RELACIÓN ENTRE LOS VIRUS INFORMATICOS (MALWARE) Y ATAQUES EN PAISES VULNERABLES DE SEGURIDAD EN INFORMATICA UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)

Relationship between computer viruses (malware) and attacks vulnerable countries using computer security with Principal Component Analysis (PCA)

Autor: Alain Donuhue Dongo Quintana
alain.dongo@gmail.com
Docente Facultad de Ingeniería UCSS
CIO Strategy TI
Lima, Febrero, 2016

RESUMEN: Utilizando información de Karpesky Lab el presente artículo se observa que a través de Análisis de Componentes Principales (ACP) existe una la relación de ciertos virus informáticos sobre todo los malware en los países más vulnerables en seguridad de informática, donde se arroja como resultados que el país Argelia es el más atacado por virus troyanos, mientras que los países como Ucrania y Uzbekistán son más propensos a infectarse con virus a través de internet, se nota también que Corea del Sur como China son más atacados por los virus con intentos de infección, finalmente Bielorrusia es el país donde a través de sus PC´s tienen demasiado riesgo a contagiarse de virus. Para este análisis se ha utilizado para el cálculo el software libre R Statistics, así como también el Rattle como una herramienta de Minería de Datos.

ABSTRACT Using data from Kaspersky Lab this article notes that through Principal Component Analysis (PCA) there is the relationship of certain computer viruses especially malware in the most vulnerable countries where it is observed that the country Algeria is the most attacked by Trojan viruses , while countries like Ukraine and Uzbekistan are more likely to become infected with the virus through internet , it also notes that South Korea and China are attacked by the virus infection attempts , Bielorusia is the last country where through their PCs have too much risk of catching viruses. For this analysis was used to calculate the free software R Statistics and Rattle as a data mining tool

Introducción:

A medida que aumentan el número y la gravedad de los crímenes cibernéticos, es importante conocer los diversos tipos de ataque como el malware (del inglés "malicious software", también llamado badware, código maligno, software malicioso o software malintencionado, es un tipo de software que tiene como objetivo infiltrarse o dañar una computadora o sistema de información sin el consentimiento de su propietario). En el presente artículo se analizan los factores que actualmente impulsan el desarrollo de malware, se exponen en detalle las características de cada uno de ellos, se describe cómo se manifiestan en la red y se explica cómo pueden remediarse.

Aunque los nombres de algunos tipos de malware pueden resultarnos familiares, estas amenazas evolucionan constantemente, obligando a todo aquel que desee garantizar la seguridad de su sistema a adaptarse a estos continuos cambios. Los principales métodos utilizados por estos cibercriminales incluyen: La extorsión. Consiste en bloquear computadoras o interrumpir su funcionamiento para posteriormente cobrar dinero por solucionar los supuestos problemas.

En este tipo de ataques, a menudo se realizan escaneos inútiles de las computadoras o se vende software "antivirus" igualmente inútil. Esta técnica puede utilizarse para obtener información de tarjetas de crédito. En ocasiones, el software que adquieren las víctimas es "scareware", que les lleva a comprar otros supuestos productos de software o a pagar servicios de suscripción inútiles. El robo de recursos electrónicos. Entre los recursos electrónicos que se roban con mayor frecuencia se encuentran los datos de identificación personal (robo de identidad) de los registros de empleados o de clientes; información y contraseñas de cuentas financieras; datos comerciales y de negocio propietarios que pueden venderse a competidores; cuentas de correo electrónico —incluidos los contactos—, que pueden utilizarse para enviar mensajes spam (desde fuentes aparentemente fiables); e incluso recursos informáticos propiamente dichos (zombies), que son controlados por los criminales para perpetrar todo tipo de ataques: desde mensajes spam hasta el hospedaje de contenidos pornográficos. El software utilizado para cometer este tipo de crímenes pertenece a la categoría de malware.

