Лекция 3.3. Асинт-разведка из публичных источников с помощью Python

В данной лекции будут рассмотрены следующие вопросы:  
1. Как работает GoogleDorking?  
2. Какую информацию можно найти через Dorks?  
3. Разбор операторов поиска.

Глава 1. Асинт-разведка из публичных источников с помощью Python

Вступление  
Google Dorks или Google Hacking — техника, используемая СМИ, следственными органами, инженерами по безопасности и любыми пользователями для создания запросов в различных поисковых системах для обнаружения скрытой информации и уязвимостях, которые можно обнаружить на общедоступных серверах. Это метод, в котором обычные запросы на поиск веб-сайтов используются в полную меру для определения информации, скрытой на поверхности.  
Как работает Google Dorking?  
Данный пример сбора и анализа информации, выступающий как инструмент OSINT, является не уязвимостью Google и не устройством для взлома хостинга сайтов. Напротив, он выступает в роли обычного поискового процесса данных с расширенными возможностями. И это не в новинку, так как существует огромное количество веб-сайтов, которым уже более десятка лет, и они служат как хранилища для изучения и использования Google Hacking.  
В то время как поисковые системы индексируют, сохраняют хедеры и содержимое страниц, и связывают их между собой для оптимальных поисковых запросов. Но, к сожалению, сетевые пауки любых поисковиков настроены индексировать абсолютно всю найденную информацию. Даже несмотря на то, что у администраторов веб ресурсов не было никаких намерений публиковать этот материал.  
Однако самое интересное в Google Dorking, так это огромный объём информации, который может помочь каждому в процессе изучения поискового процесса Google. Может помочь новичкам в поиске пропавших родственников, а может научить каким образом можно извлечь информацию для собственной выгоды. В общем, каждый ресурс интересен и удивителен по-своему и может помочь каждому в том, что именно он ищет.

Глава 2. Асинт-разведка из публичных источников с помощью Python

Какую информацию можно найти через Dorks?  
Начиная, от контроллеров удалённого доступа различных заводских механизмов до конфигурационных интерфейсов важных систем. Есть предположение о том, что огромное количество информации, выложенной в сети, никто и никогда не найдёт.  
Однако давайте разберёмся по порядку. Представьте себе новую камеру видеонаблюдения, позволяющую просматривать её трансляцию на телефоне в любое время. Вы настраиваете и подключаетесь к ней через Wi-Fi, и загружаете приложение, для аутентификации входа в систему камеры наблюдения. После этого можно получить доступ к этой же камере из любой точки мира.  
На заднем плане не всё выглядит таким простым. Камера посылает запрос на китайский сервер и воспроизводит видео в режиме реального времени, позволяя войти в систему и открыть видеотрансляцию, размещённую на сервере в Китае, с вашего телефона. Этот сервер может не требовать пароля для доступа к каналу с вашей веб-камеры, что делает её общедоступной для всех, кто ищет текст, содержащийся на странице просмотра камеры.  
И к сожалению, Google безжалостно эффективен в поиске любых устройств в Интернете, работающих на серверах HTTP и HTTPS. И поскольку большинство этих устройств содержат определённую веб-платформу для их настройки, это означает, что многие вещи, непредназначенные быть в Google, в конечном итоге оказываются там.  
Безусловно, самый серьёзный тип файлов, это тот, который несёт в себе учётные данные пользователей или же всей компании. Обычно это происходит двумя способами. В первом, сервер настроен неправильно и выставляет свои административные логи или журналы в открытом доступе в Интернете. Когда пароли меняются или пользователь не может войти в систему, эти архивы могут утечь вместе с учётными данными.  
Второй вариант происходит тогда, когда конфигурационные файлы, содержащие ту же информацию (логины, пароли, наименования баз данных и т. д.), становятся общедоступными. Это файлы должны быть обязательно скрыты от любого публичного доступа, так как в них часто оставляют важную информацию. Любая из этих ошибок может привести к тому, что злоумышленник найдёт данные лазейки и получит всю нужную информацию.  
Использование Google Dorks помогает не только находить все эти файлы, но ипоказывает насколько бывают уязвимы платформы, содержащие информацию в виде списка адресов, электронной почты, картинок и даже перечня веб-камер в открытом доступе.  
Разбор операторов поиска  
Dorking можно использовать в различных поисковых системах, не только в Google. В повседневном использовании поисковые системы, такие как Google, Bing, Yahoo и DuckDuckGo, принимают поисковый запрос или строку поисковых запросов и возвращают соответствующие результаты. Также эти же системы запрограммированы принимать более продвинутые и сложные операторы, которые значительно сужают эти условия поиска. Оператор — это ключевое слово или фраза, несущее особое значение для поисковой системы. Вот примеры часто используемых операторов: «inurl», «intext», «site», «feed», «language». За каждым оператором следует двоеточие, за которым следует соответствующий ключевая фраза или фразы.  
Эти операторы позволяют выполнять поиск более конкретной информации, например: определённые строки текста внутри страниц веб-сайта или файлы, размещённые по конкретному URL-адресу. Помимо прочего, Google Dorking может также находить скрытые страницы для входа в систему, сообщения об ошибках, выдающие информации о доступных уязвимостях и файлы общего доступа. В основном причина заключается в том, что администратор веб-сайта мог просто забыть исключить из открытого доступа.

