліки централізованих моделей. Децентралізація усуває єдині точки відмови, розподіляючи контроль та зберігання даних між усіма учасниками мережі. Прозорість гарантує, що всі транзакції та рішення є загальнодоступними та перевіряються, мінімізуючи можливості для маніпуляцій. Незмінність забезпечує, що дані, записані в блокчейн, не можуть бути змінені або видалені, створюючи надійний та аудитований реєстр. Криптографічна безпека захищає цілісність даних та автентичність учасників [3].

Ці властивості дозволяють створити надійне та верифіковане середовище для взаємодії агентів, навіть у сценаріях, що характеризуються частковою недовірою або відсутністю центрального органу. Блокчейн надає безпечний та аудитований реєстр для запису всіх транзакцій та рішень, пов'язаних з довірою. Це фундаментально переосмислює концепцію довіри: замість покладання на вразливих людських посередників або єдину довірену сутність, довіра вбудовується безпосередньо в архітектуру системи через верифікований код та розподілений консенсус. Ця форма "довіри за дизайном" усуває потребу в традиційних довірених третіх сторонах, що призводить до підвищення безпеки, прозорості та ефективності. Таке архітектурне рішення є центральним для цінності системи TrustLedger, роблячи її більш стійкою та надійною для взаємодії в мультиагентних системах, ніж традиційні централізовані підходи.

Концепція та ключові характеристики консенсусу в мультиагентних системах. У широкому сенсі, консенсус у мультиагентних системах (МАС) визначається як фундаментальний процес, за допомогою якого група автономних агентів, взаємодіючи переважно на локальному рівні, сходиться до спільної згоди щодо певного значення, рішення або синхронізує траєкторії своїх станів. Ця згода може проявлятися у формі спільного рішення, поділюваного значення або скоординованої поведінки. Основна ідея полягає у тому, щоб усі агенти дійшли згоди щодо певних величин, обмінюючись інформацією зі своїми безпосередніми сусідами.

Ключові характеристики, що визначають цей процес.

Згода. Усі агенти мають з часом зійтися до одного й того ж значення або стану, при цьому різниця між їхніми значеннями має зменшуватися.

Локальні взаємодії. Агенти приймають рішення та оновлюють свої стани на основі інформації, отриманої лише від обмеженої кількості сусідніх агентів, не маючи глобальної інформації про всю систему.

Розподілені алгоритми. Консенсус досягається за допомогою розподілених алгоритмів або протоколів, де кожен агент дотримується набору правил, заснованих на його локальних взаємодіях, без центрального органу, що диктує остаточну угоду.

Граф взаємодії (топологія). Відносини комунікації та взаємодії між агентами часто представляються графом, де вузли представляють агентів, а ребра – потік інформації. Структура цього графа відіграє вирішальну роль у визначенні того, чи може бути досягнуто консенсусу та яким чином.

Синхронізація. У ширшому сенсі, консенсус може також відноситися до повної або часткової синхронізації траєкторій станів агентів.

Фундаментальна проблема. Встановлення консенсусу вважається фундаментальною задачею у дослідженні мультиагентних систем, що допомагає зрозуміти основні принципи координації та вплив топології системи.

Асимптотична згода. Консенсус зазвичай досягається з часом, оскільки агенти ітеративно оновлюють свої значення на основі інформації, отриманої від сусідів.

Спільне значення. Кінцева мета – це досягнення згоди між усіма агентами щодо одного й того ж значення, яке може бути пов'язане з початковим станом, середнім значенням початкових спостережень або параметром у розподіленій задачі оптимізації.

Виклики у досягненні консенсусу. Досягнення консенсусу в мультиагентних системах стикається з низкою значних викликів, які можуть перешкоджати процесу узгодження:

– комунікаційні затримки: час, необхідний для обміну інформацією між агентами;

– втрата пакетів: ненадійні канали зв'язку можуть призвести до втрати інформації;

– несправні агенти (включаючи візантійських): агенти з несправностями або зловмисні агенти можуть надсилати некоректну інформацію або відхилятися від протоколів, активно перешкоджаючи досягненню консенсусу;

– динамічна топологія мережі: зміни в каналах зв'язку через мобільність агентів або збої можуть ускладнити консенсус;

– гетерогенні агенти: системи, що складаються з агентів з різною динамікою, можливостями або цілями, створюють додаткові труднощі;

– масштабованість: підтримка консенсусу у великомасштабних системах з великою кількістю агентів може бути складною через зростаючі накладні витрати на комунікацію та обчислення;

– невизначеність середовища: пепередбачувані зміни в навколишньому середовищі можуть впливати на поведінку агентів та комунікаційні патерни;

– конфліктуючі цілі: у деяких МАС агенти можуть мати індивідуальні цілі, які суперечать загальній меті консенсусу, що ускладнює досягнення згоди.