Ante ello con la información brindada por Karspesky Lab interesa investigar la relación entre los ataques de los virus troyanos y el resto de variables, utilizando Análisis de Componentes Principales (ACP) donde se ha realizado con el objetivo transformar un conjunto de variables originales, en un nuevo conjunto de variables (sin perder información), combinación lineal de las originales, denominadas componentes principales (factores) para ver la relación que hay entre los países más vulnerables frente a los ataques de los virus informáticos con la información utilizada de Kasperesky Lab del 2do Trimestre del año 2015

Estadísticas

***Todos los datos estadísticos usados en el informe se han obtenido mediante la red antivirus distribuida Kaspersky Security Network (KSN) como resultado del funcionamiento de los diferentes componentes de protección contra los programas maliciosos. Los datos se obtuvieron en los equipos de los usuarios de KSN que confirmaron su consentimiento en enviarlos. En el intercambio global de información sobre las actividades maliciosas toman parte millones de usuarios de los productos de Kaspersky Lab de 213 países del mundo.*

Amenazas para dispositivos móviles

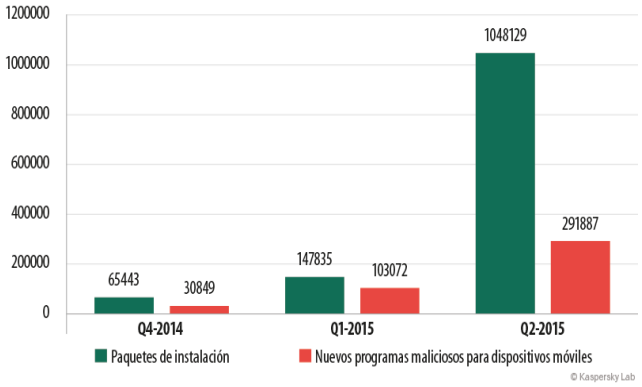
Los programas maliciosos bancarios móviles siguen siendo una de las principales amenazas móviles. En el informe del primer trimestre de 2015 según *Kaspersky Lab* sobre el troyano Trojan-SMS.AndroidOS.OpFake.cc, que sabía atacar como mínimo a 29 bancos e instituciones financieras. La versión más reciente de este troyano puede atacar a 114 instituciones bancarias y financieras. Su principal objetivo es robar el login y contraseña de la cuenta. Con el mismo fin ataca a varias aplicaciones de correo populares.

Cantidad de las nuevas amenazas móviles

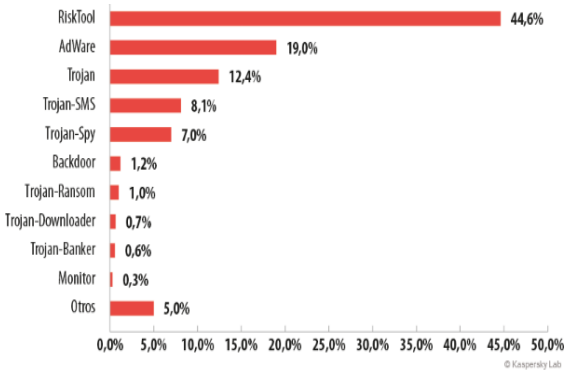
En el segundo trimestre de 2015 Kaspersky Lab ha detectado **291887** nuevos programas maliciosos móviles, 2,8 veces más que en el primer trimestre de 2015.

Con esto, la cantidad de paquetes maliciosos detectados fue de 1048129, que es 7 veces más que en el trimestre anterior.

Número de paquetes de instalación y nuevos programas nocivos móviles detectados (del cuarto trimestre de 2014 al segundo de 2015)



Distribución por tipos de los programas móviles detectados



Distribución por tipos de los nuevos programas maliciosos móviles, segundo trimestre de 2015

RiskTool (44,6%) lidera la estadística de los programas maliciosos para dispositivos móviles detectados en el segundo trimestre de 2015. Se trata de aplicaciones legales que son potencialmente peligrosas para los usuarios: su uso malicioso o irresponsable por parte del dueño del teléfono inteligente puede causar pérdidas financieras al usuario. En el segundo puesto está la categoría Adware, formada por programas publicitarios potencialmente indeseables (19%).

