Eleven Paths

TREND REPORT

What does the metadata reveal about the states in Latin America?

2018.05.17

Telefónica CYBER SECURITY UNIT

elevenpaths.com

Index

2017 © Telefónica Digital España, S.L.U. All rights reserved.

Introduction

We have conducted a study with the objective of determining the level or maturity of the security controls which the States within Latin America implement, in respect to the prevention of metadata information leaks. We have utilized public documents detected in government domains, of which can be accessed by any network user, through search engines or directly on the different entities' websites.

The study focused upon obtaining the metadata from public documents, in other words, no attempt was made to access private nor confidential documents at any time. From there, an analysis was carried out with the information that filtered from it, and that could potentially be useful for an attacker, particularly::

- Versions of the **operating systems** which could be detected; knowing that they delivered infrastructure characteristics as well as software use from inside the entities, even if these are updated or supported by the manufacturers.
- **User names** of the governmental systems in which they generated the documents, so that the analysis of these users allows the identification generic user control policies.
- **Email Addresses** that could allow them to know the origin of the files and/or determine if these emails have been exposed in some of the information leaks that they have happened to different internet companies.
- **Exposed services** so that the users can access the information, allowing them to determine the security level within the information transfer, since, the use of the transfer without encryption runs a risk of exposure leakage.
- **Location of the servers** which would allow us to determine if the policies and rules that should apply, are from the generating country or if by being in other physical locations, the States expose citizens' information.

It is worth mentioning that **in order to carry out this investigation they did not access nor take advantage of any service or server vulnerability to perform the analysis from the 20 defined governments**, but, on the contrary, they took the publicly exhibited documents from each one of the governments.

In this report the information is compiled from documents detected through the free ElevenPaths tool called FOCA OpenSource. Once the documents are downloaded, it is allowed to extract the metadata which they contain and carry out an analysis, focused on fulfilling the objective.

1. Methodology

The first stage consisted of identifying the governmental domains of the twenty Latin American States utilized for this study; whose governments display service or websites on the internet, giving its inhabitants the possibility of having access to the documentation in a public way.

The second stage is based upon the utilization of FOCA OpenSource, a free tool developed by ElevenPaths, which by using OSINT techniques then allows us to find in simple and quick way, the digital documents which are publicly exposed by the governmental domains in the study. Then once they are detected with the same tool, these documents are downloaded and they carry out an extraction of the metadata which each one of these files contains.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

By extracting the information from the metadata it is possible to identify the user names, IP addresses, email addresses, network names, storage directories and operating systems, amongst other things. From this information, it is possible to infer characteristics or unsafe processes, or those which are not considered good practices from the perspective of information security. Such as the use of generic users, obsolete operating systems, emails exposed to credential leaks, emails from non-official domains in official documents, etc.

In the third stage, by already using the IP address data from the http and https services, they utilized the whatweb tool in order to determine which of these addresses still had the services active and could control the country where these services are hosted. For the IP addresses which do not respond to the service, they utilized ip2locaction with the aim of identifying their geographical reference.

As a fourth stage, they carried out an analysis of the public and governmental domains with the emails detected during the metadata analysis; in order to determine the quantity of emails from each one of these. Additionally, they utilized the OSRFramework tool in order to determine how many of these email addresses had been exposed in the different online information leaks and which are of public knowledge.

They carried out all of the analyses using free tools and public documentation, which guarantees that it can be repeated by any entity or person.

2. Analyzed Domains

As a base for this report, we have taken the report carried out by Inter-American Development Bank on "The State of Cybersecurity in Latin America and the Caribbean" (www.observatoriociberseguridad.com), which analyzed the importance of the cybersecurity for the states, and how each one of these is deploying the controls and the policies within their respective countries. We use the list of the twenty countries, which according to the study, have made progress in one of the five rating groups, and which in turn display information for their citizens on websites. In table 1 you will find the countries which carried out the analysis, with the governmental domains which they used to detect the public documents, identifying the detected quantity, but also clarifying which of those could be finally analyzed, given that some of them did not have information, or were even downloaded corrupted.

COUNTRY	DOMAIN	DETECTED	ANALIZED
Argentina	**gob.ar**	2039	1958
Bahamas	**gov.bs**	497	496
Bolivia	**gob.bo**	1787	1597
Brasil	**gov.br**	1814	1736
Chile	**gob.cl**	1874	1832
Colombia	**gov.co**	1458	1352
Costa Rica	**go.cr**	404	396
Ecuador	**gob.ec**	1708	1415
El Salvador	**gov.sv**	1672	1517
Guatemala	**gob.gt**	1613	1564
Honduras	**gob.hn**	1454	1341
Jamaica	**gov.jm**	1125	1087

2017 © Telefónica Digital España, S.L.U. All rights reserved.

Mexico	**gob.mx**	2322	2234
Nicaragua	**gob.ni**	1522	1462
Panama	**gob.pa**	1787	1581
Paraguay	**gov.py**	1730	1665
Peru	**gob.pe**	2225	2085
Dominican Republic	**gov.do**	1802	1654
Uruguay	**gub.uy**	2401	2108
Venezuela	**gob.ve**	2081	1810

3. Initial Analysis

By determining the main governmental domain within each one of the countries, we have configured FOCA OpenSource so that it can carry out an automatic search of the main **MS Office, OpenOffice, PDF, Adobe, images,** amongst other extensions; utilizing the Google and Bing API keys, thus allowing them to find the links to the documents and download them, in order to carry out the metadata analysis. In Illustration 1 you will find the quantity of detected documents and how many were analyzed, and from there you can appreciate that it has been possible to validate the metadata from more than 90% of the detected public documents in each domain.

