УДК 004.852; 004.942

П. Кравець
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІГРОВЕ МОДЕЛЮВАННЯ МУЛЬТИАГЕНТНИХ СИСТЕМ

Досліджується проблема колективного вибору варіантів рішень у мультиагентних системах на основі стохастичних ігрових моделей. Виконано формулювання ігрової задачі в умовах невизначеності. Побудовано адаптивні рекурентні методи розв’язування ігрової задачі. Визначено умови збіжності ігрових методів до колективних станів рівноваги. Розроблено алгоритмічні та програмні засоби для розв’язування ігрової задачі.

The problem of collective alternate solutions in multiagent systems on a basis of the stochastic game models is investigated. The game problem formulation in uncertainty conditions is executed. The adaptive recurrence methods of the game solving are constructed. The convergence conditions of game methods to collective equilibrium states is defined. The algorithm and software tools for the game task solving are developed.

Вступ

Для узагальненого дослідження систем різної природи широко використовуються моделі чорної (непрозорої) або сірої (напівпрозорої) скриньки. Функціонально такі моделі визначають залежності виходів від входів системи, а структурно – описують набір інтерфейсів, через які відбувається взаємодія з навколишнім середовищем та іншими системами. Мета взаємодії визначається інформаційним агентом [1] – автономною, активною інформаційною системою з елементами штучного інтелекту. У дискретному формулюванні метою взаємодії, як правило, є вибір варіантів розв’язків для забезпечення оптимальних режимів функціонування системи. Множинність інтерфейсів системи надає можливість мультиагентного розв’язування оптимізаційної задачі [2]. У мультиагентній системі можна досягти зменшення загального часу розв’язування задачі за рахунок одночасного виконання робіт.

Сучасні теоретичні дослідження мультиагентних систем переважно спрямовані на вивчення механізмів координації, адаптації, навчання, комунікації, командної роботи агентів в умовах невизначеності [1, 2]. Зручним інструментом дослідження цих механізмів є стохастичні ігрові моделі [3]. На відміну від детермінованих матричних ігор, стохастична гра є повторюваною у часі для адаптивного формування оптимальної послідовності стратегій. Асимптотичні розв’язки стохастичної гри задовольняють одну із умов колективної оптимальності [4, 5] – за Нешем, Слейтером, Парето, Джофріоном, Байесом або інші.

Зростання розмірності гри, яка визначається декартовим добутком кількості агентів, чистих стратегій та станів системи, значно ускладнює можливості аналітичного розв’язування ігрової задачі. Тому актуальним, з теоретичного та практичного поглядів, є розроблення комп’ютерних імітаційних моделей стохастичних ігор для прийняття оптимальних рішень у мультиагентних системах.

При побудові ігрових моделей важливим є планування належної взаємодії між агентами для формування оптимізаційної стратегії. Крім цього, ефективність ігрової оптимізації значною мірою залежить від надійності роботи агентів. Ці питання є недостатньо висвітленими у науковій літературі.

Метою роботи є планування та реалізація комп’ютерного моделювання стохастичної гри в умовах невизначеності з врахуванням взаємодії та відмов агентів.

Постановка ігрової задачі

Стохастична гра задається кортежем $\left(S, U^i, \Xi^i \mid \forall i \in D\right)$, де S – множина станів середовища, U^i – вектор чистих стратегій i-го гравця, Ξ^i – функція виграшів i-го гравця, D – непорожня множина гравців. Для спрощення приймемо, що середовище перебуває в одному стані, тобто $|S|=1$.

Кожен гравець $i \in D$ здійснює у дискретні моменти часу $n = 1,2,...$ незалежний вибір однієї з власних чистих стратегій $u_n^i = u^i \in U^i = (u^i(1), u^i(2), ..., u^i(N_i))$ і до моменту часу $n+1$ спостерігає випадковий поточний виграш $\xi_n^i = \Xi^i(u_n^{D_i})$, що є функцією спільних стратегій $u^{D_i} \in U^{D_i} = \underset{j \in D_i}{\otimes} U^j$ гравців з локальної підмножини $D_i \subseteq D, D_i \neq \varnothing \, \forall i \in D$. Нехай послідовності випадкових величин $\{\xi_n^i\}$ незалежні $\forall u_n^{D_i} \in U^{D_i}$, $\forall i \in D$, $\forall n = 1,2,...$, а їхні математичні сподівання $M\{\xi_n^i(u^{D_i})\} = v^i(u^{D_i}) = const$ апріорі не відомі та мають обмежений другий момент $\sup\limits_n M\{[\xi_n^i(u^{D_i})]^2\} = \sigma_i^2(u^{D_i}) < \infty$. Матриці математичних сподівань виграшів $\left[v^i(u^{D_i})\right]$ $\forall i \in D$ назвемо середовищем гри. Якщо $\forall u_n^{D_i} \in U^{D_i}$, $\forall i \in D$ $v(u^{D_i}) > 0$, то середовище є знакододатним, якщо $v(u^{D_i}) < 0$ – знаковід'ємним, інакше – загального виду.

У грі без обміну інформацією гравці не повідомляють один одного про реалізовані стратегії та величину отриманого виграшу ξ_n^i.

У грі з обміном інформацією кожен *i*-й гравець повідомляє гравців з локальної підмножини D_i про значення поточного виграшу ξ_n^i. У результаті *i*-й гравець отримає інформацію від множини гравців $\tilde{D}_i$, виграші яких визначаються його стратегіями. Згортка поточних виграшів $\zeta_n^i = \sum\limits_{k \in \tilde{D}_i} \lambda^i(k) \xi_n^i$, де $\lambda^i = (\lambda^i(k) \mid \lambda^i(k) > 0 \;\; \forall k \in \tilde{D}_i, \sum\limits_{k \in \tilde{D}_i} \lambda^i(k) = 1)$, є оцінкою стратегії, вибраної *i*-м гравцем у момент часу *n*.

У грі з відмовами гравці характеризуються імовірностями відмов $\eta^i \in [0,1)$. Нехай $\psi^i \in \{0,1\}$ – ознака участі *i*-го гравця у грі. Якщо $\psi^i = 0$, то гравець відмовляється від поточного ходу гри з імовірністю η^i, якщо $\psi^i = 1$ – бере участь у грі з імовірністю $1 - \eta^i$. На кожному кроці гри відбувається обмін ознаками поточних станів ψ_n^i між сусідніми гравцями з множин D_i $\forall i \in D$. Відмови гравців призводять до зміни складу множин D_i $\forall i \in D$, які визначають величину поточних виграшів. Введемо повні групи подій $\Psi_i = 2^{|D_i|}$, пов'язаних із відмовами гравців з множин D_i. Нехай $D_i(\omega) \subseteq D_i$ – підмножина гравців, які залишаються у грі для події $\omega \in \Psi_i$. Тоді поточні виграші i-го гравця дорівнюють $\xi_n^i = \chi\{\sum\limits_{j \in D_i} \psi_n^j \geq \overline{\psi}\} \xi_n^i(u_n^{D_i(\omega)})$, де $\chi(\cdot) \in \{0,1\}$ – індикаторна функція події, $\overline{\psi} > 0$ – поріг гри, або мінімально допустима кількість гравців, що не відмовили.

