Дисциплина «Этичный хакинг и противодействие взлому»

Лекция 2

Сбор информации

Преподаватель: **Батыргалиев Асхат Болатканович, PhD,**
ассоц.проф. кафедры «Кибербезопасность, обработка и хранение информации»

askhat.b.b@gmail.com

Содержание

1. Использование поисковых систем
2. Техники расширенного поиска в Google
3. Google Hacking Database (GHDB)
4. Сбор данных из социальных сетей
5. Сбор информации с веб-сайта
6. Конкурентная разведка
7. Сбор данных, используя данные регистраторов
8. Сбор сетевой информации
9. Социальная инженерия для сбора данных

По завершению урока Вы будете знать:

1. Возможности использования поисковых систем
2. Техники расширенного поиска
3. Возможности сбора данных из социальных сетей
4. Возможности сбора информации с веб-сайтов
5. Понятие конкурентной разведки
6. Возможности сбора данных, используя регистраторы
7. Возможности сбора сетевой информации
8. Возможности социальной инженерии для сбора данных

Поисковые системы

Поисковая система (search engine) - алгоритмы и реализующая их совокупность компьютерных программ (в широком смысле этого понятия, включая аналоговые системы автоматизированной обработки информации первого поколения), предоставляющая пользователю возможность быстрого доступа к необходимой ему информации при помощи поиска в обширной коллекции доступных данных. Одно из наиболее известных применений поисковых систем - веб-сервисы для поиска текстовой или графической информации во Всемирной паутине. Существуют также системы, способные искать файлы на FTP-серверах, товары в интернет-магазинах, информацию в группах новостей Usenet.

Поисковый запрос - это последовательность символов, которую пользователь вводит в поисковую строку, чтобы найти интересующую его информацию.

Формат поискового запроса зависит как от устройства поисковой системы, так и от типа информации для поиска. Чаще всего, поисковый запрос задаётся в виде набора слов или фразы, иногда - используя расширенные возможности языка запросов поисковой системы.

Техники расширенного поиска в Google

- **Один из нескольких (логическое ИЛИ)**. По умолчанию Google ищет страницы, которые содержат все слова из поискового запроса, но если требуется выдать те, которые содержат хотя бы одно слово из заданного множества, можно воспользоваться логическим оператором ИЛИ, которому соответствует символ "|" (pipe symbol), например: хакинг|пентестинг|форензика.
- **Кавычки**. Если требуется найти определенную фразу дословно, можно использовать кавычки, например: “Kali Linux".
- **Исключение (логическое НЕ)**. Для того, чтобы исключить из результата поиска страницы, содержащие определенное слово, в поисковом запросе необходимо использовать символ "-". Пример: linux distrib download -suse (запрос вернет ссылки на страницы для скачивания различных дистрибутивов Linux, за исключением Suse).
- **Похожие слова**. Для того, чтобы Google искал слова, похожие на заданное, используется символ "~" (тильда). Будут найдены синонимы и слова с альтернативными окончаниями. Пример: ~hippo (по запросу будет так же найдено, например, слово hyppopotamus).
- **Маски**. Символ "*" можно использовать как маску - условное обозначение произвольного количества любых символов. Это может быть полезно, если необходимо найти сайт, домен которого запомнился только отчасти, например: *me.org
- **Расширенный поиск**. Если вы забыли какой-либо из перечисленных операторов, всегда можно воспользоваться формой расширенного поиска.