Los troyanos SMS que durante largo tiempo lideraron en esta estadística, en el segundo trimestre ocuparon sólo el cuarto puesto con un 8,1%. Esto es un 12,9% menos que en el primer trimestre. La reducción del porcentaje de este tipo de malware está condicionada por el hecho de que las personas que antes propagaban activamente los troyanos SMS empezaron a usar métodos más transparentes de monetización (de lo que da fe el crecimiento del porcentaje de RiskTool), o bien han preferido usar programas maliciosos de otros tipos. Así, la participación de los troyanos ha crecido del 9,8% en el primer trimestre al 12,4% en el segundo.

TOP 10 de países por la cantidad de usuarios atacados por troyanos bancarios móviles, del total de todos los usuarios atacados

	País*	Porcentaje de usuarios atacados en este país, del total de los usuarios atacados**
1	Corea del Sur	31,72%
2	Rusia	10,35%
3	Australia	6,62%
4	Austria	6,03%
5	Japón	4,73%
6	Uzbekistán	4,17%
7	Bielorrusia	3,72%
8	Ecuador	3,50%
9	Ucrania	3,46%
10	Suiza	3,09%

*** Porcentaje de usuarios únicos en el país atacados por troyanos bancarios móviles, en relación al total de usuarios únicos de este país atacados por malware móvil.*

Intentos de infección por programas maliciosos móviles en el segundo trimestre de 2015 (porcentaje del total de usuarios atacados)

TOP 10 de países según el porcentaje de usuarios atacados por intentos de infección por los troyanos bancarios móviles:

	País*	% de usuarios atacados**
1	China	16,34%
2	Malasia	12,65%
3	Nigeria	11,48%
4	Bangladesh	10,89%
5	Tanzania	9,66%
6	Argelia	9,33%
7	Uzbekistán	8,56%
8	Rusia	8,51%
9	Ucrania	8,39%
10	Bielorrusia	8,05%

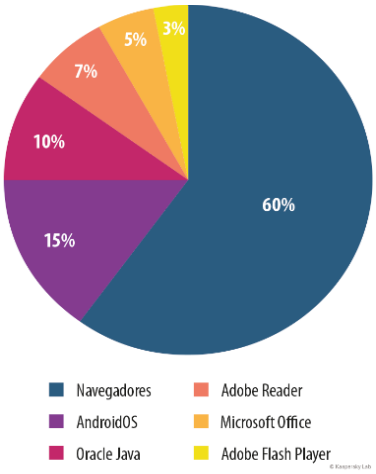
*** Porcentaje en el país de usuarios únicos atacados por los troyanos bancarios móviles, del total de usuarios del antivirus móvil de Kaspersky Lab en el país.*

El líder de esta lista es China, donde durante el trimestre el 16,34% de los usuarios de nuestro producto sufrieron ataques por lo menos una vez. En el segundo puesto está Malasia, con un 12,65%. Rusia (8,05%), Ucrania (8,39%) y Bielorrusia (8,05%) cierran el TOP 10, por debajo de los países de Asia y África.

En esta lista Corea ocupó el undécimo lugar con un 7,46%. Recordamos que en Corea los troyanos bancarios móviles son muy populares entre los delincuentes: son responsables de haber atacado al 31,72% del total de usuarios atacados por malware móvil.

Aplicaciones vulnerables usadas por los delincuentes

La siguiente clasificación de las aplicaciones vulnerables se basa en los datos de los exploits usados por los delincuentes tanto en los ataques mediante Internet como en las aplicaciones locales afectadas, entre ellas en los dispositivos móviles de los usuarios, y que bloquearon nuestros productos.



Distribución de los exploits usados en los ataques de los delincuentes, según tipos de aplicaciones atacadas, segundo trimestre de 2015

La distribución de los exploits casi no ha cambiado en comparación con el primer trimestre. Siguen liderando la estadística los exploits para navegadores de Internet (60%). En este momento la mayoría de los paquetes de exploits contienen exploits para Adobe Flash Player e Internet Explorer. Cabe destacar el aumento significativo en el número de exploits para Adobe Reader (+6 puntos porcentuales). Esto está relacionado con la gran cantidad de envíos de spam que contienen documentos PDF maliciosos.