З урахуванням відмов гравців послідовності обраних варіантів $\{u_t^{D_i(\omega)} \mid t = \overline{1,n}\}$ оцінюються поточними середніми виграшами

$$\Phi_n^i(\{u_n^{D_i(\omega)}\}) = \frac{1}{n}\sum_{t=1}^{n}\xi_t^i \quad \forall i \in D. \tag{1}$$

Метою кожного гравця є максимізація функції середніх виграшів

$$\underline{\lim}_{n\to\infty}\Phi_n^i(\{u_n^{D_i(\omega)}\}) \to \max \forall i \in D. \tag{2}$$

Розв'язки задачі векторної оптимізації (2) шукаються у множині точок рівноваги за Нешем (для гри без обміну інформацією)

$$\forall i \in D \quad \underline{\lim}_{n\to\infty}\left[\Phi_n^i(\{u_n^{D_i(\omega)}\}) - \Phi_n^i(\{\hat{u}_n^{D_i(\omega)}\})\right] \geq 0, \tag{3}$$

або оптимальності за Парето (для гри з обміном інформацією)

$$\begin{cases} \forall i \in D \quad \underline{\lim}_{n\to\infty}\left[\Phi_n^i(\{u_n^{D_i(\omega)}\}) - \Phi_n^i(\{\tilde{u}_n^{D_i(\omega)}\})\right] \geq 0, \\ \exists i \in D \quad \underline{\lim}_{n\to\infty}\left[\Phi_n^i(\{u_n^{D_i(\omega)}\}) - \Phi_n^i(\{\tilde{u}_n^{D_i(\omega)}\})\right] > 0, \end{cases} \tag{4}$$

де нерівності (3) та (4) виконуються з імовірністю 1, а $u_n^{D_i(\omega)}, \tilde{u}_n^{D_i(\omega)} \in U^{D_i(\omega)}$; $\hat{u}_n^{D_i(\omega)} = u_n^{D_i(\omega)} \setminus u_n^i + \tilde{u}_n^i \in U^{D_i(\omega)}$; $u_n^i, \tilde{u}_n^i \in U^i$.

Розв'язування ігрової задачі

Асимптотичних цілей (3) або (4) досягають за допомогою самонавчальних рекурентних методів формування векторів змішаних стратегій p_n^i, елементи яких є умовними імовірностями вибору відповідних чистих стратегій, тобто $p_n^i(u_n^i) = P\{u_n^i \mid u_t^i, \xi_t^i (t = \overline{1, n-1})\}$ $\forall u_n^i \in U^i, \forall i \in D$:

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i}\left\{p_n^i - \gamma_n R(x_n^i, p_n^i, \xi_n^i)\right\}, \tag{5}$$

де γ_n – крок методу; $R(x_n^i, p_n^i, \xi_n^i)$ – вектор руху методу; $\pi_{\varepsilon_{n+1}}^{N_i}$ – проектор на одиничний ε-симплекс [4], необхідний для нормалізації p_n^i та нагромадження повної статистичної інформації про випадкове середовище.

Вектор руху методу повинен забезпечувати виконання умови (2). Досягнення конкретного колективного розв'язку (рівноваги за Нешем, оптимальності за Парето тощо) визначається видом методу (5) та способом зміни його регульованих параметрів [4, 5].

Метод з потрібними властивостями, які визначаються умовами (3) або (4), будують на основі матричного формулювання асимптотично адекватної безкоаліційної ігрової задачі з функціями середніх виграшів гравців

$$V^i = \sum_{\omega\in\Psi_i}\prod_{j\in D_i(\omega)}\left\{(1-\eta^j)\chi(\psi^j = 1) + \eta^j\chi(\psi^j = 0)\right\}V^i(\omega), \tag{6}$$

де $V^i(\omega) = \sum_{u^{D_i(\omega)}\in U^{D_i(\omega)}} v^i(u^{D_i(\omega)}) \prod_{j\in D_i(\omega); u^j\in u^{D_i(\omega)}} p^j(u^j)$ – функція середніх виграшів, визначена для однієї із ситуацій гри з відмовами.

Парето-оптимальні розв'язки матричної гри визначають методом умовної максимізації системи локальних згорток функцій середніх виграшів $W^i = \sum_{k\in\tilde{D}_i}\lambda^i(k)V^k$ на опуклих одиничних симплексах S^{N_i} $\forall i \in D$. Для диференційованих на S^{N_i} функцій W^i оптимальні змішані стратегії

знаходяться з умови доповняльної нежорсткості

$$\nabla_{p^i} W^i = W^i e^{N_i},\ p^i \in S^{N_i}, \forall i \in D, \tag{7}$$

де $\nabla_{p^i} W^i$ – градієнт функції W^i; e^{N_i} – вектор, що складається з N_i одиниць.

Умова (7) визначає вирівнювальні за Нешем стратегії для коаліцій гравців $\tilde{D}_i \ \forall i \in D$. У загальному випадку ці стратегії є локальними (ε-оптимальними) розв'язками за Парето базової безкоаліційної гри. Вирівнювальні за Нешем стратегії безкоаліційної гри отримуються з (7) при $W^i = V^i \ \forall i \in D$.

На основі формулювань ігрової задачі в умовах невизначеності та детермінованої матричної ігрової задачі, методом стохастичної апроксимації [6] умови доповняльної нежорсткості (7) побудовано такі марківські рекурентні ігрові методи:

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i} \left\{ p_n^i - \gamma_n \zeta_n^i \left(g e^{N_i} - \frac{e(u_n^i)}{e^{\mathrm{T}}(u_n^i) p_n^i} \right) \right\}, \tag{8}$$

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i} \left\{ p_n^i - \gamma_n \zeta_n^i \left[p_n^i - e(u_n^i) \right] \right\}, \tag{9}$$

де $\pi_{\varepsilon}^{N_i}$ – оператор проектування на ε-симплекс $S_{\varepsilon}^{N_i} \subseteq S^{N_i}$; $\gamma_n \geq 0$ – параметр, що регулює величину кроку методу; $g \in \{0,1\}$; $e(u_n^i)$ – одиничний вектор-індикатор вибору варіанта u_n^i.

Метод (8) отримано за умови доповняльної нежорсткості (7), а метод (9) – на основі покомпонентного зважування векторів умови (7):

$$Z^i = diag(p_n^i)(e^{N_i} W_n^i - \nabla_{p^i} W_n^i),$$

де $Z^i \in R^{N_i}$; $diag(p_n^i)$ – квадратна діагональна матриця порядку N_i, складена з елементів вектора p_n^i. Це дає змогу на основі (9) побудувати безпроекційний метод та врахувати можливі розв'язки ігрової задачі у неповністю змішаних стратегіях на межі одиничного симплексу.

