

**А.О. Гриндей**

*Український науково-дослідний інститут спеціальної техніки та судових експертиз Служби безпеки України, м. Київ*

*<https://orcid.org/0000-0003-3576-1005>*

*AAP-1287-2021*

## **КОМПЛЕКСНА МЕТОДИКА ДІАГНОСТУВАННЯ ВУЗЛІВ МОБІЛЬНОЇ МЕРЕЖІ 5G**

*Розглянуто тенденції розвитку сучасних телекомунікаційних мереж. Показано, що розробка і активне впровадження стандарту 5G пов'язані з масштабуванням інфраструктури та розвитком функціоналу мереж центрів обробки даних. Внаслідок необхідності обробки великих обсягів даних у режимі реального часу значно зростають вимоги до обчислювального ресурсу, пропусковості каналу передачі даних та ємності інформаційного сховища загальної інформаційної системи мережевого ресурсу. Організація телекомунікаційних мереж відповідно до стандарту 5G дозволяє суттєво розширити інструментарій по виявленню причини потенційно небезпечного перевантаження каналів між парами портів комутаторів. Вказано, що у системі моніторингу потоку даних для діагностики перевантажень телекомунікаційної мережі 5G має проводитись аналіз портів комутатора фізичному рівні та у режимі реального часу з високими показниками точності оцінки і прогнозування відповідно обмежень у максимальному рівні навантаження на обчислювальний ресурс загального комплексу і вимог по реорганізації та масштабованості системи. Розроблено алгоритм, який базується на методиці обмеженого оновлення, що надає можливість збільшити точність аналізу при стабільному рівні навантаження на обчислювальний ресурс.*

*Ключові слова: стандарт 5G, телекомунікаційна мережа, мережі центр обробки даних, пропусковість каналу передачі даних, масштабованість, математична модель, цільові функції.*

**А.А. Гриндей**

## **КОМПЛЕКСНАЯ МЕТОДИКА ДИАГНОСТИРОВАНИЯ УЗЛОВ МОБИЛЬНОЙ СЕТИ 5G**

*Рассмотрены тенденции развития современных телекоммуникационных сетей. Показано, что разработка и активное внедрение стандарта 5G связанные с масштабированием инфраструктуры и развитием функционала сетей центров обработки данных. Вследствие необходимости обработки больших объемов данных в режиме реального времени значительно возрастают требования к вычислительным ресурсам, пропускной способности канала передачи данных и емкости информационного хранилища общей информационной системы сетевого ресурса. Организация телекоммуникационных сетей в соответствии со стандартом 5G позволяет существенно расширить инструментарий по выявлению причины потенциально опасной перегрузки канала между парами портов коммутаторов. Указано, что в системе мониторинга потока данных для диагностики перегрузок телекоммуникационной сети 5G должен проводиться анализ состояния портов коммутатора на физическом уровне и в режиме реального времени с высокими показателями точности оценки и прогнозирования в соответствии с ограничениями по максимальному уровню нагрузки на вычислительный ресурс общего комплекса и требованиями по реорганизации и масштабируемости системы. Разработан алгоритм, основанный на методике ограниченного обновления, который позволяет увеличить точность анализа при стабильном уровне нагрузки на вычислительный ресурс.*

*Ключевые слова: стандарт 5G, телекоммуникационная сеть, сети центр обработки данных, пропускная способность канала передачи данных, масштабируемость, математическая модель, целевые функции.*

**A. Hryndei,**

## **COMPREHENSIVE METHODOLOGY FOR DIAGNOSING OF A 5G MOBILE NETWORKNODES**

*Trends of the modern telecommunication networks development are considered. It is shown that the development and active implementation of the 5G standard is associated with the scaling of infrastructure and the development of the functionality of networks of data centers. Due to the need to process large amounts of data in real time, the requirements for computing resources, the bandwidth of the data transmission channel and the capacity of the information storage of the general information system of the network resource increase significantly. The organization of telecommunication networks in accordance with the 5G standard makes it possible to significantly expand the tools for identifying the cause of potentially dangerous channel congestion between pairs of switch ports. It is indicated that the data flow monitoring system for diagnosing congestion of the 5G telecommunications network should analyze the state of the switch ports at the physical level and in real time with high accuracy of assessment and forecasting in accordance with the restrictions on the maximum load on the computing resource of the general complex and the requirements on reorganization and scalability of the system. An algorithm based on the limited update technique has been developed, which allows to increase the accuracy of the analysis at a stable level of load on the computing resource.*

**Keywords:** 5G standard, telecommunications network, data center networks, data transmission channel capacity, scalability, mathematical model, objective functions.

