

Introduction

People talk about the Dark Web as a mysterious place, hard to find and inaccessible to normal Internet users. In this paper we argue that there is no sharp border between the Surface Web and the Dark Web, and that there are indeed links from the former to the latter. Different parts of the Web thus exhibit varying degrees of shadiness, and can even be characterized by both actual content and what it links to. Conceptually, we might distinguish three levels of the Web, each portraying different characteristics:

- › Surface Web
 - » Freely accessible
 - » Indexed by Google, Bing, and others
 - » Mostly open, but sometimes behind pay walls
 - » Fairly stable, content is available from source for a long time
 - » Language (mostly) suited for traditional Natural Language Processing (NLP), and tools exist for extracting and analyzing data
- › Deep Web
 - » Often behind logins, but accessible to anyone registering
 - » Database driven, and therefore not indexed by search engines
 - » Sometimes by invitation only
 - » Mostly un-indexed by search engines such as Google and Bing
- › Dark Web
 - » Not indexed or searchable by Google, Bing etc.
 - » Often on other networks such as TOR², Freenet³, I2P⁴, etc.
 - » Frequently behind logins, accessible by invitation only
 - » Sometimes uses special language like slang, leetspeak etc. which is not easily analyzed by normal NLP tools.
 - » Volatile, with content that sometimes only stays available for a few minutes (in one study we did more than 10% of Pastebin posts were removed within 48 hours)

Information tends to seep out even from the darkest corners of the Web, if for no other reason than because that information has a value, which cannot be realized unless it is possible to find. Therefore it has to be marketed in some way. Wikipedia lists three uses of the Dark Web⁵ (or Darknet):

1. Out of privacy concerns or for fear by dissidents of political reprisal
2. To publish for criminal gain
3. To share media files (sometimes copyrighted files)

Clearly, our argument that information needs to be made accessible outside of the Dark Web to realize its (monetary) value holds for both (2) and (3) in this list. The Surface and Deep Web contain links to the Dark Web. How frequent is such information?

² <https://www.torproject.org/>

³ <https://freenetproject.org/>

⁴ <https://geti2p.net/en/>

⁵ [http://en.wikipedia.org/wiki/Darknet_\(overlay_network\)](http://en.wikipedia.org/wiki/Darknet_(overlay_network))

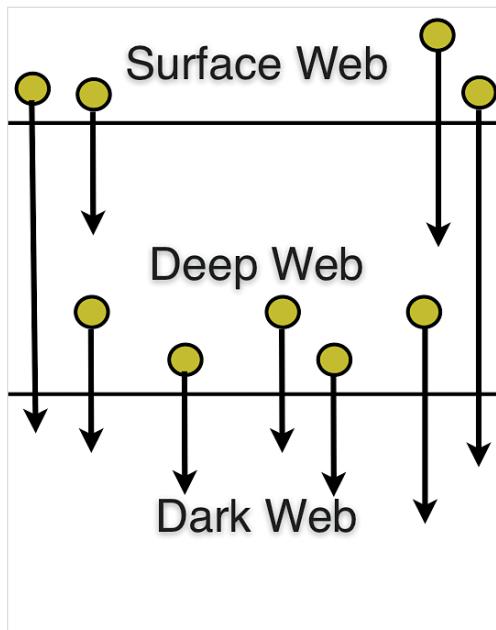
The Data Set

Recorded Future analyses more than 650,000 sources, ranging from government websites and big media to blogs, forums, paste sites, and social media. The Recorded Future index goes back more than six years in time, and has analyzed more than 8.3 billion references to facts, each an individual mention of an event in a document. Of these 8.3 billion references, more than 700 million come from paste sites (such as Pastebin, Slexy, CopyTaste), one of the source types we identify as bridging the Surface and Dark parts of the Web. Recorded Future's index contains 5 million references to malware, over 10 million references to IP addresses, over 11 million references to hashes, and 8.2 million references to cyber attack events. This is the wealth of data on which we base our threat intelligence analyses in this report.

