Ark - Archipelago Measurement Infrastructure

Rodrigo Monteiro, Diogo Abreu, and Gustavo Barros

University of Minho, Department of Informatics, 4710-057 Braga, Portugal
e-mail: {a100706,a100646,a100656}@alunos.uminho.pt

Abstract. The Ark Measurement Infrastructure Project, led by the Center for Applied Internet Data Analysis (CAIDA) at the University of California aims to build and operate an Internet measurement infrastructure to enable researchers to study the Internet's topology, routing, and performance characteristics.
Understanding how the Internet is structured and how it operates is a crucial task for researchers and engineers who are developing new technologies and applications that rely on it. However, studying the Internet is a challenging task, as it is constantly evolving and changing.
This research paper provides a comprehensive overview of Ark's sophisticated infrastructure that addresses this challenge, describing the project's goals, structure, ongoing projects, and utility.

Keywords: Measurement infrastructure · Internet security · Topology

1 Introduction

Ark is a measurement infrastructure, a globally distributed system with emphasis on security, geographical and topological diversity, communication and coordination, while being community-oriented, empowering collaborative use. Their primary objective is to make sophisticated large-scale measurements with reduced effort and novel measurement techniques more accessible.

2 Impact and Utility

The data collected since 2007 has been extensively used in many papers, and in addition to the available data, Ark provides researchers and users with access and tools to conduct their own ad hoc measurements via a web interface and even large-scale feedback-driven, dynamic measurements from the command-line through Vela: On-Demand Topology Measurement Service. This is partially possible due to a system that works with a tuple space which will be discussed later.

3 Measurements

3.1 Hosted Measurements

Ark also conducts a variety of hosted measurements such as The Spoofer Project. This project measures the Internet's susceptibility to spoofed source address IP

packets, i.e. vulnerabilities in computer networks that allow attackers to send packets with falsified source IP addresses, potentially leading to DoS, DDoS, MITM, and other attacks. Other hosted measurements include assessments of IPv4 and IPv6 stability, as well as DNS Health.

3.2 Ongoing Measurements

As for ongoing measurements, Ark runs IPv4 topology and IPv6 topology analysis. For IPv4 topology measurements, Ark performs traceroute measurements to all routed /24 networks in the IPv4 address space — the first 24 bits identify the network (subnet). For IPv6 topology measurements, given that random or sequential approaches are impractical due to the vast size of the IPv6 address space, the experiments conducted target practical methods of dividing the address space to discover active subnets, which is important to understanding IPv6 growth, structure, and evolution and relation to IPv4.

4 Structure

This infrastructure, a star topology with multiple nodes worldwide, employs a tuple space to enable communication and coordination. The tuple space used is a distributed shared memory[1] that stores tuples, which are an ordered collection of values (`string, int, float`), coupled with a small number of operations. Example of a tuple: `("RC", "TP1", "essay", 1)`. This structure is closest in concept to a database with the advantage of not having a formal schema required or declared. Since the tuple space is an associative memory, templates can be used for matching against all tuples. Example of template matching: `(("RC", *)) matches the previous example`.

5 Security

Ark also pursues security, but it's important to consider the context in which it is secure, since threats and potential vulnerabilities can vary widely. Thus, to properly mitigate security risks it is necessary to specify about the types of threats. When considering the security of the Ark project, awareness of potential threats from third parties, public users, and even collaborators is needed. Each type of threat requires a different set of mitigation strategies. For example, to protect against public users launching attacks and to prevent privilege escalation, executing in a sandbox, restricting measurement capabilities, and rate limiting can help.

[1] consists of message passing in a network of multiple nodes/ cores, each with local memory

6 Conclusions

In conclusion, exploring cooperative approaches and maximizing the collective benefit of deployed infrastructure and gathered measurements is essential for advancing the development of Internet technologies and improving the Internet's performance in the digital age. And, as shown, the Ark Measurement Infrastructure Project plays a role in achieving these goals.

References

1. Robert Beverly, William Brinkmeyer, Matthew Luckie, and Justin P. Rohrer: IPv6 Alias Resolution via Induced Fragmentation (2013)
2. Pascal Mérindol, Benoit Donnet, Jean-Jacques Pansiot, Matthew Luckie, Young Hyun: MERLIN: MEasure the Router Level of the INternet (2011)
3. CAIDA/UCSD: The 7th Workshop on Active Internet Measurements (AIMS-7) Report (2015)
4. Young Hyun: The Archipelago Measurement Infrastructure - Presentation (2006)
5. Caida - Ark: *https://www.caida.org/projects/ark* (Accessed on February 21, 2023)
6. Vela: *https://www.caida.org/projects/ark/vela* (Accessed on February 21, 2023)
7. The Spoofer Project: *https://www.caida.org/projects/spoofer* (Accessed on February 21, 2023)