УДК 004.056

М. Карапетьянц[1]*

Исследование методов сбора адресов скрытых сервисов сети TOR

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

Аннотация. Сеть Tor, обеспечивающая анонимность и конфиденциальность, стала популярной среди пользователей, но также привлекла внимание преступных субъектов для организации незаконной деятельности. Правоохранительные органы заинтересованы в мониторинге скрытых сервисов для сбора информации и прогнозирования угроз информационной безопасности. Целью данной работы является исследование методов сбора адресов скрытых сервисов. В работе рассматриваются существующие методы сбора адресов, включая ручной сбор, сканирование сети и использование публичных списков. Исследование показывает, что инструменты для сбора адресов скрытых сервисов можно разделить на несколько категорий: поисковые утилиты, инструменты для получения ссылок, инструменты для сканирования и краулеры. На основе проведенного анализа инструментов работы со скрытыми сервисами Tor можно сделать вывод, что для разработки эффективного модуля сбора адресов скрытых сервисов требуется комплексный подход, включающий сбор, обработку, проверку, агрегацию и анализ полученных данных. Дальнейшие исследования могут быть направлены на создание комплексного решения, объединяющего преимущества различных методов, и улучшение качества и достоверности собранных данных. Исследование может быть полезно для правоохранительных органов и специалистов по информационной безопасности, стремящихся улучшить мониторинг и анализ скрытых сервисов для повышения безопасности в киберпространстве.

Ключевые слова: оверлейные сети, TOR, скрытый сервис, OSINT

M. Karapetyants[1]*

Study of Methods of Collecting Addresses of Hidden Services of TOR Network

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

Abstract. The Tor network, which provides anonymity and privacy, has become popular among users, but has also attracted the attention of criminal actors for organizing illegal activities. Law enforcement agencies are interested in monitoring hidden services to collect information and predict threats to information security. The purpose of this paper is to investigate the methods of collecting addresses of hidden services. The paper reviews existing address collection methods, including manual collection, network scanning, and the use of public lists. The study shows that tools for collecting addresses of hidden services can be categorized into several categories: search utilities, link acquisition tools, scanning tools, and crawlers. Based on the analysis of tools for working with Tor hidden services, we can conclude that the development of an effective module for collecting addresses of hidden services requires a comprehensive approach that includes collection, processing, verification, aggregation and analysis of the obtained data. Further research can be aimed at creating a comprehensive solution combining the advantages of different methods and improving the quality and reliability of the collected data. The research could be useful for law enforcement and information security professionals seeking to improve the monitoring and analysis of covert services to enhance security in cyberspace.

Keywords: overlay networks, TOR, hidden service, OSINT

Введение

Сеть Tor стала широко известной среди пользователей за счет своей способности обеспечивать безопасность и конфиденциальность передаваемых данных. Одним из ключевых аспектов сети Tor являются скрытые сервисы, которые позволяют пользователям публиковать и получать доступ к ресурсам, сохраняя при этом свою анонимность [1]. Защита конфиденциальности, обеспечиваемая Tor, используется преступными субъектами, организующими незаконную деятельность в луковых сервисах. Это представляет собой значительную проблему для правоохранительных органов, которые заинтересованы в мониторинге луковых сервисов для сбора информации и прогнозирования угроз. В связи с этим возникает необходимость в исследовании особенностей первичного сбора адресов скрытых сервисов и существующих методов сбора информации для разработки эффективных решений, включающих методы OSINT [2].

Существующие методы сбора адресов скрытых сервисов Tor включают в себя ручной сбор, сканирование сети и использование публичных списков адресов. Однако каждый из этих подходов имеет свои ограничения. Ручной сбор адресов может быть трудоемким и неэффективным, поскольку он зависит от человеческого участия и может не охватить все скрытые сервисы [3]. Сканирование сети, хотя и является автоматизированным процессом, также имеет свои недостатки, поскольку оно может быть неполным и не обнаружить все активные скрытые сервисы. Кроме того, публичные списки адресов могут содержать устаревшую, неполную или недоступную информацию, что ограничивает их полезность [4].