**Вступ.** Одним з актуальних трендів у галузі інформаційних технологій (IT) стала поява і активне впровадження мереж центрів обробки даних (Data Center Networks, DCN). Відповідні системи характеризуються широкими можливостями помасштабуванню інфраструктури серверних комплексів та розширенню інструментарію по обробці великих об'ємів поточкових даних у режимі реального часу. При цьому важливо зазначити, що робота DCN полягає не тільки у виконанні запитів користувачів відповідних сервісів (як то забезпечення протоколів стабільної і конфіденційної передачі блоків даних, а також їх збереження на сервері), а і у глибокому аналізі даних (Data Mining, DM) за допомогою спеціалізованих програмних додатків. Це, у свою чергу, збільшує навантаження на обчислювальний ресурс DCN та канал передачі даних. Крім того, проблема непрогнозованості росту об'ємів даних, що поступають на мережевий ресурс, зокрема поява феномену «вибухових об'ємів даних» (Exploding Data, ED), тобто наявність значного рівня нерівномірності вхідних запитів у системі і, відповідно, поява пікових навантажень, що становлять потенційну загрозу для стабільної роботи DCN, додатково актуалізується внаслідок впровадження надзвичайно ресурсоємних DM-алгоритмів.

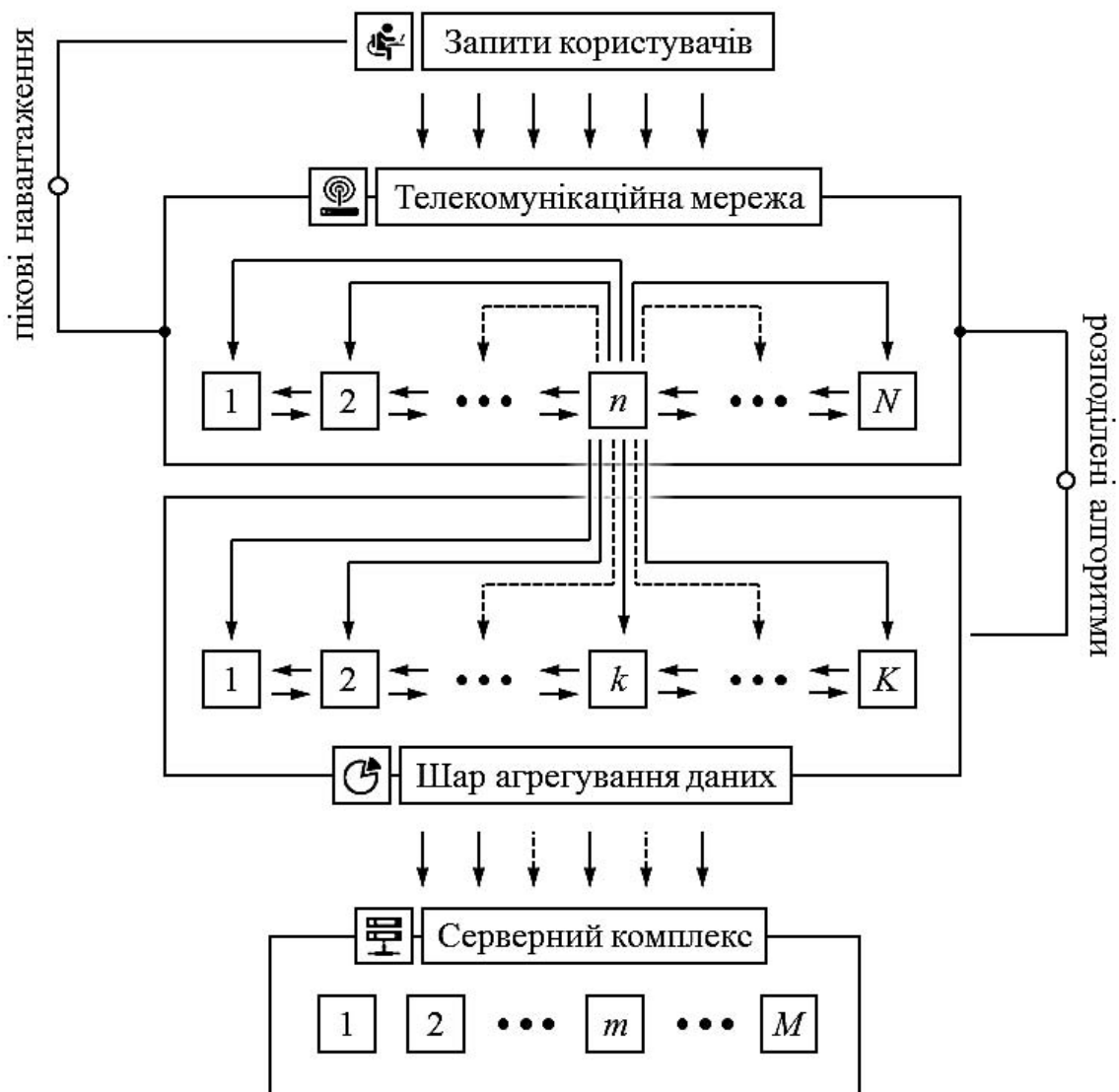


Рис. 1. Базова схема ієрархічної структури організації системи DCN

Значену проблему пропонується більш детально проаналізувати у рамках даного дослідження. Розглянемо маршрутизатор вузла DCN, що на основному рівні взаємодіє з іншими вузлами відповідно стандартів телекомунікаційної мережі. Крім того дані, які він передає та режим його роботи аналізуються системою моніторингу на базі DM-алгоритмів. Це призводить до формування трирівневої ієрархічної структури організації комплексу DCN (рис. 1):

- основний шар телекомунікаційної мережі, що складається з  $n \in [1; N]$  вузлів, структура якого активно модифікується та масштабується, а навантаження на кожен вузол є непрогнозованим;

