Identifying ASes of State-Owned Internet Operators

Esteban Carisimo[*]
Northwestern University

Alexander Gamero-Garrido
CAIDA, UC San Diego
Northeastern University

Alex C. Snoeren
UC San Diego

Alberto Dainotti
CAIDA, UC San Diego
Georgia Institute of Technology

ABSTRACT

In this paper we present and apply a methodology to accurately identify state-owned Internet operators worldwide and their Autonomous System Numbers (ASNs). Obtaining an accurate dataset of ASNs of state-owned Internet operators enables studies where state ownership is an important dimension, including research related to Internet censorship and surveillance, cyber-warfare and international relations, ICT development and digital divide, critical infrastructure protection, and public policy. Our approach is based on a multi-stage, in-depth manual analysis of datasets that are highly diverse in nature. We find that each of these datasets contributes in different ways to the classification process and we identify limitations and shortcomings of these data sources. We obtain the first data set of this type, make it available to the research community together with the several lessons we learned in the process, and perform a preliminary analysis based on our data. We find that 53% (*i.e.,* 123) of the world's countries are majority owners of Internet operators, highlighting that this is a widespread phenomenon. We also find and document the existence of subsidiaries of state-owned governments operating in foreign countries, an aspect that touches every continent and particularly affects Africa. We hope that this work and the associated data set will inspire and enable a broad set of Internet measurement studies and interdisciplinary research.

ACM Reference Format:
Esteban Carisimo, Alexander Gamero-Garrido, Alex C. Snoeren, and Alberto Dainotti. 2021. Identifying ASes of State-Owned Internet Operators. In *ACM Internet Measurement Conference (IMC '21), November 2–4, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3487552.3487822

[*]The author partially worked on this article during his time at Universidad de Buenos Aires and CONICET.

IMC '21, November 2–4, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-9129-0/21/11…$15.00
https://doi.org/10.1145/3487552.3487822

1 INTRODUCTION

In this paper, we introduce and apply a methodology to accurately identify state-owned Internet operators worldwide and their Autonomous System Numbers (ASNs). Obtaining an accurate dataset of state-owned Internet operators' ASNs enables research studies where state ownership is an important dimension, including many Internet phenomena that have salient socio-economic or political implications: Internet censorship and surveillance, cyber-warfare and international relations, Information and Communications Technology (ICT) development and the digital divide, critical infrastructure protection, and public policy. Autonomous system numbers represent a key variable to bridge these dimensions with the technical domain: not only they are central to studies involving the Internet topology, but they have an (almost) *one-to-one* mapping with IP addresses[1], which thus inherit some of their ASNs' properties and vice versa.

Despite such importance, a data set providing this information was not available before our study. Few commercial business information databases provide ownership information for companies worldwide but *(i)* they do not carry any data related to ASes and *(ii)* in this study we find that they might not be entirely accurate and comprehensive (in the context of Internet operators). Previous Internet measurement literature discussed the role of selected ASes operated by state-owned providers in one or two specific countries or regions in *e.g.,* specific Internet censorship events [23, 61] or in the development of a country's peering infrastructure [19]. Other research efforts focused on AS taxonomization and classification either from a topological perspective [25, 27] or in terms of business type (but focused on an individual country for which public databases already exist) [81]. Hard challenges include indeed the scarcity of data on a global scale and the fact that this is a multidisciplinary problem.

We tackle this research problem by proposing a multi-stage method based on datasets that are highly diverse in nature and an in-depth manual analysis that takes into account complex ownership structures. We also reduce the problem complexity by narrowing our focus to operators *(i)* of significant size, *(ii)* operating at national (*e.g.,* federal) level, and *(iii)* who do not restrict their services to only certain sectors (*e.g.,* exclusively research and education).

We obtain the first—to the best of our knowledge—data set of this type, we make it available to the research community together with the several lessons we have learned in the process, and we perform a preliminary analysis based on our data. Reinforcing the

[1]Almost all routed IP addresses are originated in the global routing system by only one AS.

relevance of this subject, we find that *state-ownership is a broad global phenomenon* but much more prevalent in Africa and Asia (blue and green countries in the heatmap in Figure 1—discussed in detail in § 8). We also find and document the existence of subsidiaries of state-owned companies operating in foreign countries (green countries in Figure 1), an aspect that touches every continent and particularly affects Africa.

Our key contributions are the following:

(1) A novel methodology to identify state-owned ASes of Internet operators worldwide, which we verify through manual analysis and cross-comparison of multiple data sources.
(2) The first publicly available data set containing the full list of state-owned ASes of Internet operators, including metadata referencing each organization to the corresponding input and confirmation sources.
(3) Being this a novel research challenge we gained significant insights, for example regarding the quality and characteristics of the data sources as well as the intricacies of the problem. We document and discuss them in detail.
(4) We find 989 state-owned ASes—including 193 ASes of state-owned providers operating abroad—of 123 countries. Combining this data set with other Internet data, we find preliminary results suggesting that the prevalence of state-owned providers in the Internet access market is substantially higher in Asia and Africa. We also find that African countries host a remarkable presence of foreign state-owned ASes and in 6 of these countries foreign state-owned ASes hold more than 50% of the estimated access market.

We hope that this work and the associated data set will inspire and enable a broad set of Internet research studies.

2 CHALLENGES

Identifying state-owned telecommunication companies and the ASes they operate is a multifaceted problem crossing various technical and administrative domains. In this section we summarize the most critical challenges, which are largely based on limitations of the available data.

Lack of public databases: There is no global public registry that indexes state-owned enterprises. Only a few countries (*e.g.,* Sweden [51], Finland [52] and Uruguay [24]) report, through public websites, which companies have state participation. Moreover, specificity and granularity of the data vary from country to country. Focusing on the topic of our study, telecommunication companies, we are unaware of any publicly available resource that lists, at a global- or regional-level, all telecommunication companies, nor one that only specifies state-owned telcos. For instance, the United Nations ITU [75] is (at least partially) aware of state participation in telecommunication companies. In fact, some ITU reports do mention the presence of state-owned telcos. However, there is no central repository or simple way to identify and access the ITU documents that include this specific information. Moreover, isolated reports are not sufficient to create a world's list of state-owned telcos.

We have identified two commercial databases that provide information about ownership of telecom enterprises: Orbis [79] and Telegeography's GlobalComms [71]. However, their methodologies and the frequency of updates are not disclosed in detail and it is unclear how accurate they are. We include Orbis data in our study and find it misses or misclassifies a few companies in terms of state-ownership. Despite following the directions on the Telegeography website in an attempt to evaluate or purchase their product, we did not receive a response and were therefore unable to evaluate this dataset.

Company-to-AS mapping: Relying on accurate company names is necessary to determine state participation. However, there is a lack of databases and methodologies that allow us to precisely map ASes to companies and vice versa. WHOIS databases from Regional Internet Registries (RIR) [54] map ASNs to the names of the companies that they were delegated to. However, WHOIS records may not be updated as an AS or company ownership/denomination changes, leading to inaccurate and obsolete information [82] (despite ICANN initiatives to enhance WHOIS such as the WHOIS Accuracy Program Specification [46]). Moreover, a company's registration name (`OrgName` field), which tends to be the company's legal name, may differ from commercial names or brand names. As an example, Colombia's state-owned *Internexa* operates AS262195 in Argentina, however LACNIC's WHOIS records report the owner's name as *Transamerican Telecomunication S.A.*

In addition, telecom companies sometimes own more than one ASN (for historical or technical reasons and because of acquisitions/mergers); these are called *sibling* ASNs. Sibling ASNs can be associated with very different names in WHOIS, making state-of-the-art WHOIS-based sibling inference methods [17, 18] unable to capture the entire of set of ASes operated by the same organization.

Complex and evolving ownership structures: Detecting state participation in a company requires checking for state presence across shareholders and through indirect chains of ownership (*i.e.,* control over a company's shareholders through state-controlled companies). The aggregated participation of multiple state-controlled bodies—such as hedge, wealth and pension funds—at companies' shareholder structures could give control to the state. For example, three of Malaysia's government-owned funds—Khazanah Nasional Berhad [10], Amanah Raya Berhad [9] and the Employees' Pension fund [5, 6, 28]—in aggregate hold more than 50% of the shares of Telekom Malaysia (AS4788) [11].

Large state-owned telecom companies also operate subsidiaries and branches abroad. E.g., Qatar's state-owned Ooredoo operates several subsidiaries across North Africa and the Middle East (Algeria, Tunisia, Kuwait, Oman *etc.*). Companies do not use homogeneous ways to report subsidiaries and branches—or vice versa if they are controlled by a parent organization—making it challenging to correctly identify relationships between parent and child companies. In addition, for legal purposes, companies sometimes register subsidiaries abroad to be able to run business activities in other countries, without necessarily associating a new ASN with them. This potential behavior adds another layer of uncertainty to *company-to-AS mapping*, since when no ASN is found for a given network operator company, it is unclear whether the mapping failed or the company actually does not own an ASN. For example, China Telecom operates subsidiaries in Brazil and Canada [69] but we believe these companies do not operate their own ASN.