Ці виклики не є ізольованими; вони часто взаємодіють, створюючи кумулятивно-негативний ефект на складність та надійність системи. Наприклад, досягнення консенсусу з візантійськими агентами в динамічній мережевій топології є значно складнішим завданням, ніж вирішення будь-якої з цих проблем окремо. Це означає, що будь-яка практична та надійна система довіри, така як TrustLedger, має бути розроблена для неявного або явного вирішення цих ускладнюючих факторів. Здатність системи функціонувати, коли агенти є зловмисними, а комунікаційні зв'язки ненадійними, є критично важливою, що підкреслює необхідність надійних рішень, здатних витримувати комбінації несприятливих факторів.

Існуючі практичні підходи та технології для консенсусу. У мультиагентних системах існує кілька практичних підходів до досягнення консенсусу, кожен з яких має свої особливості та застосовується залежно від конкретних вимог системи та її архітектури.

Один із фундаментальних підходів це використання децентралізованих методів прийняття рішень. У роботі [4] наведено теоретичні основи алгоритмів консенсусу для мультиагентних систем з акцентом на стійкість до змін топології, затримки та гарантії швидкості зближення. У таких системах кожен агент приймає рішення самостійно, базуючись на інформації зі свого оточення та від інших агентів, із якими він взаємодіє, що забезпечує гнучкість, масштабованість і стійкість до відмов.

Інший поширений метод це застосування **механізмів голосування та агрегації думок**. У цьому випадку агенти висловлюють свої переваги або думки щодо певного рішення чи значення, а потім ці думки агрегуються для визначення остаточного консенсусу. Агрегація може здійснюватися за допомогою простого голосування за більшістю, де рішення, яке отримало більше половини голосів, вважається прийнятим, або за допомогою більш складних схем, таких як зважене голосування, де голос кожного агента має певну вагу, що може залежати, наприклад, від його репутації або рівня довіри в системі. Ефективність механізмів голосування безпосередньо залежить від чесності агентів та їхньої здатності приймати обґрунтовані рішення на основі наявної інформації.

Важливу роль у досягненні консенсусу відіграє використання репутаційних систем та рівнів довіри між агентами. Один із класичних підходів – алгоритм EigenTrust, запропонований у [5]. У цьому алгоритмі загальні значення довіри обчислюються як власний вектор матриці нормалізованих локальних довір, що дозволяє агентам формувати уявлення про надійність інших на основі попередньої поведінки.

TrustLedger: Система довіри на основі блокчейну. Проєктування систем довіри на основі блокчейну, що лежить в основі TrustLedger, спирається на різні архітектурні моделі. Одна з них – абстрактна багаторівнева архітектура, яка зазвичай включає щонайменше чотири рівні: рівень застосунків, рівень блокчейну, рівень прийняття рішень та рівень ресурсів. Рівень прийняття рішень, зокрема, відіграє вирішальну роль у реалізації логіки оцінки довіри.

Блокчейн також може бути використаний для встановлення довіри між мікросервісами в архітектурі мікросервісів (MSA), особливо у відкритих системах. У таких випадках блокчейн-моделі довіри допомагають вирішити проблеми, пов'язані з довірою між незалежними сервісами. Концепція безперервної оцінки довіри під час виконання та динамічного присвоєння балів довіри, як це пропонується для мікросервісів, є дуже актуальною для мультиагентних систем, де поведінка агентів може змінюватися з часом.

Вибір між публічним (бездозвільним) та приватним (дозвільним/консорціумним) блокчейном – це важливе архітектурне рішення. Приватні блокчейни надають більший контроль над учасниками та можуть бути кращим варіантом для сценаріїв, де потрібні відомі та надійні суб'єкти. Публічні блокчейни забезпечують більшу децентралізацію та прозорість, але можуть вимагати більш надійних механізмів для боротьби з невідомими або потенційно шкідливими агентами.