Отримано ряд модифікацій методів (8) та (9): градієнтний метод (8) при $g = 0$; безпроекційний метод (9) при обмеженнях $\gamma_n \zeta_n^i \in [0,1]$; методи без обміну інформацією при $\zeta_n^i = \xi_n^i$; з обміном інформацією при $\zeta_n^i = \sum_{k \in \tilde{D}_i} \lambda^i(k) \xi_n^k$; без відмов гравців при $\xi_n^i = \xi_n^i(u_n^{D_i})$; з відмовами гравців при $\xi_n^i = \chi\{\sum_{j \in D_i} \psi_n^j > \overline{\psi}\} \xi_n^i(u_n^{D_i(\omega)})$; регуляризовані методи при $\zeta_n^i = \xi_n^i + \delta_n e^{\mathrm{T}}(u_n^i) p_n^i$, $\delta_n > 0$.

Методи (8) та (9) забезпечують адаптивний вибір варіантів розв'язків з метою досягнення стану колективної рівноваги гравців на основі динамічних векторів змішаних стратегій p_n^i, значення яких змінюється на кожному кроці гри пропорційно величині поточних виграшів. У результаті застосування адаптивних методів зростає шанс вибору тих чистих стратегій, які формують більший середній виграш.

Визначено достатні умови збіжності ігрових методів (8) та (9) до асимптотично оптимальних розв'язків у знаковизначених середовищах та середовищах загального виду.

На основі верхніх оцінок умовного математичного сподівання поточної похибки $\Delta_n = \sum_{i \in D} \left\| Z^i \right\|^2$ виконання умови доповняльної нежорсткості при фіксованій передісторії подій та

наслідків теореми Роббінса–Сігмунда [4, 7] отримано умови збіжності з імовірністю 1.

На основі усереднення отриманих оцінок по реалізаціях подій та результатів теореми про рекурентні числові нерівності отримано умови збіжності у середньоквадратичному.

Асимптотичний порядок швидкості збіжності оцінено для послідовностей величин

$$\gamma_n = \gamma n^{-\alpha};\ \varepsilon_n = \varepsilon n^{-\beta};\ \gamma, \alpha, \beta > 0;\ \varepsilon \in (0, \min_{i \in D} N_i^{-1}) \tag{10}$$

методом моментів Чжуна [6]:

$$\overline{\lim_{n \to \infty}}\ n^{\theta} M\{\Delta_n\} \le \vartheta, \tag{11}$$

де θ – параметр порядку, ϑ – величина швидкості збіжності. Більшому θ та меншому ϑ відповідає більша швидкість збіжності ігрового методу.

Доведено, що у *знакододатному середовищі* максимальний порядок середньоквадратичної швидкості збіжності методу (8) становить $n^{-1/2}$, що досягається при $\alpha \in (1/2,1]$, $\beta = \alpha - 1/2$. Максимальний асимптотичний порядок середньоквадратичної швидкості збіжності методу (9) дорівнює n^{-1}, що досягається при $\alpha = 1$, $\beta \ge 1$.

У *середовищі загального виду* метод (8) забезпечує максимальний асимптотичний порядок швидкості збіжності дорівнює $n^{-1/3}$, що досягається при $\alpha = 2/3$, $\beta = 1/3$. Для методу (9) максимальний порядок дорівнює $n^{-1/2}$, що досягається при $\alpha = 1/2$, $\beta \ge 1/2$.

Рекурентними методами (8) та (9) формують послідовність чистих стратегій, необхідної для досягнення асимптотичної мети методом проб та помилок, що загалом пояснює їх невисоку (степеневу) швидкість збіжності.

З отриманих результатів випливає, що в умовах невизначеності метод (9) забезпечує вищий порядок швидкості збіжності до множини оптимальних розв'язків, ніж метод (8), як у знаковизначених середовищах, так і у середовищах загального виду. За результатами додаткових досліджень встановлено, що відмови гравців призводять до ε-оптимальних розв'язків ігрової задачі та до сповільнення швидкості збіжності. При зменшенні порогу гри $\overline{\psi}$ швидкість збіжності ігрових методів зростає. Методи без обміну та з обміном інформацією мають однаковий асимптотичний порядок швидкості збіжності θ в афінно еквівалентних середовищах. Обмін інформацією між гравцями призводить до зростання величини швидкості збіжності (зменшення значення ϑ). Регуляризовані методи забезпечують стійкість розв'язків ігрової задачі у змішаних стратегіях.

Планування комп'ютерного експерименту

Теоретичні результати щодо умов збіжності рекурентних ігрових методів ґрунтуються на верхніх асимптотичних оцінках характеристик випадкових процесів, отриманих при $n \to \infty$.

Для практичних застосувань важливо визначити реальні характеристики поведінки ігрових методів на вибірках скінченної довжини при $n \le n_{\max} << \infty$.

Відповідність результатів теоретичним оцінкам можна визначити за допомогою комп'ютерного експерименту. Для моделювання гри на комп'ютері необхідно задати: 1) кількість гравців та структуру локальних зв'язків між ними; 2) модель середовища та його параметри; 3) модель ігрової взаємодії гравців; 4) цільові умови асимптотичної оптимальності; 5) рекурентний алгоритм зміни векторів змішаних стратегій; 6) початкові значення змішаних стратегій та параметрів алгоритму; 7) довжину вибірки, на якій відбувається навчання ігрового алгоритму.

Структура локальних зв'язків гри задається бінарною матрицею суміжностей $Q = [q]_{L \times L}$ для фіксованої кількості гравців $L = |D|$. Одиничні елементи *i*-го рядка цієї матриці визначають множину гравців D_i, в базисі векторів змішаних стратегій яких визначаються виграші *i*-го гравця.

Відповідно, одиничні елементи *i*-го стовпчика матриці визначають множину гравців $\tilde{D}_i$, виграші яких залежать від стратегій *i*-го гравця.

Кожен гравець $i \in D$ характеризується імовірністю відмови η^i від реалізації поточного ходу. Відмова *i*-го гравця моделюється виконанням умови $\omega \le \eta^i$, де $\omega \in [0,1)$ – дійсне випадкове число з рівномірним законом розподілу. Результати відмов гравців записуються у L – елементний вектор ψ поточного стану активності гравців: $\psi[i] = \{0,1\}$, де значення $\psi[i] = 0$ відповідає активному стату гравця, якщо $\omega > \eta^i$, а $\psi[i] = 1$ – пасивному стану, якщо $\omega \le \eta^i$.

Вважається, що відмови діють тимчасово до закінчення поточного кванту часу, після чого гравці знову переходять в активний стан.

Відмови гравців призводять до зміни структури гри. Оскільки *i*-й гравець представлений у грі власним вектором змішаних стратегій p_n^i, то при його відмові цей вектор вилучається з базисів гравців множини $\tilde{D}_i$, виграші яких залежать від стратегії *i*-го гравця. Іншими словами, при відмові *i*-го гравця всі елементи *i*-го стовпця матриці суміжностей набувають нульових значень.

З урахуванням відмов гравців матриця суміжностей перетворюється за правилом

$$\tilde{Q} = Q * diag(\psi),$$

де $diag(\psi)$ – діагональна квадратна матриця порядку *L*, складена з елементів вектору ψ ознак активності гравців.

