ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ВАСИЛЯ СТУСА

КАФЕДРА РАДІОФІЗИКИ ТА КІБЕРБЕЗПЕКИ

**КУРСОВА РОБОТА**

з дисципліни «Стандартизація та сертифікація в галузі інформаційної безпеки»

на тему:

Методи деанонімізації при використанні TOR технологій

Студента 3 курсу

спеціальності/напряму підготовки

б17\_д/125А Кібербезрека

Рудого Максима Вікторовича

Науковий керівник:

к.т.н. Барибін О.І.

Національна шкала \_\_\_\_\_\_\_\_\_\_\_\_

Кількість балів \_\_\_\_\_\_\_\_\_\_\_

Шкала ECTS \_\_\_\_\_\_\_\_\_\_\_

Члени комісії:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (ПІБ)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (ПІБ)

Вінниця 2020

**АНОТАЦІЯ**

Приховані сервіси Tor використовуються для надання послуг на основі TCP користувачам, при цьому не розкриваючи IP-адресу прихованого сервера, щоб отримати скритність і анти цензуру. Однак приховані сервіси використовуються різними способами. Приховані сервіси використовуються для незаконних цілей, таких як торгівля наркотиками, зброєю, розповсюдження порнографії і т. д. Через неправильне використання прихованих сервісів Tor, досліди для виявлення цих сервісів були неминучі.

У цій статті буде огляд існуючих атакуючих схем, їх порівняння, пояснення ключових ідей і показані їх взаємовідносини.

**ВСТУП**

Tor - це система анонімних комунікацій з малою затримкою, яка підтримує TCP-додатки через Інтернет, які надають користувачам такі послуги, як перегляд веб-сторінок, безпечна оболонка і обмін миттєвими повідомленнями. Він надає користувачам сервіс анонімності, допомагає обійти інтернет-цензуру і підтримує приховані сервіси для збереження анонімності веб-сервісів. В даний час Tor підтримує кілька мільйонів користувачів. Tor підтримує приховані сервіси, щоб зберегти непомітність цих веб-сервісів.

Приховані сервіси - це веб-сайти, розташовані всередині мереж Tor, які отримують вхідні повідомлення тільки через Tor. Tor необхідний для доступу до прихованим сервісів. Доступ до прихованих сервісів здійснюється через його цибульну адресу, що дозволяє користувачеві публікувати інтернет-сайти, приховуючи свою IP-адресу і місце розташування.

Однак приховані служби використовуються для незаконних сервісів, таких як інформація про торгівлю наркотиками, сайти дитячої порнографії, незаконна торгівля зброєю і т. д. На прихованих сервісах Tor існує багато чорних ринків. Шовковий шлях – був один з них, він широко займався контрабандою товарів з використанням біткоїнів в якості валюти.

Наразі існує безліч різних варіантів атак, що розкривають приховані служби Tor. Розкриваючи ці приховані сервіси з’явиться змога закрити ті приховані служби які розміщують нелегальний контент. Таким чином, у цій роботі ми зосередимось на існуючих схемах атак, їх порівнянні та принципах роботи. Порівнюючи їх, можна буде зрозуміти плюси і мінуси кожної схеми, і, таким чином, можна буде застосувати відповідну схему для деанонімізації прихованої служби.

**РОЗДІЛ 1. ЗАГАЛЬНІ ВІДОМОСТІ ПРО ТЕХНОЛОГІЮ TOR**

* 1. **Мережа Tor**

Tor – це оверлейна мережа на рівні додатків, яка дозволяє здійснювати анонімний обмін даними між її клієнтами та довільними інтернет – адресами з допомогою цибулевої маршрутизації. Клієнти здійснюють анонімний зв’язок із сервером, тунелюючи свій трафік через ланцюжок із трьох ретрансляторів Tor. У цьому розділі ми спочатку ознайомимось із мережею Tor, а потім ознайомимось із її базовими операціями та протоколами прихованих сервісів.

1. **Огляд мережі**

Тор – це оверлейна мережа для анонімного зв’язку, в якій кожен цибулевий маршрутизатор (ЦМ) працює як звичайний процес на рівні користувача без будь – яких особливих привілеїв. Це проект із відкритим вихідним кодом який забезпечує сервіс анонімності для додатків TCP. Кожний (ЦМ) підтримує TLS з’єднання із кожним (ЦМ). Кожний користувач запускає локальне програмне забезпечення яке називається onion proxy (OP), для вибору каталогів, встановлення каналів в мережі і опрацювання з’єднань із клієнтськими додатками. Ці (OP) приймають TCP потоки і мультиплесують їх по каналам. (ЦМ) на іншій стороні ланцюга з’єднуються із запрошеними адресами і передають дані. Рисунок 1.2.1 ілюструє базову архітектуру мережі Tor. Наступні компоненти беруть участь в типовому використані мережі Tor.