Смарт-контракти: фундамент довіри та автоматизації. Як описано у [6], смарт-контракти – це самовиконувані цифрові угоди на блокчейні, які автоматично забезпечують умови договору безпосередньо в коді, роблячи їх детермінованими й незмінними після розгортання.

Ключові елементи смартконтракту включають:

- середовище для укладення угоди;
- сторони смартконтракту;
- активи;
- умови;
- розподіленість;
- детермінованість;
- незмінність;
- прозорість;
- самоверифікаційні та самозастосовні;
- автоматизація.

Використання смартконтрактів у системах довіри надає значні переваги: автоматизацію, прозорість, безпеку, зниження витрат, підвищення довіри та запобігання шахрайству та помилкам. Ці переваги демонструють фундаментальне перепроєктування довіри: від довіри до традиційних посередників до довіри, вбудованої у код. Це означає, що традиційна роль довіреної третьої сторони замінюється автоматизованим, верифікованим виконанням коду. Довіра тепер покладається на математичну визначеність виконання коду та криптографічну цілісність блокчейну, а не на схильну до помилок людську установу. Це надійна форма "довіри за дизайном", вбудована у саму архітектуру системи.

Смартконтракти TrustLedger формалізують правила взаємодії та розподілу ресурсів між агентами, забезпечуючи незмінність угод та виключаючи можливості маніпуляцій даними. Агенти можуть ідентифікуватися через пари публічних/приватних ключів, що вбудовує криптографічну атестацію їх особи у систему довіри.

Функціональність смартконтракту TrustLedger: смартконтракт TrustLedger реалізує набір ключових операцій для підтримки системи довіри між агентами, забезпечуючи прозорі, автоматизовані та безпечні взаємодії без централізованого контролю.

Реєстрація агентів: кожен агент (учасник мережі) реєструється у системі через внесення свого ідентифікатора (адреси) та, за потреби, початкового стейку токенів. Ідентифікація базується на публічних ключах, що дозволяє надійно прив'язати особу агента до записів у блокчейні. Після реєстрації агент отримує можливість пропонувати та голосувати за пропозиціями.

Створення позицій (ролей/завдань): агенти або зовнішні системи можуть додавати у TrustLedger нові "позиції" – наприклад, опис вакансій чи завдань, які потрібно виконати. Такі позиції фіксуються на блокчейні і можуть слугувати підґрунтям для голосування: наприклад, визначення відповідального агента чи розподілу ресурсів.

Створення та голосування за пропозиціями: агенти формулюють пропозиції (наприклад, щодо призначення конкретного агента на позицію чи зміни політики системи). Пропозиції зберігаються у контракті і відкриваються для голосування серед зареєстрованих агентів. Кожен агент може віддати свій голос «за» або «проти» пропозиції, причому вага голосу визначається кількістю заблокованих токенів (стейком) агента. Головна мета – забезпечити демократичне ухвалення рішень, де рішення приймаються за підтримки абсолютної більшості голосів із урахуванням їх ваги.

Механізм стейкінгу токенів: для участі в голосуваннях агенти передають токени у депозит, блокуючи їх у смарт-контракті. Як доводять автори у [7], такий підхід гарантує «шкіру в грі»: більше токенів у стейку – вища вага голосу. Наявність реальної економічної зацікавленості зміцнює безпеку та стійкість голосувань.

Евакуація (видалення) агентів: за умов невиконання зобов'язань або зловмисних дій смартконтракт може вивести агента з системи. Це включає анулювання його прав участі та можливе стягнення частини заблокованих токенів як штрафу (аналог процедури слешингу в PoS-системах). Такий механізм відсторонює ненадійних агентів, захищаючи довіру решти спільноти.

Розподіл винагород і штрафів: TrustLedger автоматично розподіляє винагороди та накладає штрафи відповідно до поведінки агентів. Наприклад, за успішно виконані завдання чи участь у голосуванні добросовісні агенти можуть отримати додаткові токени, натомість невиконані або фальсифіковані результати тягнуть за собою втрату частини стейку. Всі такі транзакції відбуваються прозоро у мережі блокчейн, а правила їх нарахування зафіксовані в коді контракту. Це заохочує бажану поведінку та створює економічний стимул для утримання довіри й активної участі.