Python может быть очень полезным инструментом для сбора и анализа информации из открытых источников. Некоторые библиотеки Python, которые могут быть использованы для этой цели, включают:

* Библиотека requests: это библиотека для работы с HTTP-запросами, которая может быть использована для получения данных из различных вебсайтов и API.
* Библиотека BeautifulSoup: это библиотека для парсинга HTML и XMLдокументов, которая может быть использована для извлечения данных из вебстраниц.
* Библиотека tweepy: это библиотека для работы с API Twitter, которая может быть использована для сбора твитов и другой информации из Twitter.
* Библиотека pyquery: это библиотека для парсинга HTML-документов, которая может быть использована для извлечения данных из веб-страниц.
* Библиотека scrapy: это библиотека для сбора данных с веб-сайтов, которая позволяет автоматизировать процесс сбора данных из большого числа страниц.

Python также может быть использован для анализа собранных данных. Для этого можно использовать библиотеки, такие как pandas, numpy, matplotlib и другие.

1. Процесс сбора, анализа и интерпретации информации, полученной из публично доступных источников, таких как веб-сайты, социальные сети, блоги, форумы и другие онлайн-ресурсы:
   1. Асинт-разведка
   2. Спуфинг
   3. Google Dorking
   4. Парсинг
2. Выберите все цели, в которых может использовать асин-разведка:
   1. Выявление угроз и рисков
   2. Сбор информации о конкурентах
   3. Анализ общественного мнения
   4. Исследование рынка
3. Техника поиска, использующая специальные операторы и фразы для извлечения специфической информации из поисковой системы Google:
   1. Асинт-разведка
   2. Спуфинг
   3. Google Dorking
   4. Парсинг
4. С помощью какого оператора можно найти файлы определённого типа, такие как PDF, DOC, XLS?
   1. Оператор Intitle
   2. Оператор Filetype
   3. Оператор Site
5. Какой оператор позволяет найти информацию, связанную с определённым веб-сайтом или доменом?
   1. Оператор Intitle
   2. Оператор Filetype
   3. Оператор Site
6. С помощью какого оператора можно найти страницы, заголовки которых содержат определённые слова или фразы?
   1. Оператор Intitle
   2. Оператор Filetype
   3. Оператор Site
7. Requests, Beautiful Soup, Selenium — это:
   1. Форматы данных
   2. Протоколы python
   3. Инструменты и библиотеки python