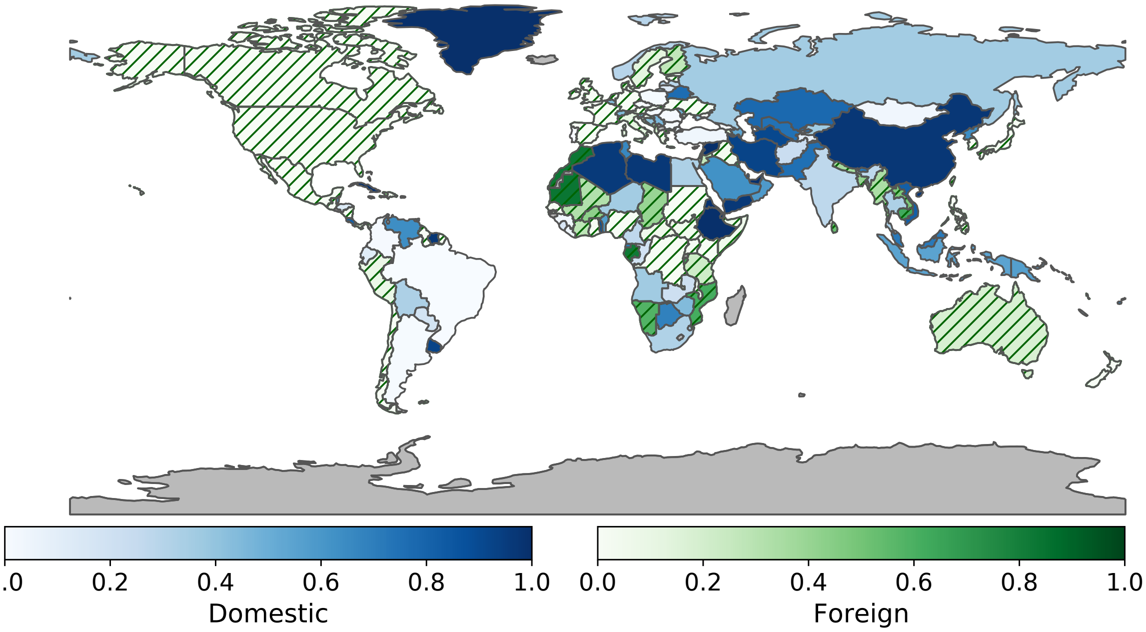


Figure 1: Footprint of state-owned Internet operators. In blue: The maximum between *(i)* the fraction of address space geolocated in the country that is originated through BGP by ASes owned by the same country and *(ii)* the fraction of eyeballs (according to APNIC Eyeballs dataset) from ASes owned by the same country. In green (lines): The same calculation but considering ASes owned by other countries.

Finally, ownership of telecommunication companies is very dynamic, and tracking these changes is challenging for three reasons. First, governments communicate privatizations and (re-)nationalizations in fairly diverse ways, including public releases, public acts, media conferences, or communications at stock exchanges. Second, variations of the state ownership of telecommunication companies happen frequently. E.g., Ucell, a subsidiary of Swedish minority-state-owned Telia's with presence in Uzbekistan, was acquired by Uzbekistan's government in 2018 [41, 77]. An example in the opposite direction is that while Angola's government has announced several times over the past 4 years its intention to privatize Angola Telecom [4, 21, 42] this transition has not yet occurred. Third, due to the size of the international telecommunication market, it is impractical to continuously monitor the shareholder structure of all telcos to detect changes on state participation.

3 METHODOLOGY OVERVIEW

Definition of state-owned AS. There is no commonly accepted definition of state-owned enterprise. The IMF, the OECD and the European Commission use different criteria [34]. However, these institutions agree on three elements to define a firm as state-owned: *(i)* the company has its own separate legal entity, *(ii)* the entity is partially controlled by a government unit, and *(iii)* the entity engages in commercial or economic activities. In this paper, we use these same criteria and, regarding point *(ii)*, we follow the IMF definition stated in the IMF's Fiscal Monitor report released in April 2020, Chapter *State-Owned Enterprises: The Other Government* [34], which considers a firm as state-owned if the government owns at least 50% of its equity[2]. In our study we specifically focus on federal-level (or equivalent) companies offering transit or unrestricted access to Internet connectivity (*i.e.,* we do not include in our definition companies operating only at subnational level). We call these companies *Internet operators* and—in the context of this paper—we define *state-owned AS* an AS owned by a state-owned Internet operator (*i.e.,* a state-owned Internet operator is in control of the AS number delegated by a Regional Internet Registry).

Data Discovery and Classification. Our process has three stages, which are depicted in the diagram in Figure 2. At a high level: we identify ASNs and company names (stage 1) to obtain a candidate list of companies to be manually verified (stage 2). During manual verification we also filter and enrich this data. An example of filtering is the exclusion of operators with minority state participation, whereas enrichment includes adding subsidiary companies. We conclude the process by obtaining a final list of ASNs operated by the selected companies and generating our final dataset (stage 3).

Reference Timeframe. We emphasize that the data we obtain as part of this process captures the state ownership of ASes during a specific time frame. For this research we generated the candidate lists and analyzed company ownership structures from June 2019 to November 2020. As mentioned in § 2, ownership structures are dynamic, and while we mitigate this risk, state ownership of these

[2]While we note that some literature suggests that a government may exercise significant influence over corporate decisions even when it owns a small number of shares [33], we cannot quantify government-exerted influence in these circumstances.

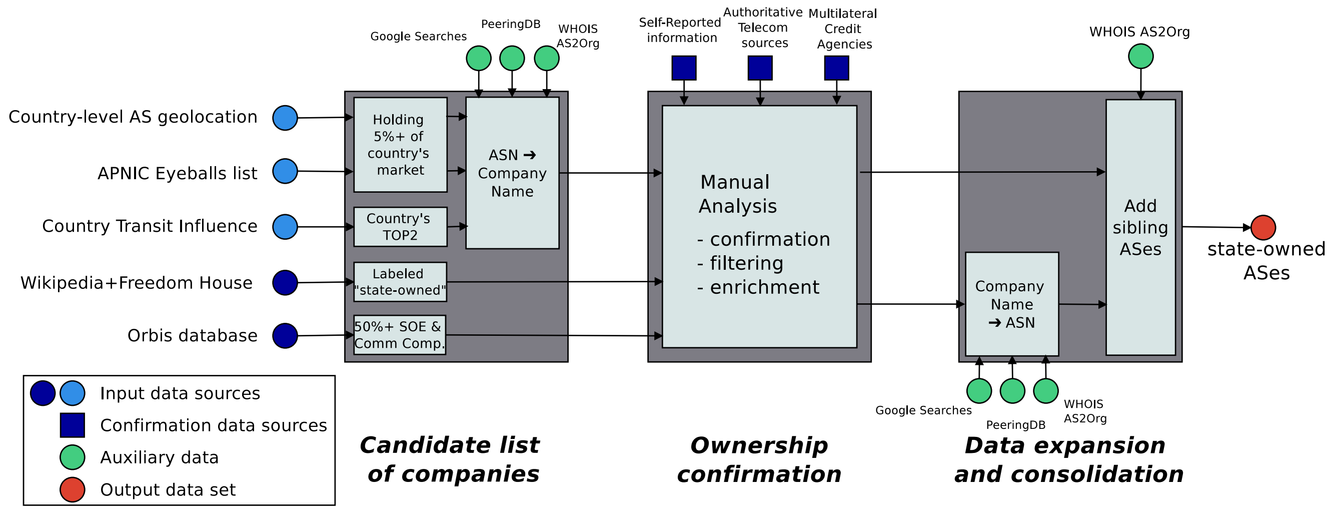


Figure 2: Block diagram of our data discovery and classification process. We describe it from left to right. Stage #1: We use five input sources (technical in light blue and non-technical in dark blue) to obtain a candidate list of potentially state-owned companies. Stage #2: For each company, we manually verify whether a federal-level government owns its majority. In addition, in this stage we filter operators with minority state participation. We also add subsidiary companies to the list of state-owned ASes. Stage #3: We add sibling ASNs obtaining a final list of ASNs operated by the selected companies. We then generate our final dataset.

companies might have changed even within the duration of this study. In § 9 we discuss the subject of dataset ageing.

4 CANDIDATE ASES AND COMPANIES

We bootstrap our process by analyzing multiple and varied data sources, represented as blue circles on the left-hand side of Figure 2, through which we select data to be manually examined in the next stage of our methodology (§ 5). These candidate data can be of two distinct types: *(i)* ASNs whose ownership we intend to verify or *(ii)* company names that according to our sources are (likely) state-owned and thus require us to verify such information. We call (Computer Networking) *Technical Sources* (§ 4.1) those from which we obtain lists of ASes, which we then map to actual company names (§ 4.2). We refer as *Non-technical Sources* (§ 4.3) to data sources from which we instead obtain names of companies reported as state-owned.

4.1 Candidate ASes

Technical Sources allow us to identify ASes providing Internet services and infer their country-specific market relevance. We use three different sources and approaches, which we describe in the following paragraphs: Country-level AS geolocation, APNIC eyeballs dataset, key transit providers in each country.

Country-level AS geolocation: We geolocate globally-routed network prefixes to identify the geographical footprint of every AS originating IP address space. We use CAIDA's prefix-to-AS list from July 1st, 2019 to obtain all pairs of BGP-routed prefixes and their correspondent origin ASes for the 68,283 visible ASes in the global routing table. We then use Digital Element's *NetAcuity Edge* IP geolocation service [26] to determine the (country-level) location of every IP address of each routed prefix.[3] The IP-level granularity of the geolocation process allows us to create a list of triplets containing *<Origin ASN (OASN), country, number of IP addresses that OASN originates in that country>*.

We limit the candidate list of providers to later examine for possible state ownership to networks with significant market share. We thus exclude ASes that originate less than 5% of a country's globally-routed IP addresses, obtaining a total of 793 ASes (≈1% of the total number in this dataset). This threshold should be sufficient to include all state-owned ASes that operate major access networks in each country.

APNIC eyeballs dataset: We rely on the number of "eyeballs" reported by APNIC to determine the most populated networks in each country. While estimating the population of Internet users of an AS is challenging due to the widespread use of NAT [65], APNIC has developed heuristics to estimate the eyeball populations leveraging web-based advertising [45]. APNIC's estimations report the eyeball population for 25,498 ASes. We use their estimates of AS eyeballs population as an additional variable to measure the market size of access networks. Similarly to what we do for the Country-level AS geolocation approach, we select only ASNs with an estimate of at least 5% eyeballs in a given country. We obtain a list of 716 unique ASes (≈3% of the total number in this dataset), which interestingly is a comparable—but smaller—number of ASes to those obtained through Country-level AS geolocation. There are 466 ASes in the intersection of both data sources and we obtain 1043 ASNs from the union of both.

[3] While geolocation databases are known to be unreliable at fine granularities, previous work has found them to be more accurate at the country level [14, 59], with Netacuity in particular having accuracy between 74–98% [38].

Countries' main upstream providers: Finally, we shift our attention from large *access* networks to key *transit* connectivity providers. Specifically, we use the Country-Level Transit Influence (CTI) metric [37] to select key transit networks providing international access in several countries. CTI is a BGP-based metric that captures the fraction of a country's IP addresses that are served by a particular transit network (AS). Formally, the transit influence of autonomous system AS on country C is the weighted fraction of C's address space for which AS is present on announced, preferred paths toward prefixes originated in C by a responsive AS that is visible in public BGP data [1, 2]. We include a more precise formulation in Appendix G.