Після корегування матриці суміжностей відбувається сканування її рядків з метою виявлення гравців, які залишаються у грі. Ознакою участі *i*-го гравця у грі є виконання однієї з таких умов: 1) $\sum_{j=1}^{L} q[i,j] \ge 1$, якщо допускається "гра з природою"; 2) $\sum_{j=1}^{L} q[i,j] \ge 2$, якщо "гра з природою" не допускається. У результаті відмов гравців можливе тимчасове порушення зв'язності структури гри.

Ознаки участі гравців у грі записуються у бінарний вектор $\lambda = (\lambda[i] |\ \lambda[i] \in \{0,1\}, i = \overline{1,L})$. Якщо $\lambda[i] = 1$, то гравець бере участь у грі, інакше – утримується від реалізації поточного ходу.

Для гравців, які беруть участь у грі, обчислюються поточні значення виграшів $\xi_n^i(u^{D_i}, \omega)$, які визначаються у скорегованих базисах множини гравців $D_i = \{ j \mid j = \overline{1,L}; q[i,j] = 1; \lambda[i] = 1\} \ne \varnothing$. У загальному випадку закон розподілу випадкових величин виграшів $\xi_n^i(u^{D_i}, \omega)$ апріорі не відомий, але для моделювання ігрових алгоритмів його необхідно задати як базову характеристику моделі середовища.

Модель середовища визначається законом розподілу поточних виграшів. Досліджуються алгоритми з дискретними та неперервними обмеженими виграшами. Дискретні виграші задаються імовірностями їх появи, а неперервні – генеруються за одним із законів розподілу. Для дослідження передбачено декілька законів формування випадкових виграшів – нормальний, експоненційний, рівномірний.

Досліджено роботу ігрових методів у знакододатному середовищі та середовищі загального виду. Матриці математичних сподівань $\left[v^i(u^{D_i})\right]_{\forall u^{D_i} \in U^{D_i}}$ та дисперсій $\left[d^i(u^{D_i})\right]_{\forall u^{D_i} \in U^{D_i}}$ виграшів для всіх гравців формуються за допомогою вбудованого генератора випадкових величин, розподілених за рівномірним законом. Для знакододатного середовища математичні сподівання виграшів набувають значення з відрізку $[\delta, 1-\delta]$, де $0 < \delta << 1$, а для середовища загального виду – з відрізку [-1,1].

Відносне порівняння ефективності роботи різних ігрових алгоритмів здійснюється у тотожних середовищах. Для цього передбачено запам'ятовування згенерованих параметрів середовища у файлах даних на зовнішньому носії інформації та відновлення параметрів середовища з файлів.

Бінарні виграші $\xi_n^i(u^{D_i},\omega)\in\{0,1\}$ формувалися так:

$$\xi_n^i(u^{D_i},\omega)=\begin{cases}0, & \text{if} \quad \omega > v^i(u^{D_i})\\ 1, & \text{if} \quad \omega \le v^i(u^{D_i})\end{cases}, \tag{12}$$

де $\omega\in[0,1)$ – дійсне випадкове число з рівномірним законом розподілу.

Рівномірно розподілені випадкові величини отримувалися за формулою:

$$\xi_n^i(u^{D_i},\omega)=v^i(u^{D_i})+2\sqrt{3d^i(u^{D_i})}(\omega-0.5). \tag{13}$$

Ширину інтервалу рівномірного розподілу прийнято $2\sqrt{3d^i(u^{D_i})}$. Середина цього інтервалу відповідає величині середнього виграшу $v^i(u^{D_i})$.

Нормально розподілені випадкові величини (за законом Гаусса) знаходилися як сума дванадцяти рівномірно розподілених на відрізку [0,1] величин:

$$\xi_n^i(u^{D_i},\omega)=v^i(u^{D_i})+\sqrt{d^i(u^{D_i})}\left(\sum_{j=1}^{12}\omega_j-6\right). \tag{14}$$

Дзеркальний експоненційний розподіл (за законом Лапласа) формувався на основі логарифмічної залежності

$$\xi_n^i(u^{D_i},\omega)=v^i(u^{D_i})+\sqrt{0.5d^i(u^{D_i})}\ln(1-|2\omega-1|)sign(2\omega-1), \tag{15}$$

де $sign(2\omega-1)=\begin{cases}1, & \text{if} \quad 2\omega-1\ge 0\\ -1, & \text{if} \quad 2\omega-1<0\end{cases}$.

Закони розподілу перевірено на вибірці 10 тис. кроків. Перевірка генераторів (12) – (15) на відповідність аналітичним рівномірному, нормальному та експоненційному законам виконана за допомогою критерію χ^2 - Пірсона на рівні значності 0.05.

Вибір чистих стратегій u_n^i гравців здійснюється на основі векторів змішаних стратегій p_n^i. Номер *k* поточної чистої стратегії $u_n^i=u^i\in U^i$ визначається з умови:

$$\sum_{j=1}^{k}p_n^i(j)\ge\omega,\ k=\overline{1,N_i}.$$

Після вибору чистих стратегій всіма гравцями з множини D_i $\forall i\in D$ визначаються математичні сподівання $v^i(u^{D_i})$ та дисперсії $d^i(u^{D_i})$ випадкових величин $\xi_n^i(u^{D_i})$, які використовуються як параметри одного з заданих законів розподілу.

Спосіб використання отриманих виграшів $\xi_n^i(u^{D_i})$ залежить від прийнятої **моделі ігрової взаємодії гравців**.

В іграх без обміну інформацією поточні виграші використовуються окремо кожним гравцем для перерахунку власних векторів змішаних стратегій за рекурентними методами (8) або (9).

В іграх з обміном інформацією відбувається обмін отриманими виграшами у межах локальних коаліцій гравців D_i, стратегії яких визначають виграші *i*-го гравця $\forall i\in D$. У результаті обміну *i*-й гравець отримає поточні виграші від всіх гравців, виграші яких залежать від стратегій *i*-го гравця. Множина таких гравців утворює локальну коаліцію $\tilde{D}_i$. Зважені виграші

$$\xi_n^i = \sum_{k \in \tilde{D}_i} \lambda^i(k) \xi_n^k$$

використовуються *i*-м гравцем для модифікації власного вектора змішаних стратегій.

Досліджено збіжність ігрових алгоритмів із однією з вирівнювальних стратегій, для яких виконується умова доповняльної нежорсткості (7).

Значення параметрів γ_n та ε_n проекційних алгоритмів змінюються у часі $n = 1,2,...$ згідно з (10).

Початкові значення параметрів рекурентних алгоритмів вибираються такими, щоб задовольнити відповідні умови тверджень про збіжність алгоритмів у середньоквадратичному.

Початкові значення елементів векторів змішаних стратегій приймаються однаковими

$$p_n^i(j) = 1/N_i, \ j = \overline{1, N_i}, \ \forall i \in D,$$

що моделює ситуацію невизначеності у початковий момент часу, коли гравці не мають інформації про середовище прийняття рішень і будь-який з допустимих варіантів рішення є прийнятним для реалізації.