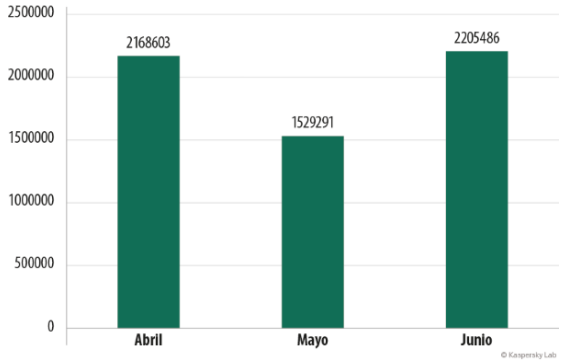
Programas maliciosos en Internet (ataques mediante la web)

Los datos estadísticos de este capítulo han sido recopilados por el antivirus web, que protege a los usuarios cuando descargan objetos maliciosos de una página web maliciosa o infectada. Los delincuentes crean sitios maliciosos adrede, pero también los sitios legítimos se pueden infectar si su contenido lo crean los usuarios (como en el caso de los foros), o si son víctimas de hackeo.

Amenazas online en el sector bancario

En el segundo trimestre de 2015 las soluciones de Kaspersky Lab neutralizaron los intentos de ejecución de programas maliciosos que roban dinero mediante el acceso a cuentas bancarias en los equipos de **755642** usuarios. En comparación con el trimestre anterior (929 082) este índice ha disminuido en un 18,7%. En el segundo trimestre de 2014 afectó a 735428 equipos.

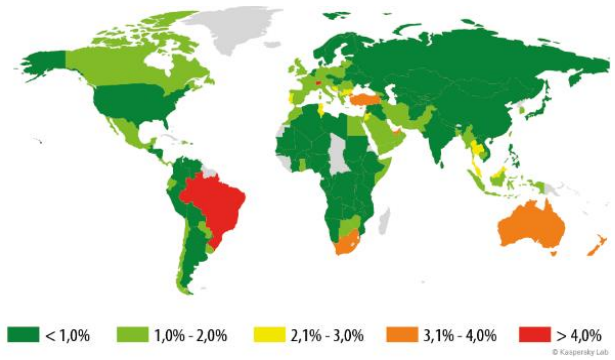
En total, los productos de Kaspersky Lab han registrado durante el año **5903377** notificaciones sobre intentos de infección con programas maliciosos que roban dinero de las cuentas bancarias online.



Número de ataques del software financiero malicioso contra los usuarios, segundo trimestre de 2015

Territorios de los ataques

Para evaluar y comparar el riesgo de infección con troyanos bancarios al que están expuestos las computadoras de los usuarios en diferentes países del mundo, ellos han calculado para cada país el porcentaje de usuarios de productos de Kaspersky Lab que se vieron afectados por esta amenaza en durante el trimestre.



Territorios afectados por los ataques del software malicioso en el segundo trimestre de 2015 (porcentaje de usuarios atacados)

TOP-10 de países según el porcentaje de usuarios atacados

	País*	% de usuarios atacados**
1	Singapur	5,28%
2	Suiza	4,16%
3	Brasil	4,07%
4	Australia	3,95%
5	Hong-Kong	3,66%
6	Turquía	3,64%
7	Nueva Zelanda	3,28%
8	África del Sur	3,13%
9	Líbano	3,10%
10	Emiratos Árabes Unidos	3,04%

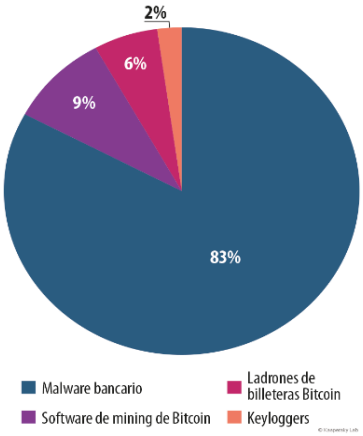
***Porcentaje de usuarios únicos que fueron víctimas de ataques web, de entre el total de los usuarios únicos de los productos de Kaspersky Lab en el país.*

En el segundo trimestre de 2015 Singapur es el país que está en el primer lugar por la cantidad de usuarios de Kaspersky Lab atacados por troyanos bancarios. Cabe destacar que en la mayoría de los países del TOP 10 el nivel de desarrollo tecnológico o de los sistemas bancarios es alto, lo que captura la atención de los delincuentes cibernéticos.