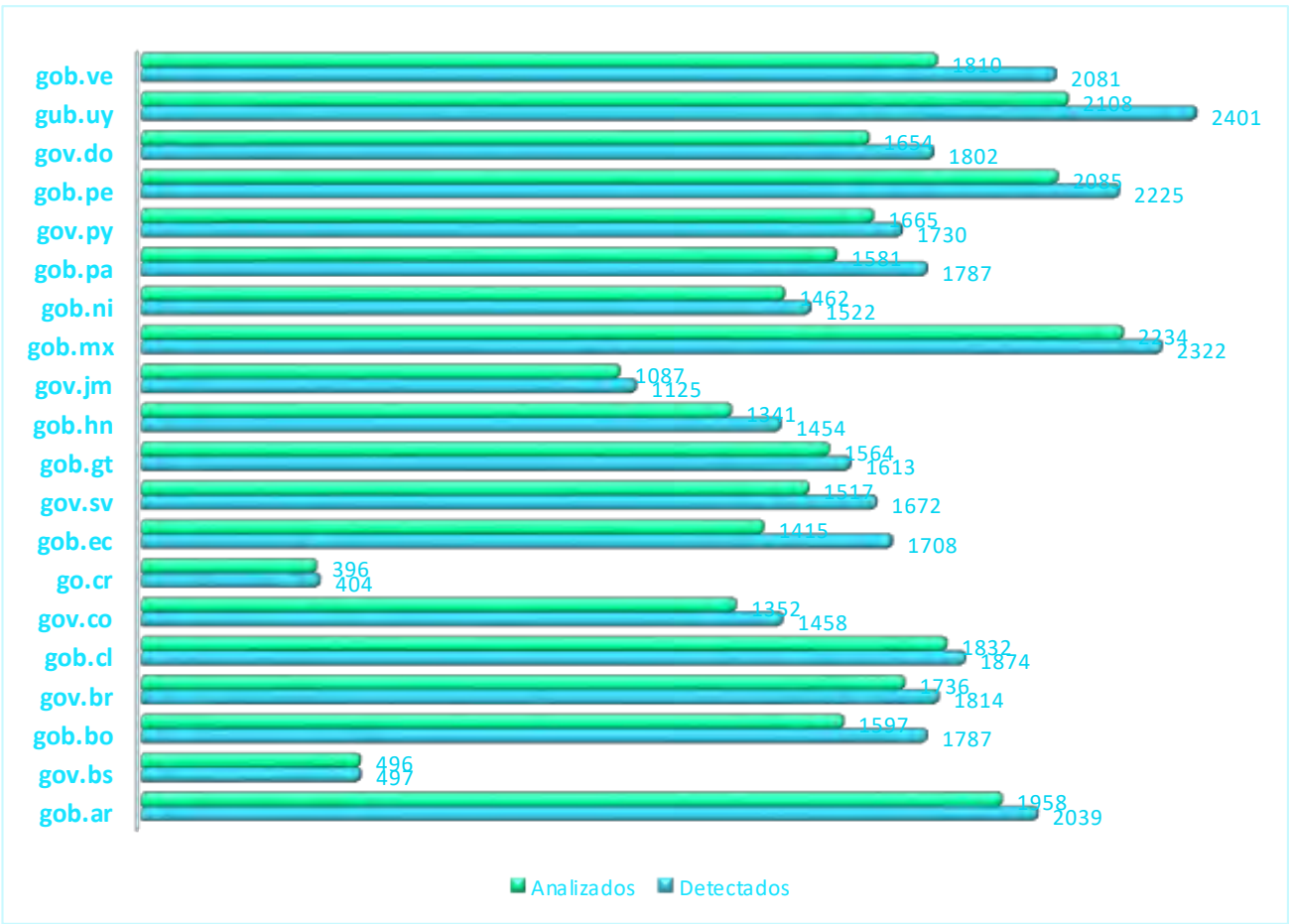
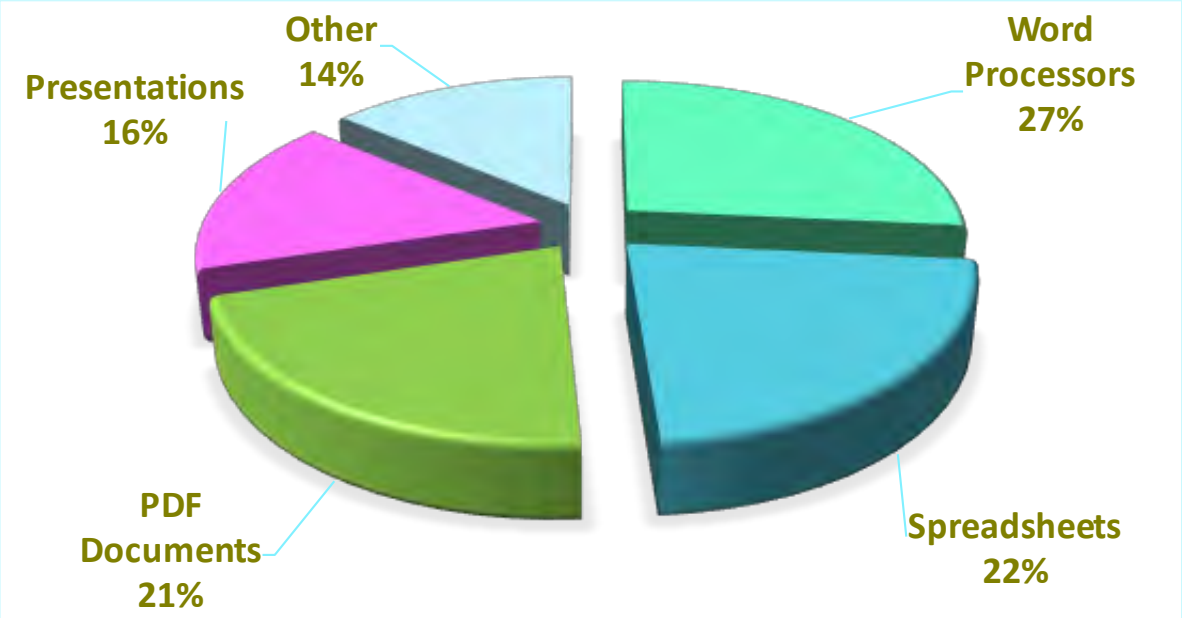


Illustration 1. Quantity of files detected and analyzed

2017 © Telefónica Digital España, S.L.U. All rights reserved.

Taking into account that more than 80% of the analyzed documents are from software related to computing, we found it interesting to be able to segment the types found within this category, and this was how we could identify that 27% belonged to the word processors, 22% to spreadsheets and 17% to presentations, as can be seen in figure 2.



In figure 3, it is possible to see the quantity of information extracted from the parameters which they considered as important in the metadata of the analyzed documents.

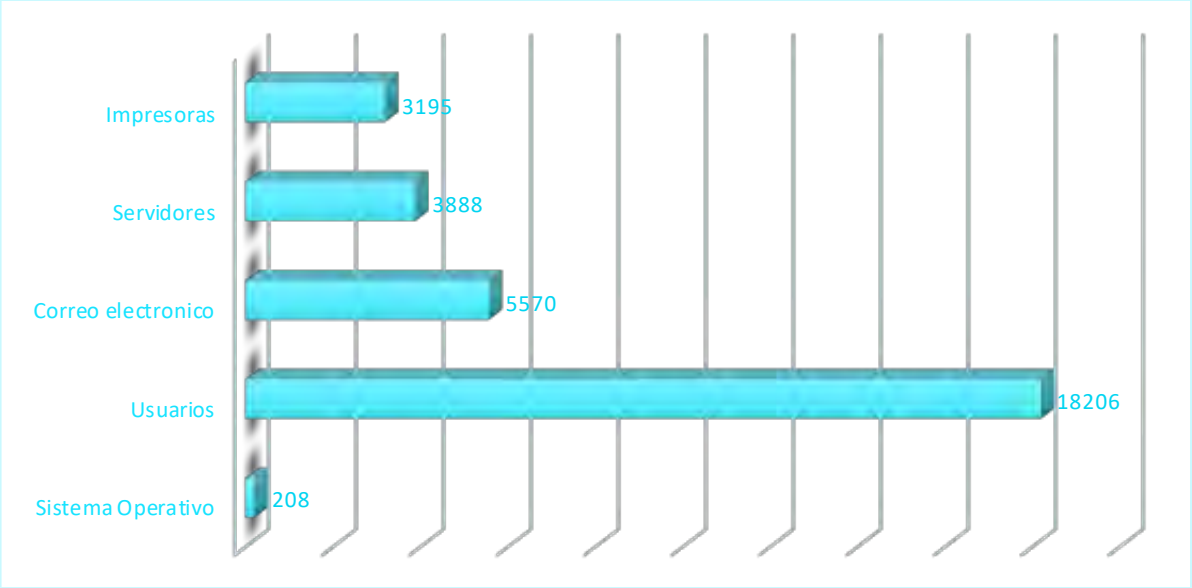


Figure 3. The quantity of detected information in the metadata.

4. Systems Analysis

As we have mentioned, the metadata delivers very valuable characteristics for the configuration of the systems in which the files were created; data such as the operating systems, printer used and the user of the machine. This information is used in computer forensics investigations to identify the device where the document was made or to determine a

2017 © Telefónica Digital España, S.L.U. All rights reserved.

timeline. However, it also allows a computing criminal to outline the characteristics of the computer or company that wants to attack; reducing the options and improving the effectiveness of the attack, so this information leakage is very sensitive for an organization, and in this case for the State; since it reveals many characteristics of its infrastructure.

4.1.Operating Systems

By extracting the metadata information which they detected in the different computers' operating systems where they generated the documents; it was shown about 10 different operating systems deployed on average from 545 computers per country. In figure 4, the number of operating systems detected in each of the domains analyzed can be seen.