Рисунок 1.2.1

* Tor клієнти: Клієнт Tor запитує дані для завантаження з сервера. Він встановлює локальне програмне забезпечення, onion proxy (OP), в якому дані додатки упаковуються в осередок однакового розміру (512 байт) і доставляють їх у мережу Tor. Осередок - це основний блок передачі Tor.
* Цибулеві маршрутизатори (onion router): OR - це реле, добровільно запропоноване різними волонтерами по всьому світу.
* Сервери каталогів: Сервери каталогів містять інформацію про операції OR і прихованих службах, таких як відкриті ключі маршрутизаторів і прихованих серверів.
* Сервери додатків: Він підтримує програми TCP, такі як веб-служба та служба IRC.

1. **Приховані сервіси**

Tor забезпечує анонімність для веб-сайтів та інших серверів. Сервери, призначені для прийому вхідних з'єднань тільки через Tor, називаються прихованими сервісами. Замість того, щоб розкривати IP-адресу сервера, до прихованого сервісу звертаються через його цибульний адрес, який є хешем його відкритого ключа. Прихований сервіс Tor складається з наступних важливих вузлів:

* Introduction Point (IP): Це реле Tor, які вибираються прихованими сервісами і використовуються для встановлення зв'язку з клієнтом.
* Rendezvous Poin (RPO): вони вибираються клієнтом Tor і використовуються для пересилання даних між клієнтом і прихованим сервером.
* Directory Server (DS): DS володіє інформацію про вузли мережі Tor, і ця інформація використовується для зв'язку з прихованими сервісами.
* Hidden server (HS): це сервер всередині мережі Tor, на якому розміщені приховані сервіси.

Принцип роботи прихованої служби заснований на підключенні двох ланцюгів, одного створеного клієнтом, а іншого - HS, на загально узгодженому реле Tor. Це реле Tor, таким чином, називається Rendezvous Poin, яка передає осередок шляхом пересилання вихідних повідомлень із схеми на стороні клієнта до ланцюга на стороні сервера та навпаки. Щоб захистити RPO від атак, прихована служба вибирає групу реле Tor як точки введення, які працюють так само, як RPO. IP-адреси використовуються виключно для передачі одного повідомлення, що містить розміщення обраного RPO. Що стосується лише прийняття запитів клієнта, прихована служба публікує прихований дескриптор служби, що містить підписаний список IP-адрес для DS, звідки клієнти можуть їх завантажувати.

Нормальне налаштування зв'язку між клієнтом та конкретним HS показано на рисунку 1.3.1. Процедура цього повідомлення полягає в наступному:

1. HS спочатку вибирає декілька OR в якості IP і будує схеми до цих IP, надсилаючи осередок RELAY\_COMMAND\_ESTABLISH\_INTRO, а IP у відповідь відповідає осередоком RELAY\_COMMAND\_INTRO\_ESTABLISHED, щоб повідомити HS про те, що схема встановлена.
2. Як тільки схеми до IP встановлені, HS встановлює ланцюг до DS та оголошує дескриптор служби DS, включаючи відкритий ключ HS та інформацію щодо IP. Тоді власник HS може розмістити цибулеву адресу у громадському місці, щоб залучити користувачів до доступу до прихованого сервісу через Tor.
3. Коли клієнт Tor отримує цибулеву адресу, клієнт створює схему DS і отримує відповідну інформацію, оголошену прихованою службою. Потім клієнт дізнається IP прихованої служби.
4. Тепер клієнт вибирає RPO і створює схему для RPO. Клієнт відправить осередок RELAY\_COMMAND\_ESTA-BLISH\_RENDEZVOUS, який містить cookie-файл рандеву, і RPO який відповідає осередоком RELAY\_COM-MAND\_RENDEZVOUS\_ESTABLISHED, щоб підтвердити успішне встановлення каналу.
5. Клієнт створює схему трьох стрибків для одного з IP і передає осередок RELAY\_COMMAND\_INTRODUCE1 на обраний IP. В команді зберігається така інформація, як RPO, cookie-файли рандеву і дані Діффі-Хеллмана які генерується клієнтом Tor.
6. Як тільки IP отримує осередок RELAY\_COMMAND\_INTRODUCE1, він відповідає клієнтові RELAY\_COMMAND\_INTRODUCE\_ACK. Після того, як клієнт отримає цей осередок ACK, він розриває цю схему до IP.
7. IP перепаковує осередок RELAY\_COMMAND\_INTRODUCE1 у осередок RELAY\_COMMAND\_INTRODUCE2, а потім відправляє осередок RELAY\_COMMAND\_INTRODUCE2 в HS. Як тільки HS отримує цей осередок, він знає інформацію про RPO, файли cookie та дані Діффі-Хеллмана . HS може генерувати дані Діффі-Хеллмана і отримати ключ K = .
8. Тоді HS будує ланцюг до RPO та посилає в RPO осередок RELAY\_COMMAND\_RENDEZVOUS1 до RPO. Коли RPO отримує цей осередок, він порівнює переглянуті cookie з осередком та те, що йде від клієнта Tor. Після того, як файли cookie, що переглядаються, збігаються, RPO видаляє файли cookie, що переглядаються, з осередоку RELAY\_COMMAND\_ RENDEZVOUS1 і перепаковує решту даних у осередок RELAY\_COMMAND\_RENDEZVOUS2, а потім пересилає осередок клієнту.
9. Коли клієнт Tor отримує осередок RELAY\_COMMAND\_RENDEZVOUS2, він може генерувати ключ K = ., використовуючи , і перевірити його на основі H (K). Таким чином клієнт та HS завершують рукостискання. Потім клієнт відправляє осередок RELAY\_COMMAND\_BEGIN для встановлення потоку в HS через шість стрибків.
10. Коли клієнт відправляє осередок RELAY\_COMMAND\_BEGIN до HS, щоб відкрити потік між клієнтом і HS, наш контрольований RPO виявляє цей спеціальний осередок на основі прихованого протоколу обслуговування

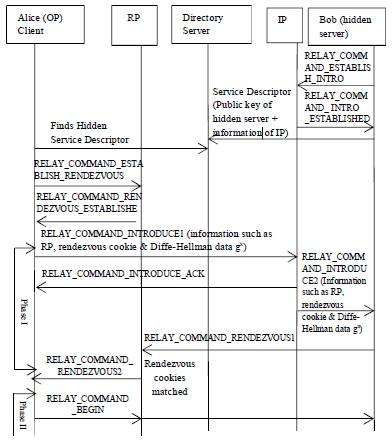


Рисунок 1.3.1

**РОЗДІЛ 2. СХЕМИ НАПАДУ**

Прихований сервіс пропонує комунікатору непомітність для постійних послуг. Користувачі можуть підключитися до сервісу через Tor, не знаючи його місцезнаходження. Ці приховані сервіси можуть розміщувати незаконні веб-сайти, наслідки яких можуть бути серйозними. Тож ці HS потрібно розкрити. Деанонімізація прихованої служби - це стратегія, в якій розкривається ip-адреса анонімних веб-сайтів. Один із способів деанонімізувати ці приховані служби - це напад на них. До цього часу дослідники розробили безліч різних схем атаки, про які йде мова нижче.

У цих атакуючих схем є деякі загальні риси:

* Усі атакуючі схеми потребують щоб клієнт – зловмисник і вузол захисту від зловмисника анонімізували прихований сервіс
* У всіх атакуючих схемах приховані служби змушені вибирати скомпрометовані охоронні вузли як свої вхідні вузли.
* Усі атакуючі схеми виявляють IP-адресу прихованої служби.

1. **Маніпуляції з осередками Tor**

Цей метод заснований на підході виявлення на рівні протоколу. Зловмисник управляє клієнтом Tor, RPO, кількома вхідними OR та центральним сервером. За допомогою цього методу легко розгорнути систему виявлення.

Порівняно з методами, заснованими на аналізі трафіку, цей підхід є значно швидшим, повністю автоматичним і може швидко знайти HS, використовуючи лише кілька осередків.

Цей підхід є точним зі спостережуваною частотою виявлення 100% та має низьку похибку 0%. Він працює на рівні протоколу і не звертає уваги на схеми трафіку, тобто він є більш загальним і може використовуватися для виявлення шкідливих прихованих служб.