In countries where transit providers (as opposed to peers) have been inferred as the dominant inbound modality [37], CTI allows us to identify a country's reliance on specific transit providers granting international connectivity. We hypothesize that in such countries the government may be engaged in deploying domestic transit connectivity. In fact, states have created diverse alternatives, such as establishing transit gateways connecting domestic ASes with international transit providers (*e.g.,* Syria's AS29386-Syrian Telecom), building national backbones (*e.g.,* Argentina's ARSAT-AS52361), or in some cases building their own submarine cable networks (*e.g.,* Angola's AS37468-ACS, Bangaldesh's AS132602-BSCCL or the WIOCC consortium—AS37662—in which some state-owned African ISPs hold shares).

CTI [3] has been applied in 75 countries comprising 1,314 unique ASes. In each of these countries we select the two highest CTI-ranked ASes for inclusion in our candidate list of ASes, resulting in 93 ASes (≈7% of the total number in this dataset).

4.2 Mapping ASes to their companies

When we combine the three technical sources we obtain a total number of 1091 ASes, which in total belong to 1023 different organizations, according to CAIDA's AS2Org data. We use WHOIS records and entries from PeeringDB [56] to identify the companies owning these ASes. We start from WHOIS records, since this is compulsory information required by RIRs from each organization requesting an AS number. Although WHOIS records have a per-RIR data structure, a few fields are common across all RIRs, such as ASN, AS name, organization, and at least one email and/or phone contact.

To mitigate errors in WHOIS records, we also use PeeringDB, a website providing a non-compulsory database of self-reported data. PeeringDB covers roughly 20% of the ASes registered in the WHOIS. Operators register their ASNs on PeeringDB to share information about peering or operational tasks (*e.g.,* how to contact NOC 24x7x365 teams in case of a failure) and to be visible on the platform in order to *e.g.,* attract more (transit) customers and peers. It is therefore in the interest of these ASes to keep their information up to date and we assume that the company names there reported are similar or identical to the brand names in order to facilitate their identification.

When unable to find any website mentioning the company names we obtained from WHOIS and PeeringDB, we Google-search for the DNS domains from the point of contacts there were listed, *e.g.,* URLs or emails. In these searches we found that the challenge in mapping these companies tends to be related to name alterations after rebrands, mergers and acquisitions. In general, AS-to-company mapping is challenging and requires further study.

4.3 Candidate Companies

We extend our data with names of telecom companies identified as state-controlled by our selected Non-Technical Sources.

Orbis: We query the commercial Orbis database to obtain a list of state-owned telecommunication companies. Orbis is produced by Bureau van Dijk's, a Moody's Analytics company that collects and distributes datasets with company information to financial risk assessors and governments [50, 80]. Orbis is a business information database containing information on more than 400 million companies and entities around the world, including their corporate ownership structures [79]. Orbis has been used in scientific publications in business and economic research fields [8, 13, 43, 49, 62] as well as in reports by national and international organizations, including the NBER [47], the OECD [36] and the World Bank [55]. Related to our work, a previous research study used Orbis to identify state-owned telecommunication companies in Africa [32] but to analyze the relationship between Internet shutdowns and the state ownership of Internet providers. Using the Orbis database engine filters we find telecommunications companies in which sovereign states own more than 50% of the equity, resulting in 994 companies.

Freedom House report & Wikipedia articles: We also add to our list companies reported as state-owned in Freedom House reports and in Wikipedia articles. Freedom House's Freedom of the Net project releases annual reports on each country's "Internet freedom". Freedom of the Net measures interventions by governments and non-state actors aimed at restricting Internet rights, and covers 65 countries. The reports are produced by in-country activists, civil society groups, academics, journalists, and tech and legal experts.

We also use Wikipedia to expand the candidate list of *candidate* state-owned companies. We find that two types of Wikipedia articles tend to include information about state-owned telcos: articles describing the country's communication landscape and articles listing the country's state-owned enterprises. We repeat this searching process for every country. We do not take state ownership in these articles at face value: we validate this information in the second stage of our process which will remove false positives (Figure 2). However, we expect Wikipedia articles and Freedom House reports to contain false negatives (*i.e.,* these reports and articles might miss some state-owned telcos): we mitigate such false negatives by including data from all the other data sources here discussed.

5 OWNERSHIP CONFIRMATION

In the second stage (Figure 2) of our process we manually examine the companies from the candidate state-owned list. We verify their ownership structure and business sector, we extend our examination to subsidiaries and parent companies, and we filter out organizations that do not strictly match our definition of state-owned Internet operator.

5.1 Confirmation data sources

To investigate the ownership structure of companies from our candidate list, we rely on manually consulting the following authoritative sources.

Self-reported information. We primarily look at company websites, government websites and corporate annual reports. In some countries (*e.g.,* Norway), transparency legislation requires their governments to fully disclose state participation in companies. When states hold shares of publicly-traded corporations, corporate ownership structures (and therefore state participation) are publicly available in the corporate annual reports. In other cases, websites of state-owned telcos explicitly declare the state control of the company, such as in Congo's CONGTEL website [68].

Authoritative telecommunication sources. We also rely on authoritative telecommunication sources such as the US SEC, the FCC, local regulators and the ITU. These regulators and organizations have purview over commercial and technical aspects of telecommunication companies, and their freely available documents may refer to the ownership structure of the company. Companies with commercial activities in the US (it may be the case of foreign state-owned companies) are subject to submit filings to US regulators such as the SEC and the FCC. Similarly, local regulators may request and disclose details of domestic companies, including state ownership. With a different set of goals, the ITU operates at international level and runs multiple commissions to promote infrastructure development and to assist developing countries [78]. These commissions regularly release documents and meeting materials which sometimes includes information about a country's telecom landscape. We use articles released by *CommsUpdate*, a well-known source of telecom news stories worldwide [20]. Its publisher, Telegeography, is a telecom market research company [20]. Several CommsUpdate articles include information about state-owned companies in their research on Internet markets.

Credit agencies. We use reports published by research and financial departments of multilateral credit agencies, such as the World Bank and the International Monetary Fund (IMF), describing countries telecom markets and including the presence of state-owned incumbents. Most of these reports are publicly available through the World Bank library, *e.g.,* [4].

5.2 Discovering state-owned subsidiaries

When manually investigating the ownership structure of the companies from our candidate list, we also look for parent state-controlled companies to identify subsidiaries. This way we actually discover *additional* companies that are not detected by our list of candidate companies.

Interestingly, we find that some state-owned subsidiaries provide Internet services in foreign countries. We define (and label in our output dataset) as *foreign state-owned subsidiary* a separate legal entity meeting two conditions: *(i)* whose state-owned parent's (or parents') holdings encompass more than 50% of the shares and *(ii)* that is registered in a foreign country. By capturing such features we hope to enable studies that cross socio-political domains: While governmental involvement in domestic Internet markets is often related to the mandate of economic prosperity, digital inclusion, and national cyber-defense, those factors do not apply when they operate abroad. A country's interest in extending operations abroad might instead be the intent of expanding into other profitable markets, or be a consequence of the nation's goals on international relations.

5.3 Excluded state-funded organizations

Following our definition in § 3, we exclude companies and ASes operated by all subnational jurisdictions: first-level (states, provinces, ...), second-level (municipalities, districts, ...), third-level (cities) or smaller administrative divisions. We remove companies belonging to lower than country-level administrative divisions to reduce the size of our problem and to avoid potential bias. To get a sense of the size, the ISO 3166-2 standard defines more than 5000 identifying codes for first-level administrative divisions (as of March 2021). Furthermore, it is unlikely that our data sources have uniform coverage across countries when considering a fine administrative granularity, which would have caused our dataset to be biased.

Based on our definition (§ 3), we also exclude state-run organizations offering access or transit services restricted to only certain sectors—*e.g.,* academic networks. While some of the excluded networks are owned by government organizations aimed at granting Internet access to certain restricted populations, *e.g.,* to close the digital divide or to connect educational institutions, they do not fall within our scope: federally-funded companies offering services to people and companies as any other commercial ISP would do. However, in Appendix E, we provide some insight about the excluded categories of companies.

6 DATA EXPANSION & CONSOLIDATION

In the last stage of our process (Figure 2), we map confirmed state-owned Internet operators to AS numbers, using—in reverse—the same methodology we apply in § 4.2. We then expand the list of confirmed state-owned ASes by including their sibling ASes using CAIDA's AS2org data [17, 18]. In conducting our analysis (§§ 4, 5), we also identified several sibling ASNs that were incorrectly not recognized as such by AS2Org (*e.g.,* because their AS names are completely different); we contributed our findings to the AS2Org project.

As the result of this process, we obtain two data products: a list of state-owned organizations and a mapping between these organizations and the ASNs they own. We save this data into an SQLite database (which we also export in JSON format) which is publicly available[4]. Listing 1 provides an example from our dataset, showing data for the Norwegian network operator Telenor in JSON format. In the list of state-owned organizations we include four types of fields:

- *Information specific to the organization*: name of the conglomerate the company belongs to, the CAIDA's AS2Org Org ID, the name of the organization, ISO-3361 country code, country name, country's RIR.
- *Confirmation sources that validated the inference*: type of confirmation source (*e.g.,* Company's website), the quote we use to determine the state ownership, the language of the quote, the URL to the confirmation data source and, in some cases, additional information (*e.g.,* specifying that a hedge

[4]The dataset is available here: https://github.com/estcarisimo/state-owned-ases

fund is state-owned). One added benefit of these fields is that they facilitate verifying in the future if the classification is still valid.

