International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 04 Issue 02 |2020

Review on OSNIT Tools & Techniques

Raj Saundatikar[1], Deepesh Seth[2]

[1, 2] Master Student, Department of MCA, Jain Deemed-to-be University, Bengaluru, Karnataka, India

ABSTRACT

Till now we've learned loads associated with searching the online in various other ways just by using the browser. During this research, we'll study some Tools and Techniques and web-based services, which automate the method of information extraction. We'll study other ways to extract data just by clicking on a graphical interface to executing commands in a very command-line interface (CLI).The tools' interface and their usage are going to be demonstrated thoroughly in order that the users can get a powerful hold of and may easily explore them further.Also we will explain some tools for scrapping information from public sources, collectivelycalled Open-source intelligence(OSNIT).

KEYWORD: Reconnaissance, IDS, IPS, Social Engineering, Foot Printing

1. INTRODUCTION

OSINT stands for Open Source Intelligence, and it is a very important aspect for understanding the cybersecurity that rules the internet these days. The term OSINT is used from many decades; in fact, US military agencies started using the term OSINT in the late 1980s as they were re-evaluating the nature of information requirements in tactical levels under battlefields. While the concept of OSINT has evolved since then, as it does not include the non-public sources, the concept originates from that time.

OSINT is the information collected from public sources such as those available on the Internet, although the term isn't only limited to the internet, rather means all publicly available sources. "OS" (from OSINT) means Open Source. It is related to any publicly available source where the user can acquire the information in their intelligence data collection. The keyword behind OSINT is information. The information which is available for free. It does not matter where the information is located, whether it is social media, images, videos, blogs, newspapers or tweets as long as it is publicly available, free and legal. With the right information, you can get a great advantage over your competition, or speed up company/people investigation.

You do not need to be a hacker to use OSINT in your daily life: you are already using it, you just might have not known it. All Internet users are using OSINT techniques in one way or another, such as when searching for something online. Whether it is a company, school, university, or person you are looking for, you are collecting some form of OSINT intelligence.

The open-source insight alludes to all data and information that can be accumulated from freely accessible sources and OSINT has moved into the front of insight gathering disciplines. Open source insight as an idea is old, as since the commencement social orders have esteemed accessible data over encompassing conditions to determine better ends. The assembled information is basic as it regularly gives a bit of leeway over another, let it involve illuminating wrongdoing, winning a fight, or succeeding better in business tasks. What has changed after some time is the measure of accessible information and the techniques to gather it. At the point when prior open-source insight concentrated on social occasion data from papers, open addresses, interviews, to name models, the information today is in the Web and approaches recovering the information are getting substantially more complex, innovatively progressed and open for all.

2. How Is Open Source Intelligence Used?

Since we've secured the essentials of open-source insight, we can take a gander at how it is usually utilized for cybersecurity. There are two basic use cases:

2.1 Ethical Hacking and Penetration Testing -:

Security experts utilize open-source insight to recognize potential shortcomings in cordial systems with the goal that they can be remediated before they are abused by danger on-screen characters. Normally discovered shortcomings include:

- Unintentional holes of delicate data, as through web-based life
- Open ports or unbound web associated gadgets
- Unpatched programming, for example, sites running old adaptations of basic CMS items
- Spilt or uncovered resources, for example, exclusive code on paste bin

2.2 Identifying External Threats -:

As we've talked about commonly previously, the web is a brilliant wellspring of bits of knowledge into an association's most squeezing dangers. From distinguishing which new vulnerabilities are as a rule effectively abused to blocking danger entertainer "jabber" about an up and coming assault, open-source insight empowers security experts to organize their time and assets to address the most huge current dangers. As a rule, this sort of work requires an investigator to distinguish and connect numerous information focuses to approve a danger before the move is made. For instance, while a solitary compromising tweet may not be cause for concern, that equivalent tweet would be seen from an alternate perspective in the event that it was attached to a danger bunch known to be dynamic in a particular industry.

One of the most significant things to comprehend about open-source knowledge is that it is frequently utilized in mix with other insight subtypes. Knowledge from shut sources, for example, inside telemetry, shut dull web networks, and outer insight sharing networks is normally used to channel and check open-source insight. There is an assortment of devices accessible to assist examiners with playing out these capacities, which we'll take a gander at somewhat later on.

3. Information Gathering Types

Three main methods to collect OSINT information: passive, semi-passive, and active. The usage of one in favour of another relies on the scenario during which the gathering process operates additionally to the type of data that you are interested in. The three gathering techniques are generally used to describe the ways during which footprinting works, in other words, acquiring technical information about target IT infrastructure (types of OS, network topology, server names, and so on).

3.1 Passive Collection

The passive collection is the most used type when collecting OSINT intelligence. All OSINT intelligence methods should use passive collection because the main aim of OSINT gathering is to collect information about the target via publicly available resources only. In this type, your target knows nothing about your intelligence-collecting activities. This sort of search is extremely anonymous and will be done secretly. From a technical perspective, this type of gathering reveals limited information about the target because you are not going to send any traffic (packets) to the target server either directly or indirectly and the main resources that you can gather are limited to archive information (mainly outdated information), unprotected files left on target servers, and content present on the target website.

3.2 Semipassive

From a technical view, this kind of gathering sends limited traffic to target servers to accumulate general information about them. This traffic tries to resemble typical Internet traffic to avoid drawing any attention to your reconnaissance activities. During this way, you're not implementing in-depth investigation of the target's online resources, but only investigating lightly without launching any alarm on the target's side. Although this kind of gathering is taken into account somehow anonymous, the target can know that there is reconnaissance happening if they investigate the issue (by checking the server or networking device logs). However, they should not be able to attribute it to the attacker's machine.