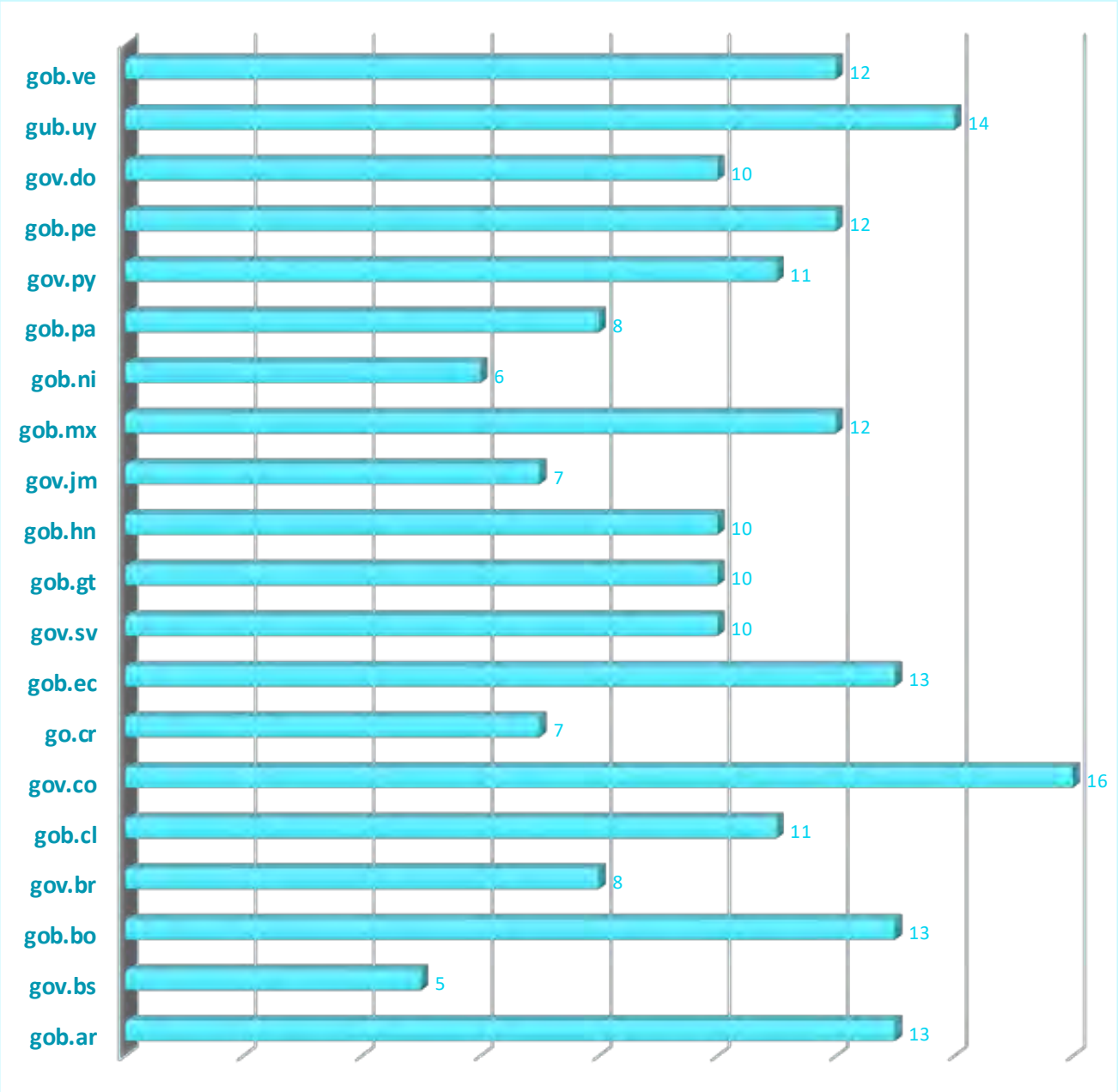
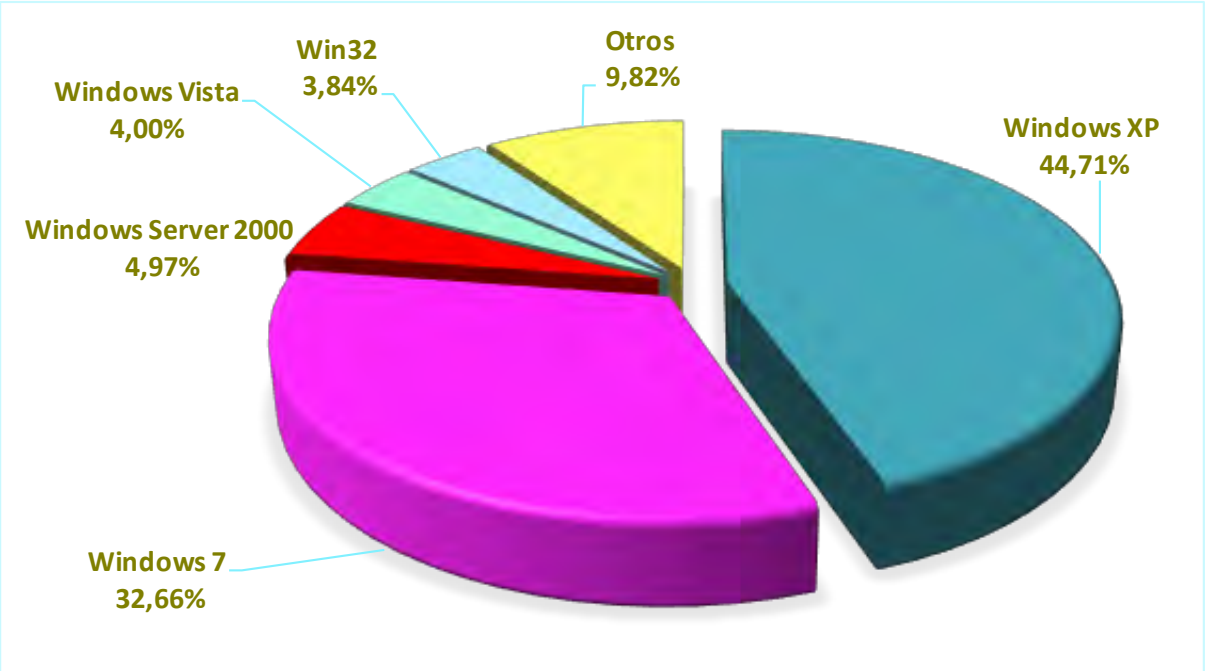


Figure. 4. Operating systems detected in each of the domains

2017 © Telefónica Digital España, S.L.U. All rights reserved.

By taking the 10,911 detected computers, it is shown that since a few years ago the most commonly used operating systems are no longer being supported by the manufacturers, which thus generates a critical risk to the information they possess; **assuming that the equipment with which those documents were made is still available**. In figure 5, the percentage of use of operating systems can be seen.