En Rusia el 0,75% de los usuarios se topó por lo menos una vez con troyanos bancarios; en EE.UU., el 0,89%; en España, el 2,02%, en Inglaterra, el 1,58% y en Alemania, el 1,16%.

Amenazas financieras

Las amenazas financieras no se limitan al malware bancario que ataca a los clientes de los sistemas de banca online.



Número de ataques del software malicioso financiero

En comparación con el trimestre anterior, ha aumentado la cantidad de software bancario malicioso, del 71% al 83%. La segunda amenaza financiera más popular en el segundo trimestre fueron los miners de Bitcoin, software malicioso que aprovecha la potencia del equipo de la víctima para generar bitcoins. El trimestre anterior esta categoría de malware estaba en el tercer puesto. Cabe destacar que algunos desarrolladores de software legítimo también instalan programas de minado de Bitcoin en sus aplicaciones.

TOP-10 de objetos detectados en Internet

Durante el segundo trimestre de 2015 el antivirus web detectó 26084253 objetos maliciosos únicos (scripts, exploits, ficheros ejecutables, etc.).

TOP 10 de objetos detectados en Internet

	Nombre*	% del total de ataques**
1	AdWare.JS.Agent.bg	47,66%

2	Malicious URL	32,11%
3	Trojan.Script.Generic	4,34%
4	AdWare.Script.Generic	4,12%
5	Trojan.Script.Iframer	3,99%
6	AdWare.JS.Agent.bt	0,74%
7	Exploit.Script.Blocker	0,56%
8	Trojan.Win32.Generic	0,49%
9	AdWare.AndroidOS.Xynyin.a	0,49%
10	Trojan-Downloader.Win32.Generic	0,37%

**Veredictos de detección pertenecientes al módulo del antivirus web. Esta información la han hecho posible los usuarios de los productos de KL que expresaron su consentimiento para la transmisión de datos estadísticos.*
*** Porcentaje del total de ataques web registrados en los computadoras de usuarios únicos.*

La agresiva propagación de programas publicitarios se ha reflejado una vez más en esta estadística: diez de las veinte posiciones las ocupan objetos que se clasifican como programas publicitarios. El primer puesto lo ocupa el script AdWare.JS.Agent.bg, que los programas publicitarios inyectan al azar en páginas web. Este último hasta ha logrado desplazar a Malicious URL, un veredicto para enlaces de la lista negra que ocupó el segundo puesto según los totales del trimestre.

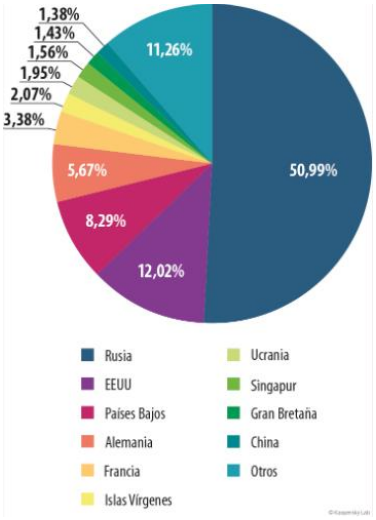
El veredicto Trojan-Ransom.JS.Blocker.a es un script que mediante la renovación cíclica de la página trata de bloquear el navegador y muestra un mensaje de que es necesario pagar una "multa" a determinada billetera electrónica por haber visto materiales indebidos. Este script se suele encontrar con mayor frecuencia en los sitios de pornografía.

Países-fuente de ataques web: Top 10

Esta estadística muestra la distribución según país de las fuentes de ataques web bloqueados por el antivirus en las computadoras de los usuarios (páginas web con redirecciones a exploits, sitios con exploits y otros programas maliciosos, centros de administración de botnets, etc.). Hacemos notar que cada host único puede ser fuente de uno o más ataques web.