3.3 Active Collection

Inactive collection, you interact directly with the system to collect intelligence about it. The target can become aware of the reconnaissance process since the person/entity collecting information will use advanced techniques to harvest technical data about the target IT infrastructure such as accessing open ports, scanning vulnerabilities, scanning web server applications, and more. This traffic will appear suspicious or malicious behaviour and will leave traces on the target's intrusion detection system (IDS) or intrusion prevention system (IPS). Conducting social engineering attacks on the target is also considered a sort of active information gathering. Active collection and semi-passive collection are types of information gathering, but we usually do not use them in OSINT gathering. The passive collection is preferred because it can harvest information from public sources secretly, and this is the essence of OSINT.

4. BENEFITS OF OSINT

In the present information age, no one can underestimate the vital role that OSINT plays in the different intelligence fields. The benefits of OSINT span many areas in today’s world. The following are the main ones:

- Less risky: Using publicly available information to gather intelligence has no risk compared with other forms of intelligence such as using spying satellites or using human sources on the ground to collect information, especially in hostile countries.

- Cost-efficient: Collecting OSINT is generally cheaper compared with other intelligence sources. For instance, using humans or spying satellite to collect data is very expensive. Small businesses with limited intelligence budgets can exploit OSINT sources with minimal costs.

- Ease of accessibility: OSINT sources are always available, no matter where you are and are always up-to-date. OSINT sources can be used by different parties in any intelligence context; all you need are the required skills/tools to harvest and analyse OSINT properly. For example, military departments can foresee future assaults by analysing activities on social networking sites, while corporations can use it to build their new market expansion strategies.

- Legal issues: OSINT assets can be shared between various parties without worrying about breaching any copyright license as these resources are already published publicly. Of course, some impediments apply when sharing grey literature.

- Aiding financial investigators: OSINT permits specialized government agencies to detect tax evaders, for instance. Many famous celebs and some huge companies are involved in tax evasion, and monitoring their social media accounts, vacations, and lifestyles has a great value for a government inspector who may be chasing them for undeclared income.

- Fighting against online counterfeiting: OSINT techniques can be used to find false products/services and direct law enforcement to close such sites or to send warnings to users to stop dealing with them. This is an incredible advantage of OSINT, especially when fighting against fake pharmaceutical and natural health products.

- Maintaining national security and political stability: This might be the most important role of OSINT; it helps governments to understand their people’s attitudes and act promptly to avoid any future clashes. Wise governments utilize OSINT in their future strategies, especially for their domestic policies.

5. THE DARK SIDE OF THE OPEN SOURCE INTELLIGENCE

Now, it's a great opportunity to address the subsequent significant issue with open source knowledge: if something is promptly accessible to insight examiners, it's additionally promptly accessible to risk on-screen characters.

Risk entertainers utilize open-source knowledge apparatuses and systems to recognize potential targets and adventure shortcomings in target systems. When powerlessness is distinguished, it is frequently an amazingly speedy and straightforward procedure to abuse it and accomplish an assortment of noxious targets.

This procedure is the primary motivation behind why such huge numbers of little and medium-sized undertakings get hacked every year. It isn't on the grounds that risk bunches explicitly check out them, but instead on the grounds that vulnerabilities in their system or site engineering are discovered utilizing straightforward open-source knowledge procedures. To put it plainly, they are obvious objectives.

Furthermore, open-source insight doesn't just empower specialized assaults on IT frameworks and systems. Risk on-screen characters likewise search out data about people and associations that can be utilized to illuminate refined social designing efforts utilizing phishing (email), vishing (telephone or phone message), and SMiShing (SMS). Frequently, apparently harmless data shared through informal organizations and online journals can be utilized to grow exceptionally persuading social designing efforts, which thus are utilized to deceive good-natured clients into trading off their association's system or resources.

This is the reason for utilizing open source knowledge for security designs is so significant. It offers you a chance to discover and fix shortcomings in your association's system and expel delicate data before a risk on-screen character utilizes similar apparatuses and methods to abuse them.

6. CHALLENGES OF OPEN SOURCE INTELLIGENCE

All intelligence gathering methodologies have few confinements, and OSINT is not exempt from this rule. In this section, we will mention some of the difficulties we face in OSINT gathering.

• Sheer volume of data: Collecting OSINT will produce a huge amount of data that must be analysed to be considered to bevaluable. Of course, many automated tools exist for this purpose, and numerous governments and big companies have developed their own set of artificial intelligence tools and techniques to filter acquired data. However, the tremendous volume of data will remain a challenge for the OSINT gatherer.

• Reliability of sources: Keep in mind that OSINT sources, especially when used in the intelligence context, need to be confirmed thoroughly by classified sources before they can be trusted. Many governments broadcast inaccurate information to mislead the OSINT-gathering process.

• Human efforts: The sheer volume of data is taken into account asthe greatest challenge for OSINT collection. Humans need to view the output of automated tools to grasp whether the collected data is reliable and trustworthy; they also need to compare it with some classified data to assure its reliability and relevance. It will consume time and precious human resources.

7. OPEN SOURCE INTELLIGENCE TOOLS

Finding a workable pace the data is accessible is a certain something. An assortment of the data is second and making an examination or insight out of them is the third. The data can be assembled physically too yet that will take the time that can rather be utilized in the later stages. Instruments can assist us with get-together the information from several destinations in minutes and in this manner facilitating the assortment stage. Let us state that the assignment is to recognize whether a username is available and assuming this is the case, on which every social medium sites. One route is to sign in to all the internet based life sites (I wager you don't have the foggiest idea about every one of them!) and afterwards testing the username in that. Another route is to utilize an open-source apparatus that is associated with different sites beyond what we can recall and checks the usernames nearness on every one of the sites without a moment's delay. This is done just in short order. Run numerous devices to assemble all objective related data that can be associated and utilized later.