Цель данной работы – исследовать методы сбора луковых адресов скрытых сервисов сети TOR и оценить их применимость для последующей разработки эффективного решения.

Теоретическая и практическая значимость данного исследования заключается в оценке применимости существующих методов сбора адресов скрытых сервисов и разработке нового подхода, основанного на использовании методов OSINT. Результаты данного исследования могут быть использованы правоохранительными органами и специалистами по информационной безопасности для улучшения мониторинга и анализа луковых сервисов, что позволит повысить безопасность в киберпространстве [5].

Для решения проблематики, связанной с аналитическими инструментами, научным сообществом разрабатывается децентрализованный подход по сбору и анализу адресов скрытых сервисов в оверлейных сетях. В данной области проводят исследования следующие ученые: Buitrago Lopez A., Pastor-Galindo J., Gomez Marmol F., Perez G. M., Holler T., Roland M., Mayrhofer R, Basynya E. [2, 6–8].

Методы и материалы

Исследователи в области информационной безопасности уделяют значительное внимание изучению луковых сервисов и сети Tor с целью оценки их без-

опасности, демонстрации потенциальных кибератак и классификации доступных ресурсов. Методы деанонимизации, направленные на раскрытие личностей пользователей в сети Tor, являются активной областью исследований, и луковые сервисы играют ключевую роль в изучении этих атак и защите конфиденциальности пользователей. Исследователи [7] показали, что около 45 % луковых сервисов были связаны с такой незаконной деятельностью, как хакерские услуги, черные рынки и другие. Однако стоит отметить, что луковые сервисы также предлагают легитимный и этичный контент, такой как обмен файлами и дискуссионные форумы, которые составляют около 47 % от общего числа сервисов.

Временные характеристики луковых сервисов имеют важное значение для правоохранительных органов и оперативных задач. Исследователи [9] показали, что луковые сервисы имеют тенденцию к непостоянству и кратковременности. Примерно 50 % наблюдаемых луковых сервисов отключились через 24 часа, а через 300 часов эта цифра сократилась до 40 %, указывая на высокую степень непостоянства долгоживущих луковых сервисов. Только 36% выявленных луковых сервисов оставались активными в течение 18-недельного временного окна. Кроме того, луковые сервисы часто меняют адреса, и около 20% из них имеют несколько URL-адресов.

В связи с отсутствием стандартного подхода к сбору адресов луковых сервисов необходимо проанализировать существующие методы сбора адресов скрытых сервисов.

Исследователи выделяют пять основных методов сбора адресов скрытых сервисов. Три из них – это методы поверхностного уровня, которые собирают адреса в открытом интернете, включая луковые поисковые системы, репозитории и общие поисковые системы. Оставшиеся два метода – это механизмы глубокого уровня, которые применяются в анонимных сетях, в частности, сканирование через Tor и ретрансляционные инъекции [2].

1. Луковые поисковые системы. Сложность поиска адресов в анонимной сети привела к созданию поисковых систем, специально разработанных для Tor и облегчающих пользователям поиск нужных ссылок. Эти системы используют автоматические пауки для постоянного индексирования и обновления своих внутренних баз данных новыми ссылками, а также возвращают результаты, основанные на определенных алгоритмах и методах ранжирования. Хотя этот метод доступен и широко используется, он может приводить к необъективным выборкам в зависимости от поисковых запросов, ранжирования и эффективности фоновых пауков. Примерами таких инициатив являются Onion City, Onion Live, Tor Link и Ahmia, причем Ahmia используется в основном как хранилище ссылок на темный интернет, а не как поисковая система [10].