Для нормування векторів змішаних стратегій p_n^i **проектують їхні рекурентні перетворення на** ε**-симплекс** $S_{\varepsilon_n}^{N_i}$ [4], що зводиться до ітераційного алгоритму проектування вектора на одиничну гіперплощину з подальшим зануленням його від'ємних компонентів.

Обчислення критеріїв. Для кожної моделі гри визначається зміна у часі випадкових процесів $\Delta_n = \sum_{i \in D} \left\| p_n^i - \tilde{p}_n^i \right\|^2$ та $\overline{\Delta}_n = \frac{1}{n} \sum_{t=1}^{n} \Delta_t$.

Для обчислення $\tilde{p}_n^i = diag(p_n^i) \nabla_{p^i} V_n^i \Big/ V_n^i$ моделювали повну групу подій, пов'язаної з відмовами гравців. Повна група подій генерувалася послідовним перебором всіх можливих значень *L*-розрядного двійкового числа $s = (s_i \mid s_i \in \{0,1\}, i = \overline{1, L})$, представленого у вигляді одновимірного масиву. Одиничне значення *i*-го розряду двійкового числа сигналізує про участь *i*-го гравця у грі, а нульове значення – про його відмову.

Можливість згенерованої ситуації для *i*-го гравця визначається значенням імовірності $p_v^i(s) = \prod_{m \in D} \left\{ (1 - s_m)[\eta^m + (1 - \eta^m)(1 - p_{im})] + s_m (1 - \eta^m) p_{im} \right\}$. Якщо $p_v^i(s) > 0$, то згенерована ситуація є допустимою для *i*-го гравця.

Додатковим обмеженням на прийнятність згенерованої ситуації для *i*-го гравця є мінімальна кількість гравців, які можуть брати участь у грі. Якщо допускаються варіанти гри з природою, то необхідно, щоб виконувалась умова $\sum_{m \in D} s_m > 0$. Якщо ж варіанти гри з природою не допускаються, то $\sum_{m \in D} s_m > 1$.

Одиничні елементи вектора *s* задають гравців, стратегії яких визначають виграші *i*-го гравця. В базисі цих стратегій визначається значення функції середніх виграшів $V_n^i(s)$ та її градієнта $\nabla_{p^i} V_n^i(s)$.

Для цього методом послідовного перебору генерувалися всі можливі комбінації спільних стратегій гравців, які визначають виграші *i*-го гравця в ситуації *s*. Функції середніх виграшів гравців $V_n^i(s)$ для ситуації *s* визначалися згідно з (6). Для повної групи подій, пов'язаних з відмовами гравців, значення функцій середніх виграшів обчислювалися так:

$$V_n^i = \sum_s p_v^i(s) V_n^i(s).$$

Довжина досліджуваної вибірки становить 10 тис. кроків. На основі оцінки швидкості збіжності (11) поведінка процесу Δ_n у часі апроксимована залежністю $\Delta_n = \vartheta / n^\theta$, де $\vartheta > 0; \theta \in (0,1]; n = 1,2,\ldots$. Після логарифмування отримуємо лінійне співвідношення

$$\lg \Delta_n = \lg \vartheta - \theta \lg n. \quad (16)$$

З врахуванням цього будується залежність $\lg \Delta_n = f(\lg n)$. Тоді параметр $\theta = \frac{\lg \Delta_n}{\lg n}$ вказує на порядок швидкості збіжності досліджуваного алгоритму.

Для визначення порядку швидкості збіжності θ виконана апроксимація випадкового процесу $\lg \Delta_n$ лінійною залежністю (16) на відрізку $\lg n \in [3,4]$ з кроком $\delta = 0.1$ за методом найменших квадратів.

Для згладжування випадкової складової швидкості збіжності та виділення порядку цієї швидкості виконується усереднення випадкового процесів Δ_n та $\overline{\Delta}_n$ за $m = 50$ реалізаціями:

$$\overline{\overline{\Delta}}_n = \frac{1}{m} \sum_{j=1}^{m} \overline{\Delta}_{n,j}. \quad (17)$$

Задачі експерименту. Під час експерименту необхідно дослідити: 1) достовірність отриманих теоретичних результатів щодо збіжності ігрових методів до оптимальних колективних розв'язків у знаковизначених середовищах та середовищах загального виду; 2) вплив початкових параметрів методів на порядок та величину швидкості збіжності; 3) вплив законів розподілу випадкових величин на швидкість збіжності ігрових методів; 4) вплив дисперсії випадкових величин виграшів на величину швидкості збіжності; 5) вплив обміну інформацією на поведінку ігрових методів; 6) вплив кількості гравців та розмірності вектору чистих стратегій на величину швидкості збіжності ігрових методів; 7) вплив відмов гравців на величину швидкості збіжності; 8) стійкість досягнутих ефективних рішень для гри з локальними зв'язками.

Простір експерименту визначимо у базисі наступних параметрів: $\Pi = (Z, v, d, \gamma, \varepsilon, \alpha, \beta, \eta)$, де Z – закон розподілу виграшів, v – значення математичних сподівань виграшів, d – значення дисперсій виграшів, $\gamma, \varepsilon, \alpha, \beta$ – параметри алгоритму; η – імовірність відмов гравців.

Ігровий алгоритм

Графічну схему алгоритму повторювальної гри мультиагентної системи зображено на рис. 1.

Вхідними даними для роботи алгоритму є кількість гравців $L = |D|$, кількість чистих стратегій N_i ($i = \overline{1, L}$), структура гри у вигляді матриці суміжностей гравців, діапазони математичних сподівань та дисперсій виграшів, закони розподілу виграшів, імовірності відмов гравців $\eta^i \in [0,1)$, початкові значення векторів змішаних стратегій $p_0^i(j) = N_i^{-1}$, $j = \overline{1, N_i}$ та початкові значення параметрів $\gamma_0; \alpha; \varepsilon_0; \beta$ для ненавченого ігрового методу.

Параметри середовища у вигляді матриць математичних сподівань та дисперсій виграшів зчитуються з попередньо підготовленого файла або генеруються у програмі у межах заданих діапазонів.

На основі імовірностей відмов гравців η^i генеруються ознаки $\psi^i \in \{0,1\}$ їх участі у грі та на основі цього формується поточна структура гри.