7.1 Maltego

Maltego is created by Paterva and is utilized by security experts and legal specialists for gathering and dissecting open-source insight. It can without much of a stretch gather Information from different sources and utilize different changes to produce graphical outcomes. The changes are inbuilt and can likewise be tweaked dependent on the prerequisite. Maltego is written in Java and comes pre-bundled in Kali Linux. To utilize Maltego, client enrollment is required, the enlistment is free. When enrolled clients can utilize this apparatus to make the advanced impression of the objective on the web.

Written in Java, this device is likewise a piece of the Kali Linux pack. Maltego is proficient in finding the impressions of any objects on the Internet. Information is gathered from different sources and showed graphically. Maltego is utilized by law requirement, legal sciences, and security experts for its speedy and effective information assortment and representation. It is accessible in a network and a business adaptation. The people group form is constrained and can't be utilized industrially and just returns a predetermined number of substances. Maltego helps discover an association between different substances associated with the Internet. The graphical design makes it simple to see these connections between two elements that could possibly be legitimately connected to one another.

What does Maltego do?

The focal point of Maltego is examining genuine connections between data that is freely available on the Internet. This incorporates footprinting Internet framework just as discovering data about the individuals and association who claim it.

Maltego can be utilised to determine the relationships between the following entities:

- People.
  Names.
  Email addresses.
  Aliases.
- Groups of people (social networks).

- Companies.
- Organizations.
- Web sites.
- Internet infrastructure such as:
  Domains.
  DNS names.
  Netblocks.
  IP addresses.
- Affiliations.
- Documents and files.

Associations between these snippets of data are discovered utilizing open-source knowledge (OSINT) systems by questioning sources, for example, DNS records, whois records, web search tools, interpersonal organizations, different online APIs and extricating meta information.

Maltego gives brings about a wide scope of graphical designs that take into account grouping of data which makes seeing connections moment and exact – this makes it conceivable to see shrouded associations regardless of whether they are three or four degrees of partition separated.

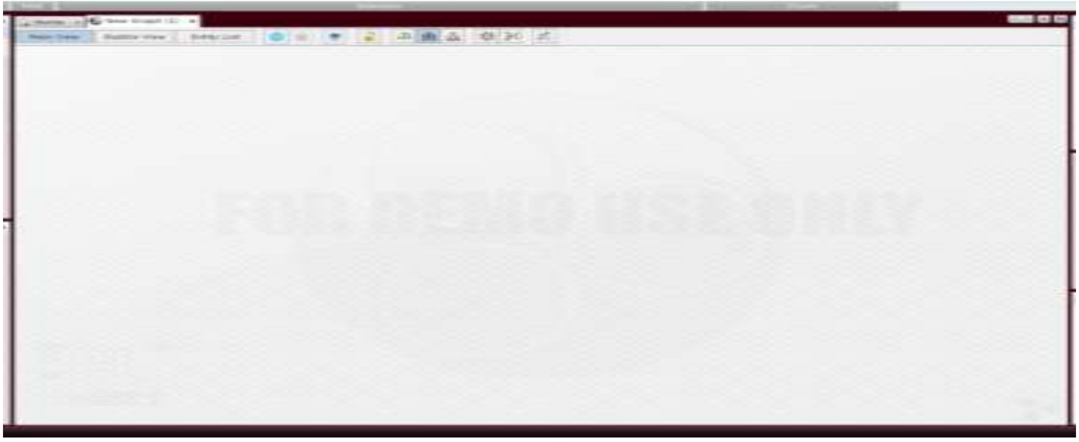
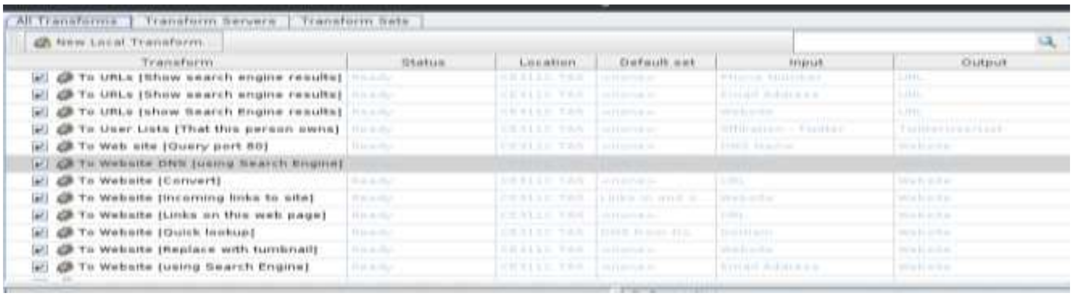


Fig 1: Maltego

The Transform group "IP proprietor detail" is likewise extremely helpful to discover pieces of information like email addresses, substances (individual names) and telephone numbers, so it is a smart thought to investigate the others changes inside it.