A Journey to the Dark Side

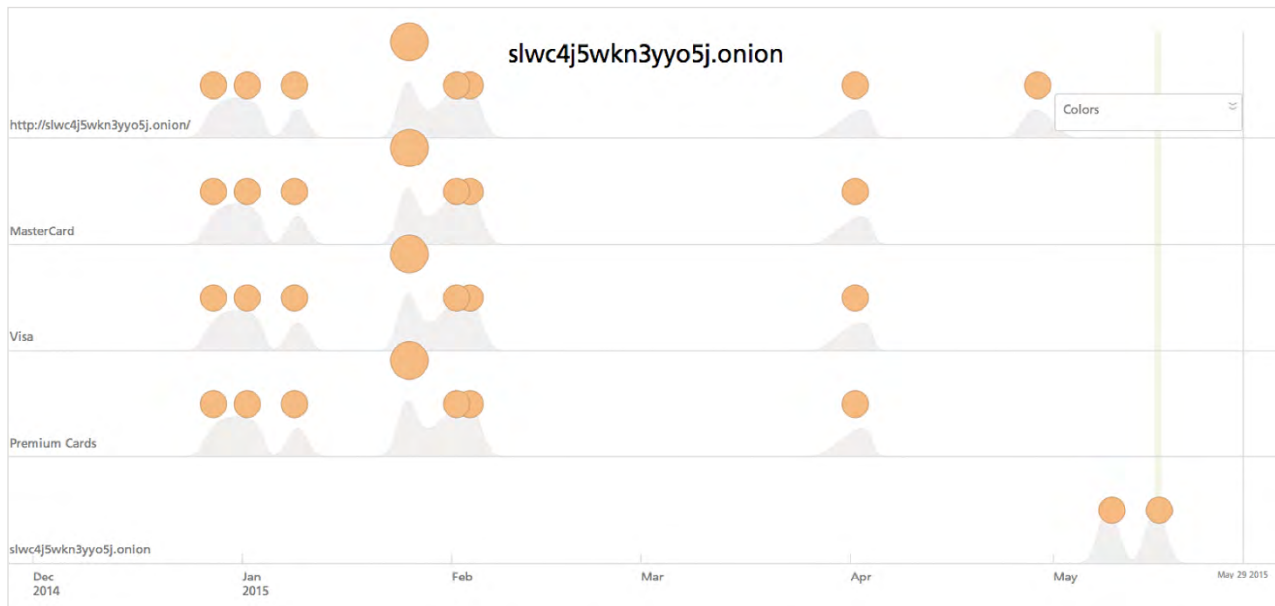
What does the linkage into the Dark Web look like, in reality? We used the Recorded Future index to investigate this. Recorded Future collects and analyzes Surface Web sources, and its index also contains data from forums, blogs, social media, and paste sites that we expect to contain both suspect or threat related content and links to other parts of the Internet (e.g., TOR sites).

As an initial example, we used the TOR Uncensored Hidden Wiki index (http://zqktlwi4fecvo6ri.onion/wiki/index.php/Main_Page) to manually locate a dubious reseller of credit cards (Premium Cards, <http://slwc4j5wkn3yyo5j.onion/>):



The screenshot shows a website with a dark blue navigation bar containing links for HOME, ABOUT, SHIPPING/PRICES, CONTACT/ORDER, and FEEDBACK. Below the navigation bar, there are logos for MasterCard, VISA, a 'SATISFACTORY 100% GUARANTEE' badge, and a 'WORLDWIDE SHIPPING' badge. The main content area features the heading 'We are offering top quality cards:' followed by text stating that all cards come with PINs and instructions, can be used at any ATM worldwide, and are equipped with magnetic strip and chip. It also mentions that a Full Guide will be emailed upon purchase. There are two sections for 'Us cards' and 'Eu cards', each listing usage options: ATM, Stores, and Online. On the right side, a vertical sidebar titled 'Payment methods we accept:' lists Bitcoin, Western Union Money Transfers, and PaysafeCard.

We then queried the Recorded Future index for the Onion link to Premium Cards, and indeed found 14 references from the last 3.5 months:

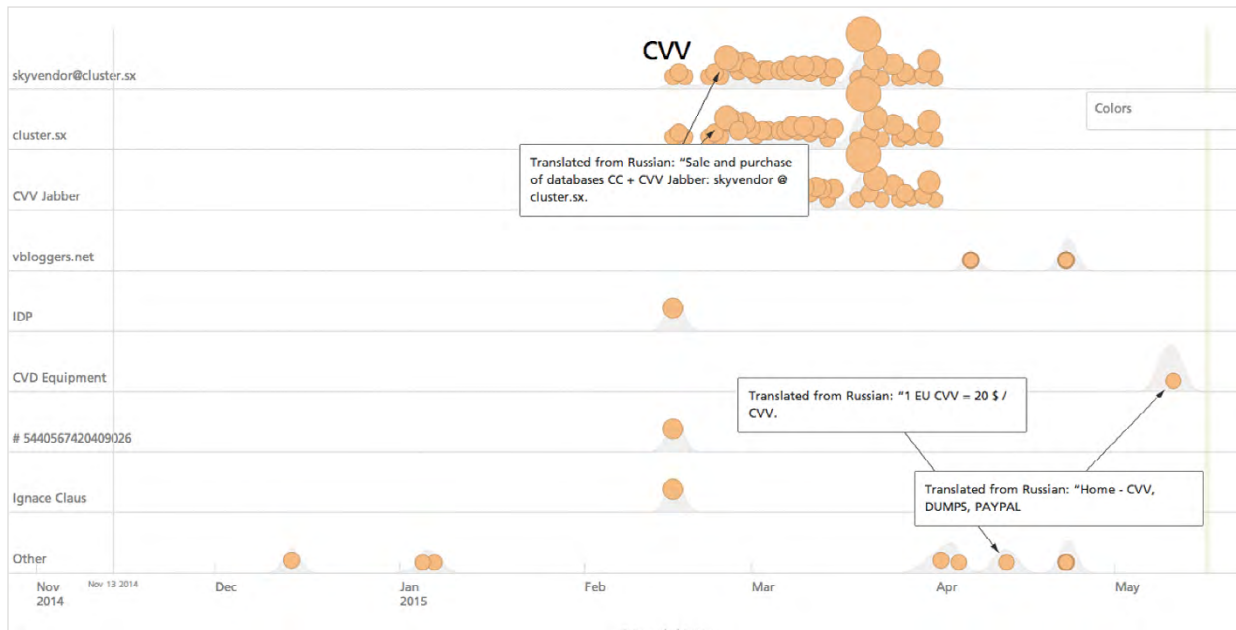


These references all come from Pastebin. One of the pastes, for example, provides an index to several useful "Financial Marketplaces":