2. Репозитории. Некоторые ресурсы специализируются на составлении списков адресов скрытых сервисов в централизованном сервисе. Хотя этот метод обеспечивает прямой способ получения ресурсов, пользователь должен знать об их существовании и полагаться на потенциально необъективные материалы от неизвестных авторов. Кроме того, часто встречаются устаревшие репозитории или списки с недоступными веб-сайтами, что приводит к значительной доле не-

активных сервисов темной паутины. Наиболее известные и используемые инструменты включают The Hidden Wiki и Ahmia, в то время как другие работы обнаруживают адреса темной паутины в таких подборках, как Deep Web Links, Dark Web Links, Reddit, проект открытых данных Tor2Web [11].

3. Общие поисковые системы. Традиционные поисковые системы также могут быть полезны для выявления адресов темного веба в поверхностном вебе. Этот метод интуитивно понятен любому пользователю или аналитику, но он требует тщательного подбора ключевых слов, чтобы избежать нерелевантных вебстраниц без адресов скрытых сервисов. Также требуется извлечение и обработка компонентов для получения чистого списка ссылок Tor из возвращаемого контента. И, наконец, результаты могут быть смещены из-за выбора ключевых слов, внутренних алгоритмов ранжирования и поиска. Помимо широко распространенного Google, в рассмотренных работах использовались также такие поисковые системы, как Bing, DuckDuckGo, Yahoo и Baidu [12].

4. Сканирование TOR. Техника сканирования Tor основана на автоматическом исследовании сети Tor, рекурсивном следовании по ссылкам и сохранении идентифицированных адресов скрытых сервисов. Эта техника требует тщательного проектирования и реализации автоматических пауков для навигации по сети Tor. Процесс должен быть правильно настроен, чтобы избежать утечек памяти и обрабатывать ошибки связи. Результаты будут зависеть от исходного списка и могут содержать только те скрытые сервисы, на которые есть входящие ссылки. Тем не менее, результаты, полученные с помощью сканирования Tor, вероятнее всего, будут активными, хотя их производительность сильно зависит от конфигурации и ресурсов [6].

5. Релейный ввод. В сети Tor любой пользователь может развернуть добровольный релейный узел, чтобы усилить инфраструктуру анонимности и улучшить скорость, стабильность и безопасность сети. Исследователи могут использовать такие узлы для получения адресов скрытых сервисов в версии Tor v2. Однако в версии 3 этот метод уже невозможен из-за использования слепых подписей [13].

Исследование

В рамках исследования методов сбора адресов скрытых сервисов сети Tor были изучены различные инструменты и методы, используемые в области OSINT (Open-Source Intelligence). Исследование показало, что существуют разные подходы к сбору данных в скрытых сервисах, и инструменты можно разделить на несколько категорий в зависимости от их специализации и целей.

Поисковые утилиты, такие как OnionSearch, Darkdump, Ahmia Search Engine, DarkSearch и Katana, играют важную роль в исследовании скрытых сервисов сети Tor. Эти инструменты выполняют функции, аналогичные традиционным поисковым системам, обеспечивая автоматизированный поиск веб-сайтов по ключевым словам. Несмотря на свою эффективность в выполнении задач поиска, эти утилиты не обладают универсальными функциями, которые присущи другим категориям инструментов, используемых в сфере OSINT. При этом про-

изводимый по ключевым словам поиск может быть нерелевантным выданному результату, что говорит о необходимости повышения точности поиска информации. Некоторые из популярных решений в этой категории включают: OnionSearch, Darkdump, Ahmia Search Engine, DarkSearch и Katana [14].

Инструменты, такие как Hunchly, H-Indexer и Tor66 Fresh Onions, специализируются на сборе и анализе ссылок со скрытых сервисов сети Tor. Эти инструменты, в основном, автоматизируют процесс сбора ссылок, значительно упрощая задачу исследования. Кроме того, они предоставляют статистические данные, которые могут помочь исследователю получить общее представление о сети Tor. Например, H-Indexer индексирует скрытые сервисы и предоставляет список недавно созданных и обновленных адресов, а также различные статистические данные. В свою очередь, Tor66 Fresh Onions предоставляет список недавно созданных скрытых сервисов, обновляемый в режиме реального времени. Проблема использования существующих инструментов заключается в низкой скорости обработки и поиска информации, что затрудняет сбор адресов скрытых сервисов.