У цьому методі процес виявлення HS поділяється на три фази. У фазі 1 клієнт, імовірно, ідентифікує HS, як показано на рисунку 1.3.1. Після ідентифікації HS, у фазі 2 перевіряється HS. Наприкінці 1 фази клієнт відправляє осередок RELAY\_COMMAND\_BEGIN в HS. Коли RPO ловить цей осередок, він змінює один біт комірки і пересилає її до HS. Модифікація комірки в RPO показана на рисунку 3.1.1. RPO також повинен надіслати часову мітку керованої комірки на центральний сервер. Через відсутність перевірки цілісності інші OR не можуть виявити керовану комірку.

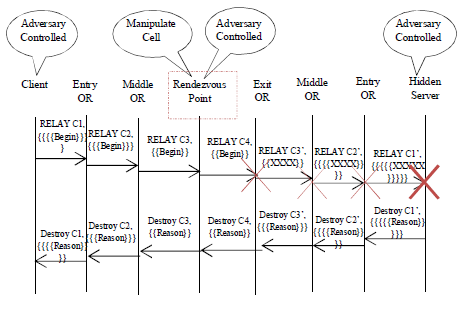


Рисунок 2.1.1

Коли маніпульована комірка досягла HS, HS не може розпізнати цю клітинку, і ланцюг між клієнтом і HS розірветься, відправивши клітинку CELL\_Destroy. Контрольований запис OR спочатку отримує цю клітинку, і вона повідомляє про тип комірки, часову мітку комірки, ідентифікатор схеми та IP-адресу джерела комірки на центральний сервер. Знищити клітинку, яка досягає клієнта через RPO. У RPO потрібно повідомити центральний сервер про часову мітку цієї комірки. Нарешті, за часовою кореляцією ми можемо знайти IP-адресу HS.

У цьому способі прихована служба повинна вибрати вузол введення противника як свій захисний вузол. Щоб це сталося, кількість компрометованих вузлів входу повинна бути дуже великою, майже 30% від всіх вхідних маршрутизаторів Tor, що з фінансової сторони дуже дорого.

1. **Метод підрахунку осередків**

У цьому методі довільні реле Tor відключаються анонімно та вибірково з дуже низькими витратами для нападника, ця атака ефективна настільки що зловмисник, зацікавлений в цензурі може вимкнути Tor замість того щоб блокувати Tor простим вимкненням всіх реле або лише орієнтуючись на ті що забезпечують найбільшу пропускну здатність чи найважливіші служби каталогів. Ця атака також відома як снайперська атака, вона класифікується як атака з підрахунком осередків, оскільки вона деанонімізує прихований сервіс, підраховуючи кількість осередків в противника.

У цьому методі будуть деанонімізовані приховані служби, вибірковим відключенням реле за допомогою снайперської атаки, сильно впливаючи на шляхи до тих, хто контролює противника. Таким чином, ця атака створює реальні, значні загрози для користувачів Tor, і це є найбільш руйнівною атакою проти мережі Tor на сьогодні.

1. **Снайперська атака**

Снайперська атака виконується з метою вбити будь-яке довільне реле. На рисунку 3.2.1.1 показані етапи цієї снайперської атаки:

* Клієнт створює схему, використовуючи ціль як запис.
* Клієнт запитує дані з сервера, а сервер відповідає.
* Клієнт перестає читати дані з цільового запису.
* Клієнт періодично надсилає пакет SENDME на сервер, проте він не може прочитати цільові записи з TCP з’єднання.
* Цільовий запис буферизує дані до тих пір, поки ОС не припинить процес Tor

Оскільки цей метод вимагає, щоб цільовий вузол входу прихованої служби припинявся буферизацією всієї вільної пам'яті, для системи з великою кількістю вільної пам'яті може знадобитися багато часу.

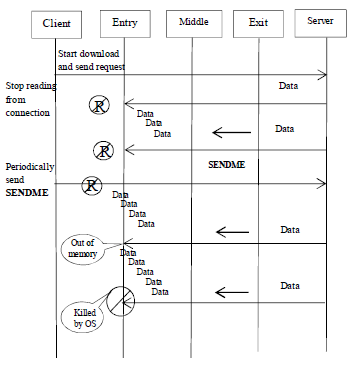


Рисунок 2.2.1.1

1. **.2 Деанонімізація прихованих сервісів**

Блок-схема процесу деанонімізації прихованих служб показана на рисунку 3.2.2.1

* Перш за все клієнт – зловмисник змушує HS будувати нові канали рандеву для того щоб визначити його захисну систему.
* Примусово змусити HS змінити захисну систему.
* Повторювати поки HS не обере ворожу систему