- проміжний шар агрегації даних (моніторингу телекомунікаційної мережі), що складається з  $k \in [1; K]$  вузлів, який працює відповідно стандартів розподілених інформаційних систем (Distributed information Systems, DIT), а отже може надсилати запити до окремого вузла телекомунікаційної мережі від кількох вузлів  $k$  одночасно;

- граничний серверний шар аналізу даних на базі DM-алгоритмів, що складається з  $m \in [1; M]$  вузлів, причому  $K < M \ll N$ .

Відповідно до цього «вузьким місцем» загальної мережі стають маршрутизатори вузлів основного шару. Рівень навантаження для  $m \in [1; N]$  залежить від кількості вузлів, що взаємодіють з ним (до  $N + K$  вузлів) та непрогнозованої складності запитів користувачів і системи моніторингу. Водночас, необхідно зазначити, що зазначених вузлів у системі набагато більше ніж інших, тому збільшення потужності відповідних маршрутизаторів значно збільшить загальний кошторис DCN, а також зменшить гнучкість і масштабованість загальної системи.

**Аналіз сучасних досліджень і публікацій** присвячених проблемам організації та діагностування вузлів сучасних DCN, зокрема мобільної мережі 5G, вказує на те, необхідність проведення комплексного аналізу потоку даних, що додатково збільшує рівень навантаження і актуально, більшою мірою, для прогнозованих систем [1-3]. Відповідно, програмні додатки в залежності від типу організації DCN та заходів по їх масштабуванню можуть як оптимізувати систему передачі даних і моніторингу шляхом визначення потенційно небезпечних для системи потоків даних (ED або «elephant» у концепції «elephant & mice»), так і призвести до додаткових затримок у процесі передачі даних [4]. Альтернативним підходом є застосування алгоритмів потокової передачі даних (Streaming Algorithms, SA), що базуються на статистичному аналізі поточних значень навантаження що виникає внаслідок потокової передачі даних між парами вузлів, що характеризуються адресами мережевого протоколу маршрутизації (Internet Protocol, IP). Актуальні на сьогоднішній день алгоритми включають у себе наступні групи:

- алгоритми розрахунку зі втратами (Lossy Counting, LC);
- економні алгоритми (Space Saving, SS);
- алгоритми з відліком мінімуму (Count-Min, CM);
- алгоритми обмеженого оновлення (Conservative Update, CU).