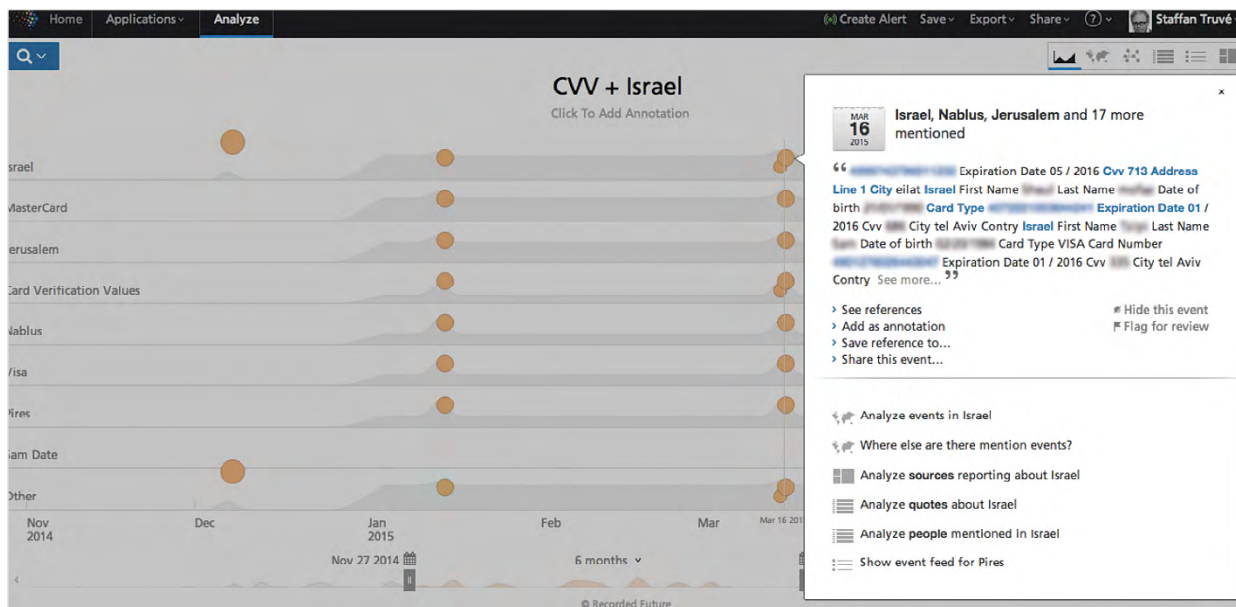
```

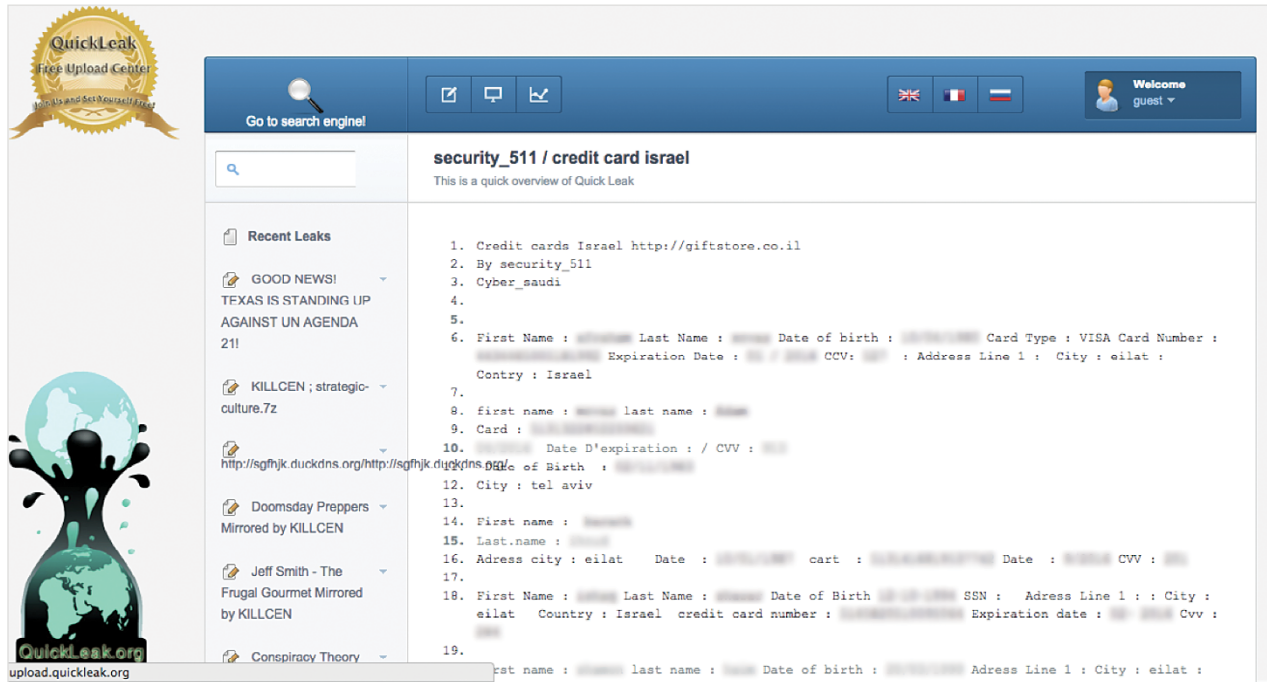
6) Financial Marketplaces
- The Green machine http://zzq7gpluliw6iq71.onion/
- Premium cards http://slwc4j5wkn3yyo5j.onion/
- Money Making Machine http://qlb3wi5ax5wjcb7y.onion/ (!REPORTED SCAM!)
- Zero Squad (Fake bills and PP accounts) http://lmyv5msldz1cp224.onion/ (!REPORTED SCAM!)
- Wallstreet http://z2hjm7uhwisw5jm5.onion
- HQER - High Quality Euro Replicas http://euronm3blw1bqh7x.onion/
- Counterfeit USD http://usdekx54qmjed3gc.onion/
- Cheap Euros http://o6klk2vx1punyqt6.onion/
- USJUD http://usjudr3c6ez6tes1.onion/
- Clone CC http://ccxdnvotswsk2c3f.onion/
- Evil Philanthropists http://mptilmknzzexj3dl.onion/
- Black&White http://blackgt4w12xgka4.onion
- CCPal store http://ccpalcrenrex5hns.onion/
- Original Skimmed Cards http://orgccd6nfdraq51j.onion/
- Automated Paypal and CC Market http://ppccfdsrcm7rf6gxf.onion/
- Ghost Squad http://paypalj6rdhgusjs.onion/
- USA Cards http://csemfkxrj4tukqq.onion/
- Credit cards http://djn4mhmbbqwjiq2v.onion/
- Card store http://7vx7qt2jyfdcvbqp.onion/
- Card store 2nd website http://khyr7fxcpslqqgo6.onion/
- SOL's USD Counterfeits http://sla2tcypjz774dno.onion/
    
```