Техники расширенного поиска в Google

- **Определения**. Используйте оператор define: для быстрого поиска определений. Пример: define:брэндмауэр.
- **Калькулятор**. Одной из полезных возможностей является вычисление арифметических выражений. В выражениях можно использовать операторы +, -, *, /, ^ (степень), sqrt (квадратный корень), sin, cos, tan, ln, lg, exp (ex), скобки и др., например: sqrt(25) * 768.
- **Числовые интервалы**. Поиск числовых интервалов, которые можно задавать с помощью крайних значений, разделенных последовательностью из двух точек. Пример: Букер 2004..2007.
- **Поиск на заданном сайте**. С помощью оператора site: можно ограничить результаты поиска определенным веб-сайтом. Именно эта возможность обычно используется при установке поисковых форм на сторонних ресурсах, например: seagate site:ixbt.com.
- **Ссылки извне**. С помощью оператора link:, можно найти страницы, которые ссылаются на заданный URL. Оператор можно использовать не только для главного адреса сайта, но и для отдельных страниц. Оператор не дает гарантии, что в результате поиска будут перечислены абсолютно все страницы. Пример: link:hackthebox.com
- **Местоположение слова**. По умолчанию Google ищет заданный текст внутри содержимого страниц. Но если есть необходимость искать в некоей определенной области, можно использовать такие операторы как «inurl:» (поиск внутри URL), «intitle:» (поиск в заголовке страницы), «intext:» (поиск в тексте страницы), и «inanchor:» (поиск в тексте ссылок).

Техники расширенного поиска в Google

- **Типы файлов**. В случае, если требуется искать, только документы в формате PDF, Word или Excel, можно использовать оператор filetype:. Список поддерживаемых форматов : Adobe Reader PDF (.pdf), Adobe Postscript (.ps), Autodesk DWF (.dwf), Google Earth (.kml, .kmz), Microsoft Excel (.xls), Microsoft PowerPoint (.ppt), Microsoft Word (.doc), Rich Text Format (.rtf), Shockwave Flash (.swf). Пример: stroustrup c++ language filetype:pdf. Update: Для выбора типа искомых файлов так же можно использовать оператор ext:.
- **Кэшированные страницы**. При поиске устаревших страниц и страниц, контент которых был обновлен, может помочь поиск в кэше поисковой машины. Для этого предназначен оператор cached:. Update: Существует так же близкий по смыслу оператор cache:, с помощью которого можно сразу получать страницы из кэша по их URL. Этой возможностью в принципе можно пользоваться как своеобразным бэкапом видимых для Google веб-страниц: даже если страница будет удалена со своего сайта, на Google может остаться ее копия.
- **Информация о сайте**. С помощью оператора info: можно получить известную Google информацию об указанном сайте. Пример: info:habrahabr.ru.
- **Похожие сайты**. С помощью оператора related: Google может выдать список сайтов, которые считает похожим на заданный. Пример: related:flickr.com.

Google Hacking DataBase (GHDB)

Google hacking (Google dorking) - это хакерский метод, который используется в Google Search и других приложениях Google для поиска дыр в конфигурации и компьютерном коде, которые используют веб-сайты.

Предполагает использование расширенных операторов в поисковой системе Google для поиска определённых строк текста в результатах поиска. Один из популярных примеров - это поиск конкретных версий уязвимостей Веб-приложений. Поисковый запрос intitle:admbook intitle:Fversion filetype:php найдет все веб-страницы, содержащие этот конкретный текст. Обычно при установке приложений по умолчанию их текущая версия указывается на каждой странице, на которой они работают, например, «Powered by XOOPS 2.2.3 Final». Можно найти устройства, подключенные к Интернету.

Строка поиска, например inurl: "ViewerFrame? Mode =", найдет общедоступные веб-камеры.

Ещё один полезный вид поиска - это intitle:index.of, за которым следует ключевое слово поиска. Так можно получить список файлов на серверах. Например, intitle: index.of mp3 предоставит все файлы MP3, доступные на различных типах серверов.

Сбор данных из социальных сетей

Социальные сети позволяют осуществить сбор большого количества пользовательских данных:

- личные данные (фамилия, имя, отчество, контактные и анкетные данные);
- метаданные (IP-адреса, геолокация, данные о пользовательском устройстве);
- информация о круге общения, друзьях, подписчиках пользователя и на кого пользователь сам подписан; группы, в которых пользователь участвует, их тематика, активность пользователя в них, его сообщения, комментарии к чужим сообщениям;
- данные с собственных страниц пользователя (подписчики этих страниц, тематика страницы, размещаемый на них контент, фотографии и собственные видеозаписи, комментарии пользователя и подписчиков).

