Information Gathering with theHarvester

Dedicated to Billy (...)

March 11, 2017 | 76.5k views

theHarvester is a tool for gathering publicly searchable information on your targets which could be anything from individuals to websites to companies. theHarvester can find e-mail accounts, subdomain names, virtual hosts, open ports and banners, and employee names from different public sources.

It is an easy-to-use open source tool built in python by Christian Martorella. It is one of the pre-installed packages in Kali Linux and a part of almost every hacker's toolkit.

Now let's stop talking about it and start using it. Since it is just a python program, theHarvester can run on anything that can run python. I recommended you use Kali Linux (since it's built in), but if you want to, you can install it on Windows as well.

How to install theHarvester on Windows?

- Install python
- Head over to the github repository and download theHarvester's source code.
- We need to install a dependency, the `requests` library which is needed by theHarvester. Run the following command:

```
pip install requests
```

- And that's it. You should now be able to run theHarvester by referencing the python file from the command line:

```
theHarvester.py
```

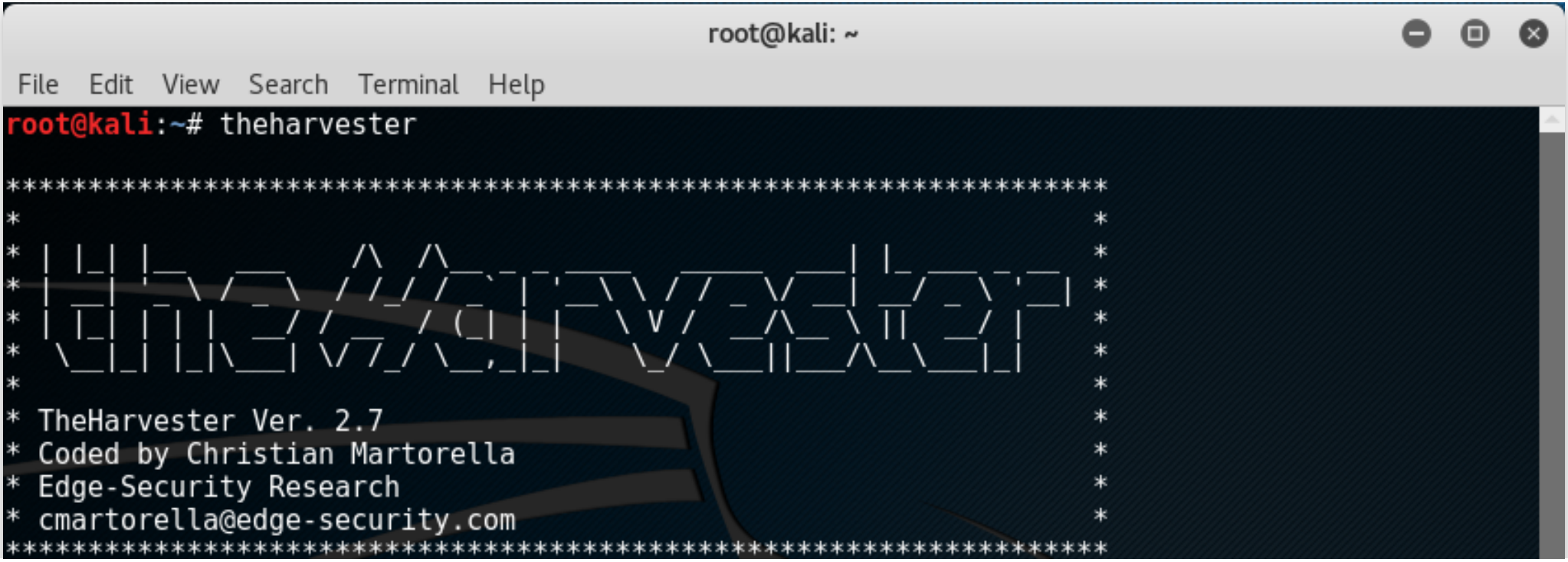
How to use theHarvester?