As a second example, we investigated if illicit material was being marketed in sources that Recorded Future does harvest. Credit card information with CVVs is a good example of such material, and we focused on material published in 2015, and only in Russian. This yielded a small but interesting set of references related to advertising content and advice on how to obtain and use the stolen credit card information:



Being even more specific, we looked for CVVs of credit cards related to Israel:



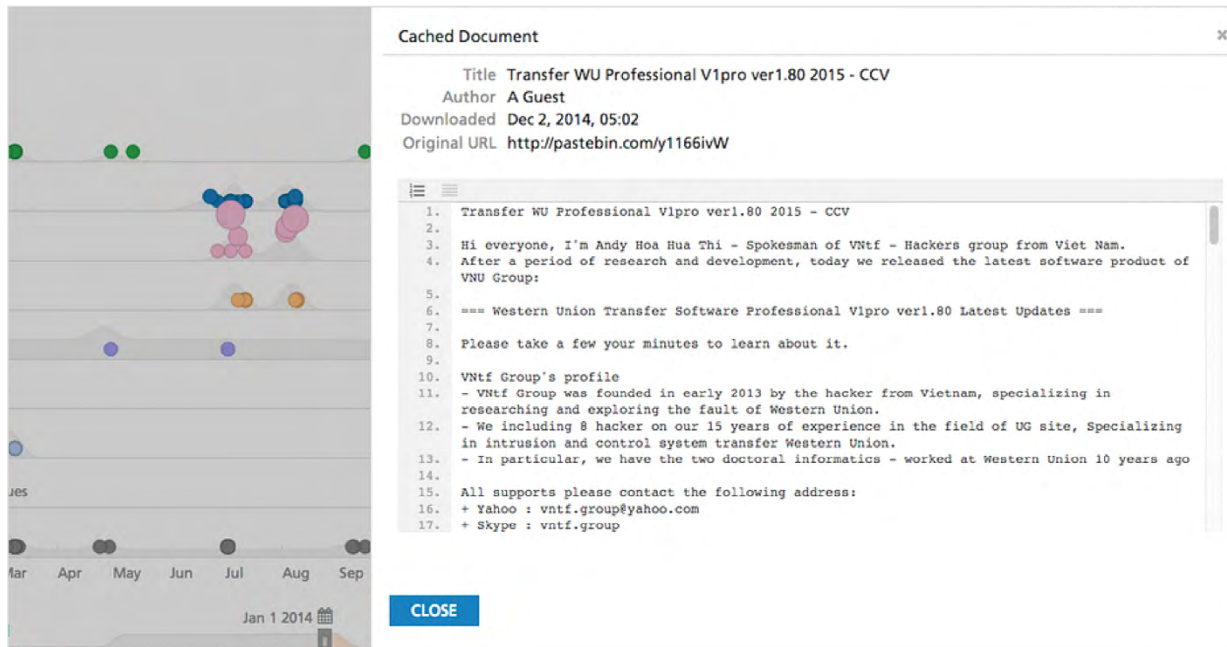


Thus, there is no doubt illicit material is being marketed not only on the Dark Web but also on other channels such as paste sites and forums.