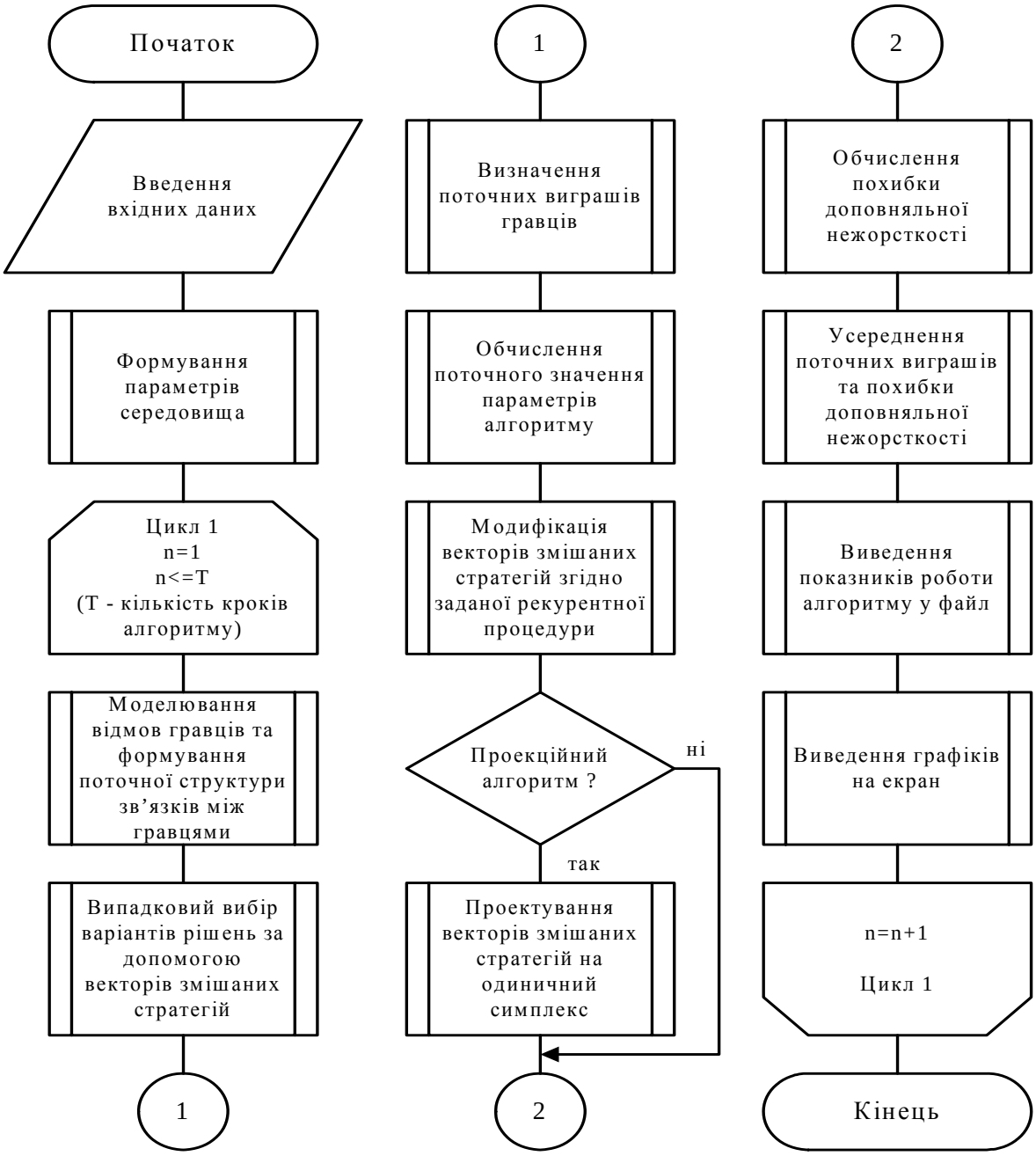


Рис. 1. Графічна схема ігрового алгоритму

Визначення номера чистої стратегії здійснюється з виконання умови

$$k = \left(K \middle| \min_{K} \sum_{j=1}^{K} p_n^i(j) > \omega \right),\ K = \overline{1, N_i},$$

де ω – випадкове число, розподілене за рівномірним законом в інтервалі [0,1]. За номером чистої стратегії визначається варіант рішення $u^i(k) \in U^i$.

Поточні виграші визначаються згідно з (14) як нормально розподілені випадкові величини з математичним сподіванням $v^i(u^{D_i})$ та дисперсією $d^i(u^{D_i})$.

Значення регульованих параметрів алгоритму γ_n та ε_n у момент часу n обчислено за правилами (10).

Нові вектори змішаних стратегій обчислюють за рекурентними перетвореннями (8) або (9) із застосуванням проектора на одиничний ε-симплекс [4].

Характеристиками гри є функції середніх виграшів та похибка умови доповняльної нежорсткості. Функції поточних виграшів (1) усереднюються по кількості гравців:

$$\overline{\Phi_n} = L^{-1}\sum_{i=1}^{L}\Phi_n^i \ .$$

Похибка виконання умови доповняльної нежорсткості обчислюється згідно з (17).

Результати імітаційного моделювання

Результати комп'ютерного моделювання підтверджують сформульовані теоретичні положення. Кількість ітерацій моделювання однієї реалізації ігрового алгоритму прийнята рівною 10 тис. кроків. Узагальнення результатів отримано усередненням 50 реалізацій кожного алгоритму для фіксованих початкових даних.

Залежність часу моделювання однієї реалізації алгоритму від розмірності ігрової задачі наведено на рис. 2 . Дані подано для ПК з процесором Intel Pentium MMX у перерахунку на тактову частоту 1 ГГц.

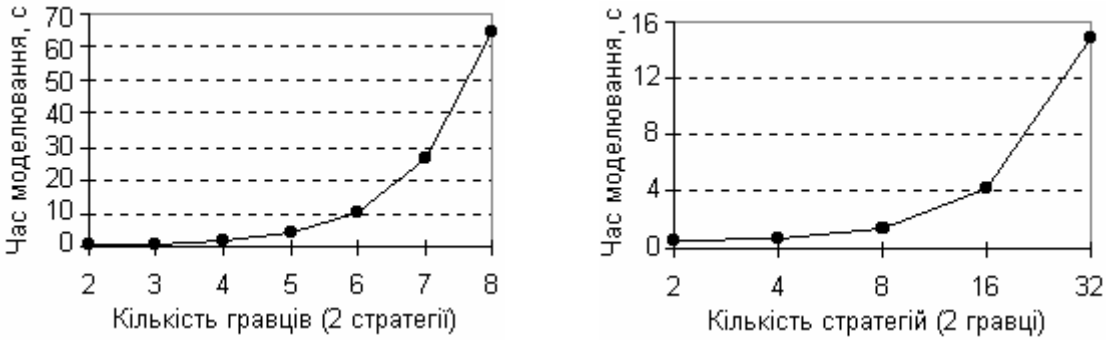


Рис. 2. Залежність часу моделювання від розмірності ігрової задачі

Експериментально встановлено, що збіжність досліджуваних адаптивних ігрових методів не залежить від початкового наближення на одиничному симплексі та від закону розподілу випадкових виграшів. При зростанні дисперсії виграшів та розмірності ігрової задачі (кількості гравців, кількості чистих стратегій, потужності множин D_i) швидкість збіжності ігрових методів зменшується. Результати моделювання стохастичної гри $|D|=|D_i|=|\widetilde{D}_i|=5,\ \ N_i=2\ \ \forall i\in D$ зображені на рис. 3 – 4.