Алгоритми потокової передачі достатньо ефективно впроваджуються у DCN стандарту 5G і характеризуються високим рівнем масштабованості [5-18], але при їх реалізації виникає одна з наступних проблем:

- високий рівень помилок [5], що є характерним для CM-алгоритмів;
- високий рівень завантаження пам'яті [6, 7], що є характерним для LC- і SS-алгоритмів;
- високий рівень затримки доступу до пам'яті [8, 9], що є характерним для CU-алгоритмів.

Проведений аналіз вказав на необхідність побудови комплексної методики побудови алгоритмів потокової передачі даних, на основі підходів, що були представлені вище. Таким чином, **невирішеною частиною загального дослідження** у рамках даної роботи пропонується обрати наступні пункти:

- розробка універсальної математичної моделі потокової передачі даних у середовищі DCN для аналізу ефективності SA-алгоритмів відповідно цільових функцій навантаження;

- побудова гібридного алгоритму обмеженого оновлення з відліком мінімуму (CU/CM-алгоритм), що базується на визначенні ковзного середнього у режимі реального часу, відповідно вимог по мінімізації рівня затримки доступу до пам'яті за умов високої точності аналізу та низького навантаження на обчислювальний ресурс.

Відповідно до цього **метою дослідження** є побудова високоєфективного алгоритму обмеженого оновлення з мінімальним навантаженням на обчислювальний ресурс апаратно-програмної платформи та канали передачі даних для моніторингу DCN стандарту 5G.

### Основна частина дослідження

Для постановки задачі оптимізації роботи системи моніторингу DCN необхідно провести аналіз статистичних даних і визначити типові причини, що призводять до виникнення ефекту перенавантаження. З цією метою обирається функція навантаження  $F_{WL}$  (Workload, WL), оптимальний часовий інтервал  $\Delta t$  і межі  $t \in [t_A; t_B]$  вимірювання  $F_{WL}$  (що визначаються на етапі налаштування системи аналізу), а також набір служб DCN.

У рамках даного дослідження набір служб представлено наступним набором:

- протокол передачі даних HTTP;
- система доменних імен DNS;
- схема уніфікованого ідентифікатору ресурсів HTTPS;
- протокол захищеного з'єднання SCSRA (Secure Computing Sidewinder RemoteAdministration);
- комунікаційний протокол SMTP.

Для аналізу потенційної загрози перевантаження актуальним є саме зміна значення  $\Delta F_{WL}$  протягом  $\Delta t$  як функція нормалізованої різниці:

$$\Delta F_{WL}(t) = \frac{|F_{WL}(t + \Delta t) - F_{WL}(t)|}{|F_{WL}(t)|} \text{ для } t \in [t_A; t_B]. \quad (1)$$

На базовому рівні CU/CM-алгоритм визначає рівень потенційної загрози від затримки виконання запиту та неточності визначення вибухових об'ємів даних через ймовірнісний параметр  $\sigma \in (0; 1]$  перемикач між CU-алгоритмом (при  $\sigma \rightarrow 1$ ) та CM-алгоритмом (при  $\sigma \rightarrow 0$ ). При цьому характерно, що як для CM-алгоритмів, так і для CU-алгоритмів використовується метод зменшення розмірності «відліковий скетч» (Count Sketch, CS), що представляє собою двовимірну матрицю-лічильник ( $M_{CM}$  та  $M_{CU}$ , відповідно) розмірності  $X \times Y$  ( $x \in [1; X], y \in [1; Y]$ ).

Точність оцінки перевантаження визначається через коефіцієнт  $\varepsilon$ , а вірогідність оцінки з точністю  $\varepsilon$  складає щонайменш  $P_\varepsilon = 1 - \delta$ , причому зазначені показники пов'язані з розмірністю матриці  $M_{CM}$  чи  $M_{CU}$  наступним чином:

У свою чергу, оцінка ефективності алгоритму визначається через показник часової складності ( $T_{CM}$  та  $T_{CU}$ , відповідно), що описує час потрібний для виконання алгоритму (на базі підрахунку кількості елементарних операцій та фіксовану кількість часу необхідну для їх виконання) шляхом застосування функції  $O(\square)$ .

