Darknet

Its evolution and its problems

Derk-Jan Karrenbeld

These days we share a lot of both public and private bits and bytes, and when that data is somewhat private, whether it is music, books, images, financial data or a simple conversation, we usually don't like people tapping in or logging that we are sharing and maybe more importantly what we are sharing.

The channels we use today make it so much easier to distribute anything we want, sadly resulting in a lot of illegal activity, which in turn is increasing both the means of copyright protection and the quality and quantity of monitoring data streams. So these channels have their downsides, such as their weaknesses when it comes to obscuring the act and the data itself or the pollution of fake files by agencies that try to stop our happy sharing activities.

This article will hopefully give you some insight in a new way of sharing and communicating with secure transfers and most notably untraceable data. Guided by the article mentioned just a few lines below, the definition, the evolution, the problems and a way to participate are all covered.

The Darknet

The term "The Darknet" is widely used to refer collectively to all covert communication networks. As stated in the article "The Darknet and the Future of Content Distribution" by Peter Bible, Paul England, Marcus Peinado and Bryan Willman the idea of the darknet is based upon three assumptions:

1. Any widely distributed object will be available to a fraction of users in a form that permits copying.
2. Users will copy objects if it is possible and interesting to do so.
3. Users are connected by high-bandwidth channels.

The four describe any darknet as a distributed network that emerges from the injection of objects according to assumption 1 and the distribution of those objects according to assumptions 2 and 3. What followed from these assumptions and definition are a set of infrastructure requirements for the network to operate efficiently. These are the means to input objects, transmit them to users, output them to rendering devices, but also the means to search for these objects and of storage, to retain objects for extended periods of time. The last functions as cache for load reducing and exposure minimizing of the source node.

The evolution

Sneaker net

Do you remember your parent(s) or yourself using tapes to copy your favourite music and distributing it with a group of friends? Prior to the mid nineties, the darknet consisted of the sneaker net. Content was transported by foot, more so people carrying physical devices such as tapes to deliver the data requested. Many small networks existed worldwide, with little legal action taken against these networks, because of the size and the (lack of) commercial impact. Although no good search engine existed and latency was high, the friend-to-friend networks were private and the database could be quite extensive when people with lot of connections were introduced (or instances such as schools).

The World Wide Web

Just before we entered this millennium, the internet had become mainstream and so the technologies on which the darknet was configured changed. Content (at least the audio) was ripped from CD's instead of being recorded by tape recorders. The latency was greatly reduced with this new distribution network and discovery was made easy by web search engines. The small sometimes greatly interconnected networks were displaced by a network with global participation. The centralized servers worked great with commercial content, whilst providing the ability of display advertisements and maintain buyer profiles. On the other hand, it wasn't so great for all illegal activity, because copyright-holders simply sent "cease and desist" letters to the web-site operators of those sites hosting content not belonging to them.

The Peer-To-Peer networks

Most of us still remember Napster, or at least the story behind its fall. In short, content was distributed, but the search index was centralized. It was only matter of time before the servers servicing the search became a legal target. With the index shrinking, so did the users. "The internet" did come up with a solution called Gnutella. The difference was a fully distributed network and search index. With an open protocol everyone could write their client to hook themselves to this open darknet. One of the reasons this darknet wasn't targeted extensively is the fact it was used for a lot of non-infringing content distribution as well.

These days we (also) have BitTorrent, which somewhat defeats the freeriding problem, by using a protocol where all users must upload in order to efficiently download. With BitTorrent, the search index is usually semi-centralized as tracker files are stored on webservers and usually not inside the network. However, where Napster failed to fight of the legal problems, the servers used for the distribution of BitTorrent trackers are great in number and the owners prove to be resourceful in relocation. Because trackers are replicated by many distributers, the data itself usually has a long retention in the darknet.

The problems

Also mentioned in the article, but also through the internet are a few problems with Gnutella-based networks. The three most imminent to the existence and usage of the networks are "Freeriding", the lack of anonymity and the validity of the data.

Figure 1: Historical evolution of the Darknet. We highlight the location of the search engine (if present) and the effective bandwidth (thicker lines represent higher bandwidth). Network latencies are not shown, but are much longer for the sneakernet than for the IP-based networks.

Freeriding

You might remember the option in your Gnutella, Kazaa-Lite or eDonkey client dealing with freeloaders (people that download but don't share). Users that just download and don't upload quickly deplete the network's bandwidth when their numbers increase. This in effect turns the Peer-To-Peer network into a unidirectional, large centralized system where some nodes, a tiny fraction of all nodes and also called the supernodes, provide the content. Not only is data not uniformly spread across the network, the supernodes are also more targeted by legal instances.

