УДК 004.056

М. Карапетьянц[1]✉

Оценка применимости инструментов анализа скрытых сервисов Tor

[1]Национальный исследовательский ядерный университет «МИФИ»,
г. Москва, Российская Федерация
e-mail: mkarapetyants@mephi.ru

Аннотация. Сеть Tor, обеспечивающая анонимность и децентрализованную маршрутизацию, активно используется как для защиты цифровых прав, так и для организации незаконной деятельности, что создает дополнительные затраты для правоохранительных органов и специалистов по кибербезопасности. Цель данной работы – оценить применимость современных инструментов анализа скрытых сервисов Tor на основе количественных критериев. В исследовании рассматриваются пять ключевых инструментов с открытым исходным кодом (TorBot, OnionSearch, Darkdump, Ahmia, Darkus), анализируется их эффективность по таким параметрам, как количество уникальных адресов, процент активных ресурсов и среднее время обработки. Результаты показали, что Ahmia демонстрирует наилучшие показатели по обнаружению уникальных адресов, а Darkdump – по актуальности данных, однако ни один инструмент не является универсальным. Исследование подтверждает необходимость разработки гибридных решений, сочетающих скорость краулеров и точность сканеров, а также стандартизации методик оценки. Работа представляет практическую ценность для правоохранительных органов, занимающихся мониторингом darknet, и может служить основой для создания более эффективных инструментов анализа скрытых сервисов. Дальнейшие исследования могут быть направлены на автоматизацию сбора данных и интеграцию методов машинного обучения для классификации контента.

Ключевые слова: оверлейные сети, TOR, скрытые сервисы, инструменты OSINT

M. Karapetyants[1]✉

Evaluating the applicability of tor hidden services analysis Tools

[1]National Research Nuclear University «MEPhI», Moscow, Russian Federation
e-mail: mkarapetyants@mephi.ru

Abstract. The Tor network, which provides anonymity and decentralized routing, is actively used both to protect digital rights and to organize illegal activities, which creates additional costs for law enforcement agencies and cybersecurity specialists. The purpose of this paper is to evaluate the applicability of state-of-the-art tools for analyzing Tor's hidden services based on quantitative criteria. The study examines five key open-source tools (TorBot, OnionSearch, Darkdump, Ahmia, Darkus) and analyzes their performance in terms of parameters such as number of unique addresses, percentage of active resources and processing time. The results show that Ahmia performs best in terms of unique address detection and Darkdump performs best in terms of data relevance, but no tool is universal. The study confirms the need to develop hybrid solutions that combine the speed of crawlers and the accuracy of scanners, as well as to standardize evaluation techniques. The work is of practical value for law enforcement agencies involved in darknet monitoring and can serve as a basis for creating more effective tools for analyzing hidden services. Further research could focus on automating data collection and integrating machine learning techniques for content classification.

Keywords: overlay networks, TOR, hidden service, OSINT tools

Введение

Сеть Tor, основанная на принципах анонимности и децентрализованной маршрутизации, продолжает оставаться ключевым инструментом как для защиты цифровых прав, так и для организации незаконной деятельности. По данным исследований, около 45 % скрытых сервисов Tor связаны с киберпреступностью, включая торговлю запрещенными товарами, распространение вредоносного ПО и утечки персональных данных [1, 2]. Это создает значительные вызовы для правоохранительных органов и специалистов по информационной безопасности, чья работа зависит от эффективного анализа содержимого луковых сервисов.

Существующие инструменты анализа сети Tor, такие как сканеры, краулеры и решения с интеграцией машинного обучения, предлагают разнонаправленные подходы к решению этой задачи. Однако их эффективность варьируется в зависимости от контекста: одни фокусируются на скорости сбора данных, другие – на глубине технического анализа, что затрудняет выбор оптимального инструментария [3]. Проблема усугубляется высокой динамичностью сети Tor, где 50 % сервисов становятся неактивными в течение 24 часов, а 20 % регулярно меняют адреса [2].

Цель данной работы – исследовать решения в области сбора и анализа данных луковых адресов скрытых сервисов сети TOR и оценить их применимость на основе количественных критериев.