Some of this content is nefarious enough to get quickly removed, even from Pastebin:



In these cases, it is convenient that Recorded Future provides cached access to paste content we have harvested (NOTE: this feature is available only to registered Recorded Future clients):



Links From the Surface to the Dark Web

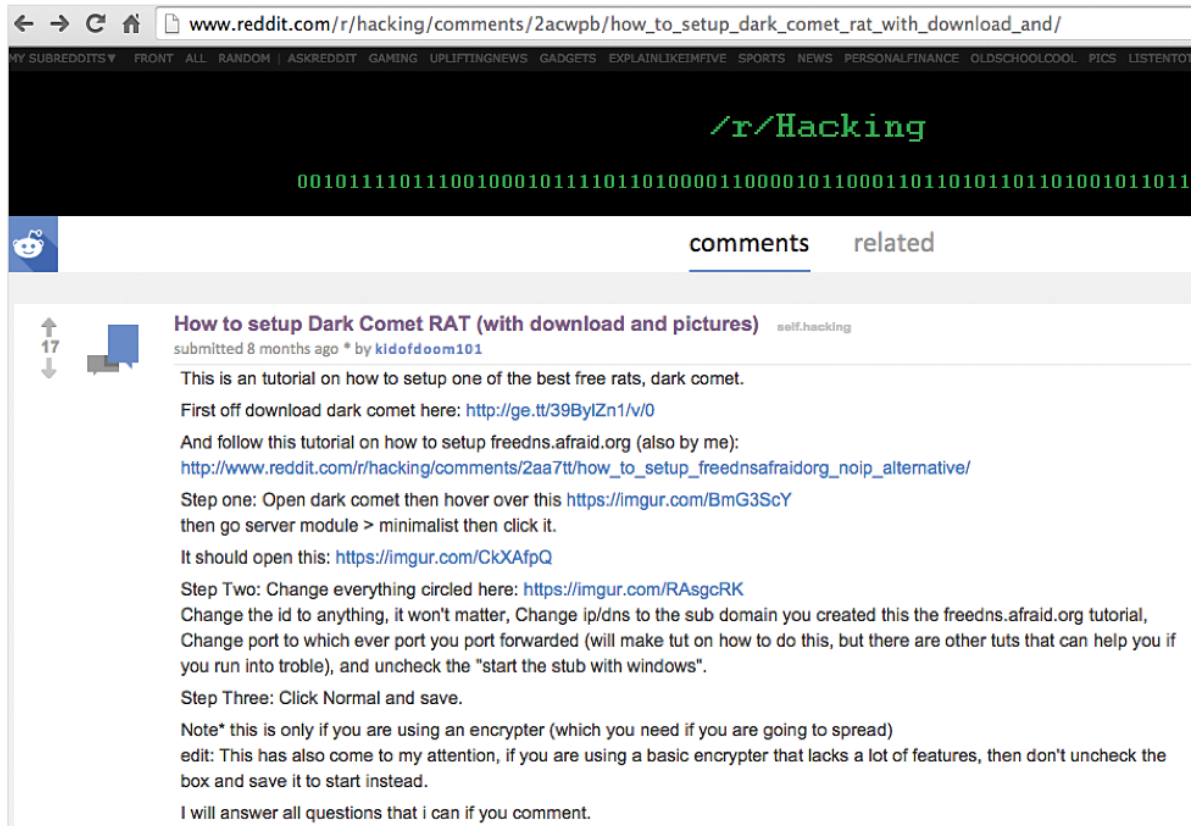
Inspired by the discoveries above, we investigated the linkage from Twitter and Pastebin to TOR/Onion links. It turns out to be fairly low volume: out of 509 million tweets, about 65 million had cyber related content published in Q1 2015⁶ there were 37 million URLs, but only 499 of those were Onion links. As another example, of 6.7 million Pastebin documents from 2015 Q1, with 226 million references in total, there were 8,316 Onion links, but only 1,036 unique links (the links with the most references were to index pages, adult comics, and sellers of cannabis, passports, and ID cards). In general, the number of links to TOR was low in volume, but some of them were high value.

The Malware Marketplace

We have shown how stolen financial credentials are marketed, but what about tools used by cyber criminals – can those also be found in this borderland?