- Which *input sources* caused this organization to be originally added to the candidate list. We abbreviate the inputs sources using the following convention: G=Country-level AS geolocation; E=APNIC eyeballs dataset; C=Country Transit Influence; O=Orbis; W=Wikipedia & Freedom House.
- If the company is a *foreign subsidiary*: the parent company's Org ID, the name ('target_country_name') and ISO-3361 country code ('target_cc') of the country where it operates; in this case the 'ownership_cc' and 'ownership_country_name' fields refer instead to the country of the parent company.

```
# Ownership details of an identified
# state-owned organization
{
  "conglomerate_name": "Telenor",
  "org_id": "ORG-NA38-RIPE",
  "org_name": "Telenor Norge AS",
  "ownership_cc": "NO",
  "ownership_country_name": "Norway",
  "rir": "RIPE",
  "source": "Company's website",
  "quote": "Major Shareholdings: Government
    of Norway (54,7%)",
  "quote_lang": "English",
  "url":"https://www.telenor.com/investors/
    share-information/major-shareholdings"
  "additional_info": "",
  "inputs": [G, E, W, O],
  "parent_org":,
  "target_cc":,
  "target_country_name":,
}
# List of ASes operated by the identified
# state-owned organization
{
  "org_id": "ORG-NA38-RIPE",
  "asn": [2119, 8210, 8394, 8786, 39197,
    197943, 200168]
}
```

Listing 1: Example from our dataset for the Norwegian network operator Telenor (AS2119).

7 TAKEAWAYS FROM THE PROCESS

Applying our semi-manual classification process we extract insight about this previously unexplored problem and make several findings which we summarize in this section.

At the end of our process we obtain 989 state-owned ASes—including 193 foreign subsidiaries—from a total of 302 state-owned companies. In aggregate, state-owned ASes originate 17% of the Internet's address space announced in BGP. This fraction increases to 25% if we exclude the US, which does not have state-owned providers and is overrepresented in the global address space due to several largely unused but announcedaddress blocks [22]. Each data source contributes a comparable number of state-owned ASes (between 500 and 640, we include details in Appendix B), except for CTI (12 ASes)—which is expected, since it selects a very specific and narrow class of ASes.

***All* sources provide a unique contribution.** Interestingly, we find[5] that *each* data source contributes a unique set of ASes that no other data source captures. The unique contribution (9 ASes of 12) from the CTI dataset is perhaps the most surprising: most of them are companies exclusively providing transit, *i.e.,* not directly serving a large user population; likely because of the nature of their business, they fly under the radar of the other sources even if they are critical for Internet communications (we provide details about these companies in the Appendix D). These findings confirm our intuition that a broad and diverse set of data sources is needed in order to comprehensively identify state-owned ASes and that existing commercial databases (like Orbis) alone are not sufficient to tackle this problem.

To provide more details on this phenomenon, in Figure 3 we show the Venn diagram of the contributions when grouping the sources in 3 categories based on their diverse nature: *(i)* popular/relevant ASes (Technical Data Sources), *(ii)* sources focused specifically on state-ownership (Wikipedia and Freedom House), and *(iii)* a commercial business information database (Orbis). While 193 ASes are

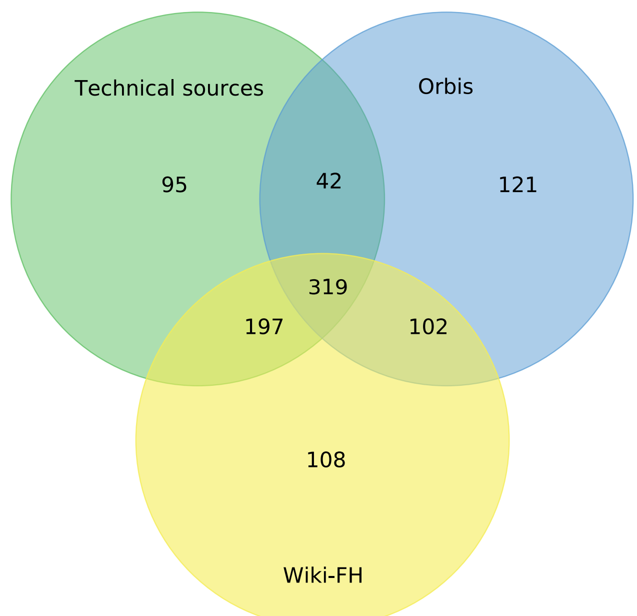


Figure 3: Venn diagram showing overlap and individual contributions of each class of data sources.

shared among all three categories, each category provides significant unique input. In particular, it is interesting that the Technical

[5]Full Venn's diagram in the Appendix C.

Confirmation source	Companies
Company's website	161
Company's annual report	44
Freedom House	33
TG's commsupdate	22
World Bank	20
ITU	6
FCC	4
News	2
regulator	2
Others	9

Table 1: Contribution of each type of confirmation data sources.

Data Sources yield a quite significant number (95) of ASNs that were not found through the others.

Company websites are the major source of confirmation. We then examine which *"confirmation sources"* allowed our manual analysis to verify the state-ownership of the 302 companies — including 84 foreign subsidiary companies— that own these 989 ASes (Table 1). The companies' websites are the most prevalent source of data, covering roughly 50% of the companies.

Freedom house is a reliable source. Freedom House is the second confirmation source: When a company is labeled as state-owned by Freedom House's *Freedom of the Net* report, in most cases we are able to manually confirm through other sources. However, for 42 companies we could not find any authoritative, external sources confirming or refuting Freedom House's assessment. Also, we did not find any false positives of Freedom House's state-ownership assessment. We believe that Freedom House is a reliable source, since it relies on countries' experts to generate their reports, and it is thus safe it to also use it at this stage of our process.

Other authoritative but non-comprehensive sources. Finally, data from reports of the world's largest multilateral credit agencies, the World Bank, and the IMF allowed us to confirm 25 companies (Table 1). However, based on our experience we believe that due to their role, these institutions are more likely to report information for the countries they provide aid to, which are mostly in the developing world.

Insights about using a commercial database. Through the confirmation process we found Orbis incorrectly labels as state-owned 12 companies (false positives) and misses to include, or does not label as state-owned, 140 companies (false negatives). Most of the false positives are foreign subsidiaries. 3 of 12 are wrongly assigned to the Colombian government: 2 are labelled as federally-owned while in fact they are owned by counties, the third one—*Comunicación Celular de Colombia* (formerly *COMCEL*, now *Claro*)—is owned by the private conglomerate América Móvil [64, 70].

Orbis' false negatives are spread across 79 countries. Most of these companies are small and/or in the developing world (Latin America, Central Asia, Southeast Asia and Africa). In the LACNIC region Orbis does not capture *any* state-owned telcos in 11 of the 14 countries in which we find state-owned telcos. For example, Argentina's ARSAT and Uruguay's ANTEL are in the Orbis database but are not labelled as state-owned. We note similar limitations in Central Asia, where Orbis reports no state-owned telcos in Iran, Kazakhstan, Uzbekistan and Tajikistan, and only partially covers Azerbaijan (*e.g.*, in Azerbaijan, Orbis did not report BakTelecom, which Freedom House correctly reports as state-owned). We also observed lack of coverage in Vietnam where for example Freedom House reports state-owned providers.

Third-party validation. We further validated our resulting dataset as much as possible with the help of local experts. We obtained feedback from two local experts. The first one has knowledge of the entire LACNIC region, being a scientist specialized in ICT development who also worked at a regional operator consortium with presence across the entire LACNIC region. The second local expert is an engineer from France specialized in computer networks and working at AFNIC [31]. The expert in Latin America validated the 35 ASNs we identified in the region (belonging to 14 countries) while the one in France validated our findings for the two French companies we have found. In both cases, the experts reported neither false positives or false negatives in our data.

Large ASes with government minority ownership. In our manual analysis, we also identified 302 minority state-owned ASes, which we excluded from our generated dataset based on our definition (§ 3): we did not specifically search for minority participation but we took note of the cases that we encountered in our process, which clearly represent a subset of all ASes with state-minority participation. Among them, we find some large players such as Deutsche Telekom (AS3320, Germany's equity: 31%) [72], Orange (AS5511, France's equity: 22.95%) [53], Telia (AS1299, Sweden's equity: 39.5%) [73] and Bharti Airtel (AS9498, Singapore's SingTel equity: 35.1%) [66, 67].

Multi-government joint ventures. Interestingly, we find instances where two governments jointly own a firm. This is the case of PTCL (AS17557) and Telkomsel (AS23693) are companies owned by two countries, Pakistan and the UAE [40, 60], and Indonesia and Singapore [74], respectively. However, in both firms one country holds the largest equity in the company, Pakistan in PTCL (70%) and Indonesia in Telkomsel (65%). This was also the case of BICS (AS6774), a long-term joint business between Belgium and Switzerland that ended in February 2021 when Belgium's Proximus acquired the rest of the shares of the company [12].

8 A FIRST LOOK AT STATE-OWNED ASES

In this section, we combine our list of state-owned ASes with other data to study the footprint of state-owned ASes in the world.

A global view. Table 2 summarizes the state ownership of Internet operators at a country-level granularity. We find that 53% (*i.e.,* 123) of the world's countries are majority owners of Internet operators, highlighting that this is a widespread phenomenon. We also find that state-owned companies of 19 countries control subsidiaries offering Internet services abroad. The table also shows that at least 24 countries have minority ownership of Internet operators (§ 7). Note that some countries may appear in multiple categories; for example, Singapore owns the majority of the equity of SingTel (AS7473), which operates Optus in Australia (AS7474) and is also a minority owner of Telkomsel (AS23693) in Indonesia.

Table 4 shows the state ownership of telcos at a RIR-level granularity. In all RIRs except ARIN, more than 40% of the country

Participation in	# of countries
state-owned operators	123
subsidiaries	19
minority state-owned operators	24
Total countries	136

Table 2: Number of countries we detected to own Internet operator businesses.

Owner country (cc)	#	Country Codes of the subsidiaries
UAE (AE)	12	AF, BF, BJ, CI, EG, GA, MA, ML, MR, NE, TD, TG
China (CN)	9	AU, GB, HK, MO, NL, PK, SG, US, ZA
Qatar (QA)	9	DZ, ID, IQ, KW, MM, MV, OM, PS, TN
Norway (NO)	9	BD, DK, FI, MM, MY, PK, SE, TH, UK
Vietnam (VN)	9	BI, CM, HT, KH, LA, MZ, PE, TL, TZ
Singapore (SG)	6	AU, HK, JP, KR, LK, TW
Malaysia (MY)	5	BD, ID, KH, LK, NP
Colombia (CO)	4	AR, BR, CL, PE
Serbia (RS)	3	AT, BA, ME
Indonesia (ID)	3	MY, SG, TL
Bahrein (BH)	3	IM, JO, MV
Tunisia (TN)	3	CY, MR, MT
Saudi Arabia (SA)	2	BH, KW
Fiji (FJ)	1	VU
Mauritius (MU)	1	UG
Belgium (BE)	1	LU
Switzerland (CH)	1	IT
Russia (RU)	1	AM
Slovenia (SI)	1	AL

Table 3: Foreign subsidiaries are a widespread phenomenon. 19 countries have subsidiaries with operations in 70 foreign countries, sometimes in an entirely different continent. The first column indicates the country where the state has a majority participation; the third column lists the countries where the subsidiaries operate.