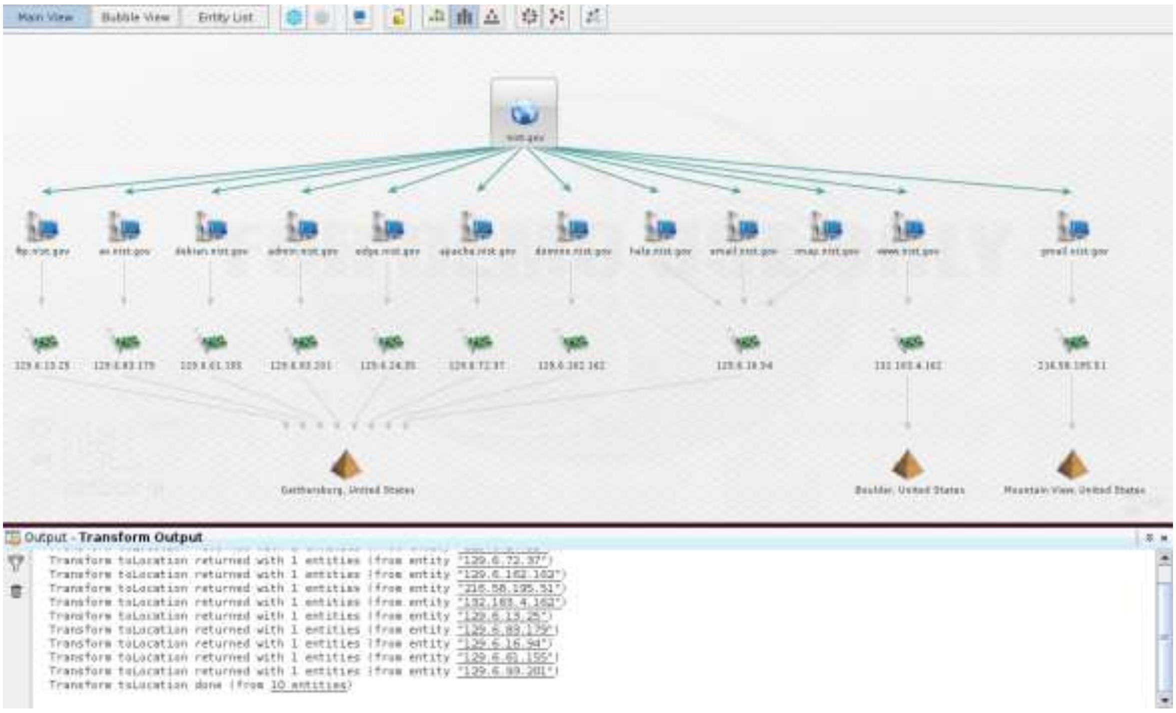
Transform	Status	Location	Default set	Input	Output
☑ To URLs [Show search engine results]	Ready	[illegible]	[illegible]	Phone Number	URL
☑ To URLs [Show search engine results]	Ready	[illegible]	[illegible]	Email Address	URL
☑ To URLs [show Search Engine results]	Ready	[illegible]	[illegible]	Website	URL
☑ To User Lists [That this person owns]	Ready	[illegible]	[illegible]	Affiliation - Twitter	TwitterUserList
☑ To Web site [Query port 80]	Ready	[illegible]	[illegible]	DNS Name	Website
☑ To Website DNS [using Search Engine]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
☑ To Website [Convert]	Ready	[illegible]	[illegible]	URL	Website
☑ To Website [incoming links to site]	Ready	[illegible]	[illegible]	Website	Website
☑ To Website [Links on this web page]	Ready	[illegible]	[illegible]	URL	Website
☑ To Website [Quick lookup]	Ready	[illegible]	[illegible]	Domain	Website
☑ To Website [Replace with tumbnail]	Ready	[illegible]	[illegible]	Website	Website
☑ To Website [using Search Engine]	Ready	[illegible]	[illegible]	Email Address	Website

Fig 2: Transform Manager

In this setup menu there are additionally revealed the various changes stacked in Maltego with their Status, Transform Server Location, Default Set, Input and Output pieces of information.

Remember that some changes are more intrusive than others: for instance, it is conceivable to find sites questioning straightforwardly port 80 utilizing the change To Web webpage [Query port 80].

This is the subsequent chart with an attention on the sites:



```
Transform toLocation returned with 1 entities (from entity "129.6.72.37")
Transform toLocation returned with 1 entities (from entity "129.6.162.162")
Transform toLocation returned with 1 entities (from entity "216.58.195.51")
Transform toLocation returned with 1 entities (from entity "132.163.4.162")
Transform toLocation returned with 1 entities (from entity "129.6.13.25")
Transform toLocation returned with 1 entities (from entity "129.6.83.179")
Transform toLocation returned with 1 entities (from entity "129.6.16.94")
Transform toLocation returned with 1 entities (from entity "129.6.61.155")
Transform toLocation returned with 1 entities (from entity "129.6.93.201")
Transform toLocation done (from 10 entities)
```

Fig 3: Graphical Chart Results

Another fascinating change is the one named Files and Documents from Domain: this will scan for records and archives inside the given space with the expansions revealed in the design menu; by tapping on the hub speaking to a record we can get information's about the inquiry used to discover it with the report download connect.

Fig 4: Domain Target using Email

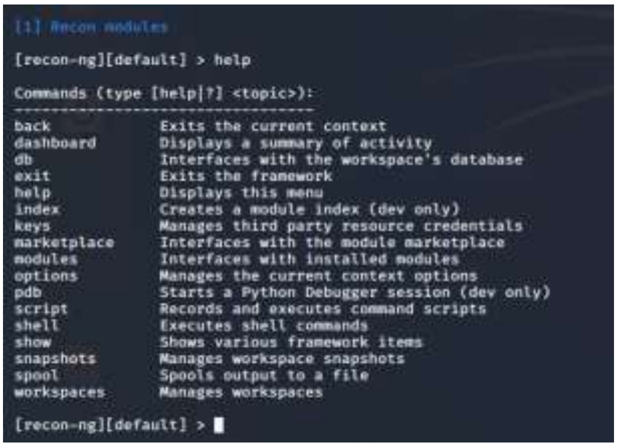
The accompanying advance could be to discover email delivers identified with the objective space by utilizing Email addresses from Domain change on the "Area" object; at that point, we could rush To Person change on the "Email" item to get individual personality identified with that email address or To Phone number [using Search Engine] change to attempt a telephone number revelation.