Para determinar el origen geográfico de los ataques web Kaspersky Lab usó el método de comparación del nombre de dominio con la dirección IP real donde se encuentra el dominio dado y la definición de la ubicación geográfica de la dirección IP (GEOIP).

El segundo trimestre de 2015 las soluciones de Kaspersky Lab han neutralizado 379 972 834 ataques lanzados desde recursos de Internet ubicados en diferentes países del mundo. El 89% de las notificaciones sobre ataques web bloqueados se obtuvo durante el bloqueo de los ataques lanzados desde recursos web ubicados en diez países.



Distribución por países de las fuentes de ataques web, segundo trimestre de 2015

El líder de nuestra estadística no ha cambiado y sigue siendo Rusia (51%), con una participación que aumentó en un 11,27%. Suiza, que estaba presente en el TOP 10 del trimestre anterior, lo ha abandonado. En el octavo puesto, con un índice del 1,56% está Singapur.

Países en los cuales los usuarios han estado bajo mayor riesgo de infectarse mediante Internet

Para evaluar el riesgo de infección a través de Internet al que están expuestos las computadoras de los usuarios en diferentes países del mundo, han calculado con qué frecuencia durante el año los usuarios de los productos de Kaspersky Lab en cada país se han topado con la reacción del antivirus web. Los datos obtenidos son el índice de la agresividad del entorno en el que funcionan las computadoras en diferentes países.

	País*	% de usuarios únicos**
1	Rusia	38,98%
2	Kazajistán	37,70%
3	Ucrania	35,75%
4	Siria	34,36%
5	Bielorrusia	33,02%
6	Azerbaiyán	32,16%
7	Tailandia	31,56%
8	Georgia	31,44%
9	Moldavia	31,09%
10	Vietnam	30,83%

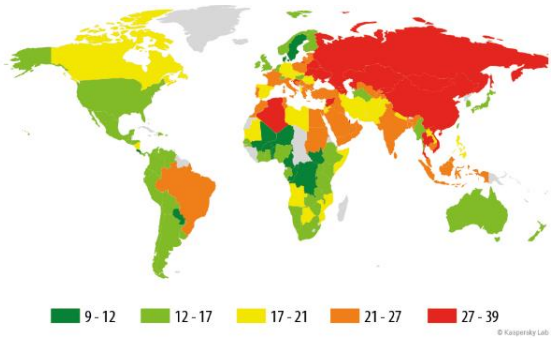
La presente estadística contiene los veredictos de detección del módulo del antivirus web que enviaron los usuarios de los

productos de Kaspersky Lab que dieron su consentimiento para el envío de datos estadísticos.

***Porcentaje de usuarios únicos que fueron víctimas de ataques web, de entre todos los usuarios únicos de los productos de Kaspersky Lab en el país.*

En el segundo trimestre de 2015 Rusia ha vuelto a ocupar el primer puesto, subiendo desde el segundo que ocupaba el trimestre anterior.

Entre los países más seguros para navegar en Internet están Argentina (13,2%), los Países Bajos (12,5%), Corea (12,4%), Suecia (11,8%), Paraguay (10,2%) y Dinamarca (10,1%).



En promedio, durante el trimestre el 23,9% de los computadoras en el mundo ha sufrido por lo menos un ataque web.

Amenazas locales

Un indicador crucial es la estadística de infecciones locales de las computadoras de los usuarios. En estos datos se enumeran los objetos que han entrado en el equipo infectando archivos o memorias extraíble, o aquellos que inicialmente entraron en forma velada (por ejemplo los programas incluidos en los instaladores complejos, archivos cifrados, etc.).

Kaspersky Lab analiza los datos estadísticos obtenidos del funcionamiento del antivirus que escanea los ficheros en el disco duro en el momento en que se los crea o cuando se los lee, y los datos del escaneo de diferentes memorias extraíbles.

El segundo trimestre de 2015 el antivirus para ficheros ha detectado 110731713 diferentes programas nocivos y potencialmente indeseables.