Спочатку захисник противника запускає одне захисне реле (). Для деанонімізації прихованого сервісу, HS повинен вибрати одне з ворожих реле як захисний вузол. Для цього HS будує нову схему в RP. Зловмисник відправляє 50 нових осередків в HS через . Якщо лічильник осередків в захисному ретрансляторі, запущеному клієнтом – противником дорівнює 52, то зловмисник може ідентифікувати вхідний вузол входу HS. Потім на вхідний вузол HS здійснюється снайперська атака, і HS змушений обрати інший вузол. Цей процес повторюється до тих пір поки HS не обере ворожий вузол. Щоб перевірити обрав HS скомпрометований вузол чи ні, противник – клієнт надсилає 50 додаткових осередків в HS. Якщо лічильник осередків в захисті противника дорівнює 53, то атакуючий може ідентифікувати приховану службу.

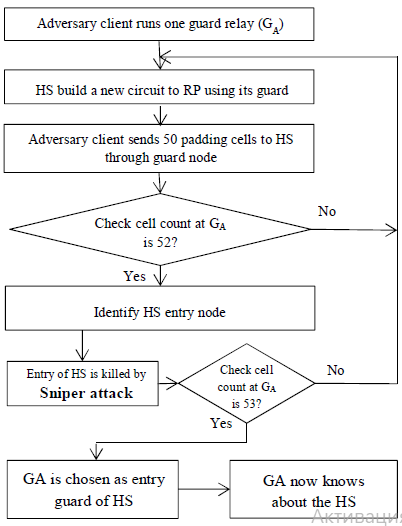


Рисунок 2.2.2.1

1. **Метод заповнення осередків**

У цій атаці прихована служба змушена встановлювати контрольовані зловмисником ланцюги зустрічі до RP. Після отримання осередку RELAY\_COMMAND\_RENDEZVOUS1 з файлом cookie зловмисника, RP генерує трафік зі спеціальною сигнатурою. Ця сигнатура може бути ідентифікована середнім вузлом нападника. Зауважте, що спеціальний механізм осередку PADDING у Tor спрощує генерацію трафіку сигнатури, який відкидається на стороні одержувача, і, таким чином, не помітний для HS.

Кроки атаки як на рисунку 2.3.1

1. Спочатку зловмисник надсилає осередок RELAY\_COMMAND- \_INTRODUCE1 до IP HS із зазначенням адреси RPO.
2. IP передає вміст у осередку RELAY-COMMAND INTRODUCE2 для прихованої служби.
3. HS встановлює трипроменеву схему до RPO, надсилаючи осередок RELAY\_COMMAND\_RENDEZVOUS1.
4. Коли RPO, керований зловмисником, отримує осередок RELAY- \_COMMAND\_RENDEZVOUS1, він надсилає 50 осередків PADDING до HS тим же шляхом.
5. Коли HS отримує цей осередок PADDING, ці осередки HS мовчки скидаються.
6. RPO посилає осередок DESTROY через ланцюг рандеву, що веде до закриття ланцюга.

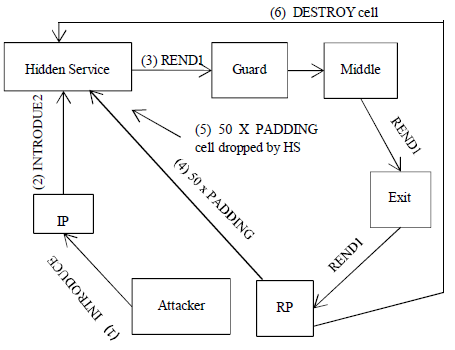


Рисунок 2.3.1

Охоронний вузол зловмисника стежить за переміщенням трафіку на ланцюгах, що проходять через нього. Щоразу, коли він отримує осередок DESTROY по ланцюгу, він перевіряє: (1) чи був осередок отриманий одразу після того, як RPO отримав осередок RELAY\_COMMAND\_REN DEZVOUS1, (2) кількість перенаправлених осередків: 3 осередка вгору по ланцюгу і 53 осередка вниз по ланцюгу. Якщо ці умови виконуються, то зловмисник розуміє, що для прихованої служби був обраний його охоронний вузол.