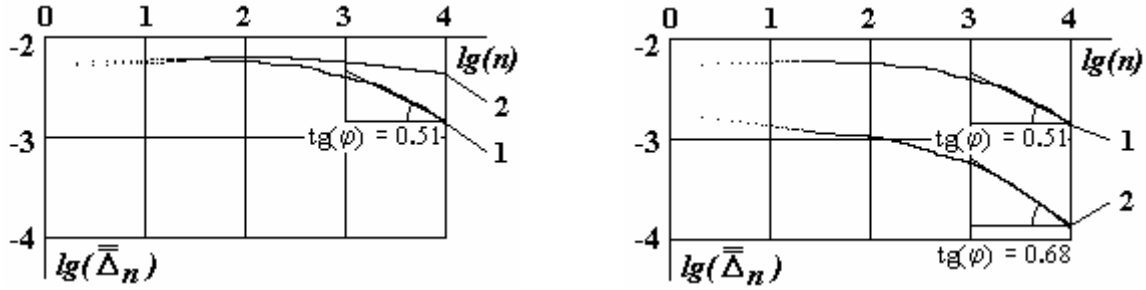
Порядок швидкості збіжності ігрових методів визначався на основі логарифмування виразу (11) та оцінки $\theta \approx tg(\varphi)$, де φ – кут нахилу прямої лінійної апроксимації функції Δ_n з координатним напрямком часу на відрізку $[10^3,10^4]$.

Графіки на рис. 3 отримано для методів (8), (9) з параметрами $\gamma_n = n^{-0.5}$, $\varepsilon_n = 0.45n^{-0.5}$, $\xi_n^i \sim Normal\left(v^i(u^{D_i})\in[0.1,0.9], d(u^{D_i})=1\right)$.

Підтверджено (рис. 3а), що у знакододатних середовищах та у середовищах загального виду ігровий метод (9) забезпечує більшу швидкість збіжності, ніж метод (8) та його частковий варіант – градієнтний метод. Методи з обміном інформацією забезпечують більшу швидкість збіжності, ніж методи без обміну інформацією (рис. 3б).

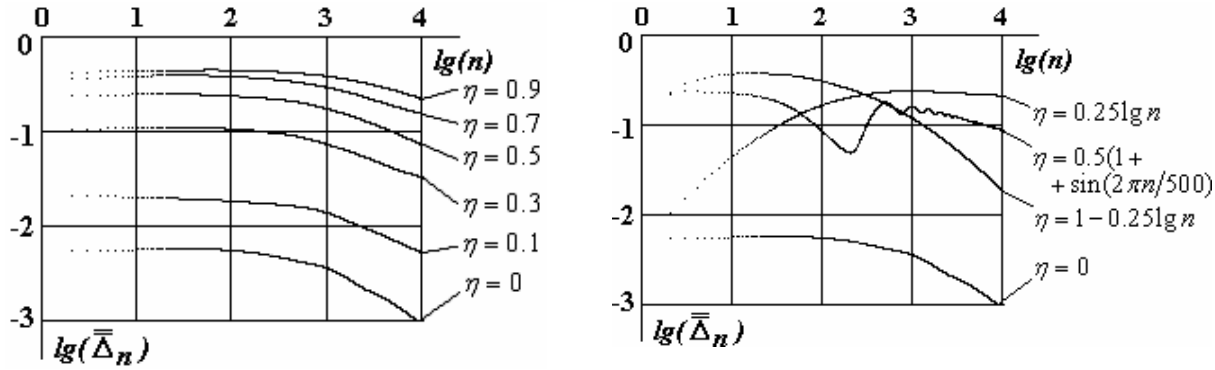
При зростанні імовірностей відмов гравців швидкість збіжності ігрових методів зменшується. Відповідні дані подано на рис. 4 для методу (9) з параметрами $\gamma_n = 0.5n^{-0.5}$, $\varepsilon_n = 0.45n^{-1}$,

$\xi_n^i \in \{0,1\}$, $v^i(u^{D_i}) \in [0.1,\ 0.9]$.



а б

Рис. 3. Збіжність ігрових методів: а – без обміну інформацією: 1 – метод (9), 2 – метод (8); б – для методу (9): 1 – без обміну інформацією, 2 – з обміном інформацією.



а б

Рис. 4. Вплив імовірностей відмов гравців $\eta^i = \eta \ \ \forall i \in D$ на збіжність ігрового методу (9): а – для стаціонарних відмов; б – для нестаціонарних відмов

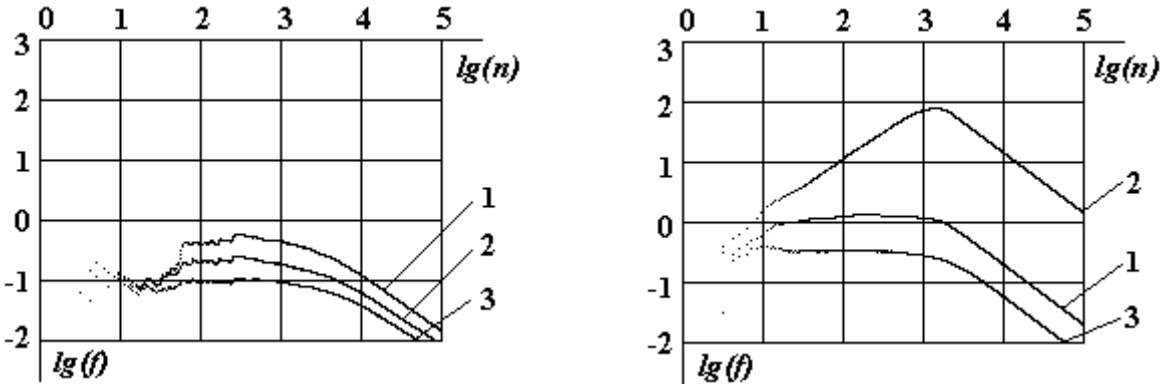
Досліджено вплив послідовностей γ_n, ε_n на швидкість збіжності ігрових методів. Виявлено незначні відмінності від зроблених теоретичних оцінок щодо можливості досягнення максимального порядку швидкості збіжності, що можна пояснити переважаючою дією на початковому відрізку часу моделювання складових вищих порядків, які не були враховані в отриманих асимптотичних оцінках, та особливостями стохастичного моделювання.

На основі розроблених адаптивних ігрових методів сформульовано та розв'язано задачу маршрутизації пакетів повідомлень у глобальних комп'ютерних мережах [8, 9].

Побудовано ігрову модель маршрутизації пакетів повідомлень у комп'ютерних мережах. Вузли комутації пакетів розглядаються як гравці, змішані стратегії яких визначають імовірності вибору вихідних каналів передавання даних. Поточні затримки передавань пакетів між вузлами комутації оформляються у маршрутні таблиці, якими сусідні гравці обмінюються між собою після завершення сеансу передавання. Після обміну інформацією гравці здійснюють перерахунок власних змішаних стратегій. Якщо передавання даних вибраним каналом призвело до зменшення середнього часу затримок пакетів, то імовірність вибору цього каналу зростає, інакше – зменшується.