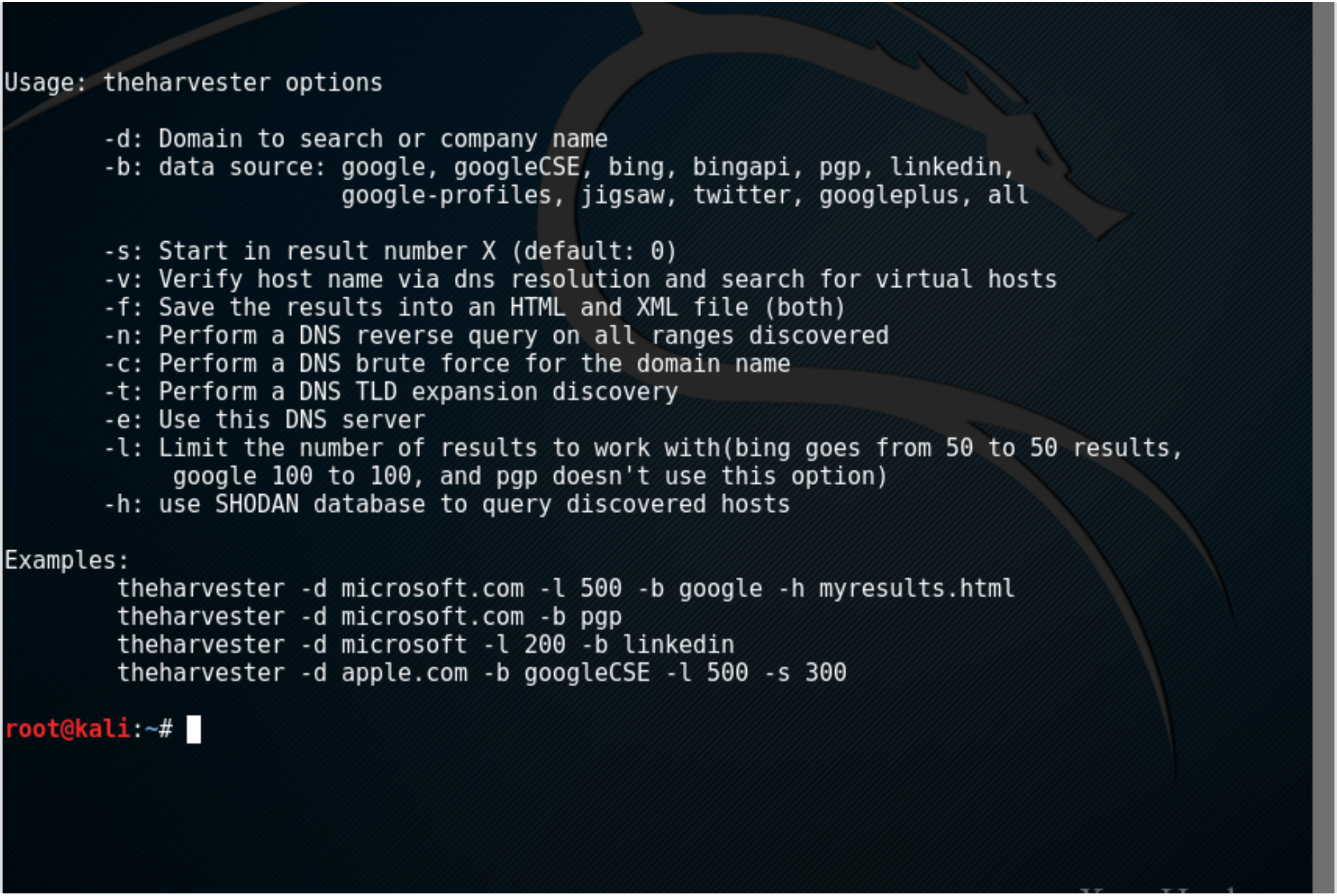
- First of all ensure that theHarvester is installed by running the following in a Kali terminal:

```
theharvester
```

P.S: For windows, use `theHarvester.py` instead of `theharvester`.

You should see something like this:

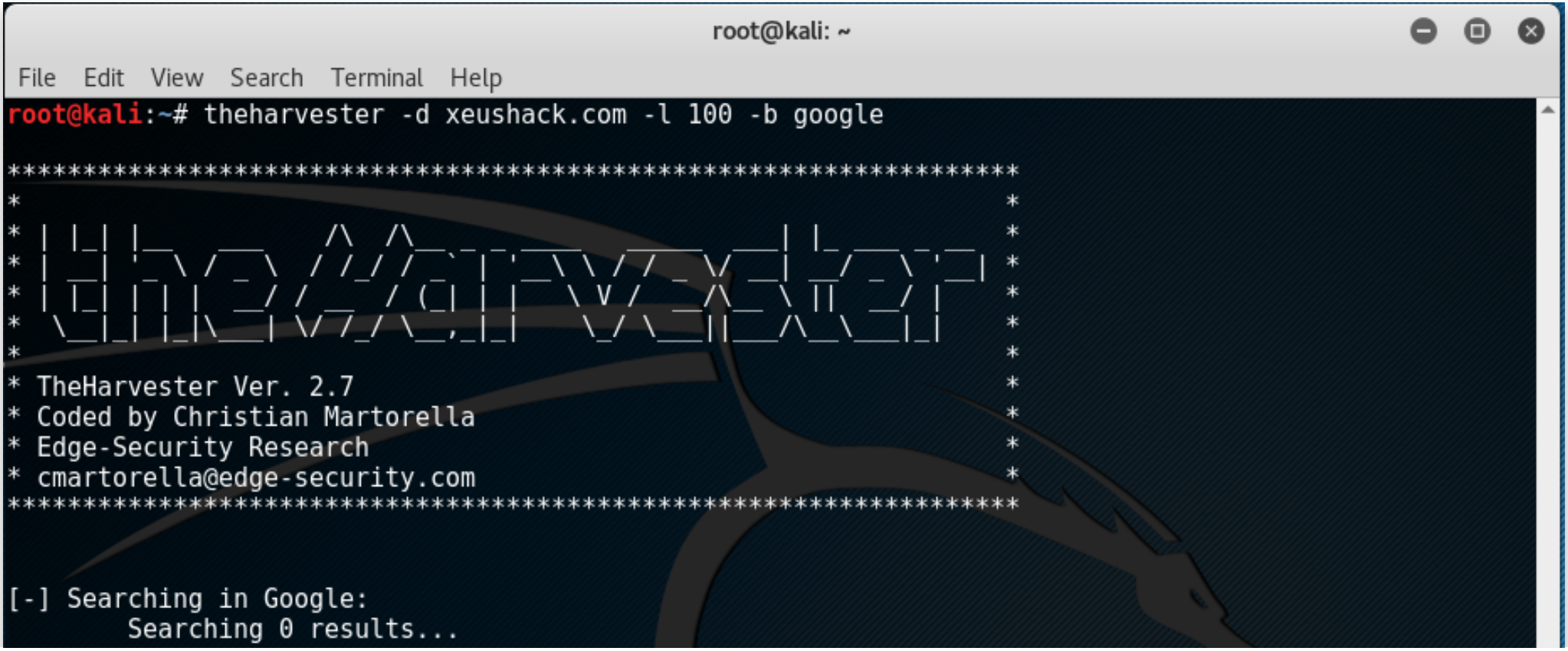


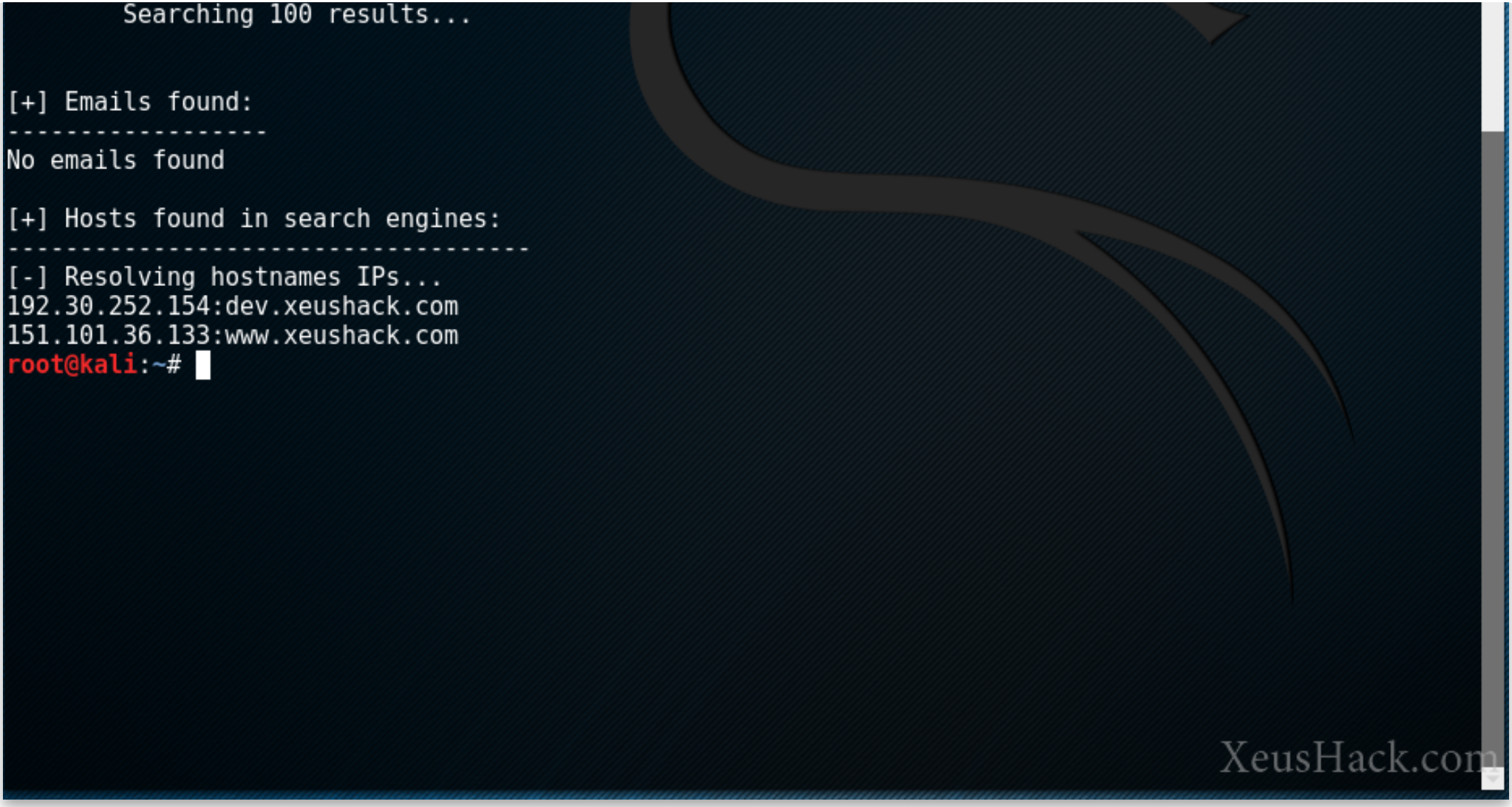


We see a bunch of options that we can use with the command along with a few examples. Let's try running a command now. I want to see what comes up for `XeusHack.com` using Google.

```
theharvester -d xeushack.com -l 100 -b google
```

This outputs:



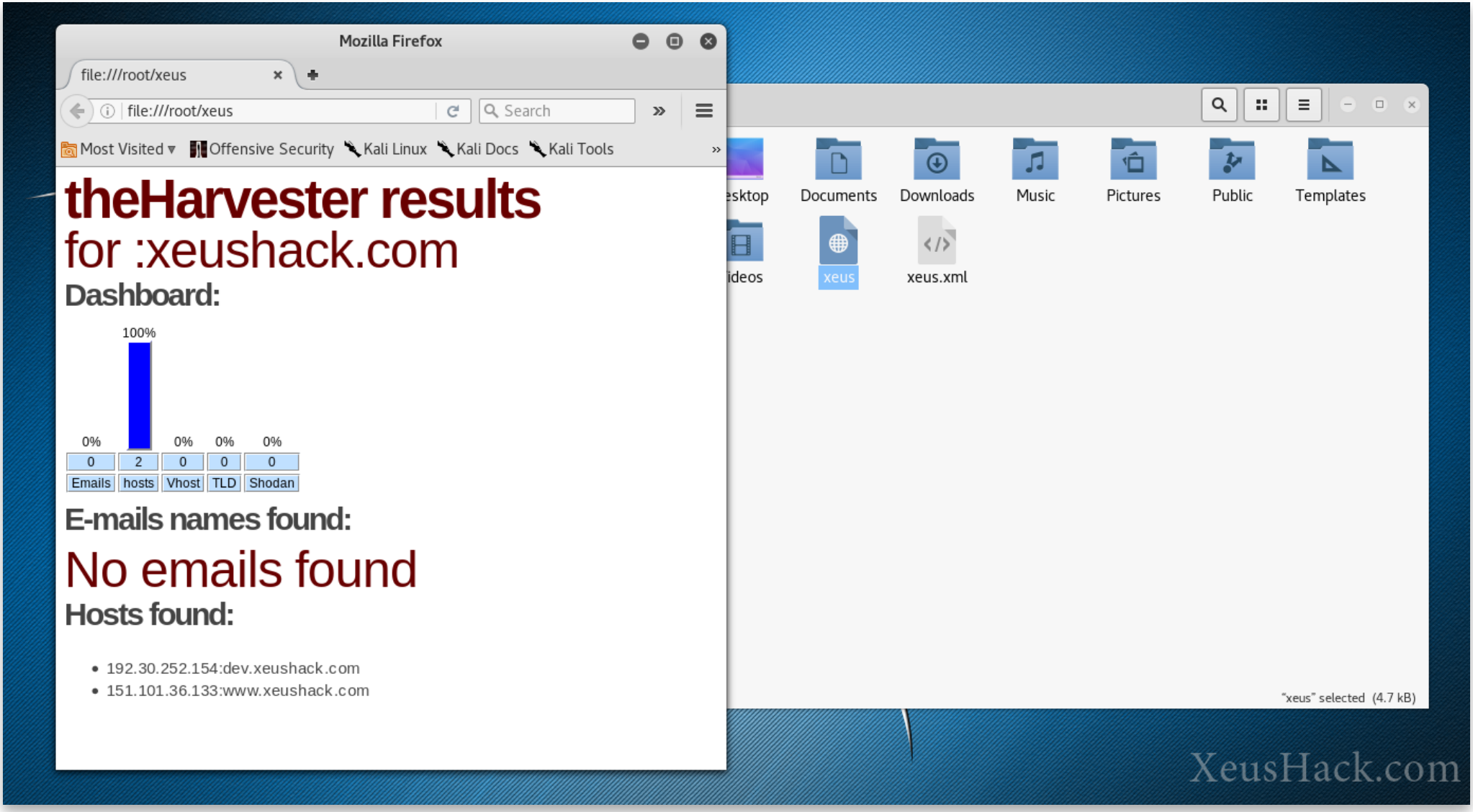


Hmm, nothing except the server's IP address. *Perfect*.

By using the `-f` parameter we can also save a