1. **Website fingerprinting атака**

Ця атака привертає велику увагу дослідників, оскільки дозволяє деанонімізувати користувачів володіючи досить скромними ресурсами, а саме - корумпованим вхідним вузлом (рисунок 2.4.1). WF атака - окремий випадок атак аналізу трафіку, під час яких локальний спостерігач намагається з'ясувати інформацію про контент (відвіданий веб-сайт), оглядаючи потоки трафіку всередині анонімного з'єднання. Атакуючий збирає метаінформацію: розмір пакету, напрямок трафіку. Збір відбувається без порушення шифрування.

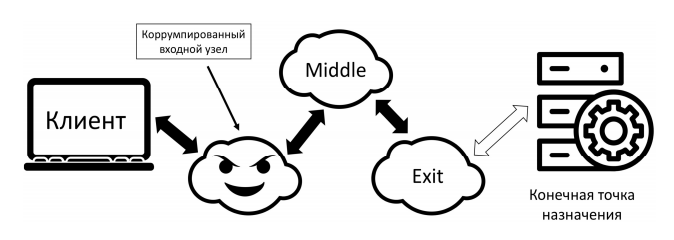


Рисунок 2.4.1

Стратегія атакуючого досить проста: він збирає мережевий трафік через вхідний вузол, а потім посилає його в класифікатор, який навчається на цих даних. Після того, як класифікатор навчився, він готовий до експлуатації. Класифікатор може бути навчений як мультикласовий (розпізнавати сайти), так і в якості бінарного: переглядає користувач сайт зі списку атакуючого або з зовнішнього світу.

Однак, потрібно зробити кілька припущень, за допомогою яких можна спростити атаку:

* Існує тільки k сторінок, які користувач може відвідати. Це дуже сильне припущення, оскільки це число завжди дуже маленьке, в порівнянні з дійсним числом існуючих сторінок.
* Поведінка користувача досить специфічна. Наприклад, він відвідує одну сторінку за однією в одній вкладці браузера.
* Атакуючий може зрозуміти, де початок і кінець завантаження сторінки в послідовності трафіку. Це досить важка задача, враховуючи трафік з реального світу.
* Атакуючий може фільтрувати весь фоновий трафік, який виробляють сторонні додатки або інші сполуки через ланцюг Tor.
* Атакуючий може тренувати класифікатор в тих же умовах, що і жертва. Наприклад, у них повинна бути однакова версія браузера Tor.

Основоположним принципом атаки є використання алгоритмів машинного навчання. Тому постає питання, що використовувати в якості навчальних даних для алгоритму, оскільки до кінця не зрозуміло, що означає «послати трафік в класифікатор». Розглянемо, що ж використовується в якості вхідних даних для алгоритму.



Рисунок 2.4.2

Найбільш важливим компонентом для будь-якого алгоритму машинного навчання є ознаки об'єктів. Дістати ознаки для них можна на різних рівнях: рівень осередків Tor, TLS, TCP. На рівні додатку, Tor передає всі дані в осередках які потім перетворюються в TLS записи як показано на рисунку 2.4.2. Примітно, що багато осередків може бути упаковано в однин TLS запис. І останній - транспортний рівень: TLS записи, як правило, фрагментируются в кілька TCP пакетів, розмір яких обмежений максимальним розміром пакету даних. Крім того, кілька TLS записів можуть бути поміщені в один TCP пакет. Однак, постає дуже важливе питання: який з шарів містить найбільше інформації, з точки зору атаки? Більшість дослідників схиляється до того, що найбільш інформативним є рівень осередків.

Однак, існує ряд проблем, які ускладнюють WF атаки:

* Розмір реального світу дуже великий, тому необхідно збирати гігантську кількість даних, щоб покривати хоча б невелику частину інтернету.
* Оскільки вихідні вузли розкидані по всьому світу, а у більшості сайтів є хороша локалізація, то можна зібрати під однією назвою кілька версій одного і того ж сайту на різних мовах і з різними контентом, що істотно знизить точність класифікатора.
* Точного способу виділяти послідовність пакетів, що стосується тільки одного запиту на сторінці досить проблематично, завдання залишається невирішеним. Якщо користувач не використовує фіксований набір вхідних вузлів, то він може просто припинити використовувати корумпований вхідний вузол

Підводячи підсумки, WF атака може бути використана як в якості моніторингового, розвідувального спостерігача (чи відвідує користувач ресурси зі списку чи ні), так і в розпізнавальному варіанті (який портал відвідує користувач зі списку).