Eleven Paths

INFORME DE TENDENCIAS

¿Qué revelan los metadatos de los estados de Latinoamérica?

17.05.2018

Telefónica CYBER SECURITY UNIT

elevenpaths.com

Índice

Introducción

Con el objetivo de determinar el nivel o la madurez de los controles de seguridad que los Estados en Latinoamérica implementan respecto a la prevención de la fuga de información a partir de los metadatos, hemos realizado un estudio sobre los documentos públicos detectados en los dominios de los gobiernos, los cuales pueden ser accedidos por cualquier usuario en la red, a través de los buscadores o directamente en los sitios web de las diferentes entidades.

El estudio se centró en obtener los metadatos de los documentos públicos, es decir que en ningún momento se intentó acceder a documentos privados ni confidenciales, y a partir de ellos, se realizó un análisis sobre la información que se filtra a partir de ellos, y que potencialmente podría ser útil para un atacante, particularmente:

- Las versiones de los **sistemas operativos** que pudieran ser detectados, sabiendo que entregan características tanto de la infraestructura como del uso de software dentro de las entidades, incluso conociendo si éstos se encuentran actualizados o soportados por los fabricantes.
- **Nombres de usuarios** de los sistemas gubernamentales con el que se generaron los documentos, por lo que el análisis de estos usuarios permite identificar las políticas de control de usuarios genéricos.
- **Direcciones de correo electrónicos** que permitieran conocer el origen de los archivos y/o determinar si estos correos han sido expuestos en algunas de las fugas de información que se han presentado a diferentes empresas de Internet.
- **Servicios expuestos** para que los usuarios puedan acceder a la información, permitiendo determinar el nivel de seguridad en la transferencia de información, pues el uso de la transferencia sin cifrado genera un riesgo de fuga de exposición.
- **Ubicación de los servidores** que permitiera determinar si las políticas y normas que se deben cumplir son las del país generador de la información o, si por estar en otras ubicaciones físicas, los Estados exponen información de los ciudadanos a las normas y controles de privacidad donde se encuentre alojada dicha información.

Cabe mencionar que **para la realización de esta investigación no se accedió ni se aprovechó de ninguna vulnerabilidad a ningún servicio o servidor de los 20 gobiernos definidos para realizar el análisis**, sino que, por el contrario, se tomaron los documentos públicos exhibidos por cada uno de los gobiernos.

En el presente informe se encuentra la información recopilada de los documentos detectados a través de la herramienta libre de ElevenPaths llamada FOCA OpenSource, que permite una vez descargados dichos documentos, que se extraigan los metadatos que contengan y se realice un análisis enfocado a cumplir el objetivo planteando.

1. Metodología

La primera etapa consistió en identificar **los dominios gubernamentales** de los veinte estados Latinoamericanos utilizados para este estudio, cuyos gobiernos exponen servicios o sitios web en Internet, entregando a sus habitantes la posibilidad de tener acceso a documentación de forma pública.

La segunda etapa, se basó en la utilización de FOCA OpenSource, herramienta libre desarrollada por ElevenPaths, que nos permite, usando técnicas de OSINT, encontrar de manera sencilla y rápida, los documentos digitales que son expuestos públicamente por los dominios de los gobiernos en estudio. Con la misma herramienta una vez detectados, estos documentos son descargados y se realiza una extracción de los metadatos que contiene cada uno de estos archivos.

Al extraer la información **de los metadatos es posible identificar** nombres de usuario, direcciones IP, correos electrónicos, nombres de red, directorios de almacenamiento y sistemas operativos, entre otras cosas. A partir de esta información, es posible inferir características o procesos inseguros, o que no son considerados como buenas prácticas desde la óptica de la seguridad de la información, tales como el uso de usuarios genéricos, sistemas operativos obsoletos, correos electrónicos expuestos a fugas de credenciales, correos electrónicos de dominios no oficiales en documentos oficiales, etc.

Como tercera etapa, ya con los datos de las direcciones IP de los servicios de http y https, se utilizó la herramienta whatweb para determinar cuáles de estas direcciones aún tienen los servicios activos y poder tener el país donde están alojados estos servicios. Para las direcciones IP que no responden al servicio, se utilizó ip2location con el fin de identificar su referencia geográfica.

Como cuarta etapa, con los correos electrónicos detectados durante el análisis de los metadatos, se realizó un análisis de dominios públicos y de los dominios gubernamentales, para determinar la cantidad de correos detectados de cada uno de estos. Adicionalmente, se utilizó la herramienta OSRFramework para determinar cuántos de estos correos han sido expuestos en las diferentes fugas de información expuestas en Internet y que son de público conocimiento.

Todos los análisis se realizaron usando herramientas libres y con documentación pública, lo que garantiza que pueda ser repetido por cualquier entidad o persona.

2. Dominios analizados

Como base para la elaboración del presente estudio, hemos tomado el informe realizado por el Banco Interamericano de Desarrollo (BID) y la Organización de los Estados Americanos (OEA) sobre "El Estado de la Ciberseguridad en América Latina y el Caribe: ¿estamos preparados en América Latina y Caribe?" (www.observatoriociberseguridad.com) que analizaba la importancia de la ciberseguridad para los Estados, y como cada uno de éstos está desplegando los controles y las políticas en sus respectivos países, y utilizamos el listado de los veinte países que según el estudio han avanzado en alguno de los cinco grupos de calificación, y que a su vez, exponen en sitios web información para sus ciudadanos. En la tabla 1 se encuentran los países sobre los que se realizó el análisis, con los dominios gubernamentales que se usaron para detectar los documentos públicos, identificando la cantidad detectada, pero aclarando cuántos de ellos pudieron ser analizados finalmente dado que algunos no tenían información, o bien, se descargaban corrompidos.

PAIS	DOMINIO	DETECTADOS	ANALIZADOS
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
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
Republica Dominica	**gov.do**	1802	1654
Uruguay	**gub.uy**	2401	2108
Venezuela	**gob.ve**	2081	1810

Tabla 1. Países y dominios analizados, mostrando archivos detectados y analizados

3. Análisis inicial

Al determinar el dominio gubernamental principal en cada uno de los países, hemos configurado FOCA OpenSource de manera que realice una búsqueda automática de las principales extensiones de **MS Office, OpenOffice, PDF, Adobe, imágenes, entre otros,** utilizando las API Keys de Google y de Bing, permitiendo así encontrar los enlaces de los documentos y descargarlos, para realizar el análisis de los metadatos. En la Ilustración 1, se encuentra la cantidad de documentos detectados y cuántos fueron analizados, y allí se puede apreciar que ha sido posible validar los metadatos de más del 90% de los documentos públicos detectados en cada dominio.