nicely formatted output to an HTML or XML file:

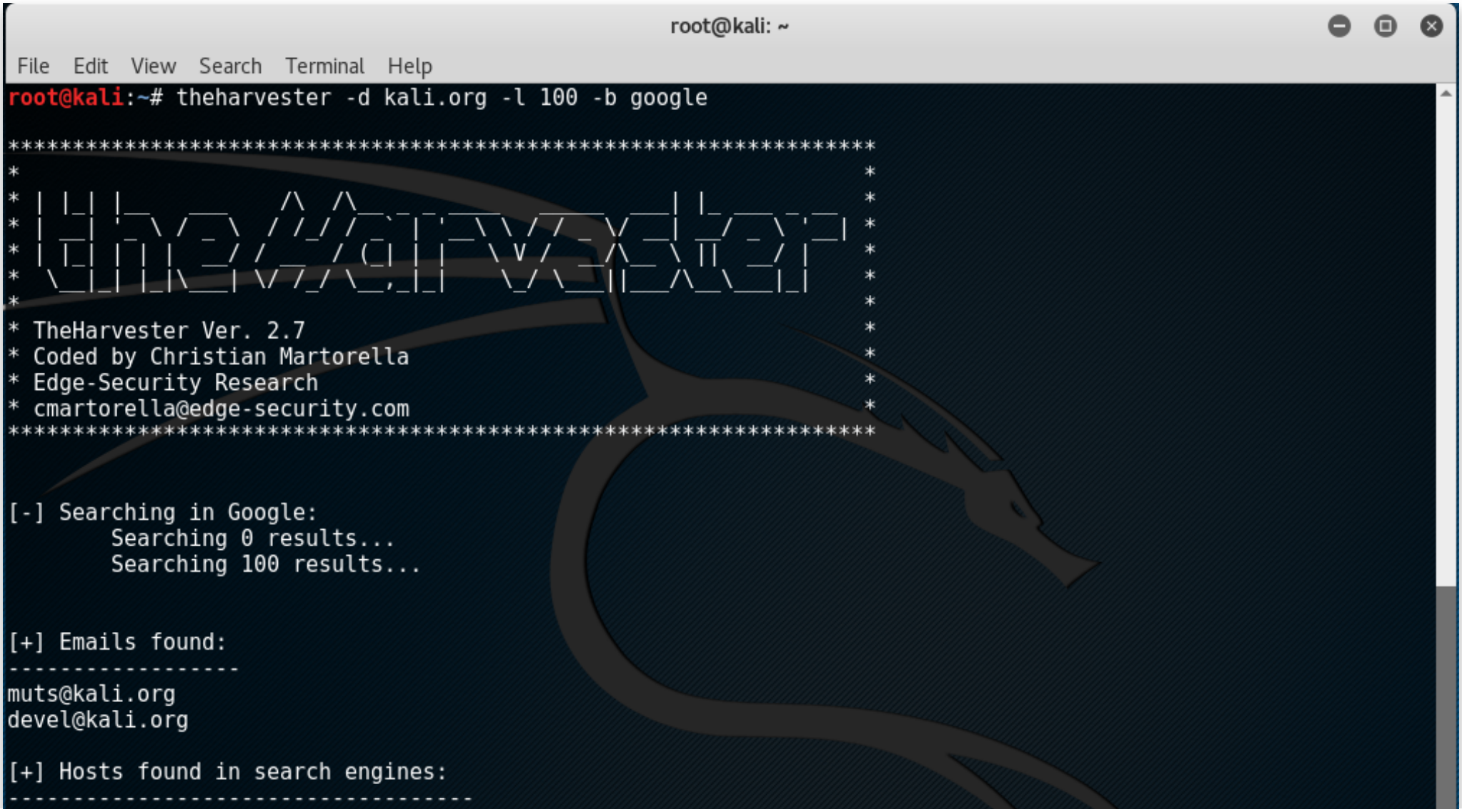
```
theharvester -d xeushack.com -l 100 -b google -f xeus
```