Инструменты сбора данных из социальных сетей

1. **Popsters**. Анализирует контент в социальных сетях. Работает со всеми популярными в России платформами: ВКонтакте, Одноклассники, Facebook, Instagram, Telegram, Twitter, YouTube, Pinterest и др. Находит самые популярные посты для любой страницы, считает ER, зависимость вовлечения от объема текста, типа контента, времени публикации поста.
2. **AgoraPulse**. Считает вовлечение в Twitter, Facebook, Instagram, Google+, LinkdIn, YouTube. Кроме стандартных метрик, сервис предлагает уведомления об упоминании вашей страницы и быстро отвечать пользователям.
3. **Hootsuite**. Кроме инструментов для контент-менеджмента, присутствует аналитика социальных сетей. Изменяет конверсии, ROI, трафик и вовлечение.
4. **Quintly**. Анализирует Facebook, Twitter, Instagram, Pinterest, LinkedIn, YouTube и Google+. Оценивает подписчиков сообществ, создает кастомные отчеты. Среди других инструментов аналитики особо выделяется возможностью находить лидеров мнений для бренда.
5. **Simply Measured**. Сервис связывается с Google Analytics и дает данные о поведении аудитории из социальных сетей на сайте. Собирает данные о продвижении и контенте конкурентов.
6. **Socialbakers**. Следит за вовлечением и другими показателями в Facebook, Twitter, Google Plus и YouTube. Отслеживает показатели не только для контента на странице, но и для рекламных объявлений. Включает инструменты аналитики, а также функционал для постинга и модерации.

Сбор информации с веб-сайта

Веб-скрейпинг (скрепинг, web scraping) - технология получения веб-данных путем извлечения их со страниц веб-ресурсов. Веб-скрейпинг может быть сделан вручную пользователем компьютера, однако термин обычно относится к автоматизированным процессам, реализованным с помощью кода, который выполняет GET-запросы на целевой сайт.

Инструменты web scraping (парсинг) разработаны для извлечения, сбора любой открытой информации с веб-сайтов. Эти ресурсы нужны тогда, когда необходимо быстро получить и сохранить в структурированном виде любые данные из интернета.

Инструменты web scraping

Инструменты web scraping:

1. **Import.io** позволяет формировать собственные пакеты данных: нужно импортировать информацию с определенной веб-страницы и экспортировать ее в CSV. Можно извлекать тысячи веб-страниц за считанные минуты, не написав ни строчки кода, и создавать API согласно требованиям.
2. **Webhose.io** обеспечивает прямой доступ в реальном времени к структурированным данным, полученным в результате парсинга тысяч онлайн источников. Парсер способен собирать веб-данные на более чем 240 языках и сохранять результаты в различных форматах, включая XML, JSON и RSS.
3. **Dexi.io** (ранее CloudScrape) способен парсить информацию с любого веб-сайта и не требует загрузки дополнительных приложений, как и Webhose. Редактор самостоятельно устанавливает своих поисковых роботов и извлекает данные в режиме реального времени. Пользователь может сохранить собранные данные в облаке, например, Google Drive и Box.net, или экспортировать данные в форматах CSV или JSON.
4. **Scrapinghub** – это облачный инструмент парсинга данных, который помогает выбирать и собирать необходимые данные для любых целей. Scrapinghub использует Crawlera, умный прокси-ротатор, оснащенный механизмами, способными обходить защиты от ботов. Сервис способен справляться с огромными по объему информации и защищенными от роботов сайтами.

Конкурентная разведка

Конкурентная разведка (Competitive Intelligence) - сбор и обработка данных из разных источников, для выработки управленческих решений с целью повышения конкурентоспособности коммерческой организации, проводимые в рамках закона и с соблюдением этических норм (в отличие от промышленного шпионажа); а также структурное подразделение предприятия, выполняющее эти функции.