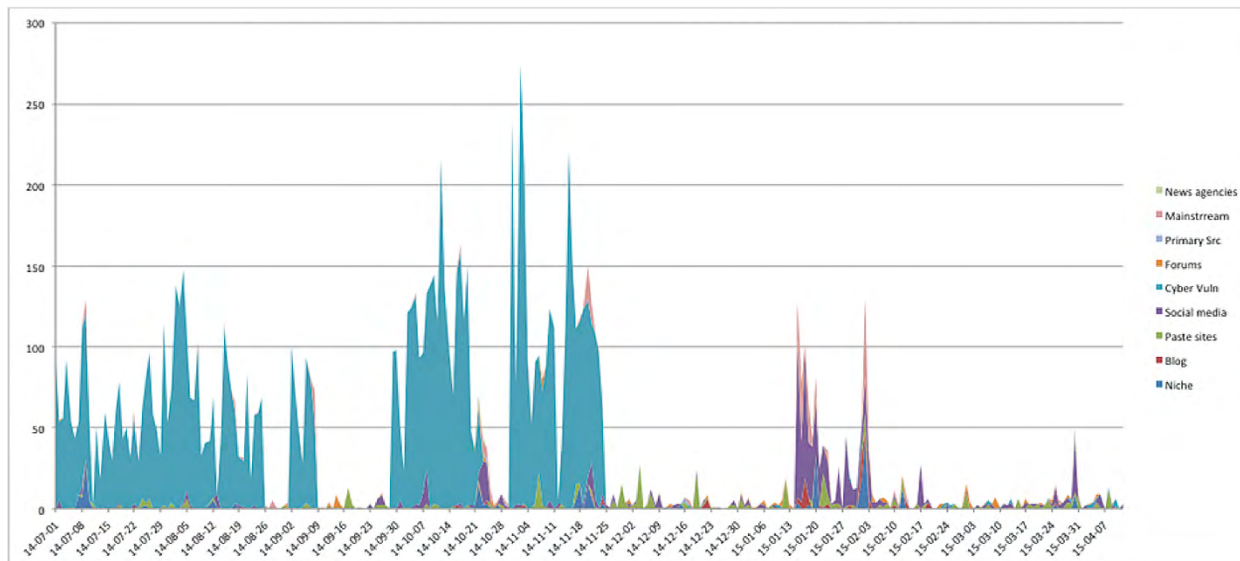
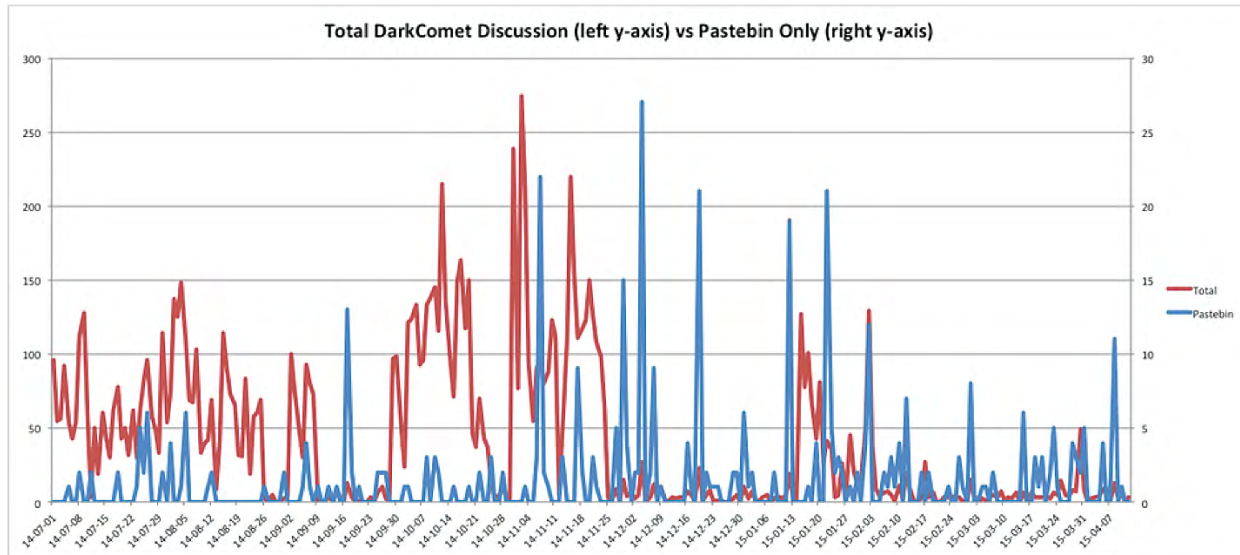
In some cases the answer is a straightforward “yes.” To download a Remote Access Trojan (RAT) like DarkComet, just Google for instructions and download sites:

⁶ These are not all cyber related tweets for that time period, but a subset selected by Recorded Future filters.