	APNIC	RIPE	ARIN	AFRINIC	LACNIC	World
# companies	56	76	29	56	31	248
# countries	30	47	2	30	14	123
% countries	54	62	7	45	50	50

Table 4: State-owned Internet operators by RIR.

members have at least one state-owned network. However, most RIRs serve a large number of countries over a vast and geopolitically diverse territory (*e.g.,* RIPE comprises 76 members spanning from Northern Europe to the Middle East); in Appendix A we show a world map with all countries that have majority participation in Internet operators.

The ability to identify foreign subsidiaries is an important feature of our dataset. We find this phenomenon is broad and touches every continent. Table 3 shows which foreign countries (third column) host Internet operators controlled by companies owned by other nation states (first column). Sometimes these relationships cross continents.

Internet access markets. We provide a more detailed geographic perspective of state-ownership, nationally and abroad, in the heatmap in Figure 1, where we look at (an approximation[6] of) Internet access market footprint. Here, for each country we calculate two numbers:

- (*blue*) The maximum between *(i)* the fraction of address space geolocated in the country that is originated through BGP by ASes owned by the same country and *(ii)* the fraction of eyeballs (according to APNIC Eyeballs dataset) associated with ASes owned by the same country.
- (*green*) The same calculation as the previous point but considering ASes owned by *other* countries.

We then select the maximum between these two numbers and color the country on the map accordingly. We find that state ownership is a phenomenon much more prevalent in Africa and Asia. In addition, in a large number (12) of African countries, operators owned by other states have a significant footprint (>5%). Specifically, in 6 of these 12 countries, our estimates suggest that foreign state-owned providers hold more than 50% of the access market. Australia is another interesting case, where Singapore's SingTel operates Optus (AS7474), one of the major providers in the country (we estimate a footprint of 18.2%).

We show the fraction of network-access markets controlled by domestic, state-owned ASes in Figure 4 (data from June 2020). These numbers reveal a significant participation of many states in their network-access markets: the state's footprint is greater than 50% of IP addresses in 49 countries. The same is true for the share of eyeballs in 42 countries. Furthermore, we find 13 and 14 countries by IP address space and eyeballs, respectively (18 countries after combining both groups) where the state footprint on access networks is over 90%. We investigated whether this phenomenon is mostly related to the size of the country in terms of land or population, but we find that only 5 of these countries have fewer than 1M people. We list these 18 countries in the Appendix F and we will publish—upon paper acceptance—the full data for each country on a dedicated website.

Interestingly, the fraction of the address space originated by state-owned ASes in AFRINIC's countries is the largest out of all the regions; AFRINIC also has the largest presence of foreign state-owned ASes. In the African continent, Ooredoo and Etisalat operate multiple subsidiaries with important market shares in various countries. Conversely, in the LACNIC region, where half of the countries have state-owned ASes, the fraction of countries' address space originated by state-owned ASes is quite small. Exceptions include Cuba, Uruguay, Venezuela and Costa Rica, where state-owned providers originate 90%, 90%, 65% and 80% of the address space. We investigate LACNIC's address space and find that Brazil, Argentina and Colombia—the first (153M), third (23.8M) and the

[6] We use the fraction of IP addresses—geolocated at country-level—and of estimated eyeballs as proxies for the fraction of Internet access market in each country.

ASN-ASname	Country (cc)	cust. cone
7473-SingTel	Singapore (SG)	4235
12389-Rostelecom	Russia (RU)	3778
20485-TTK	Russia (RU)	3171
37468-Angola Cables	Angola (AO)	1843
262589-Internexa	Colombia (CO)	1315
4809-China Telecom	China (CN)	1134
3303-Swisscom	Switzerland (CH)	702
20804-Exatel	Poland (PL)	699
10099-China Unicom	China(CN)	595
132602-BSCCL	Bangladesh (BD)	556

Table 5: Ten largest customer cones of state-owned ASes (June 2020).

fourth (19M) large address spaces in Latin-America—do have state-owned providers but these operators primarily offer transit services: Telebras (AS53237), ARSAT (AS52361) and Internexa (AS18678).

Transit connectivity market. Next, we investigate the the presence of the state-owned ASes in the Internet-wide transit ecosystem. Table 5 displays the 10 largest customer cones of state-owned ASes using CAIDA's ASRank data from June 2020. We note that some state-owned providers serve transit to a large number of ASes in their customer cone (*i.e.,* they have a relevant role in terms of connectivity) and that this phenomenon is not limited to one continent but happens in 4 of the 5 RIRs. We note some differences among these operators. Some of them have clearly a large international footprint: SingTel (AS7473), Angola Cables (AS37468) and China Telecom (AS4809) operate large networks that include submarine cables. By contrast, the Russian transit networks Rostelecom (AS12389) and TTK (AS20485) mostly serve the domestic market. Similarly, Internexa's subsidiary in Brazil (AS262589) mostly provides transit services to client within Brazil.

We also identify, among the state-owned ASes, those with the fastest growing customer cones during the past decade. To that end, we compute a temporal-linear regression based on CAIDA's ASRank data. In the top-10 we identify two interesting cases: Angola Cables (AS37468) and BSCCL (AS132602), two state-owned submarine cable networks in developing countries—Angola and Bangladesh[7]. Figure 5 shows the growth of their customer cones from January 2010 to June 2020. The establishment of these companies and the deployment of their submarine cable networks were the response of the respective governments to the limited international connectivity of these countries [16, 44].

9 LIMITATIONS

Scale. We manually investigate the company structure of thousands of telcos, however, due to the size of the Internet our research only covered a fraction of the network operators. We believe our dataset includes the largest state-owned access and transit ASes, but it is less likely to capture small state-owned telecom companies. We note that our 5% threshold for IP addresses and eyeballs is not uniformly applied, because non-technical sources such as Freedom House often do not include information regarding a state-owned company's market share. Therefore, it is possible that we included in our dataset companies with a market footprint smaller than our 5% threshold. Another limitation of our approach is that non-technical sources provide us data at a company level, which makes it difficult to quantify the sources' coverage in terms of number of ASes.

Visibility and data interpretation. We rely on online resources to confirm state participation in telecom companies which limits our inference to companies whose ownership is reported online by their own resources (e.g., websites, company's annual reports) or by authoritative third-party resources (e.g., World Bank's report repository). Our interpretation of these sources might also result in inaccuracies. This is exacerbated by our limited expertise (as computer scientists) in the fields of economics and law, given that many of the reports we study are long and complex. Our direct personal experience is also mostly limited to countries in Europe and the Americas, potentially causing us to miss crucial local context. These issues may cause both false positives and false negatives in state ownership. However, our discussions with experts revealed no such inaccuracies.

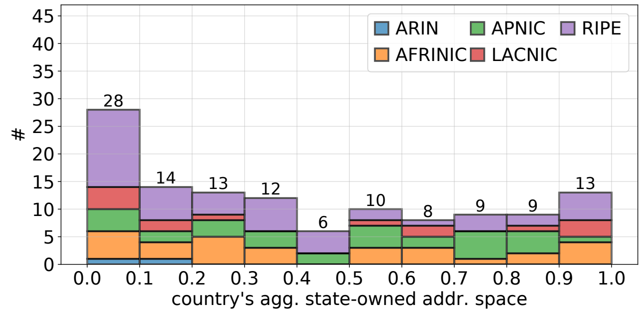
Another factor with a potential impact on our visibility is ICT adoption, as that is reflected in the number of online resources reporting state-owned enterprises (SOEs) [48]. Authoritative documents released by local institutions and websites reporting state ownership of companies are more likely to be available in countries with more mature digital ecosystems. Although this may restrict our view of a country, in this paper, we also include data from international bodies which often cover countries with presumably less mature digital ecosystems.

Misleading Company Names. Company names are central in our process of identifying state control of ASes. Misleading company names are a potential source of inaccuracies, *e.g.,* outdated names after privatizations and nationalizations. For example, the government of Fiji nationalized Vodafone Fiji in 2014 [29] but the company name has not changed. A reasonable observer might conclude, incorrectly, that the company belongs to Vodafone (a private congolomerate) when it is actually controlled by Fiji.

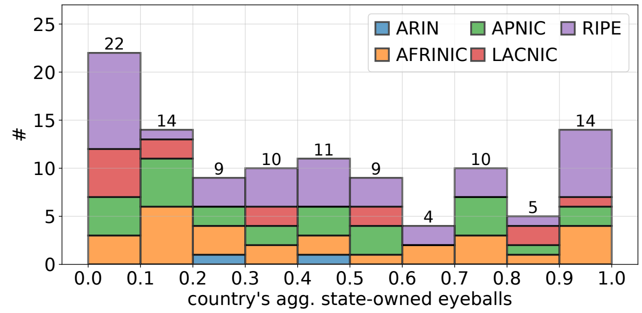
Changes in ownership over time. As we discussed in § 2, ownership structure is dynamic. As a consequence, our list would require maintenance to keep up to date. While conducting our study, we noted that privatizations are relatively rare. Another type of event that could impact the content of the list in the future is the birth of new state-owned providers. Moreover, in the future, some state-owned companies may break into new markets creating new foreign subsidiaries. In such cases, new state-owned companies or foreign subsidiaries will have to be incorporated into the list. We leave for future work a systematic study regarding the churn of Internet providers' ownership by states.

We believe that our dataset provides a valuable seed for (recurring) future work in this area: confirming the current validity of our list would be significantly less taxing than generating the initial list. For this paper, our manual inspection process took a single person approximately 4.6 months: 2.8 months for *technical sources* (1.6 months for Country-level AS geolocation, 1 month for APNIC eyeballs dataset, one week for Countries' main upstream

[7] We verified that these 2 ASes were owned by their respective states for the entire decade.



