# ArkC: Internet Circumvention Platform against Blacklist Censorship

## Introduction

Internet access is restricted in some regions due to government censorship. To help with the Internet circumvention of such restrictions, organizations provide proxy services. However, services deployed with current technologies tend to be blocked with blacklists as soon as governments discover them, which adds to the maintenance cost and causes service instability. ArkC is a circumvention platform for nonprofits or individuals to deploy free proxy services without worrying about blacklisting or disruptions. The platform employs a connection model that makes blacklisting overly costly and difficult.

## Overview of ArkC

ArkC includes software for *clients, relays, servers* and *coordinators*. Client-end is installed onto users' computers to provide circumvention in the form of an HTTP proxy. Relay utilities are installed in servers and PCs within the region of the client to receive connections from the Internet. Server-end runs on servers and PCs in regions where communication with the blocked websites and relays are allowed. Coordinator utilities run on servers and handle users' requests, coordinating servers and relays. When the client-end receives a proxy request, it sends a domain query to any DNS server outside of the restricted region and joins a Bit Torrent (BT) network. The DNS server forwards the query to a coordinator based on NS records. The coordinator decrypts the query for the client's address and authentication data. After verification, the coordinator forwards the request to a server, which initiates a TCP connection to the relay via a randomized route chosen from open proxies, Google servers, or CDN nodes, camouflaged as ordinary applications. After the relay receives a client address, it communicates with the client on the BT network with a pre-shared BT seed that the client has joined. Data are transmitted in the form of normal peer-to-peer downloading between relays and clients in the censored regions. After the connections are established the client has access to the restricted websites. In ArkC, end-to-end AES encryption and public-key authentication are employed to protect data transmission from interception or Man-in-the-Middle attack.

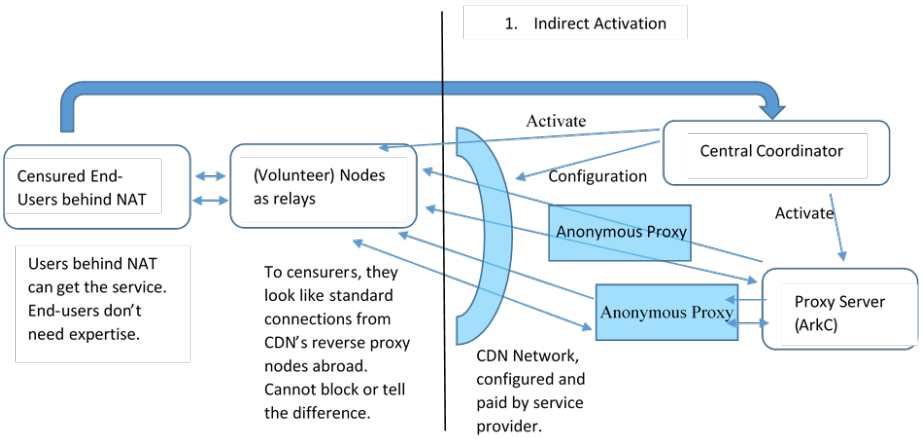


Figure 1: Connections between servers and relays, when the client runs the relay itself

## Analysis of ArkC

Blocking ArkC connections requires significant cost and involves negative impacts on the economy of the government. Thus, by employing ArkC circumvention service providers avoid the cost of replacing blocked servers, and can use their funds to support faster service and provide wider coverage.

Domain queries are fundamental to the Internet. With its high bandwidth and UDP-based attributes, the adversary (censors) is assumed unable to enforce censorship rules on every domain query, as enforcing full DNS inspection slows down Internet applications. Most of clients' domain queries are received by a foreign server, so that coordinators are able to receive the requests and assign them to servers. In addition, the adversary is assumed capable of manipulating international TCP connections, but fall short of the calculating ability to analyze all domestic ones. In this case only the processes of domain requests and obtaining connections from servers to relays are susceptible to censorship interference. However, with ArkC such connections can be initialized from any open proxy, Google servers, or other sources that are not owned or controlled by the service provider. Blocking access from so many sources demands high processing speed in the censorship system and can result in significant economic damage, for example, commercial websites within the adversary's country may not be indexed by Google, or may fail to serve foreign users. Both consequences prevent blacklisting servers. Furthermore, all relays are camouflaged to share similar attributes as normal web services so they cannot be distinguished from normal web servers. Blocking access to all of these servers could hurt domestic websites serving international users. While all relays that receive connections from servers must run ArkC, users themselves can set up relays to provide service for others. Relays can be well hidden in BT downloaders. Thus with some additional

false broadcasting, it would be hard to create a list of relays without including normal servers. Besides, the connections from across the adversary's firewall are obfuscated with camouflage such as obfs4, so they cannot be identified.

## Implementation and Testing of ArkC

Feedback and data are being collected from volunteers who use the deployed testing service and ArkC has been tested in multiple ISPs in China. Despite a 5-7% additional bandwidth due to protocol overhead, testing reports stable and ideal downloading speed. Since the development of ArkC is incomplete, no benchmark test has been carried out.

## Acknowledgement

Will Scott of University of Washington gave advice on the BT design. ArkC includes some similar attributes to that of Flashproxy[1], but was created independently. China Digital Times offered financial assistance in service deployment.

## Source Code and License of ArkC

ArkC is open source under GNU General Public License version 2. Its source code and executables may be obtained at https://github.com/projectarkc/. Update with ArkC can be found at http://arkc.org.

[1] Fifield, David, et al. "Evading censorship with browser-based proxies." *Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 2012.