7.2 Recon-ng

Recon-ng is a fully-featured Web Reconnaissance framework written in Python. Complete with independent modules, database interaction, built-in convenience functions, interactive help, and command completion, Recon-ng provides a powerful environment in which open source web-based reconnaissance can be performed quickly and thoroughly.Recon-ng has a look and feels similar to the Metasploit Framework, diminishing the learning curve for leveraging the framework. However, it is quite different. Recon-ng is designed exclusively for web-based open source reconnaissance. Some of the excellent modules, such as google-site-web and bing-domain-web, are used to find further domains related to the first initial target domain. The result of these domains will be all the indexed domains to the search engines. Another catchy module is bing_linkedin_cache that is mainly used to fetch the details of the email addresses related to the domain. This module can also be used to leverage in performing social engineering.

7.4 Using recon-ng

From the console, it is easy to get help and get started with your recon.

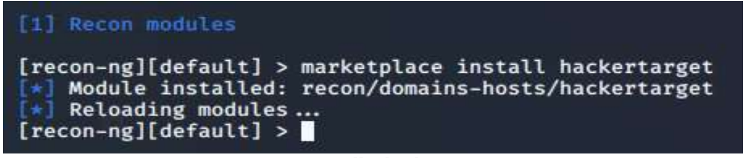


```
[1] Recon modules

[recon-ng][default] > help

Commands (type [help|?] <topic>):
---------------------------------
back            Exits the current context
dashboard       Displays a summary of activity
db              Interfaces with the workspace's database
exit            Exits the framework
help            Displays this menu
index           Creates a module index (dev only)
keys            Manages third party resource credentials
marketplace     Interfaces with the module marketplace
modules         Interfaces with installed modules
options         Manages the current context options
pdb             Starts a Python Debugger session (dev only)
script          Records and executes command scripts
shell           Executes shell commands
show            Shows various framework items
snapshots       Manages workspace snapshots
spool           Spools output to a file
workspaces      Manages workspaces

[recon-ng][default] > 
```

Fig 5: Recon Modules

How to:

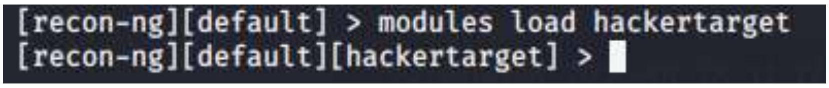
Firstly let's use the hacker target module to gather some subdomains. This uses the hackertarget.comAPI and hostname search.
Install module
Syntax to install is marketplace install hackertarget as seen below.



```
[1] Recon modules

[recon-ng][default] > marketplace install hackertarget
[*] Module installed: recon/domains-hosts/hackertarget
[*] Reloading modules ...
[recon-ng][default] > 
```

Fig 6: To install a hackertarget

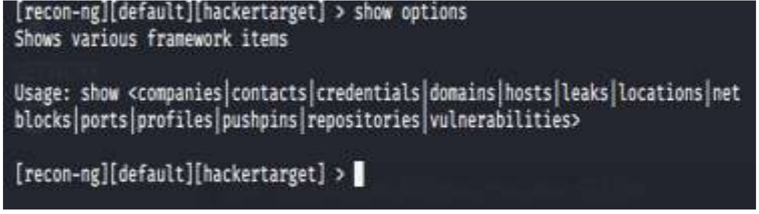
Load module



```
[recon-ng][default] > modules load hackertarget
[recon-ng][default][hackertarget] > 
```

Fig 7: To load modules

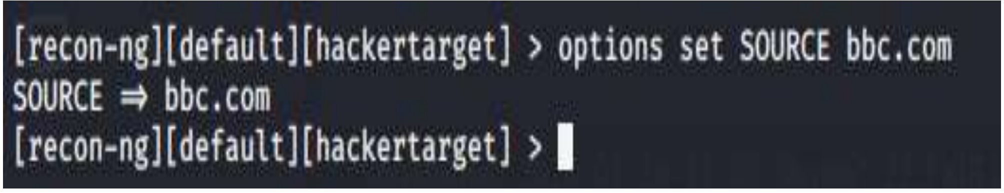
Set source
Now set the source. Currently set at default (see below)



```
[recon-ng][default][hackertarget] > show options
Shows various framework items

Usage: show <companies|contacts|credentials|domains|hosts|leaks|locations|net
blocks|ports|profiles|pushpins|repositories|vulnerabilities>

[recon-ng][default][hackertarget] > 
```

Fig 8: To show options

Syntax options set SOURCE bbc.com
I am using bbc.com as an example domain.