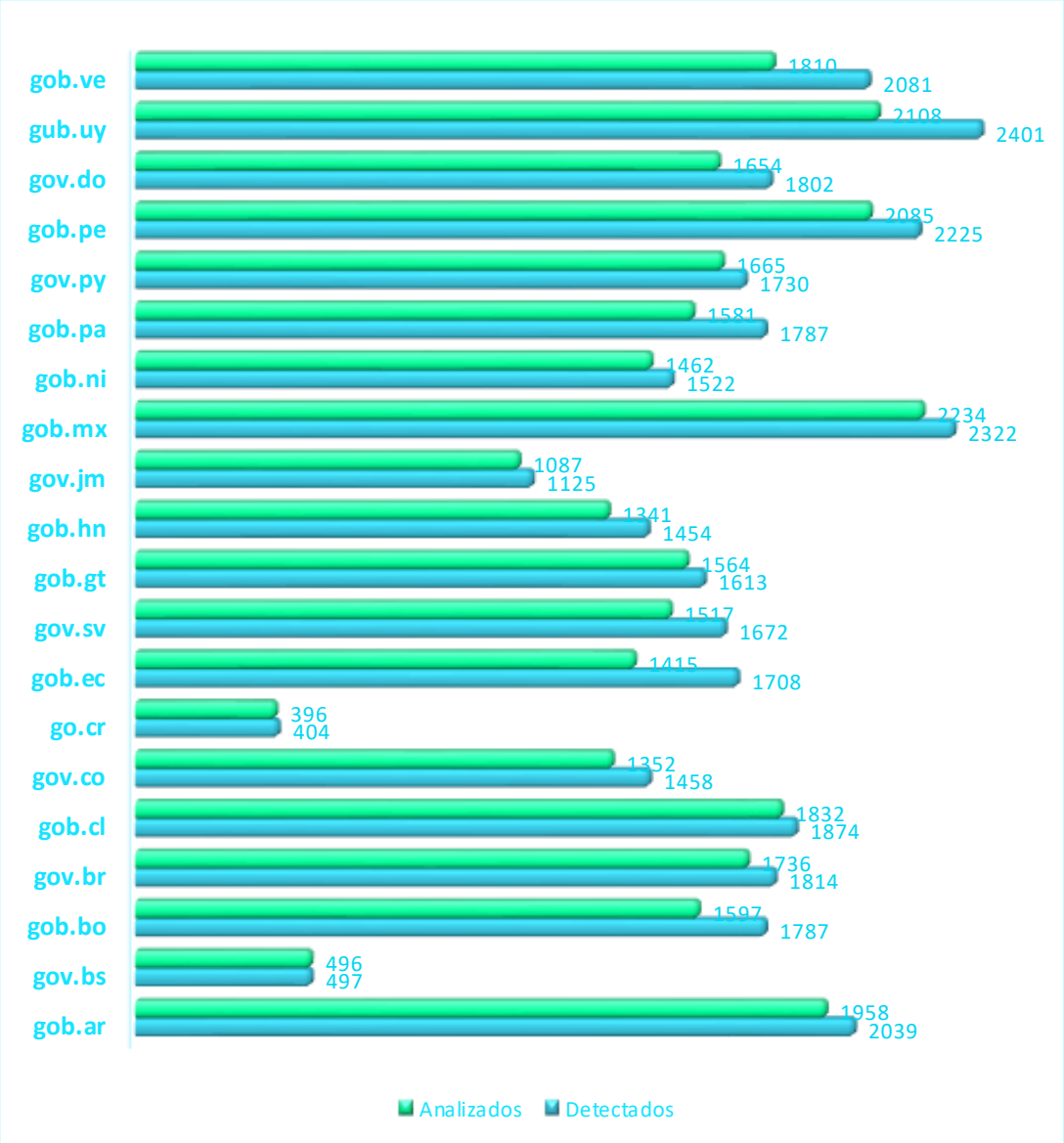


Ilustración 1. Cantidad de archivos detectados y analizados

Teniendo en cuenta que más del 80% de los documentos analizados son de software relacionado con ofimática, nos resultó interesante poder segmentar los tipos encontrados dentro de esta categoría, y fue así como pudimos identificar que el 27% pertenecen a procesadores de texto, el 22% a planillas de cálculo y un 17% a presentaciones, tal como se puede observar en la ilustración 2.

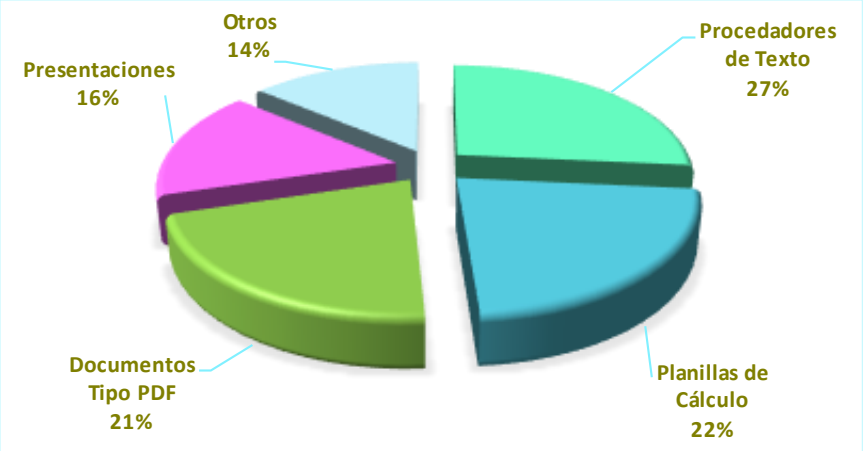


Ilustración 2. Porcentaje de tipos de archivos analizados

En la ilustración 3, es posible ver la cantidad de información extraída de los parámetros que se consideran como importantes en los metadatos de los documentos analizados.

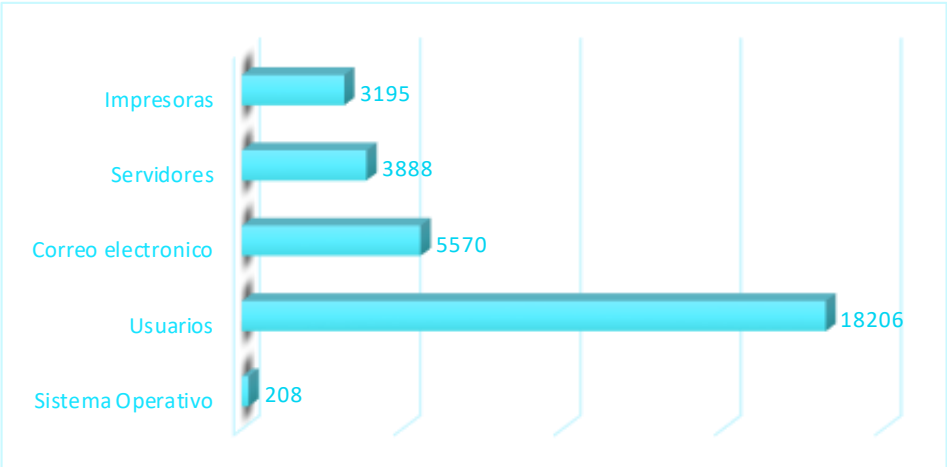


Ilustración 3. Cantidad de información detectada en los metadatos

4. Análisis de sistemas

Tal como hemos mencionado, los metadatos entregan características muy valiosas de la configuración de los sistemas en los que fueron creados los archivos, datos como el sistema operativo, impresora utilizada y el usuario de la máquina. Esta información es utilizada en investigaciones forenses informáticas para la identificación del dispositivo donde se hizo el documento o determinar una línea de tiempo, pero también permite a un delincuente informático perfilar las características del equipo o empresa que quiere atacar, reduciendo las opciones y mejorando la efectividad del ataque, por lo que esta fuga de información es muy sensible para una organización, y en este caso para el Estado, dado que revela muchas características de su infraestructura.

4.1.Sistemas operativos

Al extraer la información de los metadatos se detectaron los sistemas operativos de las diferentes máquinas en las que se generaron los documentos, alrededor de 10 diferentes sistemas operativos desplegados en un promedio de 545 máquinas por país. En la ilustración 4, se puede apreciar el número de sistemas operativos detectados en cada uno de los dominios analizados.