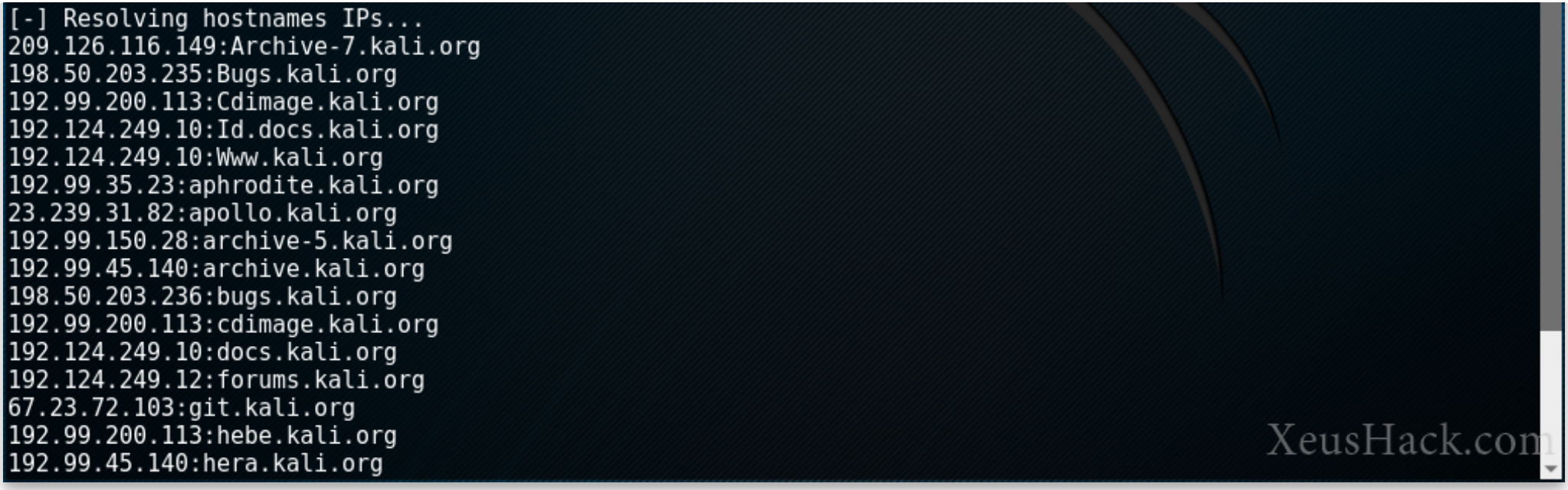
This saves a file called `xeus.html` in the same directory where the terminal is. Here's what it looks like:



Now let's try another, this time `kali.org`.

```
theharvester -d kali.org -l 100 -b google
```





Now we got a bit more data. We see a couple of email addresses and a bunch of subdomains. Now this isn't all that surprising, all of these are publicly searchable and the developers of Kali.org want them to be searchable. You can bet that `kali.org` is a pretty secure website, but the same might not be said about some small company's website. An indie developer or a small business may not have taken all the precautions to keep their website safe. In this case, the information that theHarvester turns up could be used to attack them.