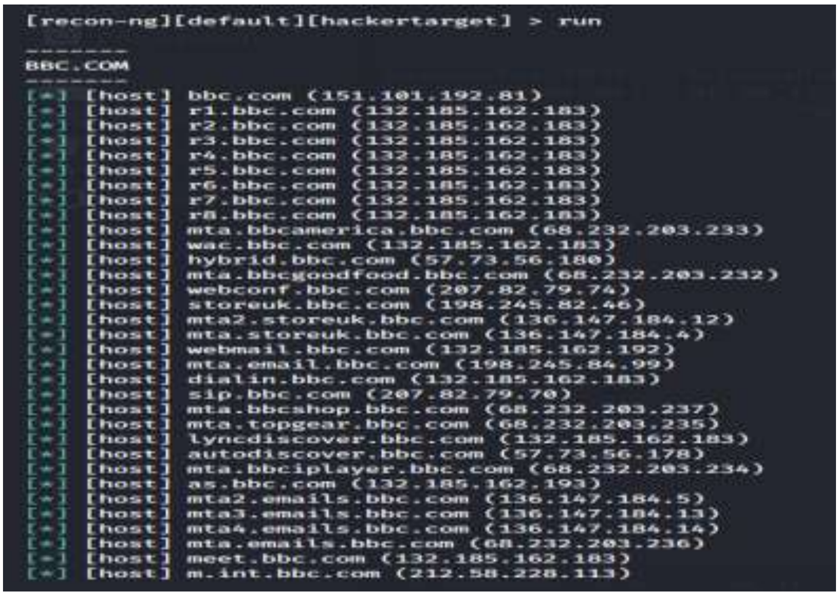


```
[recon-ng][default][hackertarget] > options set SOURCE bbc.com
SOURCE ⇒ bbc.com
[recon-ng][default][hackertarget] > 
```

Fig 9: Domain example

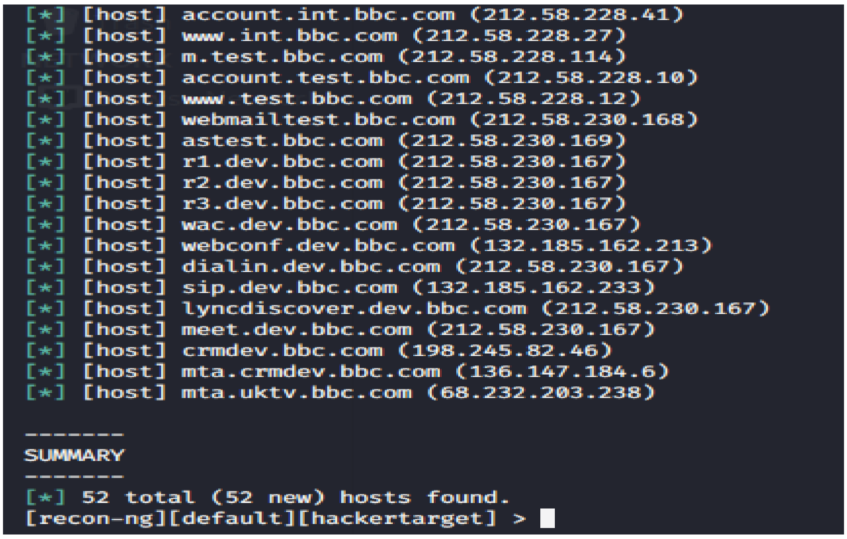
Run the module

Type run to execute the module.



```
[recon-ng][default][hackertarget] > run

-------
BBC.COM
-------
[*] [host] bbc.com (151.101.192.81)
[*] [host] r1.bbc.com (132.185.162.183)
[*] [host] r2.bbc.com (132.185.162.183)
[*] [host] r3.bbc.com (132.185.162.183)
[*] [host] r4.bbc.com (132.185.162.183)
[*] [host] r5.bbc.com (132.185.162.183)
[*] [host] r6.bbc.com (132.185.162.183)
[*] [host] r7.bbc.com (132.185.162.183)
[*] [host] r8.bbc.com (132.185.162.183)
[*] [host] mta.bbcamerica.bbc.com (68.232.203.233)
[*] [host] wac.bbc.com (132.185.162.183)
[*] [host] hybrid.bbc.com (57.73.56.180)
[*] [host] mta.bbcgoodfood.bbc.com (68.232.203.232)
[*] [host] webconf.bbc.com (207.82.79.74)
[*] [host] storeuk.bbc.com (198.245.82.46)
[*] [host] mta2.storeuk.bbc.com (136.147.184.12)
[*] [host] mta.storeuk.bbc.com (136.147.184.4)
[*] [host] webmail.bbc.com (132.185.162.192)
[*] [host] mta.email.bbc.com (198.245.84.99)
[*] [host] dialin.bbc.com (132.185.162.183)
[*] [host] sip.bbc.com (207.82.79.70)
[*] [host] mta.bbcshop.bbc.com (68.232.203.237)
[*] [host] mta.topgear.bbc.com (68.232.203.235)
[*] [host] lyncdiscover.bbc.com (132.185.162.183)
[*] [host] autodiscover.bbc.com (57.73.56.178)
[*] [host] mta.bbciplayer.bbc.com (68.232.203.234)
[*] [host] as.bbc.com (132.185.162.193)
[*] [host] mta2.emails.bbc.com (136.147.184.5)
[*] [host] mta3.emails.bbc.com (136.147.184.13)
[*] [host] mta4.emails.bbc.com (136.147.184.14)
[*] [host] mta.emails.bbc.com (68.232.203.236)
[*] [host] meet.bbc.com (132.185.162.183)
[*] [host] m.int.bbc.com (212.58.228.113)
```