Діаграма на рис. 1 показує ключові взаємозв'язки у системі TrustLedger між основними концептами управління довірою.

Репутація виступає центральним елементом, який формується на основі Довіри та впливає на неї у зворотному напрямку. Взаємовигідність – проміжна ланка, що з'єднує довіру з практичною Користю для агентів системи.

У контексті описаного смарт-контракту TrustLedger це означає, що довіра формується через прозорі механізми голосування, стейкінгу токенів та криптографічну ідентифікацію агентів. Репутація накопичується завдяки добросовісній поведінці агентів (успішне виконання завдань, участь у голосуваннях). Взаємовигідність досягається через систему винагород та штрафів, що заохочує кооперативну поведінку. Користь реалізується у вигляді токенових винагород, доступу до нових можливостей та підвищення статусу в мережі. Діаграма на цьому ж рисунку демонструє циклічну природу довіри у децентралізованій системі, де кожен елемент підсилює інші, створюючи стійку екосистему взаємодії без централізованого контролю.

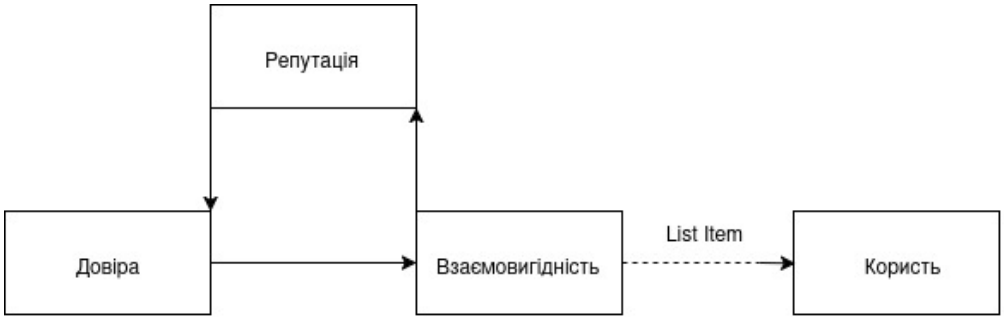

РИС. 1. Відношення репутації до довіри

Ця складна інтеграція стейкінгу, голосування, винагород та штрафів створює потужну економічну модель стимулювання. Така конструкція гарантує, що агенти фінансово мотивовані діяти чесно та позитивно впливати на систему, оскільки їхній економічний внесок (застейкані токени) безпосередньо пов'язаний з їхньою поведінкою. Цей механізм виходить за рамки простого технічного примусу, створюючи саморегульовану систему, керовану раціональним економічним інтересом, що сприяє більш надійному та довірчому середовищу.

Порівняння основних функціональних можливостей смартконтракту наведено в табл 1.

ТАБЛИЦЯ 1. Основні функціональні можливості смартконтракту TrustLedger

Функціональність	**Опис**	**Ключовий механізм/Перевага**
Реєстрація агентів	Внесення ідентифікатора та початкового стейку для участі в системі	Ідентифікація за публічним ключем, надання прав участі
Створення позицій (ролей/завдань)	Додавання описів завдань або ролей, що потребують виконання	Фіксація на блокчейні, основа для голосування та розподілу ресурсів
Створення та голосування за пропозиціями	Формулювання та голосування за рішеннями (наприклад, призначення агента)	Вага голосу залежить від стейку токенів, демократичне ухвалення рішень
Механізм стейкінгу токенів	Депозит токенів для участі в голосуванні	Економічна зацікавленість ("шкіра в грі"), підвищення ваги голосу, мотивація чесної участі
Евакуація (видалення) агентів	Виведення агента з системи за невиконання зобов'язань або зловмисні дії	Анулювання прав участі, можливе стягнення штрафу (слешинг)
Розподіл винагород і штрафів	Автоматичне нарахування винагород та накладення штрафів	Заохочення бажаної поведінки, економічний стимул для підтримки довіри та активної участі

Токени ERC-20 та голосування, зважене токенами. Для реалізації механізмів голосування та репутації у TrustLedger застосовується стандарт ERC-20 через допоміжний контракт Vote. ERC-20 є уніфікованим інтерфейсом для створення взаємозамінних токенів на платформі Ethereum, що забезпечує стандартизацію та сумісність з різними додатками та обліковими записами. Контракт Vote відповідає за емісію, зберігання та передачу цих голосувальних токенів, при цьому кожна операція прозоро записується у блокчейн.