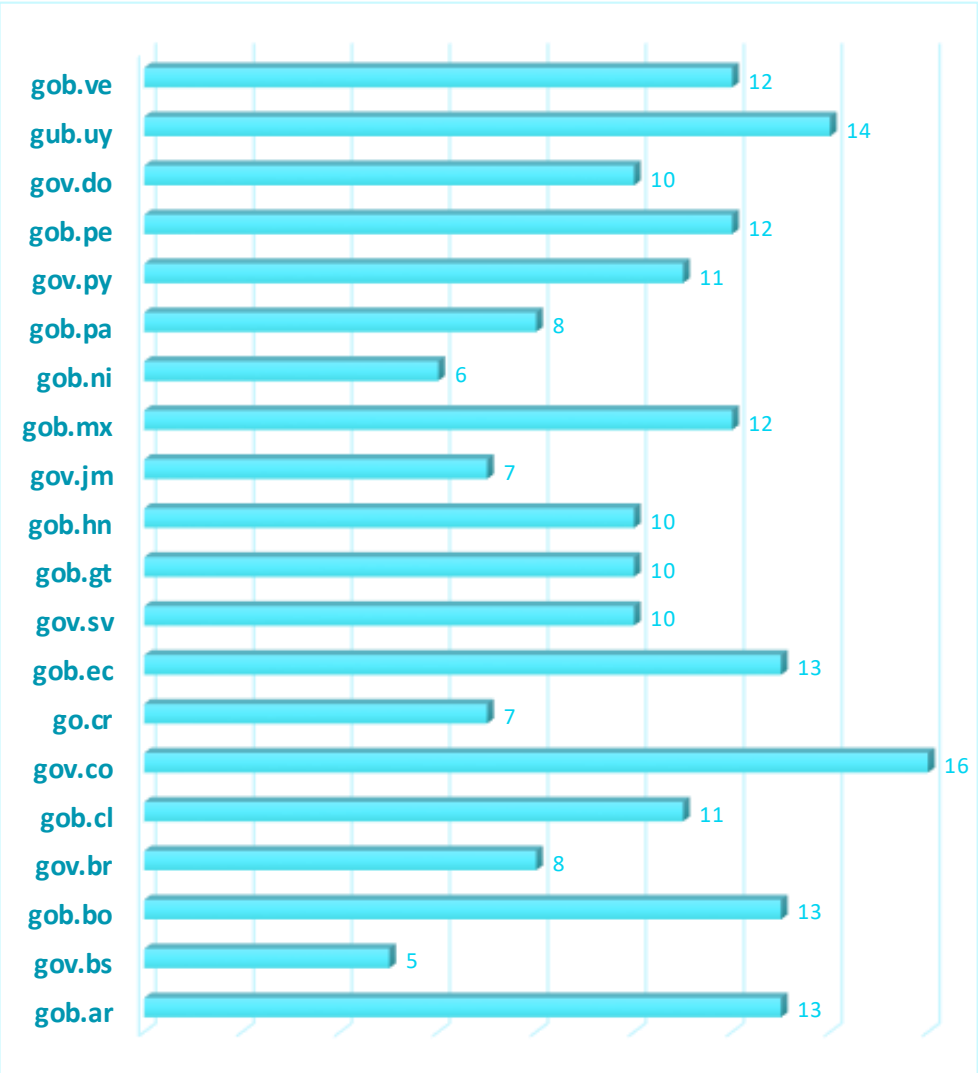


Ilustración 4. Sistemas operativos detectados en cada uno de los dominios

Al tomar las 10.911 máquinas detectadas, los sistemas operativos más utilizados ya no tienen soporte del fabricante hace algunos años, lo que genera un riesgo crítico a la información contenida en los mismos **suponiendo que los equipos con los que fueron hechos esos documentos aún sigan disponible**s. En la Ilustración 5, se puede apreciar el porcentaje de uso de los sistemas operativos.

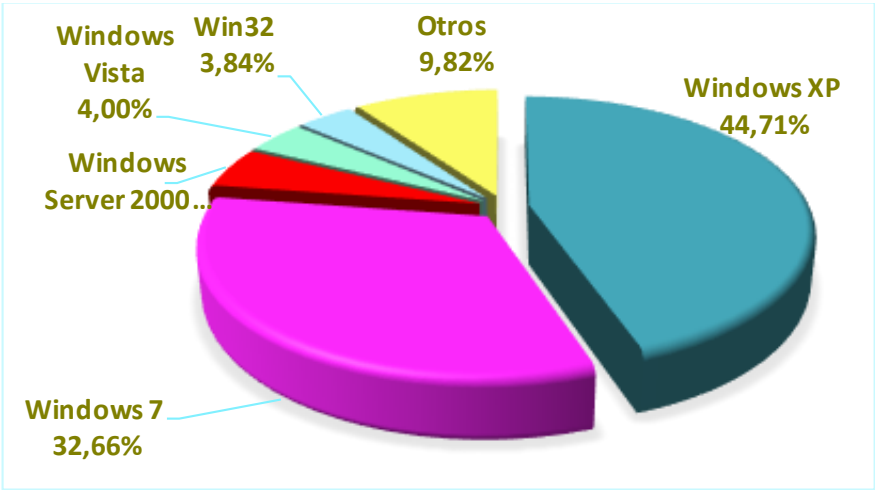
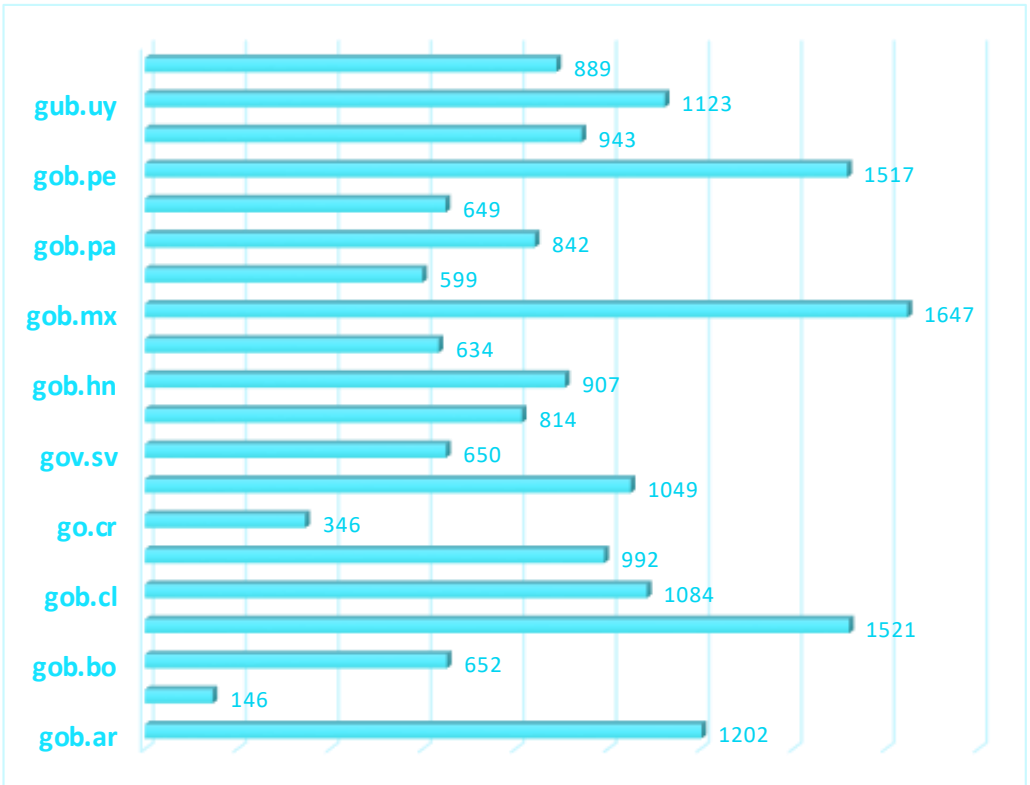


Ilustración 5. Porcentaje de uso de los sistemas operativos detectados.