Якщо потік даних протягом  $t \in [t_A; t_B]$  складає  $N_a^b$  блоків, часову складність CU/CM-алгоритму можна визначити наступним чином:

$$T_{CM} = O(N_a^b \cdot Y \cdot (1 + \sigma)) \text{ при } \sigma \in (0; 1] \text{ для } t \in [t_A; t_B]. \quad (3)$$

Відповідний підхід, тим не менш, не враховує нерівномірність навантаження в залежності від особливостей функціонування окремих служб сервісу і представляє собою компроміс між недоліками CU- та CM-алгоритмів. Тому подальше вдосконалення підходу полягає у застосуванні по відношенню до відлікових скетчів функцій розбиття  $F_{PT}$ . Це надає можливість зменшити обчислювальну складність CU-алгоритмів через корегування розміру ескізу індивідуально по відношенню до окремої служби. При цьому функція розбиття виконується як для рядків, так для стовпчиків матриці (відповідно визначається набір значень функції  $F_{PT}(x)$  для  $x \in [1; X]$  при  $y = const$  і набір значень функції  $F_{PT}(y)$  для  $y \in [1; Y]$  при  $x = const$ ).

Для оцінки зазначеного підходу необхідно порівняти матриці відлікових скетчів (до розбиття та після розбиття) однієї розмірності. Для  $I$  операцій розбиття можна розрахувати значення  $X$ ,  $Y$  та  $N$ :

$$\begin{cases} X = \sum_{i=1}^I (x_i) \\ Y = \sum_{i=1}^I (y_i) \\ N_a^b = \sum_{i=1}^I (n_i) \end{cases} \quad (4)$$

У такому разі показник часової складності визначається як час виконання  $N$  оновлень:

$$T_{PT} = O\left(2 \sum_{i=1}^I (n_i \cdot Y)\right) = O(2N_a^b Y), \quad (5)$$

$$NY = \sum_{i=1}^I (n_i \cdot y_i) + \sum_{i=1}^I (n_i \cdot (Y - y_i)), \text{ причому } N_a^b Y \gg \sum_{i=1}^I (n_i \cdot y_i). \quad (6)$$

Для CU/CM-алгоритму з функцією розбиття  $F_{PT}$  і розмірністю матриці відлікового скетчу  $X \times Y$  може бути досягнуто мінімум обчислювальної складності  $O(2NY/I)$  при розбитті матриці на  $I$  ескізів однієї розмірності незалежно від типу розбиття. При цьому повний набір значень  $\{y_i\}$  як цільових функцій, що відповідає мінімізації часу оновлення, може бути розраховано через функцію аргументу мінімізації:

$$y_i = \arg\left(\min_{y_i} \left(\sum_{i=1}^I (n_i \cdot y_i)\right)\right) \text{ для } \forall i \in [1; I]. \quad (7)$$

Відповідно задача оптимізації алгоритмів може бути визначена через математичну задачу оптимізації з обмеженнями, що вирішується через метод множників Лагранжа:

$$\frac{\partial L(\lambda)}{\partial y_i} = 0, \text{ де } L(\lambda) = \sum_{i=1}^I (n_i \cdot y_i) - \lambda \left(\sum_{i=1}^I y_i - Y\right), \quad (8)$$

Рівень помилок при цьому також залежить від розмірності матриці відлікового скетчу: (i)  $X$  обернено пропорційна ймовірності того, що у карті хеш-функції кілька об'єктів знаходяться у одній комірці; (ii) збільшення  $Y$  розподіляє об'єкти, що зменшує вірогідність відповідної похибки.

**Висновки.** В результаті проведеного дослідження було розглянуто тенденції розвитку сучасних мереж центрів обробки даних, що базуються на стандарті 5G. Показано, що при обробці великих обсягів даних у режимі реального часу значно зростають вимоги до обчислювального ресурсу, перепускності каналу передачі даних та ємності інформаційного сховища. Розглянуто методи, що дозволяють розширити інструментарій по виявленню причини потенційно небезпечного для мережі перевантаження каналів. Розроблено алгоритм, що базується на методиці обмеженого оновлення, відліку мінімуму та функція розбиття, що характеризується більшою точністю аналізу при стабільному рівні навантаження на обчислювальний ресурс. Показано, що задача оптимізації алгоритмів може бути визначена через математичну задачу оптимізації з обмеженнями і вирішена через метод множників Лагранжа.