Переваги використання ERC-20 такі:

– **прозора емісія:** початкова кількість токенів та всі подальші зміни (винагороди чи штрафи) видно всім у блокчейні;

– **стандартизована передача:** будь-яка операція голосування чи стейкінгу це проста транзакція ERC-20, яку легко відстежувати у загальнодоступному реєстрі;

– **масштабованість:** ERC-20 добре підтримується екосистемою інструментів Ethereum (наприклад, бібліотеками OpenZeppelin), що спрощує розгортання та аудит коду контракту.

TrustLedger реалізує модель **token-weighted voting** (голосування, зважене токенами), де вага голосу агента прямо пропорційна кількості токенів, які він застейкав. Цей підхід ґрунтується на принципі "шкіри в грі": агенти, які інвестували значну кількість токенів у систему, вважаються більш зацікавленими в її успішній роботі, тому їм надається більша вага при голосуванні. Це сприяє тому, що власники великих стейків мають економічний стимул давати якісні голоси та протидіяти фальсифікаціям, оскільки їхні кошти перебувають під загрозою.

Мотивуюча роль токенів реалізована через систему винагород і штрафів. По-перше, сам факт стейкінгу може оподатковуватися бонусом: агенти отримують додаткові токени за участь у голосуваннях. По-друге, успішне виконання умов (наприклад, набрання більшості голосів на пропозиції) стимулюється винагородою токенами, тоді як небажана поведінка (наприклад, голосування з фальшивого акаунта або бездіяльність) карається втратами стейку. Ця багаторівнева система стимулів робить мережу безпечнішою та демократичнішою, оскільки всі зацікавлені в прийнятті рішень різними групами учасників.

Стратегічний вибір стандарту ERC-20, який є широко прийнятим та перевіреним стандартом токенів у екосистемі Ethereum, це важливе дизайнерське рішення. Це означає, що TrustLedger отримує переваги від притаманної безпеки, можливості аудиту та інтероперабельності, що надаються зрілим стандартом. Такий вибір значно знижує ризики розробки та підвищує потенціал системи для інтеграції з існуючою блокчейн-інфраструктурою, що безпосередньо сприяє її надійності та майбутній життєздатності.

Результати: динаміка та кінцевий розподіл довіри. На графіках (рис. 2) та у табл. 2 показано результат симуляції мультиагентної системи, в якій беруть участь 7 агентів: 5 чесних та 2 нечесних.