Fig 10a: Run the module



```
[*] [host] account.int.bbc.com (212.58.228.41)
[*] [host] www.int.bbc.com (212.58.228.27)
[*] [host] m.test.bbc.com (212.58.228.114)
[*] [host] account.test.bbc.com (212.58.228.10)
[*] [host] www.test.bbc.com (212.58.228.12)
[*] [host] webmailtest.bbc.com (212.58.230.168)
[*] [host] astest.bbc.com (212.58.230.169)
[*] [host] r1.dev.bbc.com (212.58.230.167)
[*] [host] r2.dev.bbc.com (212.58.230.167)
[*] [host] r3.dev.bbc.com (212.58.230.167)
[*] [host] wac.dev.bbc.com (212.58.230.167)
[*] [host] webconf.dev.bbc.com (132.185.162.213)
[*] [host] dialin.dev.bbc.com (212.58.230.167)
[*] [host] sip.dev.bbc.com (132.185.162.233)
[*] [host] lyncdiscover.dev.bbc.com (212.58.230.167)
[*] [host] meet.dev.bbc.com (212.58.230.167)
[*] [host] crmdev.bbc.com (198.245.82.46)
[*] [host] mta.crmdev.bbc.com (136.147.184.6)
[*] [host] mta.uktv.bbc.com (68.232.203.238)

-------
SUMMARY
-------
[*] 52 total (52 new) hosts found.
[recon-ng][default][hackertarget] >
```

Fig 10b : Output

Recon-ngis an effective tool to perform reconnaissance on the target.

7.5 Google Dorks

A Google Dork, otherwise called Google Dorking or Google hacking, is a significant asset for security analysts. For the normal individual, Google is only an internet searcher used to discover content, pictures, recordings, and news. Be that as it may, in the infosec world, Google is a helpful hacking instrument.

Google "Dorking" is the act of utilizing Google to discover powerless applications and servers by utilizing local Google web index abilities. Except if you square explicit assets from your site utilizing a robots.txt record, Google lists all the data that is available on any site. Coherently, after some time any individual on the planet can get to that data in the event that they comprehend what to look for.while this data is freely accessible on the Internet, and it is given and urged to be utilized by Google on a legitimate premise, individuals with an inappropriate goal could utilize this data to hurt your online nearness.

Google's search engine has its own built-in query language. This is the following list of queries can be run to find a list of files, and find the information, track people, get information about SEO backlinks, build email lists, and of course, discover web vulnerabilities.

Google Dorks Operators

This is the most popular Google Dorks and what they do.
cache: this dork will show you the cached version of any website, e.g. cache: securitytrails.com
allintext: searches for specific text contained on any web page, e.g. allintext: hacking tools
allintitle: exactly the same as allintext, but will show pages that contain titles with X characters, e.g. allintitle:"Security Companies"
allinurl: it can be used to fetch results whose URL contains all the specified characters, e.g: allinurl client area
filetype: used to search for any kind of file extensions, for example, if you want to search for jpg files you can use: filetype: jpg
inurl: this is exactly the same as allinurl, but it is only useful for one single keyword, e.g. inurl: admin
intitle: used to search for various keywords inside the title, for example, intitle:security tools will search for titles beginning with “security” but “tools” can be somewhere else in the page.
inanchor: this is useful when you need to search for an exact anchor text used on any links, e.g. inanchor:"cybersecurity"
intext: useful to locate pages that contain certain characters or strings inside their text, e.g. intext:"safe internet"
link: will show the list of web pages that have links to the specified URL, e.g. link: microsoft.com
site: will show you the full list of all indexed URLs for the specified domain and subdomain, e.g. site:securitytrails.com

7.6 The Harvester

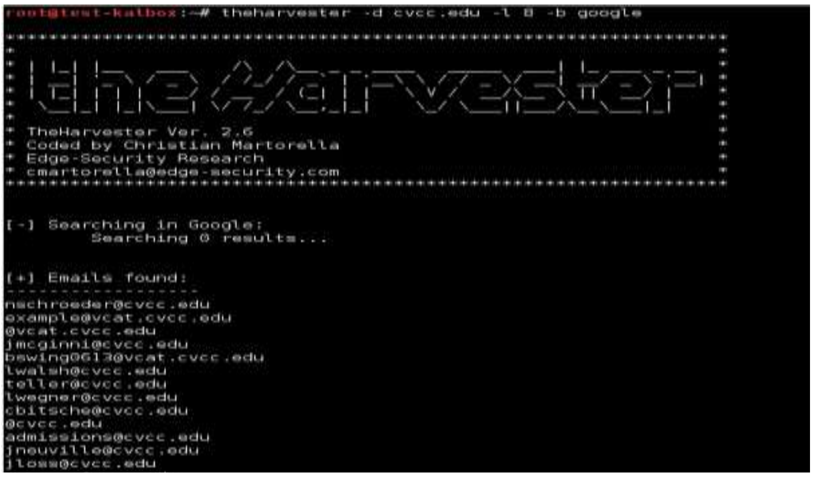
The Harvester is an astonishing tool for discovering emails, subdomains, IP, etc. from different public data. The goal of this program is to assemble emails, subdomains, hosts, employee names, open ports and banners from different public sources like search engines, PGP key servers and SHODAN computer database.

This tool is expected to help Penetration testers in the early periods of the penetration test in order to understand the customer footprint on the Internet. It is also useful for anyone that wants to know what an attacker can see about their organization.