Другое определение понятия - это особый вид информационно-аналитической работы, позволяющий собирать обширнейшую информацию о юридических и физических лицах без применения специфических методов оперативно-розыскной деятельности, являющихся исключительной прерогативой государственных правоохранительных органов и спецслужб. Синонимы конкурентной разведки - бизнес-разведка, деловая разведка, аналитическая разведка, экономическая разведка, маркетинговая разведка, коммерческая разведка.

Основная цель конкурентной разведки заключается в исследовании рынка для развития бизнеса и разработки стратегии его дальнейшего продвижения. Чтобы собрать как можно больше информации и получить максимально полную картину, работу проводят в трех основных направлениях.

Конкурентная разведка предполагает сбор информации из множества разных источников, включая отраслевые выставки, конференции, экспертные интервью и публичные отчеты.

Конкурентная разведка. Сбор информации

Сайт конкурентов. Один из первых источников информации, который используют для конкурентной разведки. Сайт предоставляет информацию о ценах, акциях, качестве работы службы поддержки и многое другое. С помощью SEMRush и Ahrefs можно провести SEO-аудит, отследить рейтинг в поисковой выдаче, изучить топ ключевых слов, по которым продвигаются конкуренты. Чтобы узнать интересы целевой аудитории и также проанализировать поисковый и социальный трафик, используют сервис SimilarWeb.

Социальные сети. Анализ страниц в Instagram, Facebook, ВКонтакте позволяет изучить язык бренда, проанализировать частоту публикаций, охват, уровень вовлеченности подписчиков, а также узнать размер целевой аудитории.

Форумы и блоги. Мониторинг сторонних ресурсов позволяет найти вопросы, которые волнуют потенциальных покупателей и клиентов, узнать их мнение о конкурентах, качестве обслуживания, удовлетворенности после покупки и многое другое. Форумы и блоги могут предоставить подсказки, позволяющие улучшить торговое предложение и повысить конкурентоспособность.

Отзывы клиентов. Покупатели могут оставлять их на сайте, в социальных сетях и на других ресурсах таких как Google Карты, IRecommend.ru, Отзовик. Исследование отзывов из разных источников помогает увидеть реальную картину удовлетворенности покупателей.

Конкурентная разведка. Сбор информации

Интервью. Этот вид маркетингового исследования подразумевает изучение конкурентов через интервьюирование клиентов, экспертов отрасли, лидеров мнений. Такой метод помогает узнать о тенденциях развития отрасли и проблемах в ней. Интервью позволяет построить прогнозы касательно продвижения бизнеса и расширения рынка.

Ивенты. Выставки, конференции, семинары и другие мероприятия помогают собрать дополнительную информацию о позиционировании конкурентов, направлениях их развития, новых товарах и услугах. Посещение мероприятий конкурентов помогает получить личный опыт взаимодействия с другими компаниями и их продуктами.

Пресс-релизы и анонсы. Отслеживание новостей конкурентов помогает держать руку на пульсе рынка и узнавать о мероприятиях, презентациях товаров и услуг, изменениях PR-стратегии, рекламной политики, а также о появлении новых партнеров, поставщиков, дистрибьюторов.

Квартальные и годовые отчеты. С их помощью вы получите информацию о прибыли, убытках и финансовом положении конкурентов. Посмотреть отчеты можно на официальных сайтах компаний и на специальных ресурсах вроде Opendatabot и YouControl.