Теоретическая значимость работы заключается в систематизации метрик, которые могут быть использованы для стандартизации тестирования инструментов OSINT. Практическая ценность проявляется в рекомендациях для интеграции решений в системы мониторинга darknet, используемые правоохранительными органами и SOC-командами.

Работа продолжает исследования, начатые в [4, 5], где были систематизированы методы сбора адресов, и дополняет их анализом этапа обработки данных. Данное исследование фокусируется на этапе анализа, что позволяет выявить оптимальные комбинации инструментов для конкретных сценариев. Результаты также актуальны для разработки гибридных решений, сочетающих скорость краулеров и точность сканеров, что особенно важно в условиях роста количества onion-сервисов.

Для решения проблематики, связанной с разработкой универсального инструмента для сбора и анализа данных в сети TOR, научным сообществом исследуются и разрабатываются новые подходы в рамках проактивного мониторинга скрытых сервисов в оверлейных сетях. В данной области проводят исследования следующие ученые: Sarda T., Pastor-Galindo J., Marmol F. G., Perez G. M., Butler, Wardman B., Pratt N., Basynya E. [1, 2, 5, 6]

Методы и материалы

Современные исследователи в области информационной безопасности уделяют значительное внимание разработке и совершенствованию инструментов для

анализа скрытых сервисов сети Tor. Как показали исследования [7], эффективный мониторинг луковых сервисов требует комплексного подхода, сочетающего автоматизированные методы сбора данных с глубоким анализом контента. Особую актуальность эта задача приобретает в свете того, что около 45 % скрытых сервисов используются для организации незаконной деятельности, в то время как 47 % предоставляют легитимные услуги, такие как анонимные форумы и платформы для обмена файлами.

Для проведения исследования были отобраны пять ключевых инструментов с открытым исходным кодом [8], представляющих различные подходы к работе со скрытыми сервисами.

1. TorBot – краулер с возможностью рекурсивного обхода ссылок. Как отмечают исследователи [9], подобные инструменты полезны для первичного сбора данных, однако требуют тщательной настройки.

2. OnionSearch – инструмент поиска скрытых сервисов, использующий режим многопроцессорной обработки. По данным [10], такие системы позволяют сократить время поиска релевантных сервисов на 30–40 % по сравнению с ручными методами.

3. Darkdump – инструмент для агрегации данных из открытых источников, включая форумы и репозитории. Его главное преимущество, как подчеркивается в работе [11] – способность работать с устаревшими и малопопулярными источниками информации.

4. Ahmia Search Engine – одна из немногих поисковых систем, официально одобренных проектом Tor. Исследование [12] показало, что Ahmia охватывает около 60 % активных англоязычных сервисов.

5. Darkus – модульный фреймворк для глубокого анализа содержимого скрытых сервисов. Его архитектура, описанная в [13], позволяет гибко настраивать параметры сканирования в зависимости от конкретных задач.

Для тестирования использовался набор 10 ключевых слов, характерных для различных категорий скрытых сервисов.

Критерии оценки были выбраны на основе анализа современных исследований [14]:

1. Количество уникальных адресов – ключевой показатель для оценки полноты охвата. Как показали эксперименты [15, 16], профессиональные инструменты обнаруживают в несколько раз больше уникальных сервисов по сравнению с ручными методами.

2. Процент активных ресурсов – показатель, непосредственно влияющий на достоверность результатов. Методика [17] рекомендует сочетать автоматические проверки с выборочной ручной верификацией.

3. Среднее время обработки – критически важный параметр для оперативного мониторинга. Оптимальные значения, согласно [18, 19], должны находиться в диапазоне 2–5 секунд на адрес.

Экспериментальная часть исследования проводилась на специализированном стенде, включающем:

– серверное оборудование с процессорами Intel и 32 ГБ оперативной памяти;

– выделенный канал связи с пропускной способностью 1 Гбит/с;
– изолированную среду выполнения на базе Docker;
– систему мониторинга ресурсов и сбора статистики.

Все тесты проводились с соблюдением норм и рекомендаций проекта Tor [20], включая ограничение скорости запросов и минимизацию нагрузки на сеть.