This is a complete rewrite of the tool with new features like:

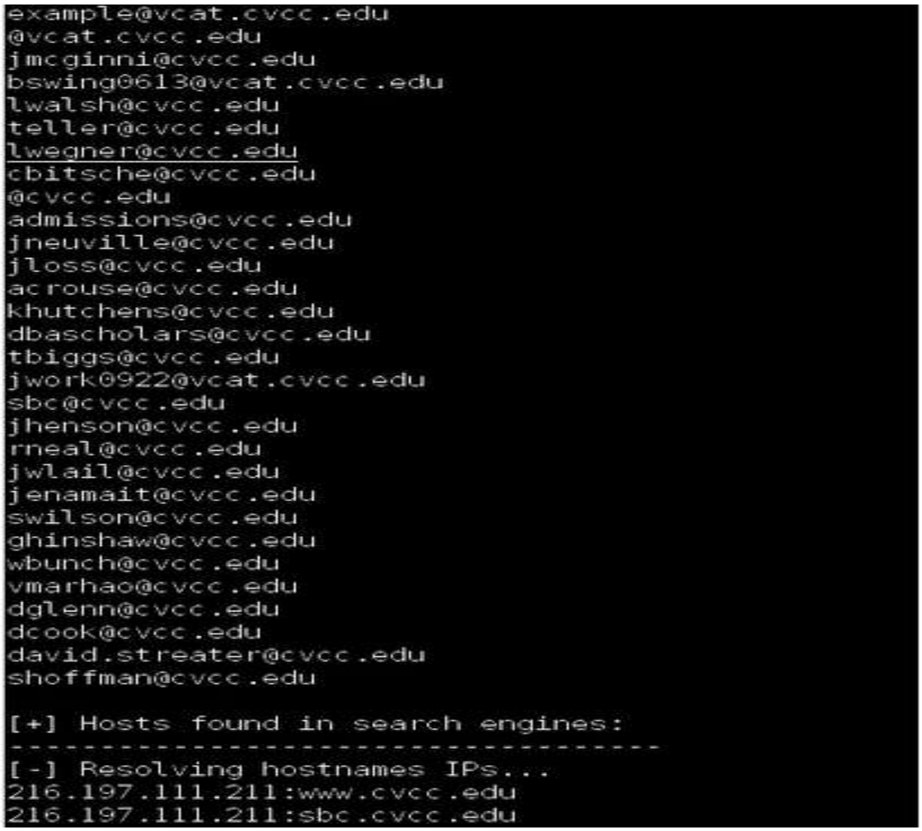
- Time delays between request
- All sources search
- Virtual host verifier
- Active enumeration (DNS enumeration, Reverse lookups, TLD expansion)
- Integration with SHODAN computer database, to get the open ports and banners
- Save to XML and HTML
- Basic graph with stats
- New sources

How to Find Email ID’s in Domain:



```
root@test-kalbox:~# theharvester -d cvcc.edu -l 8 -b google

* TheHarvester Ver. 2.6                                         *
* Coded by Christian Martorella                                 *
* Edge-Security Research                                        *
* cmartorella@edge-security.com                                 *


[-] Searching in Google:
        Searching 0 results...


[+] Emails found:
------------------
nschroeder@cvcc.edu
example@vcat.cvcc.edu
@vcat.cvcc.edu
jmcginni@cvcc.edu
bswing0613@vcat.cvcc.edu
lwalsh@cvcc.edu
teller@cvcc.edu
lwegner@cvcc.edu
cbitsche@cvcc.edu
@cvcc.edu
admissions@cvcc.edu
jneuville@cvcc.edu
jloss@cvcc.edu
```

Fig 11a: Enter a domain name



```
example@vcat.cvcc.edu
@vcat.cvcc.edu
jmcginni@cvcc.edu
bswing0613@vcat.cvcc.edu
lwalsh@cvcc.edu
teller@cvcc.edu
lwegner@cvcc.edu
cbitsche@cvcc.edu
@cvcc.edu
admissions@cvcc.edu
jneuville@cvcc.edu
jloss@cvcc.edu
acrouse@cvcc.edu
khutchens@cvcc.edu
dbascholars@cvcc.edu
tbiggs@cvcc.edu
jwork0922@vcat.cvcc.edu
sbc@cvcc.edu
jhenson@cvcc.edu
rneal@cvcc.edu
jwlail@cvcc.edu
jenamait@cvcc.edu
swilson@cvcc.edu
ghinshaw@cvcc.edu
wbunch@cvcc.edu
vmarhao@cvcc.edu
dglenn@cvcc.edu
dcook@cvcc.edu
david.streater@cvcc.edu
shoffman@cvcc.edu

[+] Hosts found in search engines:
------------------------------------
[-] Resolving hostnames IPs...
216.197.111.211:www.cvcc.edu
216.197.111.211:sbc.cvcc.edu
```

Fig 11b: Email ID’s found

8. CONCLUSION

In this paper several prospects about OSNIT tools like Maltego, Recon-ng, TheHarvester and Google Dorks and some basic Techniques which we can gather some information using open source tools.And by using some vulnerable website, Mail-id to collect basic information by using OSNIT tools. Also, we can use these tools in cybersecurity and ethical Hacking to gather information.

9. REFERENCES

1. "Open Source Intelligence" (PDF).

2. ^ McLaughlin, Michael (June 2012). "Using open-source intelligence software for cybersecurity intelligence". ComputerWeekly.com. Retrieved 2018-06-29.

3. https://securitytrails.com/blog/what-is-osint-how-can-i-make-use-of-it

4. https://www.sentinelone.com/blog/what-is-osint-how-is-it-used/

5. https://geekflare.com/osint-tools/

6. https://www.csoonline.com/article/3445357/6-top-osint-tools-find-sensitive-public-info-before-hackers-do.html

7. https://books.google.com/books?id=AqNiDwAAQBAJ&pg=PA16&lpg=PA16&dq=%22However,+the+tremendous+volume+of+data+will+remain+a+challenge+for+the+OSINT+gatherer.%22&source=bl&ots=fIPzWDMSCi&sig=ACfU3U2gSkrNWUOAcbk2xyS0gUty6lJq7A&hl=en&sa=X&ved=2ahUKEwiO3PXi9cPnAhXxoFsKHUCbAbsQ6AEwAHoECAcQAQ.