Сбор данных, используя данные регистраторов

Регистратор данных (даталоггер) - это электронное устройство, которое записывает во внутреннюю память, на внешнее хранилище или передаёт в облачный сервис данные. Данные сохраняются с течением времени или по отношению к местоположению. Данные могут поступать от встроенного в прибор сенсора или датчика или от внешних приборов и датчиков. Они, как правило, небольшие, на батарейках, портативные, и снабжены микропроцессором, внутренней памятью для хранения данных и различными встроенными датчиками. Некоторые регистраторы данных обладают специальным интерфейсом для подключения к персональному компьютеру, а также используют программное обеспечение, чтобы активировать регистратор данных, просматривать и анализировать собранные данные. В то же время некоторые устройства могут иметь локальный интерфейс (клавиатура, дисплей) и могут быть использованы в качестве автономного устройства.

Сбор сетевой информации

Исследование сети необходимо для планирования атаки. Для этого используется сканирование, пробы, определение используемых приложений и их версий, установление топологий сети, таких как локальная, сеть провайдера, телекоммуникационная или корпоративная. Во время исследования сети наиболее важными являются IP-адреса, сетевые адреса, списки доступа, точки доступа, настройки межсетевого экрана, версии приложений и имена пользователей.

Большинство инструментов для сбора информации используется удаленно. Команды traceroute, ping и сканирование сети позволяет составить физическую/логическую структуру сети, в то время как пакетный сбор данных, например снифферы, позволяют получить данные о запущенных приложениях, именах пользователей и паролей.

Сервис WHOIS (SamSpade) поможет определить, кому принадлежит конкретный IP-адрес, в то время как NetBIOS и RPC dumps позволят проанализировать трафик локальной сети.

После того, как получена вся ценная информация о пользователе, в web crawlers (нерезидентные вирусы) позволят просмотреть не удаленные данные с любой страницы. И в конце, IRC клиент поможет получить доступ и собрать сведения о любом пользователи в сети.

При этом, используется весь диапазон сканирования. Сюда входит сканирование портов, IP-адресов, маршрутизаторов (фальсификация BGP - протокол IP-маршрутизации, и OSPF - протокол предпочтения кратчайшего пути), сканирование SNMP (протокол сетевого управления), исследование портов и приложений, проверка межсетевых экранов, наличия беспроводных точек, операционной системы, установленного оборудования и стека протоколов IP.

Социальная инженерия для сбора данных

Социальная инженерия - психологическое манипулирование людьми с целью совершения определенных действий или разглашения конфиденциальной информации.

Техники социальной инженерии:

- **фишинг (phishing)** - вид интернет-мошенничества, целью которого является получение доступа к конфиденциальным данным пользователей - логинам и паролям. Это самая популярная схема социальной инженерии на сегодняшний день;
- **несуществующие ссылки**. Атака, которая заключается в отправлении письма с соблазнительной причиной посетить сайт и прямой ссылкой на него, которая лишь имеет сходство с ожидаемым сайтом;
- **мошенничество с использованием брендов известных корпораций**. В таких фишинговых схемах используются поддельные сообщения электронной почты или веб-сайты, содержащие названия крупных или известных компаний. В сообщениях может быть поздравление с победой в каком-либо конкурсе, проводимом компанией, о том, что срочно требуется изменить учётные данные или пароль. Подобные мошеннические схемы от лица службы технической поддержки также могут производиться по телефону;
- **подложные лотереи**. Пользователь может получить сообщения, в которых говорится о том, что он выиграл в лотерею, которая проводилась какой-либо известной компанией. Внешне эти сообщения могут выглядеть так, как будто они были отправлены от лица одного из высокопоставленных сотрудников корпорации;