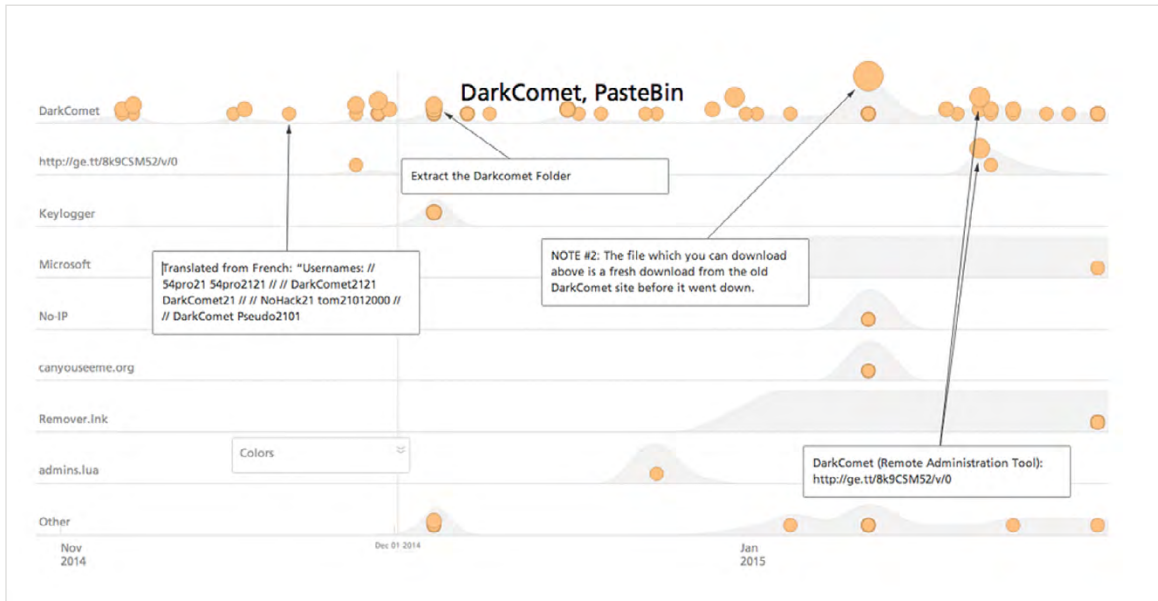


To get a bigger picture of where `DarkComet` is being distributed and discussed, we extracted all links in documents related to it for a three-month period, using the Recorded Future API, and visualized the resulting links using the open source graph visualization tool Gephi⁷:

⁷ <http://gephi.github.io/>



During 2014, the discussion was mostly active on sites related to cyber vulnerability conversations. In January 2015 the discussion shifted over to social and mainstream media, mostly due to the discussions around the use of this Malware in connection to the [Charlie Hebdo events](#). There was actually an increase in mentions of DarkComet on Pastebin in late November and December 2014. They are small in number, but the mentions which do exist are very instructive, as the following screenshot illustrates:



Here are a few of the sites linked to from Pastebin - note that these are instructions for how to download and set up DarkComet:

10100001 create new paste trending pastes sign up login my a

Want more features o

DarkComet RAT 531 VIP Version

BY: A GUEST ON DEC 16TH, 2014 | SYNTAX: NONE | SIZE: 0.25 KB | VIEWS: 125 | EXPIRES: NEVER

DOWNLOAD | RAW | EMBED | REPORT ABUSE | PRINT

1. This is the easiest way of setting up and using DARKCOMET RAT, which is TROJAN creating software for controlling remotely, THIS IS EDUCATIONAL PURPOSES ONLY.
- 2.
3. Link for darkcomet vip version, click here: <http://www.mediafire.com/download/bwx4gwxda9ca2fu>

LOGIN SI

GET.TT Say thanks to o [Click here](#)

DarkComet.zip

This file:

- 27 Nov 2014
- 5.9 MB
- 99 downloads

[Download](#)

No preview available for this type of file.
[Download DarkComet.zip](#)

In addition to showing increased interest in DarkComet, the growing amount of mentions also indicates usage migrating from higher risk threat actors to “garden variety” threat actors who source their malware tools from Pastebin.