## Список літератури

1. Liu, C. H., Kind, A., & Vasilakos, A. V. (2013). Sketching the data center network traffic. *IEEE Network*, 27(4), 33–39. <https://doi.org/10.1109/mnet.2013.6574663>.
2. Muller, N. J. (2017). Securing Distributed Data Networks. Handbook of: Data Center Management, 377–388. <https://doi.org/10.1201/9780203712726-36>.
3. Al-Fares, M., Loukissas, A., & Vahdat, A. (2008). A scalable, commodity data center network architecture. *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication - SIGCOMM '08*. <https://doi.org/10.1145/1402958.1402967>.
4. M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, and A. Vahdat, (2010) “Hedera: Dynamic flow scheduling for data center networks”. *In Proc. NSDI*, p. 19.
5. Cormode, G., & Muthukrishnan, S. (2005). An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms*, 55(1), 58–75. <https://doi.org/10.1016/j.jalgor.2003.12.001>.
6. Metwally, A., Agrawal, D., & El Abbadi, A. (2004). Efficient Computation of Frequent and Top-k Elements in Data Streams. Database Theory - ICDT 2005, 398–412. [https://doi.org/10.1007/978-3-540-30570-5\\_27](https://doi.org/10.1007/978-3-540-30570-5_27).
7. GuoliangNie,&Zhengding Lu. (n.d.). Approximate frequency counts in sliding window over data stream. Canadian Conference on Electrical and Computer Engineering, 2005. <https://doi.org/10.1109/ccece.2005.1557432>.
8. Estan, C., & Varghese, G. (2003). New directions in traffic measurement and accounting. *ACM Transactions on Computer Systems*, 21(3), 270–313. <https://doi.org/10.1145/859716.859719>.
9. Liu, C. H., & Fan, J. (2014). Scalable and Efficient Diagnosis for 5G Data Center Network Traffic. *IEEE Access*, 2, 841–855. <https://doi.org/10.1109/access.2014.2349000>.
10. B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom (2002) Models and issues in data stream systems. *in Proc. 21st ACM SIGMOD-SIGACTSIGART Symp. Principles Database Syst.*, pp. 1-16.
11. B. Babcock and C. Olston (2003) Distributed top-k monitoring. *in Proc. ACM SIGMOD Int. Conf. Manage. Data*. pp. 28-39.
12. M. Roughan, S. Sen, O. Spatscheck, and N. Duffeld (2004) Class-of-servicemapping for QoS: A statistical signature-based approach to IP traffic classification. *In Proc. ACM Conf. IMC*. pp. 135-148.
13. N. Farrington et al. (2010) Helios: A hybrid electrical/optical switch architecture for modular data centers. *In Proc. ACM SIGCOMM Conf., 2010*. pp. 339-350.
14. C. Barakat, G. Iannaccone, and C. Diot (2005) Ranking flows from sampled traffic. *In Proc. ACM CoNEXT*. pp. 188-199.
15. L. Pietronero, E. Tosatti, V. Tosatti, and A. Vespignani (2001) Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf. *Phys. A, Statist. Mech. Appl.*, vol. 293, no. 1, pp. 297-304.
16. Goyal, J. Jagarlamudi, H. Daumé, III, and S. Venkatasubramanian (2010) Sketching techniques for large scale NLP. *In Proc. NAACL HLT 6th WebCorpus Workshop*. pp. 17-25.
17. E.D. Demaine, A. López-Ortiz, and J.I. Munro (2002) Frequency estimation of internet packet streams with limited space. *In Proc. 10th Annu. Eur. Symp. Algorithms*. pp. 348-360.
18. G. Cormode and S. Muthukrishnan (2005) What's hot and what's not: Tracking most frequent items dynamically. *ACM Trans. Database Syst.*, vol. 30, no. 1, pp. 249-278.