Lack of anonymity

Users of BitTorrent and Gnutella services who share data are not anonymous. In order to transfer data, the endpoints have to be determined. And if a client can, so can a lawyer or government agency. There are some workarounds to the absence of endpoint anonymity, but analysis of the robustness of any of these workarounds is worth an article on its own. One workaround however is obscuring the data sent by encrypting the data stream. This brings at least three problems that apply when the data sent is not legal:

1. If the originating endpoint is a legal instance, the data and the infringer (destination) are known
2. If the tracker's origin is a legal instance, the data and the infringer (destination) are known
3. Because BitTorrent doesn't have the encryption by default, some users will not be able to connect and share data, limiting bandwidth.

Validity of the data

Because the servers providing the trackers keep on running, legal instances are trying new ways to take down the BitTorrent (and Gnutella) darknet. One of these ways is to inject the network with fake or invalid data. This pollutes both search results and data streams. Because anyone can freely contribute to the network, no one can pre-emptively protect the database from invalid data. Algorithms are designed to flag this data, but over time, the injectors overcome these measures. The only trusted data comes from trusted nodes (off course identity theft can reduce this trust) and nodes that would be available in the sneaker net, also known as friends or friends of friends.

Attacks

The image shows the weaknesses on each connection point, intermediary node and environment. In Gnutella-based solutions, the centralized nodes and data are usually targeted by attackers. Other darknets can be removed when they can be monitored, so it's key that the users are trustfully, the connection is secure and the data is real.

Today's Darknet

If all above is summarized we retain some positive features out of each network.

- A distributed network works best in terms of availability (more replication), sustainability (no supernodes needed to maintain database) and as a darknet in general.
- If access is limited to trusted users, the chance of pollution is greatly reduced.
- If data streams are encrypted and destinations are obfuscated (by tunnelling data/routing data streams through nodes), transportation is secure and data stays private.

A way to incorporate all these points is by forcing encryption and using a friend-to-friend based distributed network. Several clients already exist to incorporate all these. One of the clients is called RetroShare and uses GPG keys

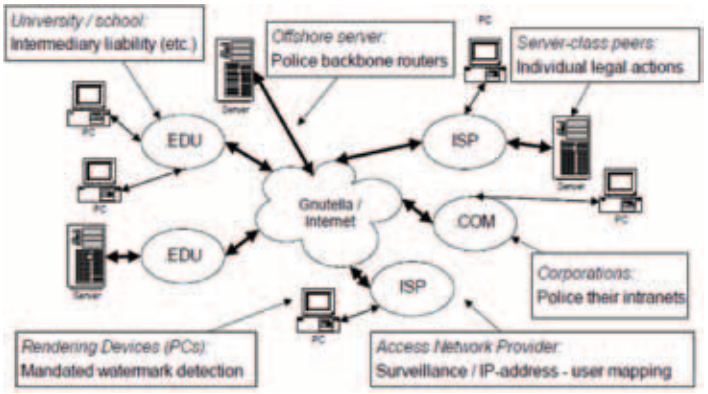

Figure 2: Policing the darknet. Gnutella-style networks appear hard to police because they are highly distributed, and there are thousands or millions of peers. Looking more closely there are several potential vulnerabilities.

for authentication. The friend-to-friend structure makes it hard to intrude or monitor from the outside and anonymity can be increased by turning off DHT and IP/cert. exchange services turning it into a real darknet.

The article mentioned earlier discusses the future of the Darknet sketched just now. Their conclusion is that in terms of interconnectability it will survive great lengths and in time will have a huge amount of data. Legal instances won't always be as happy as can be, since DRM content is unlikely to be injected because of the impeding features of DRM content.

Why should I participate?

The more people are on any, or even better multiple darknets, the more interconnected all these networks become. With features such as tunnelling it is possible to find all the data in the world, if everyone would be connected (according to the "Six degrees of separation"). Tunnels are anonymous and the actual endpoints are almost always untraceable. It is more secure than BitTorrent in terms of data validity and tracing and could be the future of content distribution.

And who doesn't like to download humongous amounts of untraceable possible sensitive data…

Where do I start?

1. Download a client such as RetroShare and find friends to use it as well.
   - If you don't have friends, you can always try to find some at Christiaan Huygens'. Some people might be very eager to share with you.
   - You can always try to throw a key sharing party!
2. Create a key pair and physically exchange keys. These people you fully trust.
3. Add friends of friends (but don't sign their keys) for more data.
4. Use search to find data on any connected node.
5. Most clients have secure messaging and/or email as well. Some also have boards and distribution channels, all secure.
6. If a 'friend' goes corrupt, simply revoke the friendship.
7. Inject content and retrieve content! Happy sharing.

There is a lot written about everything described and mentioned in this article. The Wikipedia page links to numerous darknet clients and articles can be found on all previous darknets such as Gnutella. More information is freely available online.