В рамках исследования скрытых сервисов сети Tor инструменты для сканирования, такие как Onionscan, Onioff и Onion-nmap, играют важную роль, позволяя проводить анализ уязвимостей и потенциальных угроз. Эти инструменты имеют функциональность, дающую возможность обнаруживать проблемы в настройках сервиса и собирать информацию о сервисах. Например, Onionscan – это инструмент для сканирования скрытых сервисов на предмет уязвимостей и конфигурационных ошибок, что помогает выявить потенциальные проблемы. Onion-nmap, адаптированная для работы с сетью Tor версия популярного инструмента для сканирования сетей Nmap, позволяет сканировать скрытые сервисы на предмет открытых портов. Главным недостатком является невысокая точность определения статуса веб-сайта, особенно если ресурс использует сложные механизмы для обеспечения анонимности.

Краулеры (или веб-скраперы) позволяют автоматически собирать данные со скрытых сервисов, включая ссылки, тексты и другие метаданные. Они могут быть полезны для создания баз данных или для сбора информации о конкретных сервисах, а точнее ссылок, текстов и других метаданных. В качестве популярных инструментов можно выделить TorBot, TorCrawl, VigilantOnion, OnionIngestor. Основная проблема существующих решений заключается в отсутствии сбора и обработки изображений, что возможно при использовании моделей машинного обучения [15].

На основе проведенного анализа, можно сделать вывод, что каждый из рассмотренных инструментов эффективно работает в своей определенной области. Поисковые утилиты, инструменты для получения onion ссылок, инструменты для сканирования и краулеры – все они являются важными и незаменимыми в своем контексте использования.

Однако, на текущий момент не существует единого инструмента, который бы мог совмещать в себе все эти функции. Поэтому для проведения исследования скрытых сервисов сети Tor, приходится использовать несколько инструментов параллельно. Такой подход может приводить к увеличению времени и усилий, затрачиваемых на процесс исследования, и к потере ценной информации из-за недостатка централизации данных [16].

Таким образом, обнаруживается значительная потребность в разработке универсального комплексного инструмента, который бы обеспечивал сбор, обработку, проверку, агрегацию и анализ адресов скрытых сервисов сети Tor в одном месте. Такой инструмент значительно улучшил бы процесс исследования, увеличил бы его эффективность и позволил бы получать более полные и точные результаты.

Обсуждение и заключение

Основываясь на проведенном исследовании, можно сделать ряд выводов и предположений.

Все рассмотренные методы сбора адресов скрытых сервисов сети Tor обладают различными функциональными возможностями и подходами к решению задачи.

Существующие методы сбора адресов Tor включают в себя ручной сбор, сканирование сети, использование публичных списков, обычные и луковые поисковые системы, а также релейный ввод. Однако каждый из этих методов имеет свои ограничения, такие как трудоемкость, неполнота данных или зависимость от внешних факторов.

Инструменты для сбора адресов скрытых сервисов можно разделить на несколько категорий: поисковые утилиты, инструменты для получения ссылок, инструменты для сканирования и краулеры. Каждый тип инструментов имеет свои преимущества, но основными недостатками являются низкая скорость и точность сбора и обработки данных. Рассмотренные решения применимы в своей определенной области по взаимодействию со скрытыми сервисами Tor, но для разработки эффективного модуля сбора адресов скрытых сервисов требуется комплексное решение, включающее в себя сбор, обработку, проверку, агрегацию и анализ полученных данных.

Результаты исследования подтверждают важность разработки гибких и масштабируемых решений для сбора адресов скрытых сервисов, которые могут адаптироваться к различным задачам и обеспечивать достоверные и актуальные данные.

Дальнейшие исследования в этой области могут быть направлены на разработку комплексного решения для сбора адресов скрытых сервисов, которое объединит преимущества различных методов и инструментов и при этом позволит улучшить качество и достоверность собранных данных.