Исследование

В рамках исследования была проведена комплексная оценка применимости пяти современных инструментов для анализа скрытых сервисов сети Tor. Экспериментальная работа осуществлялась на специально подготовленном наборе данных, включающем 10 ключевых запросов различного направления, что позволило объективно сравнить эффективность каждого решения по установленным критериям (табл. 1).

Таблица 1

Сравнительный анализ инструментов

Инструмент	Уникальные адреса	Активность (%)	Время обработки (с)
TorBot	820	62	4,8
OnionSearch	750	58	3,2
Darkdump	945	68	5,0
Ahmia	977	70	5,4
Darkus	900	65	6,3

По количеству уникальных адресов безусловным лидером стал Ahmia. Однако, как показали дополнительные проверки, около 12 % обнаруженных им сервисов оказались дубликатами или зеркалами, что указывает на необходимость дополнительной фильтрации результатов.

По проценту активных ресурсов выделяется Ahmia (70 %), что объясняется его постоянным мониторингом статусов ресурсов и обновлении БД.

Скорость обработки варьировалась от 3,2 до 6,3 с/адрес. Как отмечается в работах [18, 21], оптимальным для оперативного мониторинга считается диапазон 2–5 с/адрес. OnionSearch отличается высокой скоростью обработки, что соответствует его многопроцессорной работе.

Сравнительный анализ позволяет заключить, что наиболее сбалансированным инструментом для исследования скрытых сервисов по ключевым запросам является Darkdump, сочетающий в себе необходимые количественные и качественные характеристики в рамках исследования сети TOR.

При этом для решения специализированных задач, например, проверки активности или глубокого сканирования, рекомендуется комбинировать с утилитами ска-

нирования и краулерами. Как справедливо отмечают исследователи [2], подобный гибридный подход позволяет компенсировать ограничения отдельных инструментов и добиваться наиболее полных и достоверных результатов.

Полученные данные подтверждают выводы предыдущих исследований [2] о необходимости разработки универсального решения, интегрирующего преимущества краулеров, поисковых систем и сканеров уязвимостей. В перспективе это позволит существенно упростить процесс мониторинга скрытых сервисов и повысить эффективность выявления противоправного контента.

Как отмечается в современных исследованиях [22], ни один из существующих инструментов не может полностью удовлетворить все требования анализа скрытых сервисов. Наиболее перспективным направлением представляется разработка гибридных решений, сочетающих скорость Darkdump с точностью Prying Deep и функциональностью TorCrawl.

Обсуждение и заключение

Основываясь на проведенном исследовании, можно сделать ряд ключевых выводов о современных инструментах анализа скрытых сервисов сети Tor.

Все рассмотренные инструменты демонстрируют различную эффективность в зависимости от конкретных задач исследования. Ahmia показал наилучшие результаты в обнаружении уникальных адресов, в то время как Darkdump обеспечил оптимальное значение в совокупности показателей.

Проведенное исследование подтвердило, что ни один из существующих инструментов не может считаться универсальным решением для анализа скрытых сервисов. С другой стороны, стоит отметить, что текущее исследование позволило выявить ключевые механизмы в виде рекурсивного обхода ссылок, глубокого анализа контента и интеллектуальной фильтрации результатов. Грамотное комбинирование их сильных сторон для первичного сбора и валидации данных позволяет достигать высоких результатов при мониторинге сети Tor.

Важным выводом является также необходимость дальнейшей стандартизации методик оценки и разработки комплексных решений, что соответствует тенденциям, отмеченным в работах [23]. Это особенно актуально для правоохранительных органов и специалистов по кибербезопасности, нуждающихся в эффективных инструментах противодействия цифровой преступности.

В дальнейшем планируется использовать текущие наработки для автоматизации процесса развертывания и тестирования инструментов по сбору данных из скрытых сервисов, что позволит проводить комплексную оценку существующих решений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Sardá T. The dark side of the internet: A study about representations of the deep web and the Tor network in the British press : дис. – Loughborough University, 2020.,

2. Pastor-Galindo J., Mármol F. G. Pérez G. M. On the gathering of Tor onion addresses //Future Generation Computer Systems. – 2023. – Т. 145. – С. 12-26.