(a) Countries' address space



(b) Countries' eyeballs

Figure 4: Aggregated footprint of all state-owned ASes in their home countries in terms of address space and eyeballs.

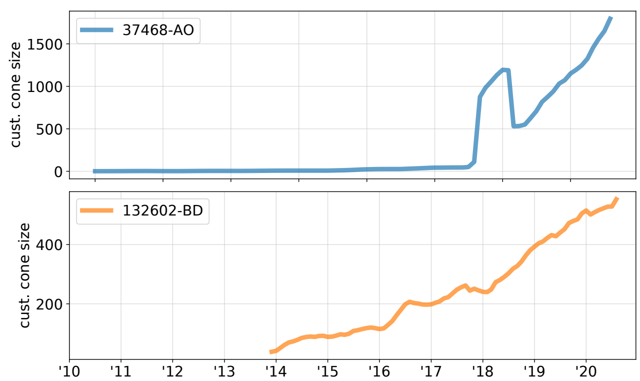


Figure 5: Growing pace of the customer cones of Angola Cables (AS37468) and BSCCL (AS132602) in the past decade.

providers) and 1.8 months for *non-technical sources* (0.8 months for Wikipedia/Freedom House and 1 months for Orbis).

Subsequently, future studies would expand the list to introduce nationalizations and known company expansions, which year by year is likely to be fractional in size compared with the preceding year's aggregate list. In the future we will try to expand this research into a cooperative project in which volunteers, from any community, could audit and suggest updates to our dataset. Making the dataset publicly available will also help with soliciting feedback.

Further, we only have evidence that the state ownership of the ASes we found in this research is valid for the reference timeframe we mention in § 3. Our list of state-owned companies does not capture companies that were state-owned at some point in the past but then were privatized (*e.g.,* Telecom Italia [39]). Moreover, the current data structure of our list does not report since when a company is under state control. Expanding the list with a field for timestamps would allow researchers to conduct more precise longitudinal analyses.

Language. We confirm the state ownership of a company by manually studying authoritative documents. Globally investigating state ownership of providers, then, might involve documents in many languages. However, in applying our methodology we noticed that the vast majority of such sources—including company websites and reports—have versions in English or Spanish. International organizations such as the IMF also publish reports primarily in English. However, we acknowledge that our view of some countries could have been compromised if for some of their companies the documents in English or Spanish we accessed were not sufficiently detailed.

10 RELATED WORK

Previous research efforts also shared the goal of classifying ASes but most are focused on classifying *all* ASes on the Internet and on assigning categories designed for purposes that are different from our study (*e.g.,* role in the AS-level Internet topology). Dhamdhere *et al.* [25] use decision trees to classify ASes—based on their peering and customer degree—into *enterprise customers, small transit providers, large transit providers, and content/hosting/access providers.* A similar study by Dimitropoulos *et al.* [27] also apply machine learning to the problem of classifying ASes using RIR registration and routing data. Wahlisch *et al.* [81] present a classification of the economic purpose of ASes in Germany by extending a taxonomy created by the German Government, and using partially-verified assignment using keyword search on AS names, descriptions, and address fields. The verification is by manual inspection.

In the Internet measurement literature, some studies leveraged the concept of state-owned Internet operator to better analyze events and extract insight: Focusing on Internet censorship, Dainotti *et al.* [23] and Raman *et al.* [61] analyzed the role of state-owned providers in country-wide Internet connectivity shutdowns and interception of HTTPS traffic, respectively. Carisimo *et al.* [19] argued the absence or lack of development of IXP in Latin American countries with almost concentrated Internet markets in the hands of state-owned ISPs. Fontugne *et al.* [30] studied the structure of the Internet in Crimea noting the role of Russian state-owned operators providing connectivity to the region.

11 CONCLUSIONS & DISCUSSION

In this paper we proposed a novel methodology to identify state-owned ASes of Internet operators using multiple and varied data sources, including both technical and non-technical sources. We found that each data source provides a unique contribution to the identification process, which indicates that our approach is well conceived but also highlights the challenging nature of the problem and the possibility that false negatives are still present.

We have also learned about the various shortcomings and limitations of a prominent commercial business information database when utilized for this research problem. Interestingly, we would have not detected some influential transit providers neither through this database or the other input sources if we did not specifically include in our input a selection of transit operators.

Through our method we identified 989 state-owned ASes, including 193 foreign subsidiaries, from 123 and 18 countries, respectively. We discovered a higher prevalence of state-owned companies across Africa and Asia. In addition, when considering IP addresses and eyeballs as rough estimates of Internet access market share, we found that in 18 countries state-owned operators hold at least 90% of the estimated access market in their home countries. Interestingly, we also found that in 12 African countries, foreign subsidiaries hold a larger share of the access market than the country's state-owned ASes (if they were to have one). Moreover, in 6 out of these 12 countries, foreign subsidiaries originate more than 50% of the estimated access market.

As a result of applying our methodology, this work also contributes a new publicly available data set containing the state-owned organizations and ASes we have identified, as well as additional metadata reporting confirmation sources and flags identifying foreign subsidiaries. In addition, this paper contributes a new dimension to the broader topic of AS classification.

We hope that this work and the associated data set will inspire and enable a broad set of Internet measurement studies and interdisciplinary research. For example this data could be correlated with evidence about network interference for state-sponsored censorship or surveillance. Topological studies could better understand the role of autocratic governments in shaping the national Internet infrastructure. Moreover, this data could also help gaining new insight about the digital divide, often seen as a problem of infrastructure deployment in rural and urban communities [7, 15, 35, 57, 58, 63, 76]. In many countries the government's response to digital divide and specifically lack of Internet infrastructure is investing in state-owned operators. Our data set of state-owned providers could be a valuable resource to comprehensively track and measure the impact of governments' policy, for example to understand whether more ASes located in remote non metropolitan areas appear in the customer cone of state-owned providers.

12 ACKNOWLEDGEMENTS

This study was supported in part by the National Science Foundation, Grant No. 1705024. Author Gamero-Garrido was supported in part by the Microsoft Research Dissertation Grant (2019).

REFERENCES

[1] 2020. RIPE Routing Information Service (RIS). https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris. (2020).

[2] 2020. RouteViews. http://www.routeviews.org/routeviews/. (2020).

[3] Anonymous Authors. 2021. Quantifying Nations' Exposure to Traffic Observation and Selective Tampering. In *(Concurrent submission) ACM SIGCOMM Conference on Internet Measurement (IMC '21)*. ACM, New York, NY, USA, 13.

[4] World Bank. 2021. ANGOLA: SYSTEMATIC COUNTRY DIAGNOSTIC. CREATING ASSETS FOR THE POOR. https://openknowledge.worldbank.org/bitstream/handle/10986/31443/angola-scd-03072019-636877656084587895.pdf. (2021).

[5] World Bank. 2021. Case Study on the Employees Provident Fund of Malaysia. http://documents1.worldbank.org/curated/en/197861540400101962/pdf/131289-WP-WorldBankReport-PUBLIC.pdf. (2021).

[6] World Bank. 2021. CORPORATE GOVERNANCE COUNTRY ASSESSMENT: Malaysia. http://documents1.worldbank.org/curated/en/216211468088741775/pdf/908220ROSC0Box0laysia0201200PUBLIC0.pdf. (2021).

[7] Karine Barzilai-Nahon. 2006. Gaps and bits: Conceptualizing measurements for digital divide/s. *The information society* 22, 5 (2006), 269–278.

[8] Sebastian Beer and Jan Loeprick. 2015. Profit shifting: drivers of transfer (mis) pricing and the potential of countermeasures. *International Tax and Public Finance* 22, 3 (2015), 426–451.

[9] Amanah Raya Berhad. 2021. Corporate Information. https://www.amanahraya.my/profile/. (2021).

[10] Khazanah Nasional Berhad. 2021. About Us. https://www.khazanah.com.my/who-we-are/about-us/. (2021).

[11] Telekom Malaysia Berhad. 2021. Capital Structure. https://tm.listedcompany.com/capital_structure.html. (2021).

[12] BICS. 2021. Proximus acquires full ownership of BICS, securing the flexibility to execute the development and growth path of BICS and TeleSign. https://bics.com/news/proximus-acquires-full-ownership-of-bics-securing-the-flexibility-to-execute-the-development-and-growth-path-of-bics-and-telesign/. (2021).

[13] Nicholas Bloom, Aprajit Mahajan, David McKenzie, and John Roberts. 2010. Why do firms in developing countries have low productivity? *American Economic Review* 100, 2 (2010), 619–23.

[14] Bradley Huffaker and Marina Fomenkov and kc claffy. 2011. Geocompare: a comparison of public and commercial geolocation databases - Technical Report. Cooperative Association for Internet Data Analysis (CAIDA), May 2011. (2011).

[15] Francesco Bronzino, Nick Feamster, Shinan Liu, James Saxon, and Paul Schmitt. 2021. Mapping the Digital Divide: Before, During, and After COVID-19. *During, and After COVID-19 (February 17, 2021)* (2021).

[16] Angola Cables. 2021. COMPANY PROFILE. https://www.angolacables.co.ao/wp-content/uploads/2020/08/RC_Angola-Cables-2013_EN-1.pdf. (2021).

[17] Xue Cai, John Heidemann, Balachander Krishnamurthy, and Walter Willinger. 2010. Towards an AS-to-organization map. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 199–205.

[18] CAIDA. 2021. Inferred AS to Organization Mapping Dataset. https://www.caida.org/data/as-organizations/. (2021).

[19] Esteban Carisimo, Julián M Del Fiore, Diego Dujovne, Cristel Pelsser, and J Ignacio Alvarez-Hamelin. 2020. A first look at the Latin American IXPs. *ACM SIGCOMM Computer Communication Review* 50, 1 (2020), 18–24.

[20] commsupdate. 2021. About. https://www.commsupdate.com/about/. (2021).

[21] Telegeography's commsupdate. 2021. Angola Telecom finishes restructuring ahead of privatisation. https://www.commsupdate.com/articles/2019/10/10/angola-telecom-finishes-restructuring-ahead-of-privatisation/. (2021).

