

Classifying the Types of ASes in the Internet

Xenofontas Dimitropoulos, Dmitri Krioukov, George Riley, kc claffy

Introduction

The AS-level topology of the Internet has attracted significant research attention the last few years. Measuring the topology of the Internet, analyzing the properties of Internet topology graphs, and generating Internet-like synthetic graphs are prominent research topics in the field. Although the AS topology of the Internet has been studied extensively, most of the studies have focused on logically abstracted AS-level graphs, ignoring individual properties and characteristics of different types of ASes.

In this work we seek to develop a taxonomy of the types of organizations that own AS numbers. We pose and provide a first answer to the following fundamental question: **What types of organizations own AS numbers?**

Motivation

Statistical knowledge of the types of organizations that own AS numbers is essential to modeling the structure and evolution of Internet topology:

1) Knowledge of AS types is important for augmenting AS-level topologies with realistic intra-AS and inter-AS router-level topologies. For example, we expect that the network of a dual-homed university is vastly different from that of a dual-homed small company. Without knowing the AS type, we cannot associate an appropriate router-level topology type with the given AS.

2) Different types of ASes are associated with different growth patterns. For example, Internet Service Providers (ISP) grow by attracting new customers and by engaging into business agreements with other ISPs. On the other hand, the node degree of an AS owned by a small company is not expected to grow significantly over time. Thus, categorizing different types of ASes will help to identify network evolution patterns and develop accurate Internet evolution models.

Methodology

We employ information retrieval techniques to analyze data from the Internet Routing Registries (IRRs).

1) We first extract organization description records from the IRRs.

2) Then we perform *extraction indexing* on the extracted records and determine the highest frequency key-words (key-phrases).

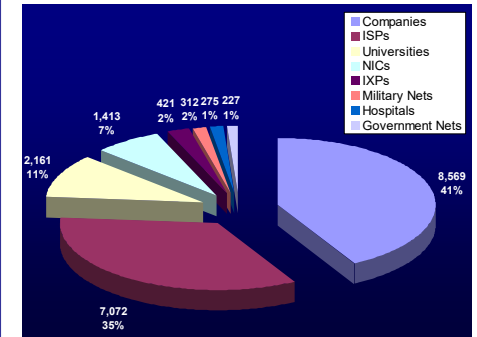
3) Next, we examine the extracted index and group together keywords (or key-phrases) of similar semantic content that characterize distinct AS types. Using this method we identify the following categories:

1. Companies.
2. ISPs.
3. Universities, research centers, colleges and schools.
4. Network Information Centers (NICs).
5. Internet Exchange Points (IXPs).
6. Military-related networks.
7. Hospitals and health centers.
8. Government-related networks.

4) Finally, we develop a *text classification expert system* that classifies organization description records into the above categories.

Results

We effectively classify 20,598 (63.01%) out of a total of 32,689 AS records we extracted from the IRRs.



Per-category breakdown of classified ASes.

Future Work

Different types of ASes have different network properties and infrastructures. Development of accurate topology generators and evolution models requires independent analysis of the network properties of these diverse networks. In the future we intend to: 1) further develop our classification methodology to incorporate additional data sources (advertised prefixes, geographical span); 2) identify network topology patterns of different AS types.