4.2.Usuarios

La identificación de los usuarios a través de los metadatos es una información que permite a un atacante tener una lista de usuarios validos dentro de la infraestructura de la organización, en el proceso de análisis que se desarrolló en los dominios gubernamentales en Latinoamérica, se encontraron en promedio, 911 usuarios por cada uno de los dominios. Con un total de 18.206 usuarios identificados en los metadatos de los documentos analizados, distribuidos por dominio como se puede observar en la ilustración 6.



Las buenas prácticas sugieren que el uso de usuarios genéricos en los equipos de las organizaciones no debería estar permitido, pues esto genera que no sea posible tener una trazabilidad de los sucesos en los diferentes sistemas que un

usuario pueda realizar, aumentando el riesgo de una potencial perdida de confidencialidad de la información por un empleado desleal debido a la falta de controles.

En los procesos de investigación de riegos cibernéticos, siempre se tiene establecida una lista de usuarios que son típicamente utilizados en las infraestructuras empresariales, al realizar la búsqueda de estos usuarios, se detectaron 2.194 máquinas donde dicho usuario se ha identificado. A su vez, podemos identificar que el 12% de los usuarios detectados son genéricos, tal como se ve en la ilustración 7.

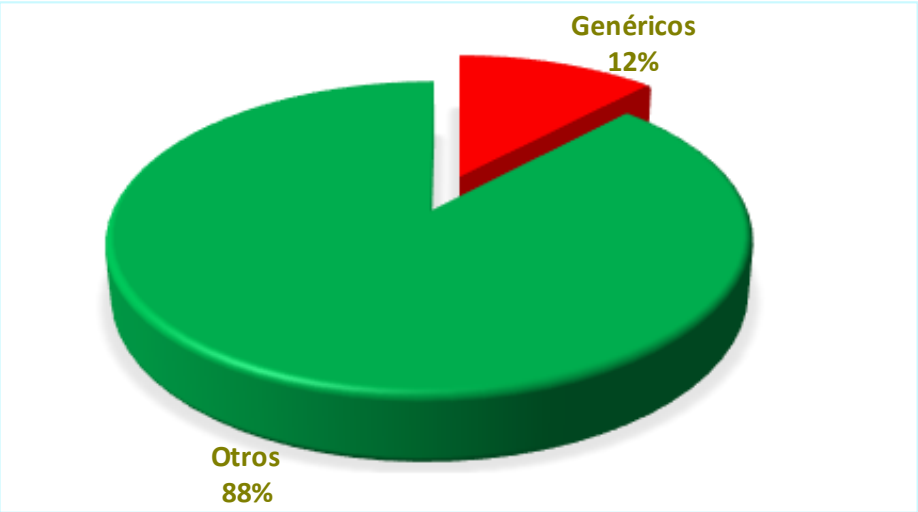


Ilustración 6. Porcentaje de usuarios genéricos detectados

Podemos observar también, que uno de los más utilizados es el "administrador" del sistema, lo que aumenta el nivel de criticidad a la amenaza generada por el uso de este tipo de usuarios. En la ilustración 8 se puede apreciar los usuarios genéricos más usados.

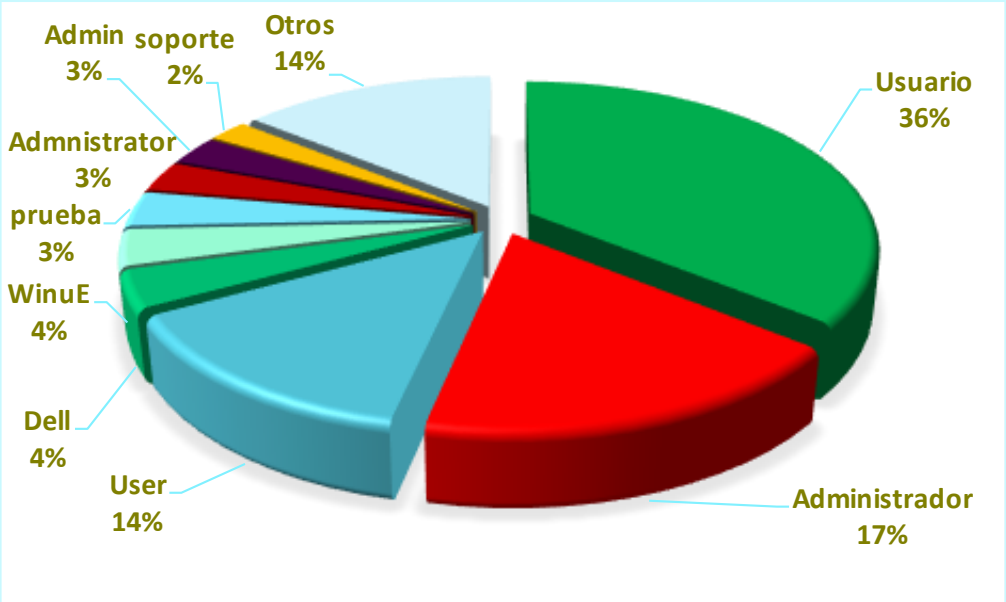


Ilustración 7. Top de usuarios genéricos detectados

5. Análisis de servicios

Los documentos analizados se obtienen de los servicios web de las entidades gubernamentales, las cuales a través de los protocolos de *http* y *https* permiten a los ciudadanos acceder a servicios digitales del Estado. Estos servicios deben cumplir con todas las normas de seguridad y de privacidad del país. Como parte del estudio, se realizó un proceso de análisis sobre los servicios detectados a fin de determinar cuántos realizan un aseguramiento del tráfico de la información utilizando cifrado en la conexión, se analizó la ubicación física del servidor basado en la dirección IP, y por último se intentó identificar por medio de técnicas de *fingerprinting* no invasivas, el tipo de software con el que se brindan dichos servicios.

5.1.Servidores detectados

Los documentos descubiertos se encuentran disponibles por parte de los Estados en Latinoamérica, mayormente bajo el uso del protocolo *http*, lo que se ve reflejado en la ilustración 9.

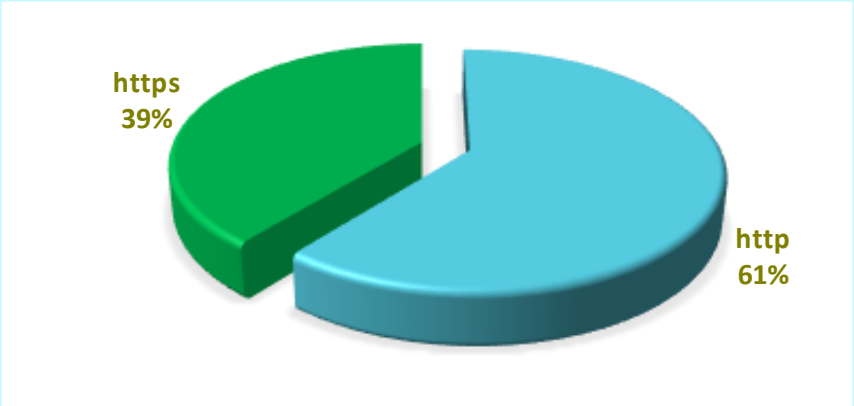


Ilustración 8. Porcentaje de los servicios detectados

En promedio cada Estado tiene 92 servidores de *http* y 59 de *https*. En la ilustración 10, se puede apreciar la cantidad de servicios detectados en cada uno de los dominios en ambos protocolos.