3. Alkhatib B., Basheer R. Crawling the dark web: A conceptual perspective, challenges and implementation //J. Digit. Inf. Manag. – 2019. – Т. 17. – №. 2. – С. 51.

4. Исследование методов сбора адресов скрытых сервисов сети TOR // Интерэкспо Гео-Сибирь, 2024г. Т. 7, Вып. 1 Стр. 133-139

5. Basynya E. A., Karapetyants N., Karapetyants M. Bitcoin Transaction Analysis System //Programming and Computer Software. – 2024. – Т. 50. – №. Suppl 2. – С. S104-S112.

6. Butler, B. Wardman, and N. Pratt, "REAPER: an automated, scalable solution for mass credential harvesting and OSINT," 2016 APWG Symp. Electron. Crime Res., pp. 1–10, 2016.

7. Garcia-Alfaro J. , Kozik R., Choras M., Katsikas S. Computer Security–ESORICS 2024: 29th European Symposium on Research in Computer Security, Bydgoszcz, Poland, September 16–20, 2024, Proceedings, Part IV. – Springer Nature, 2024. – Т. 14985.

8. Cracking the Dark Web: Essential OSINT Tools for Investigators. — URL: https://www.osintteam.com/osint-tools-for-the-dark-web/ (дата обращения 13.04.2025).

9. Dutta N. , Jadav N., Tanwar S., Sarma HKD, Pricop E. Cyber security: issues and current trends. – Princeton, NJ : Springer, 2022.

10. Машинное обучение в поиске — URL: https://habr.com/ru/companies/reksoft/articles/865452/ (дата обращения 13.04.2025).

11. Aittokumpu M. Organisaation käyttämien kyberturvallisuusjärjestelmien etsiminen OSINT-tiedustelun kautta. – 2024.

12. Huete Trujillo D. L., Ruiz-Martínez A. Tor hidden services: A systematic literature review //Journal of Cybersecurity and Privacy. – 2021. – Т. 1. – №. 3. – С. 496-518.

13. Darkus is a Onion websites searcher — URL: https://github.com/Lucksi/Darkus (дата обращения 13.04.2025).

14. Pascale D.D., Cascavilla G., Tamburri D. A., Heuvel W. J. V. D.. CRATOR: a Dark Web Crawler //arXiv preprint arXiv:2405.06356. – 2024.

15. Khder M. A. Web scraping or web crawling: State of art, techniques, approaches and application //International Journal of Advances in Soft Computing & Its Applications. – 2021. – Т. 13. – №. 3.

16. Rawat R., Rajawat A.S, Mahor V, Shaw S.N., Ghosh A. Dark web—onion hidden service discovery and crawling for profiling morphing, unstructured crime and vulnerabilities prediction //Innovations in electrical and electronic engineering: proceedings of ICEEE 2021. – Springer Singapore, 2021. – С. 717-734.

17. Stafeev A., Pellegrino G. {SoK}: State of the Krawlers–Evaluating the Effectiveness of Crawling Algorithms for Web Security Measurements //33rd USENIX Security Symposium (USENIX Security 24). – 2024. – С. 719-737.

18. Sharma A. K., Shrivastava V., Singh H. Experimental performance analysis of web crawlers using single and Multi-Threaded web crawling and indexing algorithm for the application of smart web contents //Materials Today: Proceedings. – 2021. – Т. 37. – С. 1403-1408.

19. Yang D., Thiengburanathum P. Scalability and robustness testing for open source web crawlers //2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. – IEEE, 2021. – С. 197-201.

20. Research Tools— URL: https://research.torproject.org/tools/ (дата обращения 13.04.2025).

21. Еще меньше данных и больше смысла: как ещё можно оптимизировать затраты на мониторинг? — URL: https://habr.com/ru/articles/884058/ (дата обращения 13.04.2025).

22. Xu Y. et al. Research on dark web monitoring crawler based on TOR //2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). – IEEE, 2021. – Т. 2. – С. 197-202.

23. Bergman J., Popov O. B. Exploring dark web crawlers: a systematic literature review of dark web crawlers and their implementation //IEEE Access. – 2023. – Т. 11. – С. 35914-35933.