4.2.Users

The identification of users through metadata provides information that allows an attacker to get a list of valid users within the infrastructure of the organization. In the analysis process that was developed in government domains in Latin America, they found on average 911 users per domain. With a total of 18,206 identified users in the metadata from the analyzed documents, which were distributed by domain as shown in figure 6.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

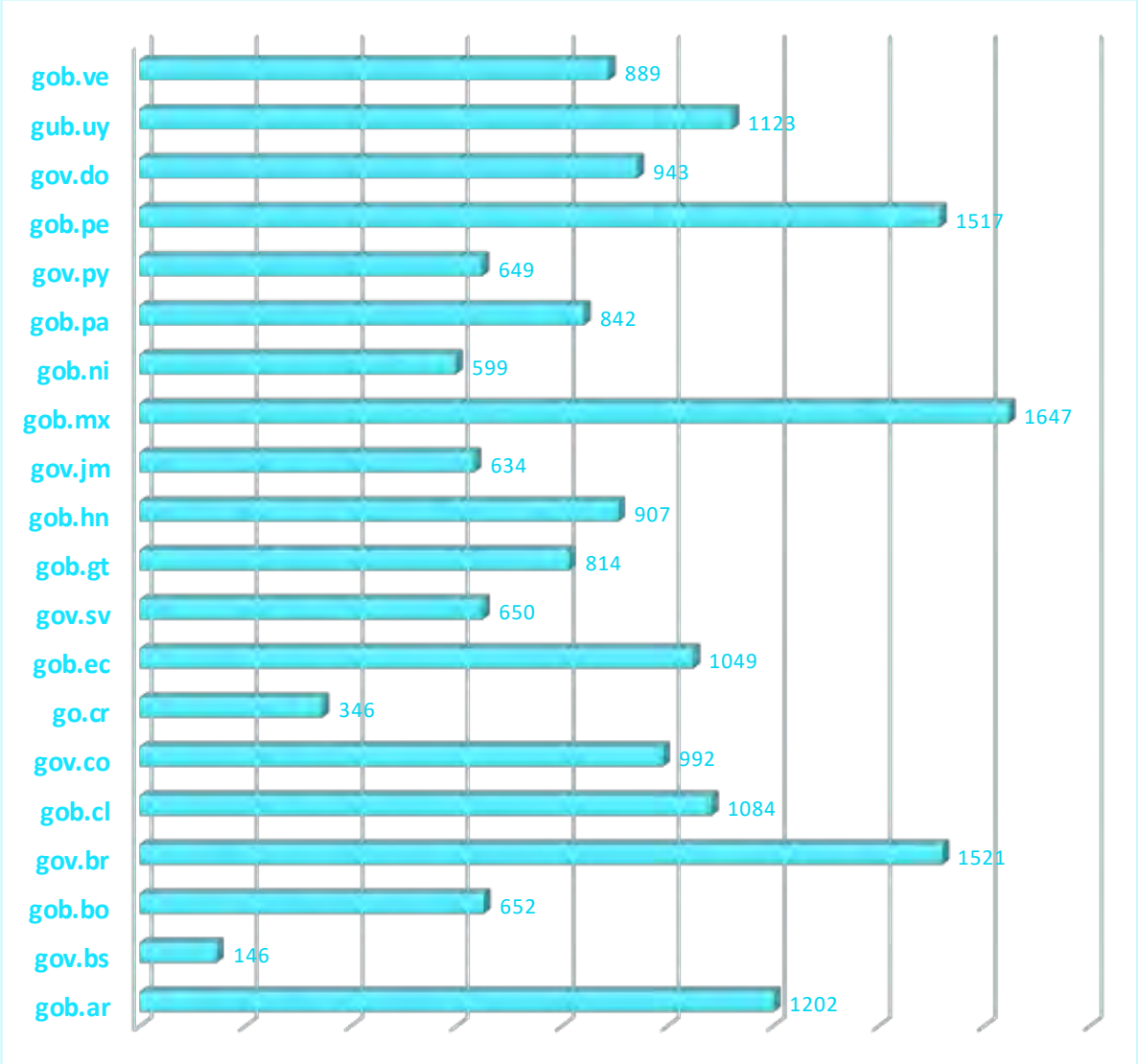


Figure 1. Number of users detected in each government domain.

Good practices suggest that generic user use of the organizations' computers should not be allowed, as this makes it impossible to trace the events within different systems to a specific user, increasing the risk of potential information confidentiality loss by a disloyal employee.

In the process of the cyber-irrigation investigation, a list of users that are typically used in business infrastructures has always been established. By carrying out a search for these users, they detected 2,194 computers where they identified the user. In turn, we can identify that 12% of the detected users are generic, as shown in Figure 7.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

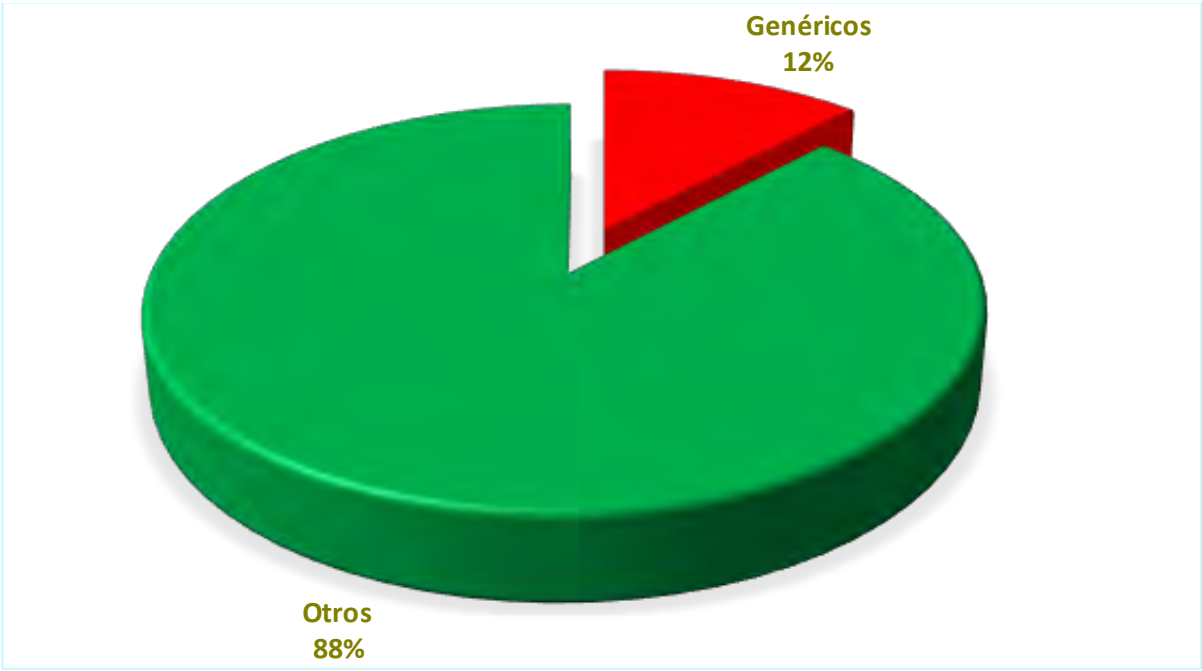


Figure 2. Percentage of generic users detected

We can also observe that one of the most commonly used is the "administrator" of the system, which increases the level of threat criticality generated by this type of user use. Figure 8 shows the most commonly used generic users.