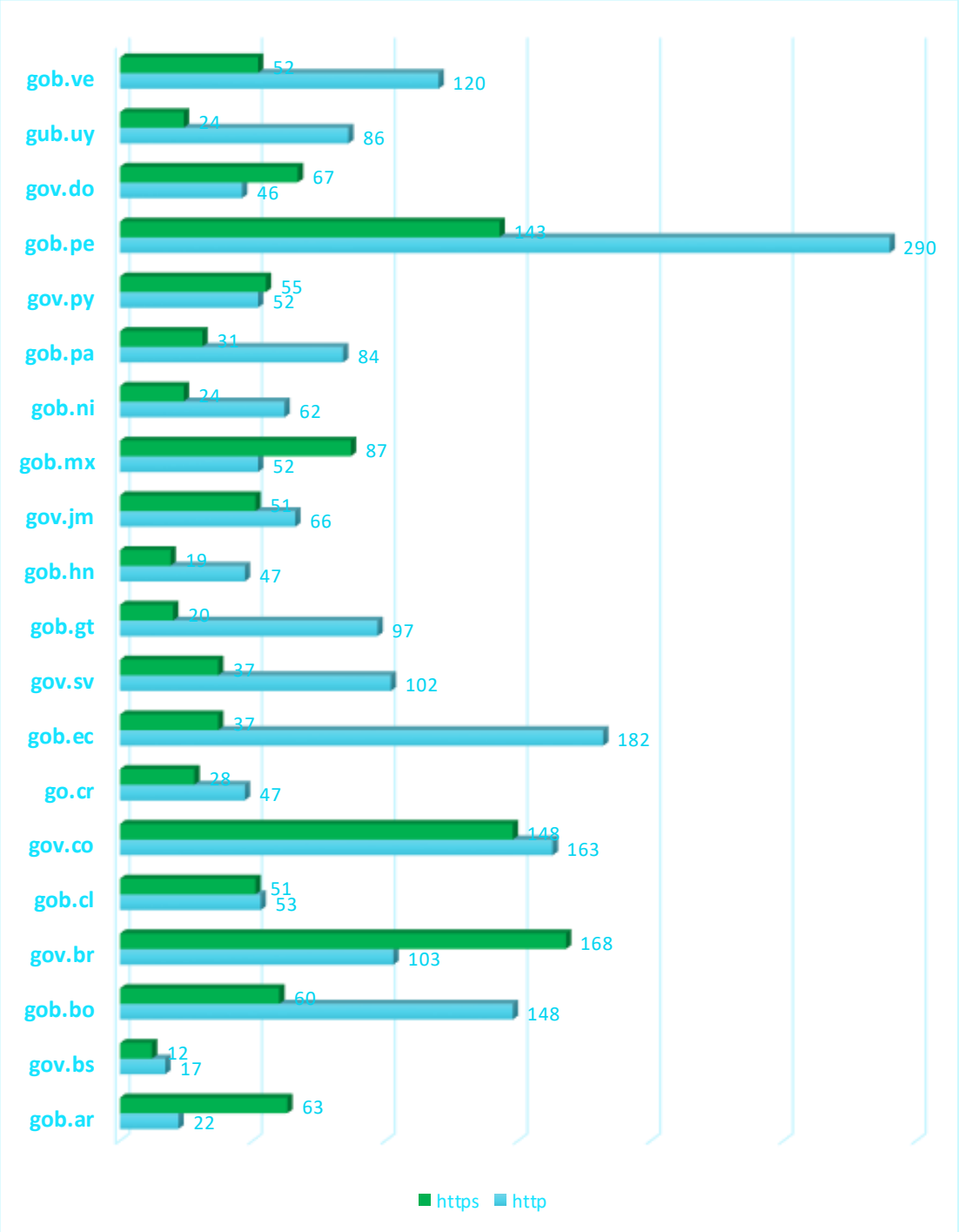


Ilustración 9. Cantidad de servicios de http y https detectados

5.2.Ubicación de los servicios

Para una correcta gestión de la seguridad de la información, es necesario conocer las políticas y normas de seguridad de la información que se deben cumplir en los países, por lo que para los Estados es vital poder tener el control de estas políticas sobre su información, para ello la ubicación de los servidores es de suma importancia, pues es donde físicamente se encuentran los datos, y las leyes de cada país tratan o exigen diferentes cuestiones relacionadas incluso

a este punto. Es por esto que se incluyó el análisis de la ubicación de los diferentes servidores detectados. En la ilustración 11, se puede ver la localización de los servidores de *http* y en la ilustración 12 se puede apreciar la localización de los servidores de *https*.

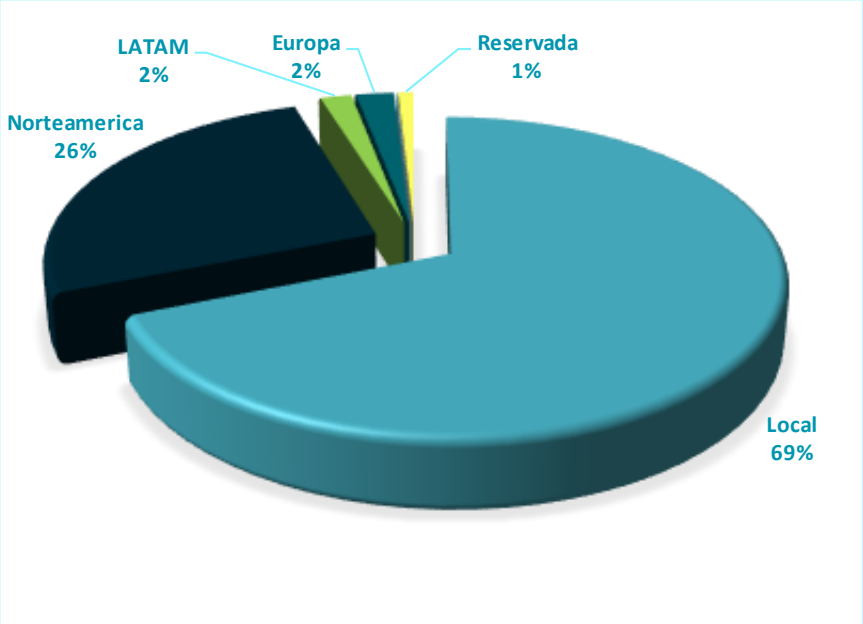


Ilustración 10. Ubicación de los servidores de HTTP

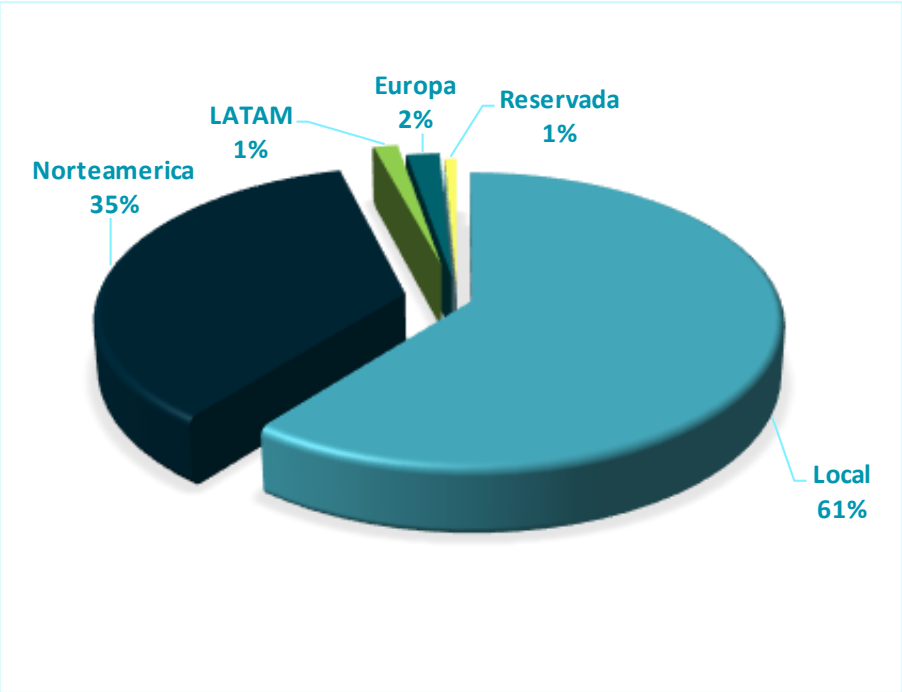


Ilustración 11. Ubicación de los servidores HTTPS.