[22] Alberto Dainotti, Karyn Benson, Alistair King, Bradley Huffaker, Eduard Glatz, Xenofontas Dimitropoulos, Philipp Richter, Alessandro Finamore, and Alex C Snoeren. 2016. Lost in space: improving inference of IPv4 address space utilization. *IEEE Journal on Selected Areas in Communications* 34, 6 (2016), 1862–1876.

[23] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. 2011. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 1–18.

[24] Oficina de Planeamiento y Presupuesto del Uruguay. 2021. Portal de Transparencia Presupuestaria. https://transparenciapresupuestaria.opp.gub.uy/inicio/empresas-públicas. (2021).

[25] Amogh Dhamdhere and Constantine Dovrolis. 2008. Ten Years in the Evolution of the Internet Ecosystem. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*. ACM, New York, NY, USA, 183–196. https://doi.org/10.1145/1452520.1452543

[26] Digital Element. 2021. NetAcuity. https://www.digitalelement.com/solutions/. (2021).

[27] X. Dimitropoulos, D. Krioukov, G. Riley, and k. claffy. 2006. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *Passive and Active Network Measurement Workshop (PAM)*. PAM 2006, Adelaide, Australia.

[28] KWSP EPF. 2021. Organisation Structure: Board Members. https://www.kwsp.gov.my/about-epf/corporate-profile/organisation-structure#boardMembers. (2021).

[29] Vodafone Fiji. 2021. About Vodafone Fiji. https://www.vodafone.com.fj/about/about-us. (2021).

[30] Romain Fontugne, Ksenia Ermoshina, and Emile Aben. 2020. The Internet in Crimea: a Case Study on Routing Interregnum. In *2020 IFIP Networking Conference (Networking)*. IEEE, 809–814.

[31] Association francaise pour le nommage Internet. 2021. https://www.afnic.fr/en/. (2021).

[32] Tina Freyburg and Lisa Garbe. 2018. Blocking the bottleneck: Internet shutdowns and ownership at election times in sub-Saharan Africa. *International Journal of Communication* 12 (2018), 3896–3916.

[33] International Monetary Fund. 2014. GOVERNMENT FINANCE STATISTICS MANUAL 2014. https://www.imf.org/external/Pubs/FT/GFS/Manual/2014/gfsfinal.pdf. (2014).
[34] International Monetary Fund. 2020. STATE-OWNED ENTERPRISES: THE OTHER GOVERNMENT. https://www.imf.org/~/media/Files/Publications/fiscal-monitor/2020/April/English/ch3.ashx. (2020).
[35] Bjørn Furuholt and Stein Kristiansen. 2007. A rural-urban digital divide? Regional aspects of Internet use in Tanzania. *The Electronic Journal of Information Systems in Developing Countries* 31, 1 (2007), 1–15.
[36] Peter N Gal. 2013. Measuring total factor productivity at the firm level using OECD-ORBIS. (2013).
[37] Alexander Gamero-Garrido. 2021. *Transit Influence of Autonomous Systems: Country-Specific Exposure of Internet Traffic*. Ph.D. Dissertation. UC San Diego.
[38] Gharaibeh, Manaf and Shah, Anant and Huffaker, Bradley and Zhang, Han and Ensafi, Roya and Papadopoulos, Christos. 2017. A look at router geolocation in public and commercial databases. In *ACM Internet Measurement Conference (IMC)*.
[39] Andrea Goldstein. 2003. Privatization in Italy 1993-2002: Goals, institutions, outcomes, and outstanding issues. *Institutions, Outcomes, and Outstanding Issues (April 2003)* (2003).
[40] Etisalat Group. 2021. 2020 Etisalat Group Annual Report (page 46). https://etisalat.com/en/system/com/assets/docs/annual-report/2020/en-2020-etisalat-group-annual-report.pdf. (2021).
[41] Freedom House. 2019. Freedom of the 2019: Uzbekistan. https://freedomhouse.org/country/uzbekistan/freedom-net/2019. (2019).
[42] Freedom House. 2021. Freedom of The Net: Angola 2018. https://freedomhouse.org/country/angola/freedom-net/2018. (2021).
[43] Harry Huizinga and Luc Laeven. 2008. International profit shifting within multinationals: A multi-country perspective. *Journal of Public Economics* 92, 5-6 (2008), 1164–1182.
[44] Faheem Hussain. 2011. ICT sector performance review for Bangladesh. *Available at SSRN 2013707* (2011).
[45] Geoff Huston. 2014. How Big is that Network? https://labs.apnic.net/?p=526. (2014).
[46] ICANN. 2013. 2013 Registrar Accreditation Agreement. https://www.icann.org/resources/pages/approved-with-specs-2013-09-17-en#whois-accuracy. (2013).
[47] Sebnem Kalemli-Ozcan, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas. 2015. *How to construct nationally representative firm level data from the Orbis global database: New facts and aggregate implications*. Technical Report. National Bureau of Economic Research.
[48] Raul Katz and Fernando Callorda. 2018. Accelerating the development of Latin American digital ecosystem and implications for broadband policy. *Telecommunications Policy* 42, 9 (2018), 661–681.
[49] Loet Leydesdorff and Ping Zhou. 2014. Measuring the knowledge-based economy of China in terms of synergy among technological, organizational, and geographic attributes of firms. *Scientometrics* 98, 3 (2014), 1703–1719.
[50] Moody's. 2021. Moody's Completes Acquisition of Bureau van Dijk. https://ir.moodys.com/news-and-financials/press-releases/press-release-details/2017/Moodys-Completes-Acquisition-of-Bureau-van-Dijk/default.aspx. (2021).
[51] The Government of Sweden. 2021. State-owned enterprises. https://www.government.se/government-policy/state-owned-enterprises/. (2021).
[52] Prime Minister's Office. 2021. State majority-owned companies . https://vnk.fi/en/state-majority-owned-companies. (2021).
[53] Orange. 2021. Consolidated financial statements 2019. https://www.orange.com/sirius/derniers_resultats/en/Consolidated%20financial%20statements%202019.pdf. (2021).
[54] The Number Resource Organization. 2021. Regional Internet Registries. https://www.nro.net/about/rirs/. (2021).
[55] Alberto Osnago, Nadia Rocha, and Michele Ruta. 2017. Do deep trade agreements boost vertical FDI? *The World Bank Economic Review* 30, Supplement_1 (2017), S119–S125.
[56] PeeringDB. 2021. PeeringDB. https://www.peeringdb.com. (2021).
[57] Lorna Philip, Caitlin Cottrill, John Farrington, Fiona Williams, and Fiona Ashmore. 2017. The digital divide: Patterns, policy and scenarios for connecting the 'final few'in rural communities across Great Britain. *Journal of Rural Studies* 54 (2017), 386–398.
[58] Lorna J Philip, Caitlin Cottrill, and John Farrington. 2015. 'Two-speed'Scotland: Patterns and implications of the digital divide in contemporary Scotland. *Scottish Geographical Journal* 131, 3-4 (2015), 148–170.
[59] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP Geolocation Databases: Unreliable? *SIGCOMM Comput. Commun. Rev.* 41, 2 (April 2011), 5356. https://doi.org/10.1145/1971162.1971171
[60] PTCL. 2021. PTCL Annual Report (page 233). https://ptcl.com.pk/uploads/Annual%20Report-2020.pdf. (2021).
[61] Ram Sundara Raman, Leonid Evdokimov, Eric Wurstrow, J Alex Halderman, and Roya Ensafi. 2020. Investigating Large Scale HTTPS Interception in Kazakhstan. In *Proceedings of the ACM Internet Measurement Conference*. 125–132.
[62] Francesco Rosati and Lourenço Galvão Diniz Faria. 2019. Business contribution to the Sustainable Development Agenda: Organizational factors related to early adoption of SDG reporting. *Corporate Social Responsibility and Environmental Management* 26, 3 (2019), 588–597.
[63] Koen Salemink, Dirk Strijker, and Gary Bosworth. 2017. Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas. *Journal of Rural Studies* 54 (2017), 360–371.
[64] SEC. 2021. LIST OF CERTAIN SUBSIDIARIES OF AMERICA MOVIL, S.A.B. DE C.V. https://www.sec.gov/Archives/edgar/data/1129137/000119312511138519/dex81.htm. (2021).
[65] Justine Sherry, Shaddi Hasan, Colin Scott, Arvind Krishnamurthy, Sylvia Ratnasamy, and Vyas Sekar. 2012. Making middleboxes someone else's problem: Network processing as a cloud service. *ACM SIGCOMM Computer Communication Review* 42, 4 (2012), 13–24.
[66] SingTel. 2021. Annual Report 2019. https://cdn.aws.singtel.com/annualreport/2019/files/Singtel-AR2019-FULL_LR.pdf. (2021).
[67] SingTel. 2021. Singtel to subscribe to Bharti Airtel's rights issue. https://cdn.aws.singtel.com/annualreport/2019/files/Singtel-AR2019-FULL_LR.pdf. (2021).
[68] Congo Telecom. 2021. About. http://www.congotelecom.cg/about.html. (2021).
[69] China Telecom. 2021. China Telecom Will Achieve 800M Complete Coverage, Launch Pre-commercial VoLTE in 2017. https://www.ctamericas.com/china-telecom-to-achieve-800m-complete-coverage-launch/. (2021).
[70] Telegeography. 2021. Comcel board approves Telmex Colombia stake purchase. https://www.commsupdate.com/articles/2010/12/30/comcel-board-approves-telmex-colombia-stake-purchase/. (2021).
[71] Telegeography. 2021. GlobalComms Database Service. https://www2.telegeography.com/globalcomms-database-service. (2021).
[72] Deutsche Telekom. 2021. Shareholder structure. https://www.telekom.com/en/investor-relations/company/shareholder-structure. (2021).
[73] Telia. 2021. Shareholders. https://www.teliacompany.com/en/about-the-company/corporate-governance/shareholders/. (2021).
[74] Telkomsel. 2021. 2017 Annual Report (page 47). https://www.telkomsel.com/download?type=annual&category=report&file=TSEL_AR2017_Web.pdf. (2021).
[75] the International Telecommunications Union. 2021. https://www.itu.int. (2021).
[76] Leanne Townsend, Arjuna Sathiaseelan, Gorry Fairhurst, and Claire Wallace. 2013. Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. *Local Economy* 28, 6 (2013), 580–595.
[77] Ucell. 2021. About company. https://ucell.uz/en/myucell/about_company. (2021).
[78] International Telecommunications Unit. 2021. Our Mandate, Mission and Strategy. https://www.itu.int/en/ITU-D/Capacity-Building/Pages/MandateStrategy.aspx. (2021).
[79] Bureau van Dijk. 2014. Orbis Overview. https://www.bvdinfo.com/en-us/our-products/data/international/orbis. (2014).
[80] Bureau van Dijk. 2021. About us. https://www.bvdinfo.com/en-us/about-us. (2021).
[81] Wahlisch, Matthias and Schmidt, Thomas and de Brun, Markus and Haberlen, Thomas. 2012. Exposing a Nation-Centric View on the German Internet - A Change in Perspective on AS-level. In *Passive and Active Measurement Conference (PAM)*.
[82] Paul A Watters, Aaron Herps, Robert Layton, and Stephen McCombie. 2013. ICANN or ICANT: Is WHOIS an Enabler of Cybercrime?. In *2013 Fourth Cybercrime and Trustworthy Computing Workshop*. IEEE, 44–49.