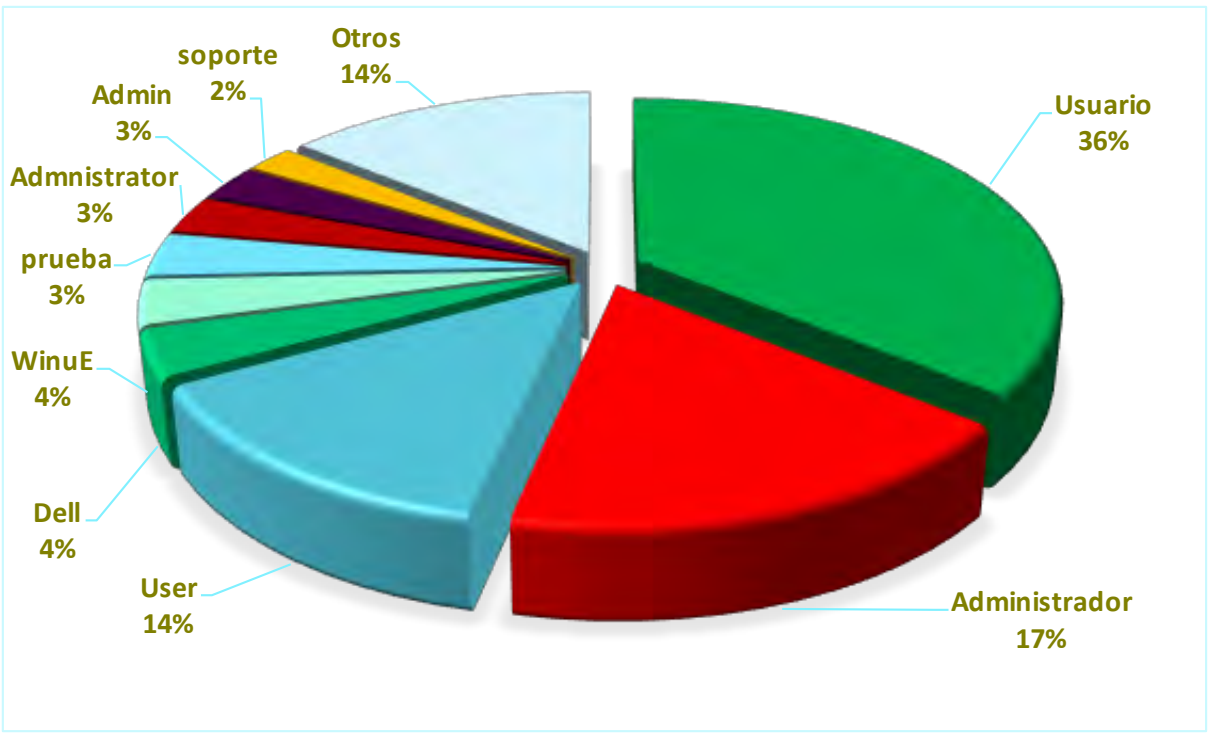


Figure 3. Top generic users detected

2017 © Telefónica Digital España, S.L.U. All rights reserved.

5. Services Analysis

The documents analyzed are obtained from the government entities' web services, which through the **http** and **https** protocols allow citizens to access the State's digital services. These services must comply with all security and privacy regulations in the country. As part of the study, an analysis process was carried out on the services detected in order to determine how many perform an information traffic assurance using encryption at the connection; the physical location of the server based on the IP address was analyzed, and finally, an attempt was made to identify, by means of non-invasive fingerprinting techniques, the type of software with which these services are provided.

5.1. Detected Servers

The documents discovered are available to the States in Latin America, mostly through the use of the http protocol, which is reflected in figure 9.

Figure 9. Percentage of detected services

On average, each state has 92 http and 59 https servers. In figure 10, the number of services detected in each of the domains in both protocols can be seen.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

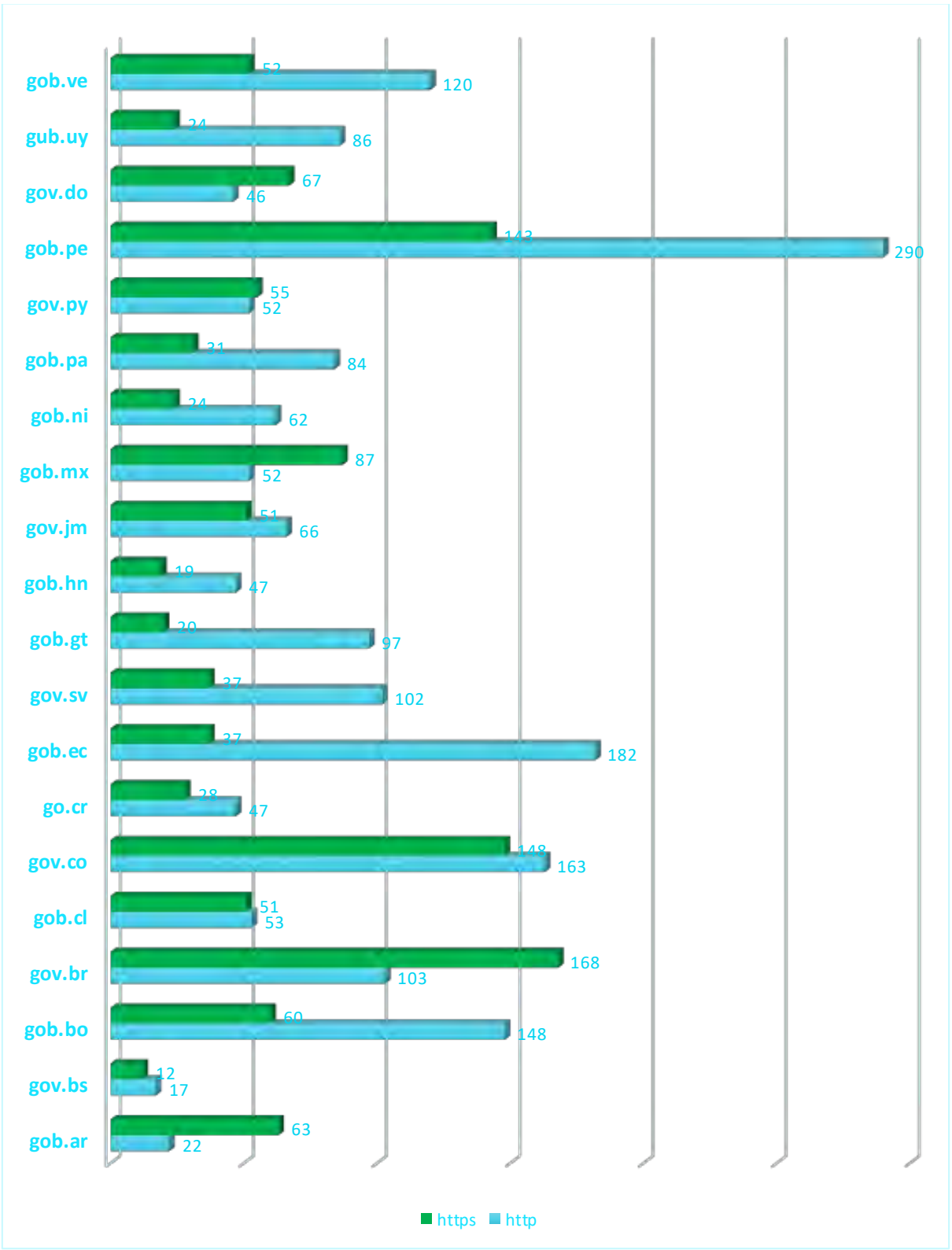


Ilustración 4. Cantidad de servicios de http y https detectados

2017 © Telefónica Digital España, S.L.U. All rights reserved.

5.2.Location of Services

For a correct management of information security, it is necessary to know the security policies information and regulations that must be complied with in the countries; so for the States it is vital to be able to have control over their information policies. Therefore, the location of the servers is of utmost importance, since it is where the data is physically located, and the laws of each country deals with or demands different things in relation to this point. This is why the analysis of the location of the different servers detected was included. In figure 11, you can see the location of the http servers and in figure 12 you can see the location of the https servers.