Las ubicaciones se agrupan por regiones para facilitar el análisis y determinar las zonas donde se encuentra la mayor concentración de la información y los equipos que si están en cada uno de los países está calculado dentro del grupo que se denomina local.

6. Análisis de correos electrónicos

En los metadatos de los documentos de ofimática es usual encontrar direcciones de correo electrónico asociadas, o a datos de las maquinas donde fueron creados estos archivos, .por lo que en el estudio, uno de los parámetros en los que nos enfocamos, es la cantidad de correos electrónicos detectados y los dominios a los cuales pertenecen, pues en algunos casos permiten no solo revelar el editor original del archivo, sino también sus posibles receptores.

Adicionalmente, los correos electrónicos son usualmente utilizados para la creación de diferentes perfiles en los servicios de Internet, por lo que es posible validar si los correos electrónicos detectados en los archivos han sido expuestos en alguna fuga de información de estos servicios que han resultado de conocimiento público.

6.1.Cantidad de correos detectados

Al extraer los metadatos de los documentos, se encuentra que en todos los dominios estatales existen correos electrónicos divulgados, siendo un promedio de casi 279 correos electrónicos por cada dominio y se han encontrado un total de 5.570 diferentes cuentas de correo electrónico, distribuidas como se puede apreciar en la ilustración 13.

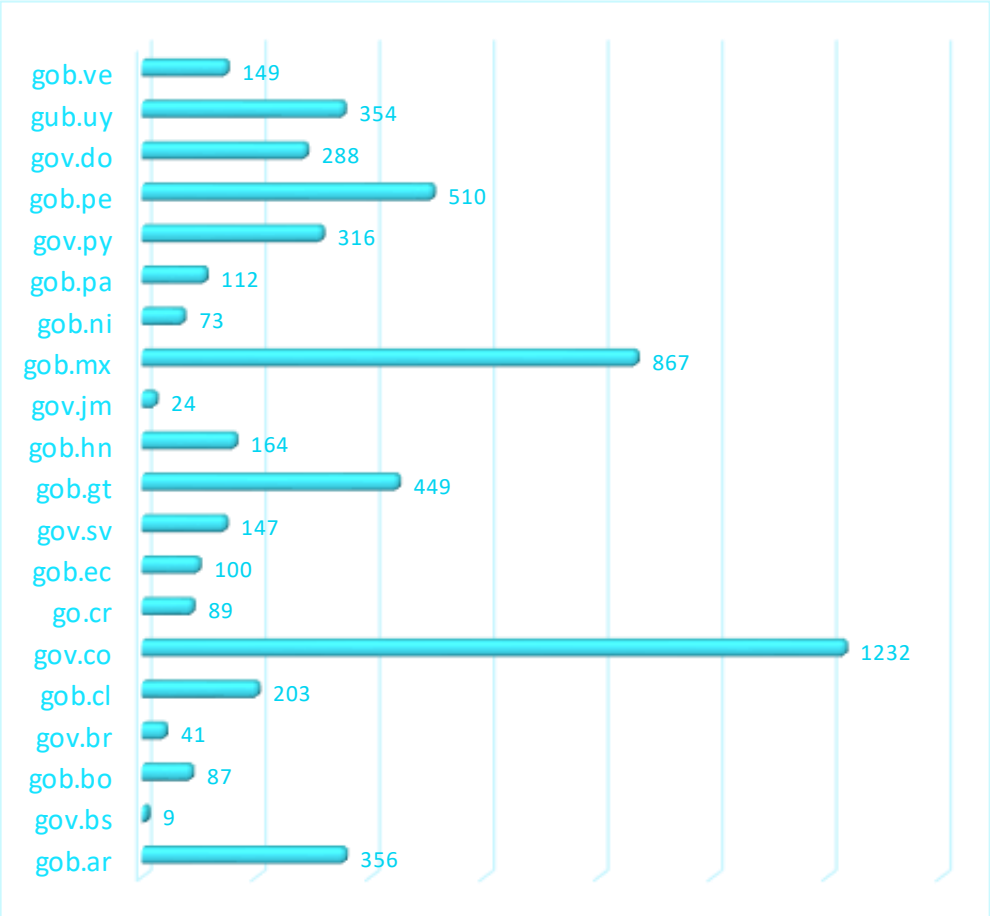


Ilustración 12. Cantidad de correos electrónicos detectados por dominio.

Con estos correos electrónicos se distribuyen en 1.462 dominios diferentes, evidenciando que los documentos expuestos en los servicios de los Estados Latinoamericanos provienen de diversas fuentes. Al hacer el análisis de la

cantidad de correos electrónicos agrupados por dominios se encuentra que el 36% pertenece a servicios públicos de correo electrónico en vez de estar relacionados a las entidades de manera directa. En la ilustración 14 se puede ver el top de los dominios detectados.

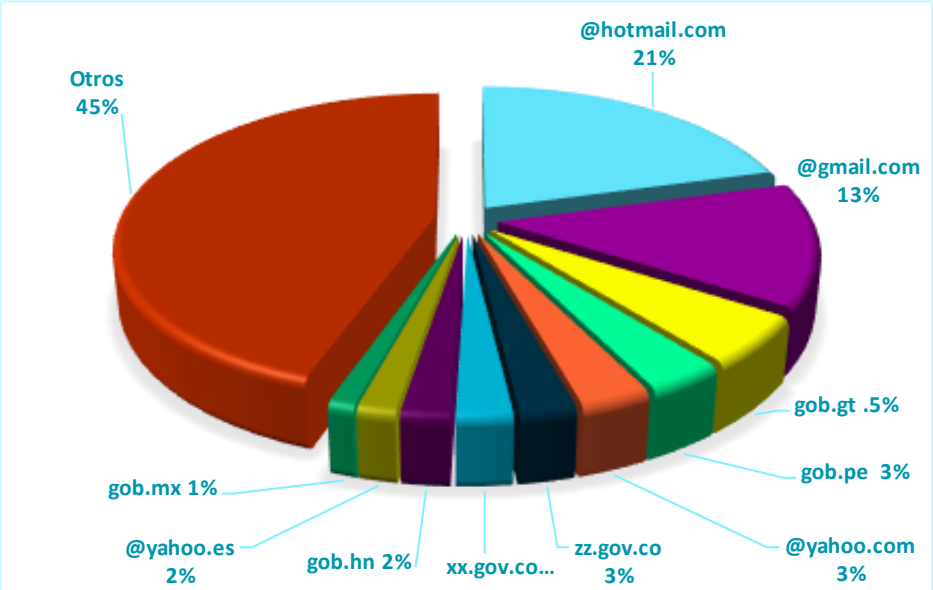
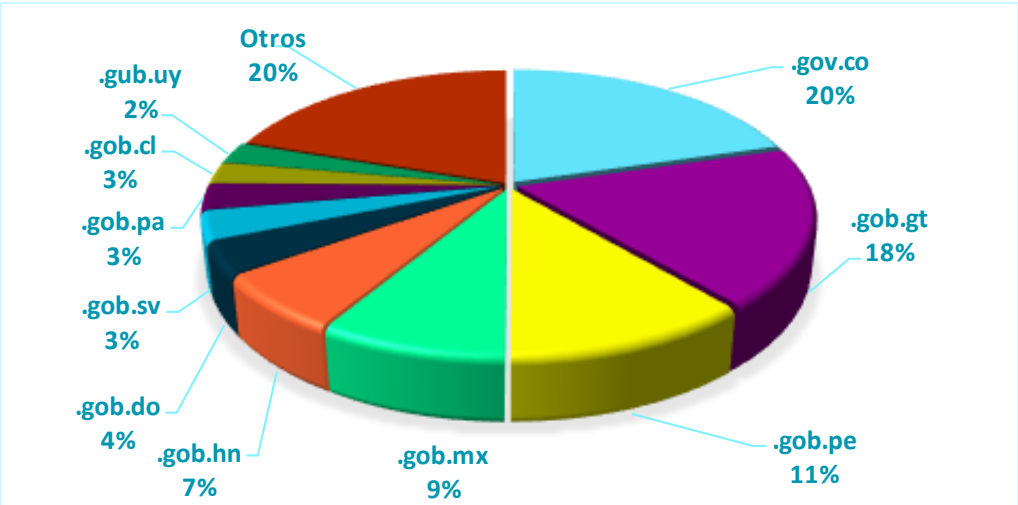