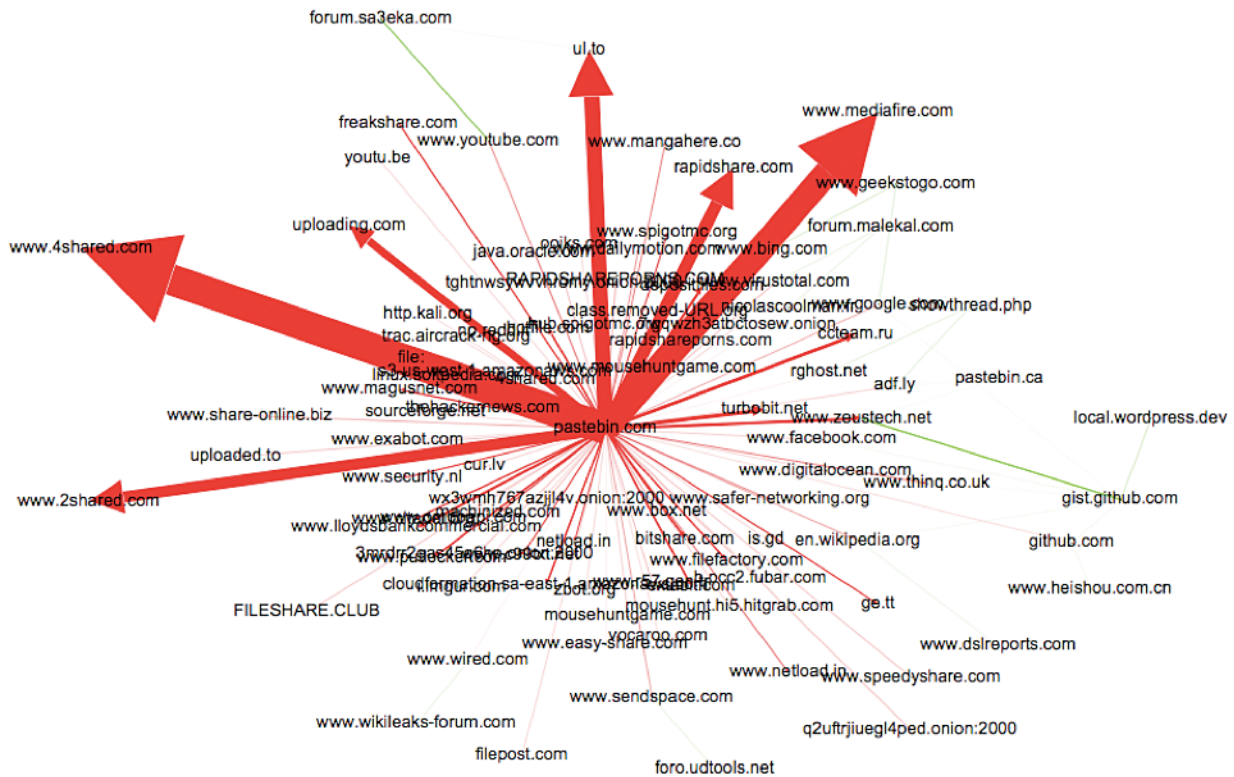
Link Patterns

Next, we examined all links from texts on paste sites and forums for a period of 3.5 months that contained a reference to malware and had a link to some other site, which we evaluated to see where the link was directed. Below are the top link targets. If we compare this list with a list of popular file sharing sites for general content, such as <http://www.ebizmba.com/articles/file-sharing-websites>, we see a mix of “general” file sharing sites and some clearly more focussed on shady material. We also note that some very popular file sharing sites, like Dropbox, are missing from the top link list.

Destination Site	Count	Position On Top List
www.4shared.com	3469	3
www.mediafire.com	2463	2
rapidshareporn.com	1239	
www.2shared.com	1206	
uploading.com	1153	
turbobit.net	898	
ul.to	824	
rapidshare.com	709	14
www.easybytez.com	646	
fileshare.club	547	
hotfile.com	547	
www.jeuxvideo.com	329	
bitshare.com	327	
www.juanmata10.com	260	
depositfiles.com	245	
salefiles.com	220	
www.example.com	215	
www.netload.in	212	
noc.yartv.ru	201	
netload.in	196	
pastebin.com	179	
extabit.com	173	
www.putlocker.com	163	
www.exploit4arab.net	157	

tech4all.criativin.com.br	156	
www.youtube.com	154	
github.com	150	
www.gov.ai	147	
www.owasp.org	143	
filepost.com	140	
hosting.risp.ru	138	
pan.baidu.com	132	
www.exploit-db.com	130	
freakshare.com	120	
www.zeustech.net	119	
www.voxility.com	117	

As seen, again a majority of the link destinations are file sharing sites of different kinds, showing that discussions around malware on these sites tend to be accompanied with links where other content can be downloaded. This graph illustrates the link pattern, and Pastebin is the main source of links:



Conclusions

There are clear borders between the Surface, Deep, and Dark Web in terms of accessibility and tools, but there exists information on the Surface Web and on the Deep Web that can be used to gain important understanding of what is happening on the Dark Web. Simple marketing mechanics underlies this – when something needs to be sold, prospective customers need to be able to find information about it quickly. The available information includes topics, link patterns, and activity levels.

As illustrated by the study of mentions of the DarkComet malware, sites such as Pastebin act as a marketing channel by providing a fairly unregulated place for posting both instructions and links to download sites for malware. Using a threat intelligence platform to monitor the activity on paste sites can therefore be a good way to get early warning signals for increased use of certain kind of malware and stolen data or credentials.

Topics also tend to migrate over time, from Dark to Surface Web, and analyzing these patterns allows us to understand when high end malware tools are becoming commodity malware. Such a shift means the volume of attacks using the commodity malware will increase, but the average skill level of attackers will go down - and the highly skilled attackers will have moved on to using another tool.

About Recorded Future

We arm you with real-time threat intelligence so you can proactively defend your organization against cyber attacks. With billions of indexed facts, and more added every day, our patented Web Intelligence Engine continuously analyzes the entire Web to give you unmatched insight into emerging threats. Recorded Future helps protect four of the top five companies in the world.

Recorded Future, 363 Highland Avenue, Somerville, MA 02144 USA | © Recorded Future, Inc. All rights reserved. All trademarks remain property of their respective owners. | 06/15