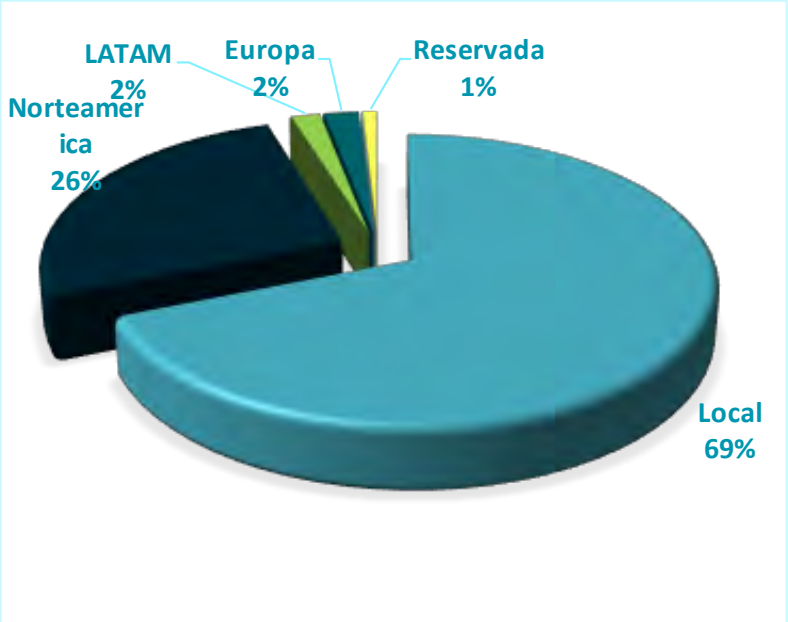


Figure 5. Locations of HTTP servers

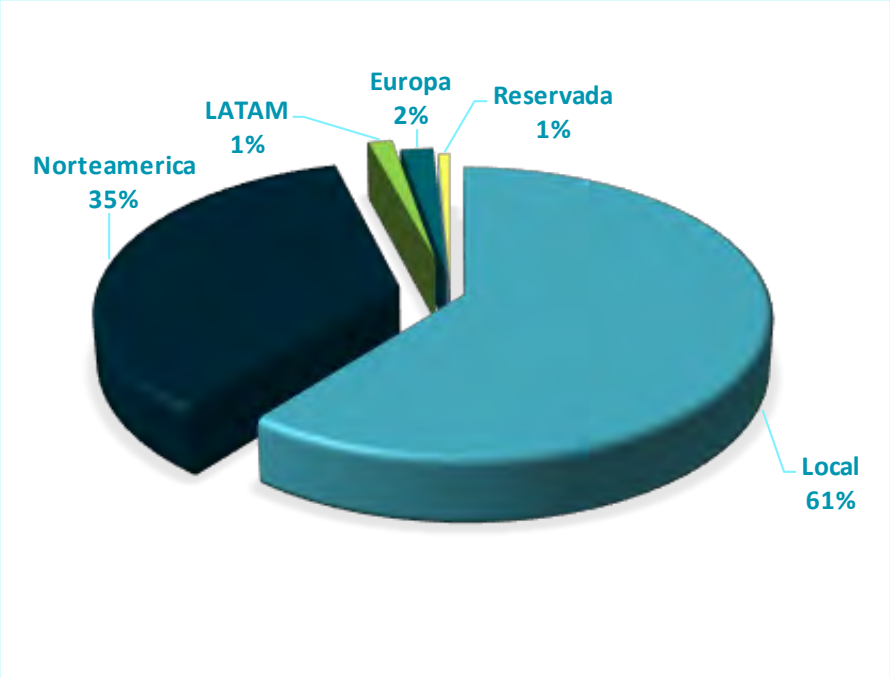


Figure 6. Locations of HTTPS servers.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

The locations are grouped by region in order to facilitate the analysis and to determine the areas where the greatest concentration of information and equipment is located, and if they are in each of the countries, this is calculated within the group that is called local.

6. Analysis of Email Addresses

In the office documents' metadata, it is normal to find associated email addresses, or data from the computers where these files were created. Consequently, in the study, one of these parameters is what we focus upon; which is the quantity of detected email addresses and the domains which they belong to. Well in some cases they allow them not only to reveal the original editor of the file, but also their possible receptors.

Additionally, the email addresses are usually utilized for the creation of different profiles within Internet services. Thus, it is possible to be validated if the detected email addresses in the files have been exposed within an information leak from these services which have then become public knowledge.

6.1. Quantity of Detected Emails

By extracting the documents' metadata, it is found that within all of the state-run domains there are disclosed e-mails, with an average of almost 279 emails for each domain and they have found a total of 5570 different email accounts, which are distributed as you can see in figure 13.

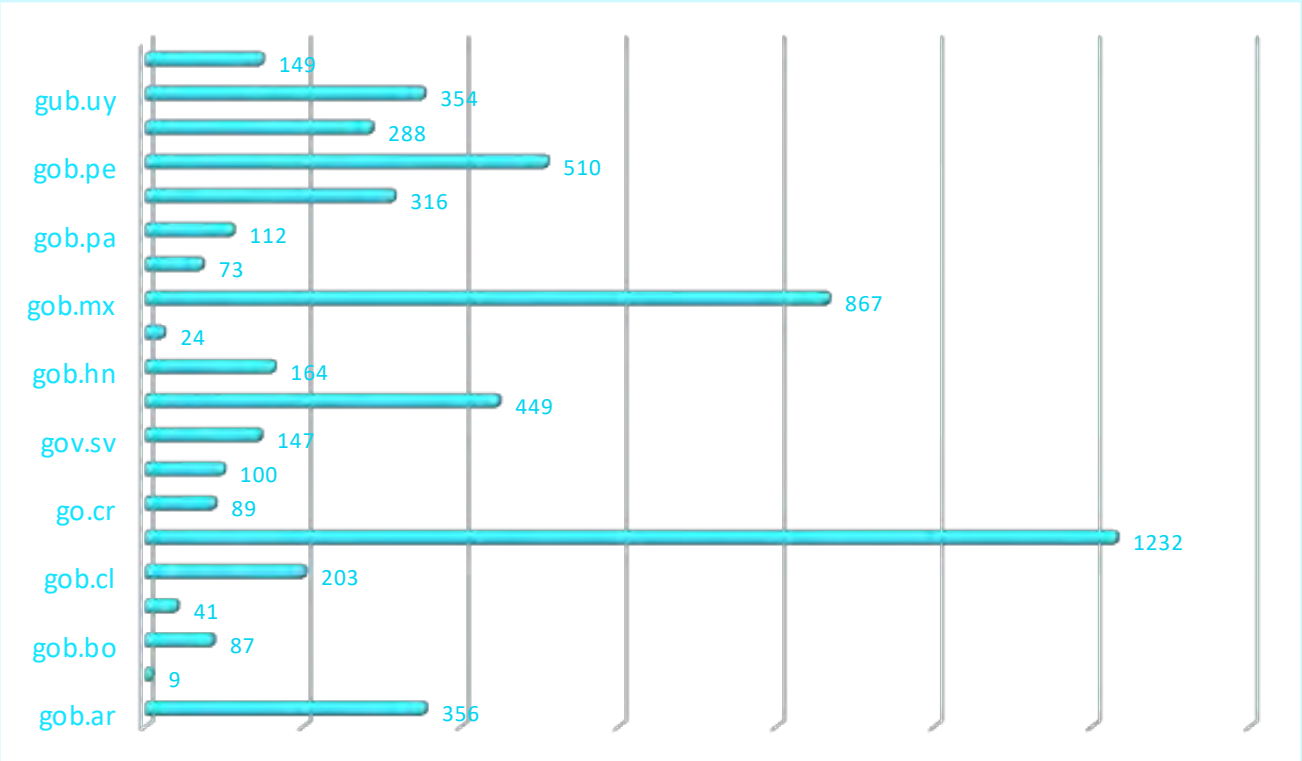


Ilustración 7. Cantidad de correos electrónicos detectados por dominio.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

With these emails they are distributed in 1462 different domains, showing that the exposed documents in the Latin American States' services come from different sources. When analyzing the number of emails grouped by domains, it is found that 36% belong to public email services instead of being directly related to the entities. Figure 14 shows the top of the detected domains.