Ilustración 13. Top de dominios de correo electrónico detectados.

En el análisis se extraen los dominios que poseen alguna relación con los gobiernos, como son los educativos (.edu) y los de las fuerzas armadas (.mil), adicionalmente de los que estamos analizando que son propiamente los gubernamentales (.gov, gob, .gub, .go), encontrando un total de 2013 correos electrónicos. Donde los dominios más recurrentes se ven en la ilustración15.



6.2. Correos expuestos en fugas de información

En Internet todos los días son expuestas bases de datos donde credenciales de muchísimos dominios quedan a disposición de cualquier persona que la descarga de Internet, y expone la identidad de sus propietarios y la seguridad de las empresas u organizaciones ya que muchos usuarios utilizan la misma contraseña para diferentes servicios. Un

ejemplo de estas fugas, fueron las credenciales de acceso de más de 60 millones de usuarios de Dropbox que fueron filtrados abiertamente.

Dicho lo anterior, dentro del análisis se realizó la validación de las 5.570 cuentas de correos electrónicos dentro de las bases de datos expuestas, usando la herramienta de *mailfy*, que se encuentra en el marco de OSRFramework, y se detectó que **el 23% de las cuentas encontradas en los metadatos de los archivos han sido al menos una vez expuestas en alguna de estas fugas**. En la ilustración 16 podemos observar cuantas cuentas han sido expuestas por cada uno de los dominios analizados.

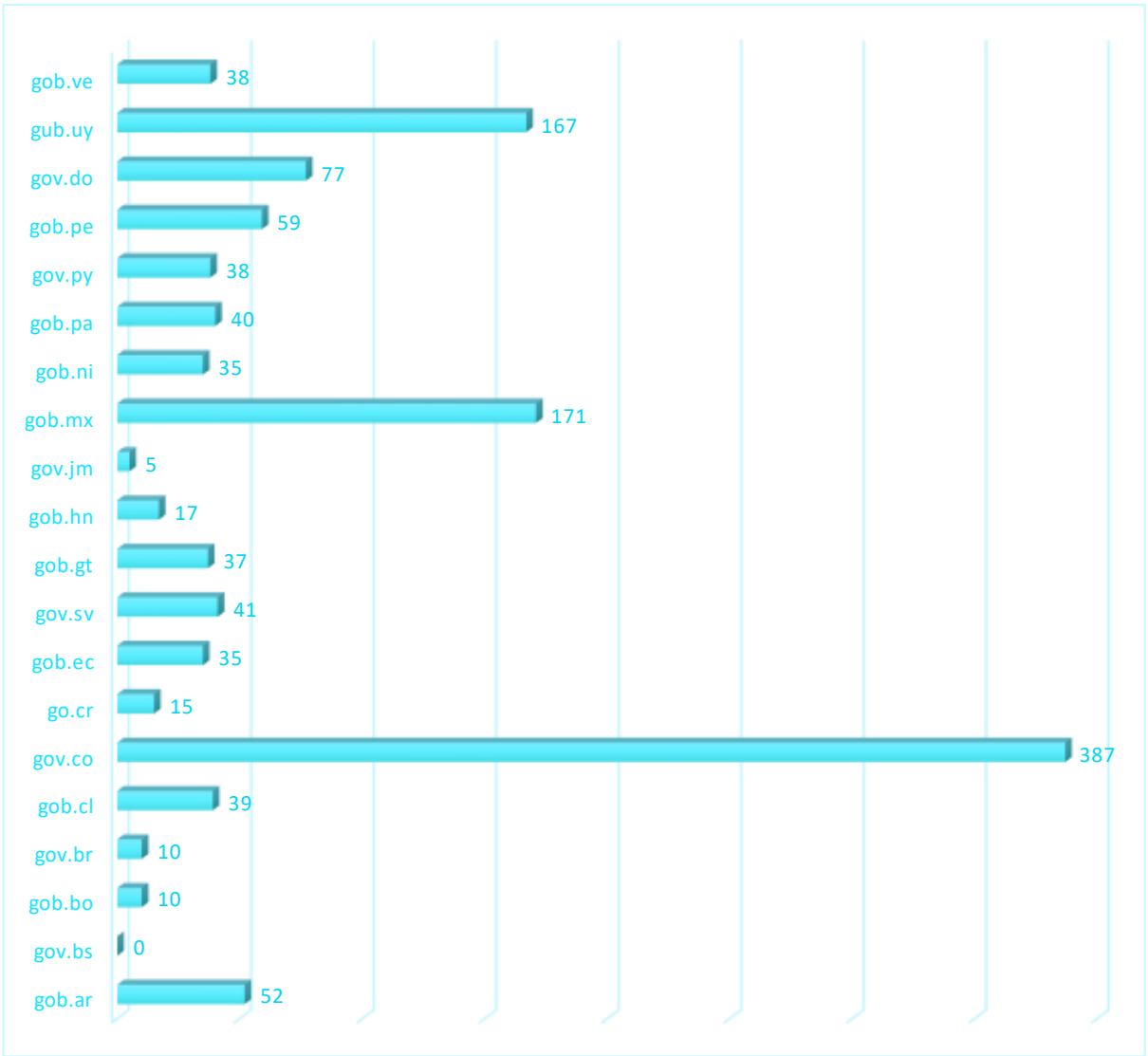


Ilustración 14. Cantidad de cuentas de correo electrónico expuestas por dominio.

El *framework* nos permite evidenciar en que fuga de información fue expuesta cada cuenta, encontrando que **en los 114 millones de cuentas expuestas en la fuga a LinkedIn está la mayor cantidad de cuentas asociadas a los metadatos de los archivos analizados.**

En la ilustración 16 se encuentra el porcentaje de cuentas encontradas en el top 10 de las fugas de información.