Социальная инженерия для сбора данных

- **ложные антивирусы и программы для обеспечения безопасности**. Подобное мошенническое программное обеспечение, также известное под названием «scareware» - это программы, которые выглядят как антивирусы, хотя, на самом деле, все обстоит совсем наоборот. Такие программы генерируют ложные уведомления о различных угрозах, а также пытаются завлечь пользователя в мошеннические транзакции. Пользователь может столкнуться с ними в электронной почте, онлайн объявлениях, в социальных сетях, в результатах поисковых систем и даже во всплывающих окнах на компьютере, которые имитируют системные сообщения;
- **телефонный фрикинг** (phreaking) - термин, описывающий эксперименты и взлом телефонных систем с помощью звуковых манипуляций с тоновым набором. Техника появилась в конце 50-х в Америке. Телефонная корпорация Bell, которая тогда покрывала практически всю территорию США, использовала тоновый набор для передачи различных служебных сигналов. Энтузиасты, попытавшиеся повторить некоторые из этих сигналов, получали возможность бесплатно звонить, организовывать телефонные конференции и администрировать телефонную сеть;
- **претекстинг (pretexting)** - атака, в которой злоумышленник представляется другим человеком и по заранее подготовленному сценарию узнает конфиденциальную информацию. Эта атака подразумевает должную подготовку, как то: день рождения, ИНН, номер паспорта либо последние цифры счета, для того, чтобы не вызвать подозрений у жертвы. Обычно реализуется через телефон или электронную почту;

Социальная инженерия для сбора данных

- **IVR или телефонный фишинг (вишинг, vishing - voice fishing)**. Данная техника основана на использовании системы предварительно записанных голосовых сообщений с целью воссоздать «официальные звонки» банковских и других IVR систем. Обычно жертва получает запрос (чаще всего через фишинг электронной почты) связаться с банком и подтвердить или обновить какую-либо информацию. Система требует аутентификации пользователя посредством ввода PIN-кода или пароля. Поэтому, предварительно записав ключевую фразу, можно выведать всю нужную информацию. Например, любой может записать типичную команду: «Нажмите единицу, чтобы сменить пароль. Нажмите двойку, чтобы получить ответ оператора» и воспроизвести её вручную в нужный момент времени, создав впечатление работающей в данный момент системы предварительно записанных голосовых сообщений;
- **квид про кво (Quid pro quo)** - в английском языке это выражение обычно используется в значении «услуга за услугу». Данный вид атаки подразумевает обращение злоумышленника в компанию по корпоративному телефону (используя актёрское мастерство) или электронной почте. Зачастую злоумышленник представляется сотрудником технической поддержки, который сообщает о возникновении технических проблем на рабочем месте сотрудника и предлагает помощь в их устранении. В процессе «решения» технических проблем злоумышленник вынуждает цель атаки совершать действия, позволяющие атакующему запускать команды или устанавливать различное программное обеспечение на компьютере жертвы;

Социальная инженерия для сбора данных

- **«Дорожное яблоко»**. Этот метод атаки представляет собой адаптацию троянского коня, и состоит в использовании физических носителей. Злоумышленник подбрасывает «инфицированные» носители информации в местах общего доступа, где эти носители могут быть легко найдены, такими как туалеты, парковки, столовые, или на рабочем месте атакуемого сотрудника[5]. Носители оформляются как официальные для компании, которую атакуют, или сопровождаются подписью, призванной вызвать любопытство. К примеру, злоумышленник может подбросить CD, снабжённый корпоративным логотипом и ссылкой на официальный сайт компании, снабдив его надписью «Заработная плата руководящего состава». Диск может быть оставлен на полу лифта или в вестибюле. Сотрудник по незнанию может подобрать диск и вставить его в компьютер, чтобы удовлетворить своё любопытство;
- **плечевой серфинг** (shoulder surfing) включает в себя наблюдение личной информации жертвы через её плечо. Этот тип атаки распространён в общественных местах, таких как кафе, торговые центры, аэропорты, вокзалы, а также в общественном транспорте;
- **обратная социальная инженерия** - жертва сама предлагает злоумышленнику нужную ему информацию. Лица, обладающие авторитетом в технической или социальной сфере, часто получают идентификаторы и пароли пользователей и другую важную личную информацию просто потому, что никто не сомневается в их порядочности. Например, сотрудники службы поддержки никогда не спрашивают у пользователей идентификатор или пароль; им не нужна эта информация для решения проблем. Однако, многие пользователи ради скорейшего устранения проблем добровольно сообщают эти конфиденциальные сведения.