Кроме того, важным аспектом дальнейших исследований может стать исследование методов анализа и классификации данных из скрытых сервисов для выявления незаконной деятельности и улучшения безопасности в киберпространстве.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Дутта Н., Джадав Н., Танвар С., Сарма Х.К.Д., Прикоп Е. TOR – луковый маршрутизатор // Cyber Security: Issues and Current Trends. – 2022. – С. 37-55.

2. Пастор-Галиндо Х., Мармол Ф. Г., Перес Г. М. О сборе адресов тор-лука // Future Generation Computer Systems. – 2023. – Т. 145. – С. 12-26.

3. Хуете Трухильо Д. Л., Руис-Мартинес А. Скрытые сервисы Tor: Систематический обзор литературы //Journal of Cybersecurity and Privacy. – 2021. – Т. 1. – №. 3. – С. 496-518.

4. Штайнебах М., Шефер М., Каракуз А., Брандл К. и Янникос Ю. Обнаружение и анализ луковых сервисов tor // Proceedings of the 14th International Conference on Availability, Reliability and Security. – 2019. – С. 1-10.

5. Басыня Е. А., Малышев Е. А. Обеспечение достоверности результатов научно-практических изысканий с применением программной инженерии // Защита информации. Инсайд. – 2023. – Т. 112 – №. 4. – С. 14-21.

6. Хёллер Т., Роланд М., Майрхофер Р. О состоянии луковых сервисов V3 // Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet. – 2021. – С. 50-56.

7. Буитраго Лопес А., Пастор-Галиндо Х., Гомес Мармол Ф. Обновленное исследование сети Tor: реклама, доступность и протоколы луковых сервисов // Wireless Networks. – 2024. – С. 1-15.

8. Басыня Е. А., Карапетянц Н., Карапетянц М., Когос К. Г. Модуль для сбора, обработки и агрегирования данных системы анализа транзакций в сети Bitcoin // 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE). – IEEE, 2023. – С. 980-986.

9. Забихимайван М., Доран Д. Первый взгляд на ссылки из темного в поверхностный веб-мир: исследование на примере Tor // International Journal of Information Security. – 2022. – Т. 21. – №. 4. – С. 739-755.

10. Бергман Я., Попов О. Б. Распознавание вредоносных программ tor и луковых сервисов // Journal of Computer Virology and Hacking Techniques. – 2023. – Т. 20. – С. 1-15.

11. Bian J., Cao C., Wang L., Ye J., Zhao Y., Tang C. Обнаружение и анализ скрытых сервисов Tor: обзор литературы // Journal of Physics: Conference Series. – IOP Publishing, 2021. – Т. 1757. – №. 1. – С. 012162.

12. Пастор-Галиндо Х., Несполи П., Мармол Ф. Г., Перес Г. М. Еще не освоенная золотая жила OSINT: возможности, открытые проблемы и будущие тенденции // IEEE Access. – 2020. – Т. 8. – С. 10282-10304.

13. Чжан З., Чжоу В., Шерр М. Обход блокировки выхода из тор с помощью услуг exit bridge onion // Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. – 2020. – С. 3-16.

14. Bernaschi M., Celestini A., Cianfriglia M., Guarino S., Lombardi F., Mastrostefano E. Лук под микроскопом: Глубокий анализ Tor Web //World Wide Web. – 2022. – Т. 25. – №. 3. – С. 1287-1313

15. Наир В., Каннимула Дж. М. Инструмент для извлечения луковых ссылок из скрытых сервисов Tor и выявления незаконной деятельности //Inventive Computation and Information Technologies: Proceedings of ICICIT 2021. – Singapore : Springer Nature Singapore, 2022. – С. 29-37.

16. Стрельцов А. С., Французова Г. А., Басыня Е. А. Разработка системы сбора, обработки, анализа, идентификации корреляции событий информационной инфраструктуры предприятия // Системы анализа и обработки данных. – 2023. – №. 1 (89). – С. 101-113.