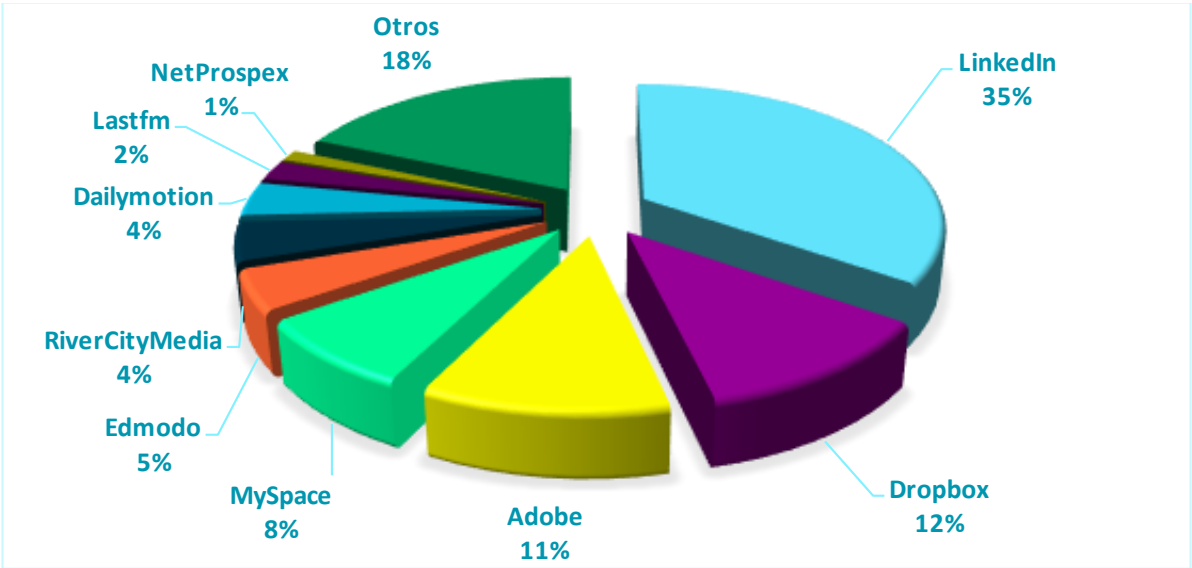


Ilustración 15. Porcentaje de cuentas expuestas en las fugas de información.

Dentro de lo correos detectados en las fugas de información se encuentran **dominios gubernamentales**, por lo que se toman los 2.013 correos detectados de gobiernos y se analizan con *mailfy*, encontrando que **102 de estas cuentas han sido expuestas en alguna fuga**. En la ilustración 18 se pueden apreciar las fugas de información donde se detectaron cuentas gubernamentales.

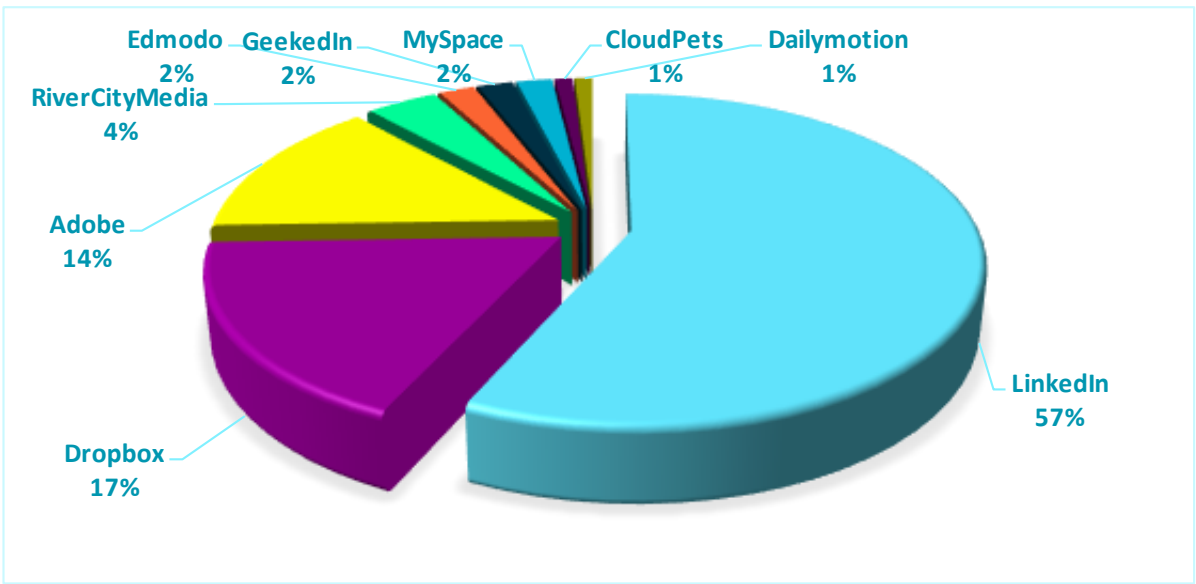


Ilustración 16. Porcentaje de cuentas gubernamentales detectadas en las fugas de información públicas.

7. Conclusiones

Del análisis presentado sobre los metadatos de los documentos expuestos por los gobiernos de Latinoamérica, podemos interpretar que, **al momento de realizar el análisis, aún es posible mejorar las acciones relacionadas a la contención de la fuga de este tipo de información, ya sea por medio de la implementación de procedimientos y controles manuales, o automatizados,** y así evitar que se expongan datos sobre la infraestructura tecnológica, sobre los usuarios de los sistemas y sobre cuentas de correo electrónico de los gestores de los archivos.

A partir de los resultados obtenidos, se podría inferir que **la infraestructura tecnológica no se encuentra actualizada** ya que en su mayoría se encontraría basada en sistemas operativos que ya no son soportados por sus fabricantes, **sin embargo, esto no podemos asegurarlo dado que no hemos realizado ese tipo de análisis, y los archivos podrían estar desde hace tiempo publicados mientras que la tecnología pudo ser actualizada**.

En el análisis de los usuarios detectados se encuentra que **el 12% son usuarios genéricos de los sistemas** y que en una gran mayoría tienen características de administradores, **lo cual supone una posible falta de controles y políticas en la gestión de usuarios en los Estados**.

Al realizar la validación de la ubicación de los servidores **se detecta que el 30% no se encuentra alojado en el mismo país productor de la información**.

El análisis de los metadatos basados en los correos electrónicos detectados **evidencia una tercerización en la elaboración de la documentación de los Estados**.

Las fugas de información de las grandes empresas de servicios o redes sociales, como Linkedin, Dropbox y Adobe, son las que contienen la mayoría de las cuentas de correo electrónico gubernamentales evidenciadas en los metadatos analizados.

Es preciso destacar que para la elaboración del presente informe **no se ha realizado una labor exhaustiva de análisis de aspectos distintos a los mencionados en la introducción**: sistemas operativos, ubicación de direccionamiento IP, usuarios de sistemas operativos y cuentas de correo asociadas a los metadatos de los documentos públicos que se descargaron. A pesar de ello, y sin necesidad de analizar en profundidad otras características, se han descubierto posibles problemas de seguridad relevantes para los gobiernos Latinoamericanos, y esperamos que el presente estudio sea de utilidad para cada gobierno, y sirva para apoyar los análisis de seguridad que seguramente cada uno de ellos realiza periódicamente.

Acerca de ElevenPaths

En ElevenPaths, la Unidad de Ciberseguridad de Telefónica, creemos en la idea de desafiar el estado actual de la seguridad, característica que debe estar siempre presente en la tecnología. Nos replanteamos continuamente la relación entre la seguridad y las personas con el objetivo de crear productos innovadores capaces de transformar el concepto de seguridad y de esta manera, ir un paso por delante de nuestros atacantes, cada vez más presentes en nuestra vida digital.

Más información

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