A COUNTRIES WITH STATE-OWNED INTERNET PROVIDERS AROUND THE WORLD

Figure 6 shows a world's heatmap of countries having majority state-owned Internet providers (blue) and the minority state-owned Internet providers (orange) we found during our data discovery process.

B INDIVIDUAL CONTRIBUTION OF EACH DATA SOURCE TO THE FINAL LIST OF STATE-OWNED ASES

Table 6 shows the individual contribution of each data source (fraction of subsidiaries in parentheses) at the end of the process. In addition, the third row indicates the number of minority state-owned that we also found with each data source.

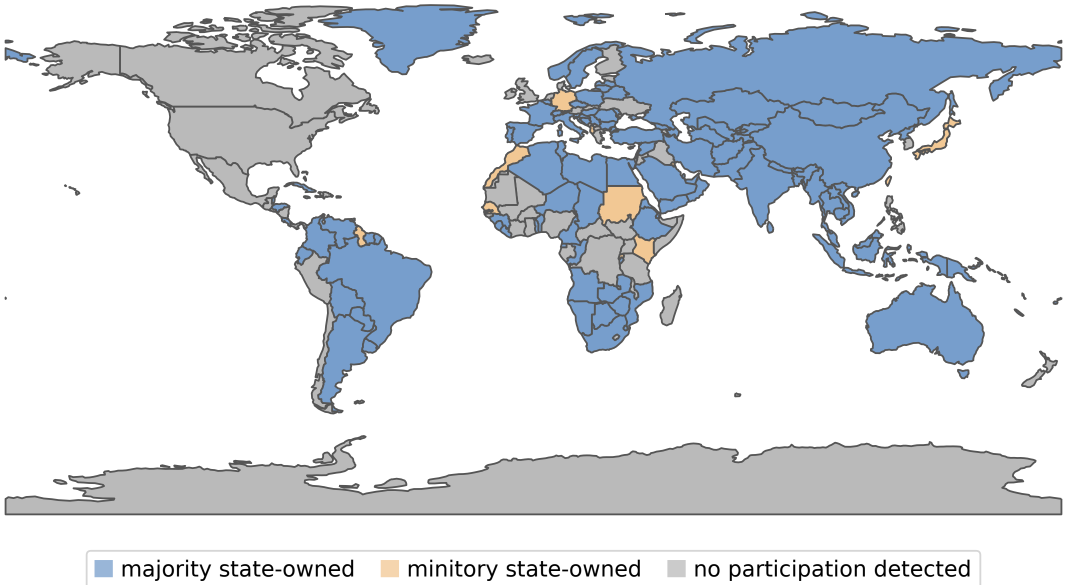


Figure 6: World map of countries having majority state-owned Internet providers (blue). Minority state-owned Internet providers found during the data discovery process are indicated in orange.

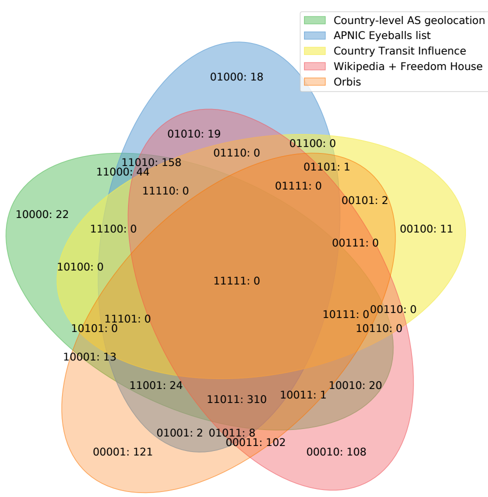


Figure 7: Venn Diagram of the contribution of each data source to the list of state-owned ASes

C CONTRIBUTION AND OVERLAP OF DATA SOURCES TO THE FINAL LIST OF STATE-OWNED ASES

Figure 7 shows a Venn diagram with the contribution of each data source to the final list of state-owned ASes.

Data source	State-owned ASes	Minority state-owned
Geolocated addresses	593 (126)	253
APNIC's Eyeballs list	586 (151)	288
CTI	15 (0)	7
Wikipedia+FH	728 (126)	4
Orbis	587 (123)	0
TOTAL	984 (193)	302

Table 6: Individual contribution of each data source (fraction of subsidiaries in parentheses). Note that the total is not the sum of the rows since the individual contributions partially overlap.

D STATE-OWNED ASES ONLY DISCOVERED BY CTI

Table 7 shows the list of state-owned ASes that are only captured by CTI.

E EXCLUDED STATE-FUNDED ORGANIZATIONS

Academic networks: We remove university networks and academic backbones from the candidate list. We exclude government-funded university networks (*e.g.,* Universidad de Buenos Aires-AS3449) and academic backbones (*e.g.,* Germany's DFN-AS680) from the list because we assume that governments do not uses such networks to participate and promote Internet markets. In addition, despite campus residents access to the Internet by using university networks and universities rely on academic backbones as upstream

Country name (cc)	ASN	AS name
Vietnam (VN)	45895	MOBIFONEGLOBAL-AS-VN
Vietnam (VN)	45896	MOBIFONEGLOBAL-AS-VN
Vietnam (VN)	45897	MOBIFONEGLOBAL-AS-VN
Bangladesh (BD)	132602	BSCCL
Cuba (CU)	11960	ETECSA
Belarus (BY)	60330	BCTBY-AS
Belarus (BY)	205475	BECLOUD-RDC-AS
Belarus (BY)	35647	BYIX-AS
Belarus (BY)	60280	NTEC

Table 7: List of state-owned ASes only covered CTI.

providers, none of these networks compite in general and open access and transit markets.

Governments's bureaucratic networks: We exclude all ASes providing connectivity to government offices, secretaries or any other government-dependent institutions. Again, we remove these government-funded networks since they just serve to these offices and not to the resident, for instance the State of California-AS2642 or the European Central Bank-AS31614. Another prominent example is the US DoD-AS721 announcing a customer cone of nearly 80M IP address in March 2021, however, this network only connects institutions related to the DoD.

Government's Internet administrative organizations: We exclude from the list governments organizations that support and enable the functioning of the country's Internet but do not provide Internet services such as transit or access. Some country-level organizations, for examples the Bolivian Agency for the Development of the Information Society, ADSIB-AS52250, are relevant to the country's Internet functioning since these institutions delegated Internet resources such as domain names, and sometime, prefixes and ASNs too. In other countries, this role is reserved to the NIC, which tends to have a more complex legal structures (non-profit, under university managements or multistakeholder boards) that could not be classified as state-owned. In either case, these organizations sometimes host the ccTLD serves, which is a key component in a country's Internet infrastructure. However, we exclude these organizations because they do not offer broadband subscriptions or transit services to the general public. Although these institutions are beyond the scope of this paper, we would like to acknowledge that these institutions play an important role in closing the digital divide by providing resources, knowledge and services to operators and companies in their countries.

Unrelated to Internet services: We filter out hardware manufacturers or telecommunication companies that do not provide Internet services.

F STATE-OWNED INTERNET PROVIDERS WITH MORE THAN 0.9 OF THE ESTIMATED ACCESS MARKET

Table 8 shows the list of countries (19 in total) in which our estimated Internet access market footprint of state-owned ASes is over 0.9.

Country (cc)	Approx. Internet access market footprint
Ethiopia (ET)	1.00
Tuvalu (TV)	1.00
Cuba (CU)	1.00
Greenland (GL)	1.00
Djibouti (DJ)	1.00
Syria (SY)	1.00
United Arab Emirates (AE)	0.99
Eritrea (ER)	0.99
Suriname (SR)	0.97
China (CN)	0.97
Libya (LY)	0.97
Yemen (YE)	0.97
Algeria (DZ)	0.96
Macao (MO)	0.96
Andorra (AD)	0.94
Iran (IR)	0.92
Uruguay (UY)	0.92
Turkmenistan (TM)	0.91

Table 8: Countries with over 0.9 Approx. Internet access market footprint in their home countries.

G COUNTRY-LEVEL TRANSIT INFLUENCE

The transit influence $CTI_M(AS, C) \in [0,1]$ is calculated using a set of monitors[8] M as

$$\sum_{m \in M} \left(\frac{w(m)}{|M|} \cdot \sum_{p|\text{onpath}(AS,m,p)} \left(\frac{a(p,C)}{A(C)} \cdot \frac{1}{d(AS,m,p)} \right) \right), \quad (1)$$

where $w(m)$ is monitor m's weight, calculated as the inverse of the number of monitors available from its same AS; onpath(AS, m, p) is true if AS is present on a preferred path observed by monitor m to a prefix p originated by a probed and responsive origin network, and m is not contained within AS itself; $a(p, C)$ is the number of addresses in prefix p geolocated to country C that are not covered by a more specific prefix; $A(C)$ is the total number of IP addresses geolocated to country C; and $d(AS, p, m)$ is the number of AS-level hops between AS and prefix p as viewed by monitor m. Please refer to Gamero-Garrido's doctoral dissertation [37] for a more detailed definition of the CTI metric.

[8] A BGP monitor is an operational border router that forwards announcements to a collection database. These routers are hosted by cooperative ASes who voluntarily disclose their routing information.