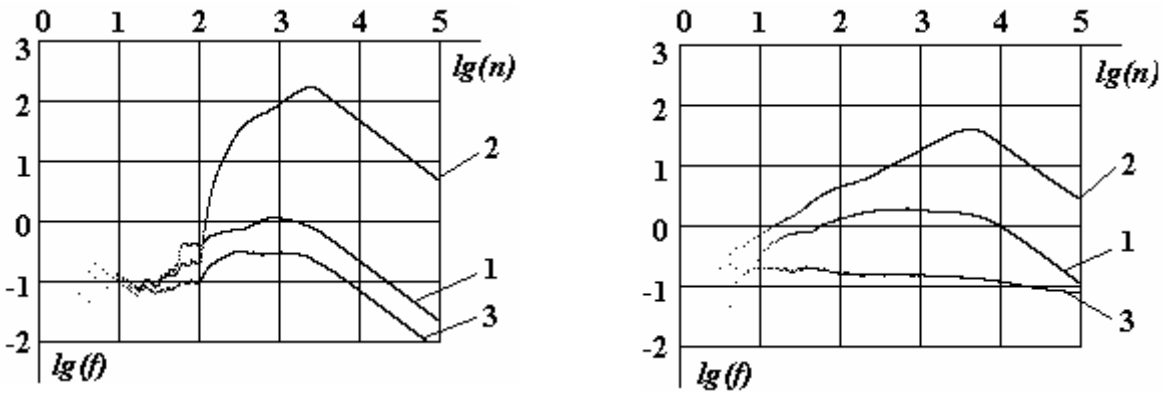
Працездатність ігрових методів маршрутизації перевірено на основі імітаційної моделі роботи повнозв'язної та варіантів неповнозв'язної мережі, яка складається з п'яти вузлів комутації пакетів. Досліджено ефективність ігрових методів маршрутизації при різних значеннях інтенсивностей вхідних безпріоритетних та пріоритетних потоків пакетів, різних режимах їх адресування, без відмов та з відмовами вузлів комутації, без обмеження та з обмеженням на довжини черг. Результати моделювання ігрового методу (9) з обміном службовою інформацією для

послідовностей $\gamma_n = 0.1n^{-0.1}, \varepsilon_n = 0.2n^{-1}$ зображено на рис. 5, 6 у вигляді графіків зміни у часі середніх значень довжин черг та кількості транзитних передавань пакетів у повнозв'язній мережі. Графік 1 показує зміну середньої кількості транзитних передавань пакета, графік 2 – середньої довжини черги, графік 3 – усередненої у часі функції програшів.



$\mu^i = 0.1$ $\mu^i = 0.5$

Рис. 5. Ефективність ігрової маршрутизації для стаціонарних імовірностей генерування пакетів



а б

Рис. 6. Ефективність ігрової маршрутизаці: а – для нестаціонарних імовірностей генерування пакетів $\mu^i = random \in [0,1)$ *з періодом зміни* $t = 100$ *кроків; б – для відновлювальних відмов вузлів з імовірностями* $\eta^i = 0.5$ *та джерелами пакетів* $\mu^i = 0.1$

Ефективність алгоритмів маршрутизації характеризується параметрами $f = \left(\overline{k_n},\ \overline{O}_n,\ \overline{\Phi}_n\right)$, де $\overline{k}_n = L^{-1}\sum_{i\in D}\sum_{j=1}^{O_n^i} k_{ij} \Big/ O_n^i$ – середня кількість транзитних передавань пакетів (O_n^i – довжина черги *i*-го вузла, k_{ij} – кількість транзитних передавань *j*-го пакета *i*-ї черги); $\overline{O}_n = L^{-1}\sum_{i\in D} O_n^i$ – середня довжина черг пакетів; $\overline{\Phi}_n = L^{-1}\sum_{i\in D}\sum_{t=1}^{n} \xi_t^i \Big/ n$ – середня у часі функція програшів.

Графіки рис. 5 отримано для стаціонарних імовірностей генерування пакетів повідомлень. Графіки рис. 6, а отримано для нестаціонарних імовірностей генерування пакетів, а на рис. 6, б – для відновлювальних відмов вузлів.

Виявлено, що поведінка адаптивних ігрових методів маршрутизації переважно визначається інтенсивностями та структурою потоків пакетів у мережі, а також способом зміни у часі

регульованих параметрів методів. Адаптивні ігрові методи, завдяки їх властивості самонавчання, забезпечують ефективне керування як стаціонарними (рис. 5), так і нестаціонарними (рис. 6) потоками пакетів в умовах відновлювальних відмов вузлів комутації.

Встановлено, що ігровий метод з обміном службовою інформацією забезпечує працездатність мереж різної топології при удвічі більших інтенсивностях вхідних потоків, ніж ігровий метод без обміну інформацією та обрані контрольні методи маршрутизації, що виявляється у зменшенні середніх довжин черг та середньої кількості транзитних передавань пакетів.

Висновки

Розроблені ігрові моделі дозволяють дослідити процеси колективного формування варіантів рішень в мультиагентних системах. У загальному такі рішення є компромісними щодо можливості отримання максимально можливих індивідуальних виграшів. Побудова компромісних рішень здійснюється шляхом координації дій агентів, яка досягається методами самонавчання та адаптації до невизначеностей системи прийняття рішень.

Для експериментального дослідження збіжності ігрових методів розроблено програмні інструментальні засоби моделювання стохастичної гри в умовах невизначеності. Проведено моделювання розроблених ігрових методів для перевірки отриманих теоретичних результатів на вибірках скінченної довжини. Показано, що за достатньо великої кількості кроків (декілька тисяч) досягається близька до теоретичної швидкість збіжності ігрових методів.

На основі математичного та програмного моделювання адаптивних ігрових методів розроблено алгоритми та програми, призначені для керування потоками інформації в комп'ютерних мережах. Встановлено, що ігрові методи з обміном інформацією забезпечують ефективну маршрутизацію пакетів повідомлень у мережах різної топології при нестаціонарних вхідних потоках та відновлювальних відмовах вузлів комутації пакетів. Розроблені ігрові методи є простими для програмування, не вимагають значних затрат машинних ресурсів, є стійкими до дій неконтрольованих випадковостей та похибок обчислень.

1. Gerhard Weiss and Sandip Sen, editors. Adaptation and Learning in Multiagent Systems. *Springer Verlag, Berlin, 1996. 2. Stone P. Layered Learning in Multiagent Systems. – MIT Press, 2000. 3. Fudenberg D., Levine D. K. The Theory of Learning in Games. – Cambridge, MA: MIT Press, 1998. 4. Назин А.В., Позняк А.С. Адаптивный выбор вариантов: Рекуррентные алгоритмы. – М.: Наука, 1986. 5. Воробьев Н.Н. Основы теории игр: Бескоалиционные игры. – М.: Наука, 1984. 6. Вазан М. Стохастическая аппроксимация. – М.: Мир, 1972. 7. Невельсон М.Б., Хасьминский Р.З. Стохастическая оптимизация и рекуррентное оценивание. – М.: Наука, 1972. 8. Nelson Minar, Kwindla Hultman Kramer, and Pattie Maes. Cooperating Mobile Agents for Dynamic Network Routing. In Alex Hayzelden, editor,* Software Agents for Future Communications Systems, *chapter 12. – Springer-Verlag, 1999. 9. Littman M. and Boyan J. A Distributed Reinforcement Learning Scheme for Network Routing,TR CS-93-165,CMU, 1993.*