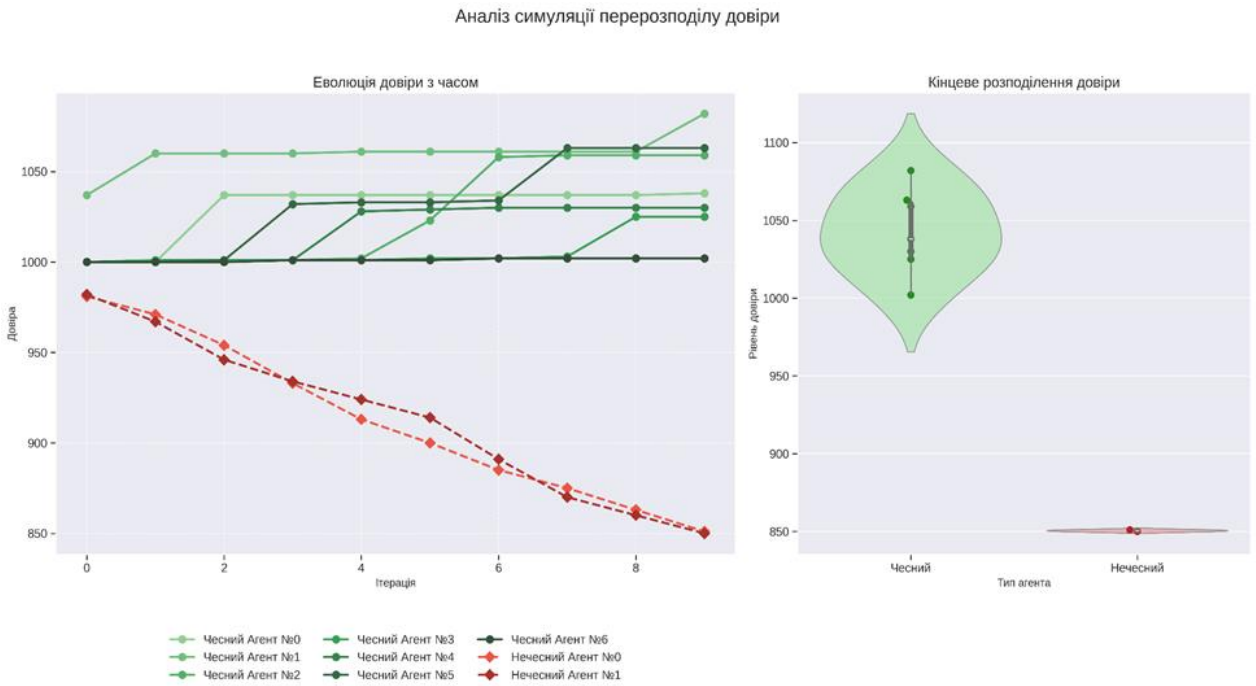

РИС. 2. Еволюція довіри чесних та нечесних агентів з часом

ТАБЛИЦЯ 2. Рівні довіри впродовж симуляції

	Ітерація 1	Ітерація 2	Ітерація 3	Ітерація 4	Ітерація 5	Ітерація 6	Ітерація 7	Ітерація 8	Ітерація 9	Ітерація 10
Чесний 0	1000	1000	1037	1037	1037	1037	1037	1037	1037	1038
Чесний 1	1037	1060	1060	1060	1061	1061	1061	1061	1061	1062
Чесний 2	1000	1001	1001	1001	1002	1023	1058	1059	1059	1059
Чесний 3	1000	1001	1001	1001	1001	1002	1002	1030	1025	1025
Чесний 4	1000	1000	1000	1001	1028	1029	1030	1030	1030	1030
Чесний 5	1000	1000	1001	1032	1033	1033	1034	1034	1063	1063
Чесний 6	1000	1000	1000	1001	1001	1001	1002	1002	1002	1002
Нечесний 0	981	971	954	933	913	900	885	875	860	851
Нечесний 1	982	967	946	934	924	914	891	870	860	850

Можна побачити, що добросовісні агенти зберігають відносно високий рівень довіри протягом усіх раундів: їхні криві залишаються стабільними на високих значеннях. Натомість, шкідливі агенти демонструють поступове та значне зменшення голосувальної сили з кожним раундом. Після кількох раундів їхні показники майже падають до мінімуму, що свідчить про відчутне покарання системою тих, хто регулярно голосує проти більшості. Таким чином, з часом шкідливі агенти практично втрачають можливість віддавати вагомі голоси. Ця динаміка узгоджується з результатами, отриманими в подібних довірчих моделях, які показують, що правильно спроектовані довірчі механізми дозволяють ідентифікувати та виключати зловмисників.

Висновки. У цій роботі було розглянуто та реалізовано систему довіри TrustLedger для мультиагентних середовищ на основі технології блокчейн. Основні досягнення включають розробку системи, яка ефективно виявляє та ізолює недобросовісних агентів, мінімізуючи їхній вплив на колективні рішення. Проведена симуляція у сценарії управління розподілом електроенергії підтвердила ефективність запропонованого механізму довіри, де недобросовісні агенти поступово втрачали свою голосувальну силу, а їхній вплив знижувався практично до нуля. Ці спостереження узгоджуються з висновками сучасних досліджень, які підтверджують, що довіра є ключовим фактором у мультиагентних системах, а впровадження довірчих механізмів дозволяє знецінити неправомірні дії агентів.

Авторські внески: Довженко А.О. – дослідження, програмна реалізація, написання – оригінальна, чернетка; Яременко В.С. – формулювання ідеї, методологія, формальний аналіз, написання – рецензування та редагування.

Наявність та доступність даних для повторення та перевірки результатів розрахунків: зовнішні дані не використовувались, всі описані агенти є програмною симуляцією.

Джерела фінансувань: без зовнішнього фінансування.