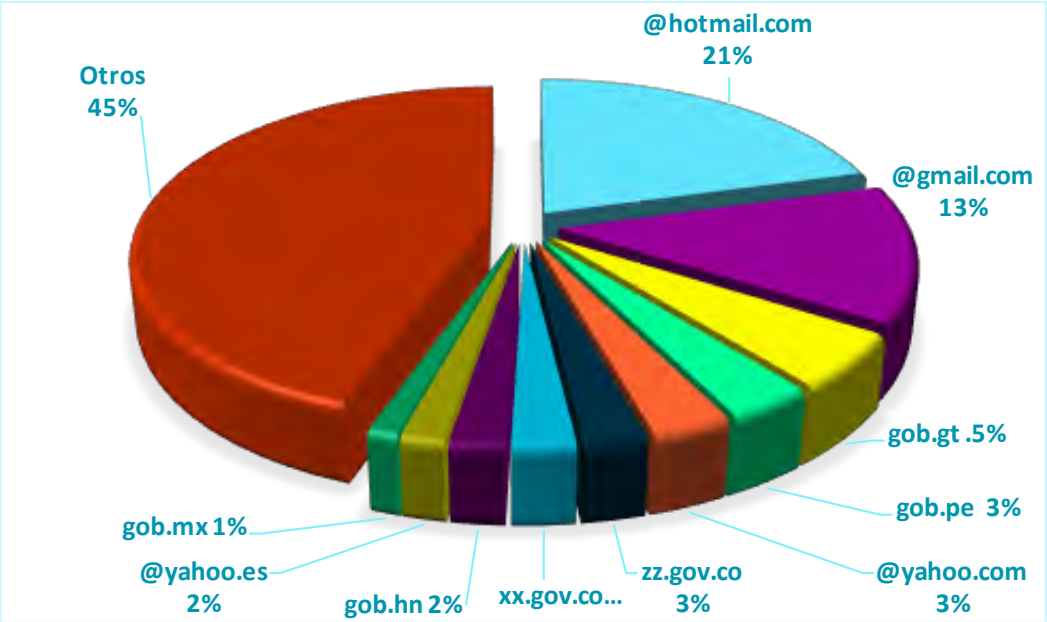


Figure 8. Top of the detected email domains.

In the analysis they extract the domains which they possess some relationship to the government, as they are educational (edu.) and from the armed forces (.mil), additionally of those which we are analyzing which are strictly governmental (.gov, .gob, .gub, .go), finding a total of 2013 email addresses. Where the most recurrent domains can be seen in figure 15.

Figure 9. Top of the detected state email domains

2017 © Telefónica Digital España, S.L.U. All rights reserved.

6.2.Emails Exposed in Information Leaks

Every day, databases are exposed on the Internet where credentials of many domains are available to anyone who wants to download them, and expose the identity of their owners and the security of companies or organizations, since many users use the same password for different services. An example of these leaks was the access to credentials of more than 60 million Dropbox users who were then openly leaked.

That being said, within the analysis, the validation of the 5,570 e-mail accounts within the exposed databases was performed using a mailfy tool, which is found in the OSRFramework framework. It detected that 23% of the accounts found in the files' metadata have been found at least once before, through being exposed in any of these leaks. In Figure 16 we can see how many accounts have been exposed by each of the domains analyzed.

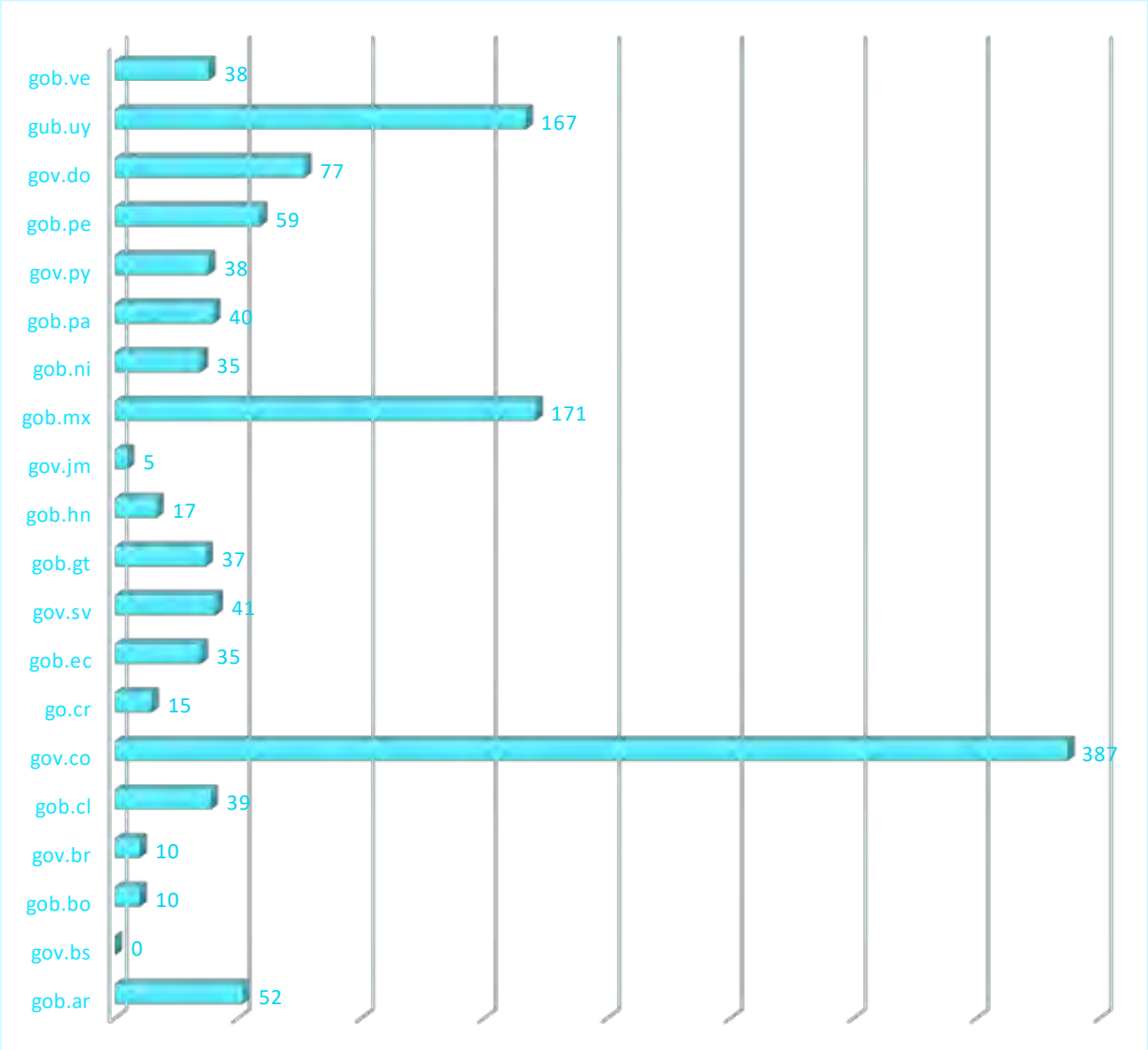


Figure 16. Number of email accounts exposed per domain.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

The framework allows us to show in which information leakage each account was exposed, we found out that within **114 million accounts exposed in the leak, LinkedIn has the largest number of accounts associated with the metadata of the analyzed files.**

Figure 16 shows the percentage of accounts found in the top 10 of data leaks.