To understand how theHarvester could be used in a real-life setting, let us drift away for a moment and read a short story. This is the story of **Billy**.

The Misadventures of Billy

Who's Billy you say? Billy was just an average guy with big dreams. Billy was interested in learning how to set up his own website. So he googled "how to create a website" and arrived at WordPress.org. He got a domain `something.com` along with hosting. But Billy was just starting out and didn't want to make the website live until he was satisfied with it. So instead he sets up WordPress on a subdomain `test.something.com`.

A few weeks go by and Billy is now happy with his website. Ready to take the world by storm, Billy copies his WordPress installation over to `something.com` and lives happily ever after.

Not so fast. If you're familiar with WordPress you'll know that it's often plagued by numerous security holes. To keep your WordPress website safe, you need to constantly update the core as well as all the plugins you use.

But Billy forgot about the instance of wordpress still running on `test.something.com`.

Now a hacker comes along and runs:

```
theharvester -d something.com -l 100 -b google
```

He sees Billy's suspicious subdomain, goes to it and finds a now outdated and unsecure installation of WordPress. And the rest is history. Because of one small mistake, the hacker gained access to the server and wiped away everything.

All of Billy's hard work was gone. Just like that.

Back to theHarvester…

The moral of the story is that all a hacker needs is one tiny oversight, the smallest of security holes, and a system is rendered defenseless. This is what makes theHarvester useful. The reconnaissance stage of hacking is devoted to following this trail of breadcrumbs that lead to a vulnerability.

Information is our best weapon. Information about a target can mean the difference between quickly exploiting a system and a long and fruitless hacking attempt. theHarvester can also help us with social engineering. For example, it could give us the private email addresses of a company. This could be used for a phishing campaign or even for a direct attack if the company isn't so strict on password strength.

I recommend you play around with theHarvester, try out all the options, look up your favorite websites and see what you find. Get used to it and make it a part of your go-to hacking toolset. It should definitely be one of the first tools you run when getting to know a new target.

Below I've included some information about the sources of data used by theHarvester, you'll likely find this useful when working with it:

Passive Sources

- **Google:** Google search engine - www.Google.com

- **GoogleCSE:** Google custom search engine
- **Google-Profiles:** Google search engine, specific search for Google profiles
- **Bing:** microsoft search engine - www.bing.com
- **Bing API:** microsoft search engine, through the API (you need to add your Key in the discovery/bingsearch.py file)
- **DogPile:** Dogpile search engine - www.dogpile.com
- **PGP:** pgp key server - mit.edu
- **Linkedin:** Google search engine, specific search for Linkedin users
- **vhost:** Bing virtual hosts search
- **Twitter:** twitter accounts related to an specific domain (uses Google search)
- **Google+:** users that works in target company (uses Google search)
- **Yahoo:** Yahoo search engine
- **Baidu:** Baidu search engine
- **Shodan:** Shodan Computer search engine, will search for ports and banner of the discovered hosts (ShodanHQ.com)

Active Sources

- **DNS brute force:** this plugin will run a dictionary brute force enumeration
- **DNS reverse lookup:** reverse lookup of ip´s discovered in order to find hostnames
- **DNS TDL expansion:** TLD dictionary brute force enumeration

Modules that need API keys to work:

Since theHarvester makes use of third party information sources, some of these require you to have API keys to work. That is, you need to go and sign up for the specific service, register your app with them and they provide you with a key that lets you access the service. Only the following two need API keys:

- **GoogleCSE**: You need to create a Google Custom Search engine(CSE), and add your Google API key and CSE ID to the file: `discovery/GoogleCSE.py`
- **Shodan**: Add your API key in `discovery/shodansearch.py`

networking , attack , recon

Trending

WiFi android attack batch-file
books fun general government kali
legacy linux metasploit mobile
networking online osx
password privacy recon review

Get the **latest** content

Email address

Subscribe

security social social-engineering tools

tricks virus windows