Список літератури

1. Buterin V. ERC-20 Token Standard. Ethereum Foundation. 2014. https://eips.ethereum.org/EIPS/eip-20
2. Leonardos S., Reijsbergen D., Piliouras G. Weighted Voting on the Blockchain: Improving Consensus in Proof of Stake Protocols. 2019. *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. Seoul, Korea (South), 2019. P. 376–384. https://doi.org/10.1109/BLOC.2019.8751290
3. Zheng Z., Xie S., Dai H., Chen X., Wang H. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. *In 2017 IEEE International Congress on Big Data (BigData Congress)*, 2018. P. 557–564. https://doi.org/10.1109/BigDataCongress.2017.85
4. Olfati-Saber R., Fax J.A., Murray R.M. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 2007. **95** (1). P. 215–233. https://doi.org/10.1109/JPROC.2006.887293

5. Kamvar S.D., Schlosser M.T., Garcia-Molina H. The EigenTrust Algorithm for Reputation Management in P2P Networks. *In Proceedings of the 12th International Conference on World Wide Web*. 2003. P. 640–651. https://doi.org/10.1145/775152.775242
6. Grishchenko I., Maffei M., Schneidewind C. A Semantic Framework for the Security Analysis of Ethereum Smart Contracts. *In Proceedings of the 30th IEEE Computer Security Foundations Symposium (CSF 2018)*, 2018. P. 33–47. https://doi.org/10.1109/CSF.2018.00010
7. Kiayias A., Russell A., David B., Oliynykov R. Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol. *In Annual International Cryptology Conference (CRYPTO 2017)*. Lecture Notes in Computer Science, Springer, 2017. Vol. 10401. P. 357–388. https://doi.org/10.1007/978-3-319-63688-7_12

Одержано 02.06.2025

Довженко Андрій Олегович,
студент кафедри системного проєктування навчально-наукового інституту прикладного системного аналізу
Національного технічного університету України «КПІ ім. Ігоря Сікорського»,
https://orcid.org/0009-0009-1648-7283

Яременко Вадим Сергійович,
доктор філософії у галузі комп'ютерних наук, старший викладач кафедри системного проєктування навчально-наукового інституту прикладного системного аналізу
Національного технічного університету України «КПІ ім. Ігоря Сікорського».
https://orcid.org/0000-0001-8557-6938

UDC 004.8, 004.4'2, 004.738.5

Andrii Dovzhenko, Vadym Yaremenko *

A Blockchain-Based Approach to Trust System Design in Multi-Agent Environments

National technical university of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv
* *Correspondence: yaremenko.v.s@gmail.com*

Introduction. This paper presents a decentralized approach to building a trust system for distributed multi-agent environments using blockchain technology. The developed system, TrustLedger, is based on Ethereum smart contracts, ERC-20 tokens, and token-weighted voting mechanisms. This approach enables transparent, secure, and scalable interactions between agents, even in scenarios with partial distrust and without centralized authority. TrustLedger includes mechanisms for agent registration, role and task creation, proposal submission, voting, token staking, and the automatic distribution of rewards and penalties. Agent reputation is built through honest participation in the system, while the influence of dishonest participants decreases as their voting power is gradually reduced. A simulation of an energy distribution scenario demonstrated the system's ability to effectively identify and isolate malicious agents while maintaining high levels of internal trust. In contrast to centralized trust models, which are vulnerable to failures and abuse, the proposed solution provides resilience to network dynamics, latency, communication disruptions, and adversarial behavior. The advantages of smart contracts – automation, transparency, and immutability – enable "trust by design," reducing dependence on human intervention or centralized intermediaries. The described system is a promising tool for use in smart grids, IoT environments, and collaborative robotics, where autonomous agents must interact without prior trust.

The purpose of the paper. To develop and evaluate the decentralized trust system TrustLedger for multi-agent distributed environments using blockchain technology.

Results. A prototype of the TrustLedger system was implemented, demonstrating the ability to isolate dishonest agents and maintain consensus stability. Simulation results showed a consistent reduction in the influence of malicious agents over several rounds of interaction. Agents that regularly acted against collective interest lost their reputation and voting weight, confirming the effectiveness of the incentive and penalty mechanisms.

Conclusions. The TrustLedger system has proven to be an effective decentralized solution for managing trust in multi-agent systems. By utilizing smart contracts and token-weighted voting, it enables transparency, security, and self-regulation without the need for centralized control. This approach provides a resilient environment for agent interaction, even in dynamic and potentially adversarial settings.

Keywords: multi-agent systems, distributed systems, TrustLedger, blockchain, trust system, smart contracts, ERC-20, decentralized governance, agent reputation, voting.