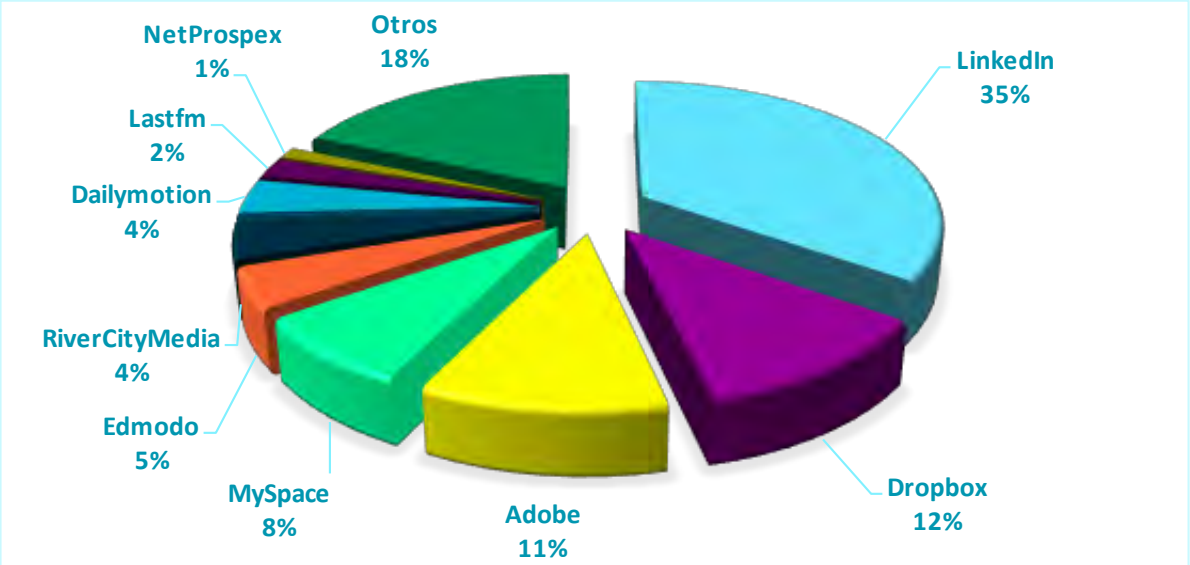


Figure 17. Percentage of accounts exposed to the information leaks.

Among the emails detected in the information leaks there are **government domains**. So they took 2,013 emails detected from the governments and analyzed them with **mailfy**, finding that **102 of these accounts have been exposed in some leakage. Figure 18 shows the information leaks** where the government accounts were detected.

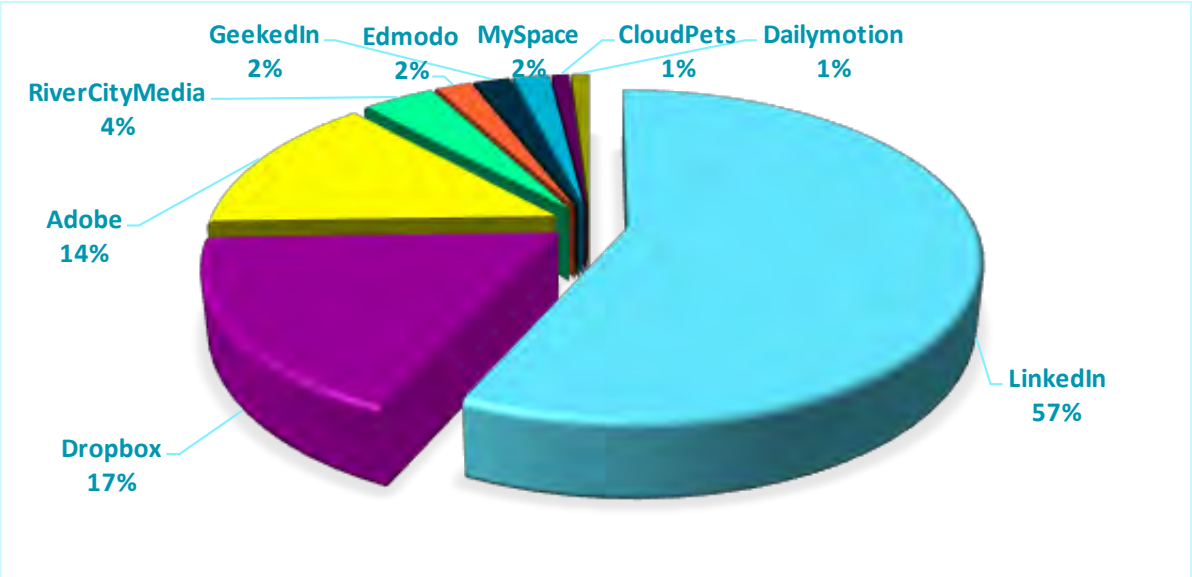


Figure 18. Percentage of government accounts detected in the public information leaks.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

7. Conclusion

From the analysis carried out on the metadata of the documents presented by the Latin American governments, we can interpret that, at the time of carrying out the analysis, it is still possible to improve the actions related to the containment of the leakage of this type of information, either through the implementation of manual or automated procedures and controls, and thus avoid exposing data on the technological infrastructure, on the users of the systems and on the e-mail accounts of the file managers.

From the results obtained, it could be inferred that the technological infrastructure has not been updated because most of it is based on operating systems that are no longer supported by their manufacturers, however, we cannot say for sure since we have not done this type of analysis, and the files may have been published for some time whilst the technology could have been updated.

In the analysis of the detected users, 12% are generic system users and the vast majority have the characteristics of administrators, which implies a possible lack of controls and policies in the user management of the States.

By validating the location of the servers, it is detected that 30% are not hosted in the same country as the producer of the information.

The analysis of the metadata based on the e-mails detected shows that the States have outsourced the preparation of their documentation.

Information leaks from large social networking or service companies, such as Linkedin, Dropbox, and Adobe, are the ones that contain most of the government email accounts, which was revealed in the metadata analyzed.

It should be noted that for the preparation of this report, an exhaustive analysis of aspects other than those mentioned in the introduction has not been carried out: the operating systems, IP address locations, users of operating systems and e-mail accounts associated with the metadata of the public documents that were downloaded. In spite of this, and without the need to analyze other characteristics in depth, possible relevant security problems in Latin American governments have been discovered, and we hope that this study will be useful for each government, and will help to support the security analyses that each of them will most likely carry out periodically.

2017 © Telefónica Digital España, S.L.U. All rights reserved.

About ElevenPaths

At ElevenPaths, Telefónica Cyber Security Unit, we believe in the idea of challenging the current state of security, an attribute that must always be present in technology. We're always redefining the relationship between security and people, with the aim of creating innovative security products which can transform the concept of security, thus keeping us one step ahead of attackers, who are increasingly present in our digital life.

More information

www.elevenpaths.com
@ElevenPaths
blog.elevenpaths.com

2017 © Telefónica Digital España, S.L.U. All rights reserved.

The information disclosed in this document is the property of Telefónica Digital España, S.L.U. ("TDE") and/or any other entity within Telefónica Group and/or its licensors. TDE and/or any Telefonica Group entity or TDE'S licensors reserve all patent, copyright and other proprietary rights to this document, including all design, manufacturing, reproduction, use and sales rights thereto, except to the extent said rights are expressly granted to others. The information in this document is subject to change at any time, without notice.

Neither the whole nor any part of the information contained herein may be copied, distributed, adapted or reproduced in any material form except with the prior written consent of TDE.

This document is intended only to assist the reader in the use of the product or service described in the document. In consideration of receipt of this document, the recipient agrees to use such information for its own use and not for other use.

TDE shall not be liable for any loss or damage arising out from the use of the any information in this document or any error or omission in such information or any incorrect use of the product or service. The use of the product or service described in this document are regulated in accordance with the terms and conditions accepted by the reader.

TDE and its trademarks (or any other trademarks owned by Telefonica Group) are registered service marks.

2017 © Telefónica Digital España, S.L.U. All rights reserved.