Objetos detectables encontrados en los computadoras de los usuarios: Top 10

	Nombre*	% de usuarios únicos atacados**
1	DangerousObject.Multi.Generic	22,64%
2	Trojan.Win32.Generic	15,05%
3	Trojan.WinLNK.StartPage.gena	8,28%
4	AdWare.Script.Generic	7,41%
5	Adware.NSIS.ConvertAd.heur	5,57%
6	WebToolbar.Win32.Agent.azm	4,48%
7	WebToolbar.JS.Condonit.a	4,42%
8	Trojan-Downloader.Win32.Generic	3,65%
9	Downloader.Win32.MediaGet.elo	3,39%
10	Trojan.Win32.AutoRun.gen	3,29%

**Veredictos de detección de los módulos OAS y ODS del antivirus, que fueron proporcionados por los usuarios de los productos de Kaspersky Lab que dieron su consentimiento para la transmisión de datos estadísticos. **Porcentaje de usuarios únicos en cuyos computadoras el antivirus detectó este objeto, del total de usuarios únicos de los productos de Kaspersky Lab y en los que ocurrió la detección.*

Países en los que las computadoras de los usuarios han estado bajo mayor riesgo de infección local

Para cada uno de los países hemos calculado qué porcentaje de usuarios de los productos de Kaspersky Lab se ha topado con las reacciones del antivirus de ficheros durante el periodo que abarca el informe. La presente estadística refleja el nivel de infección de las computadoras personales en diferentes países del mundo.

TOP 10 de países según su cantidad de computadoras infectados

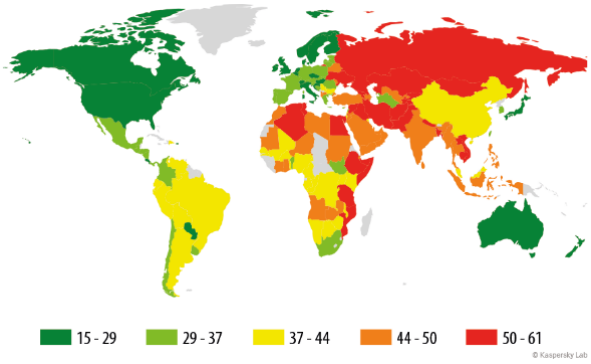
	País*	% de usuarios únicos**
1	Bangladesh	60,53%
2	Vietnam	59,77%
3	Pakistán	58,79%
4	Mongolia	58,59%
5	Georgia	57,86%
6	Somalia	57,22%
7	Nepal	55,90%
8	Afganistán	55,62%
9	Argelia	55,44%
10	Armenia	55,39%

*La presente estadística contiene los veredictos de detección del antivirus que proporcionaron los usuarios de los productos de Kaspersky Lab que dieron su consentimiento para el envío de datos estadísticos. Han tomado en cuenta los programas maliciosos encontrados directamente en los computadoras de los usuarios o en las memorias extraíbles conectadas a éstos: memorias USB, tarjetas de memoria de cámaras, teléfonos y discos duros externos. ** Porcentaje de usuarios únicos en cuyos computadoras se detectaron amenazas locales, de entre la cantidad total de usuarios de productos de Kaspersky Lab en el país.*

Este trimestre Bangladesh ocupa el primer puesto con un 60,53%, desplazando al segundo puesto a Vietnam, que fue el líder durante los dos últimos años. Pakistán (58,79%) subió del puesto 13, que ocupaba el trimestre pasado, al tercero.

Los nuevos de la estadística son Georgia, que de entrada ocupó el quinto lugar (57,8%); Rusia, que ocupa el undécimo puesto (55%); Túnez el decimosexto (53,7%) y Ucrania el decimonoveno (53%).

Los países con el menor nivel de infección son Suecia (19,7%), Dinamarca (18,4%) y Japón (15,5%).



En promedio, en el mundo por lo menos una vez durante el trimestre se detectaron amenazas locales en el 40% de los equipos de los usuarios, un 0,2% más que en el primer trimestre.

Análisis:

Para el análisis se ha considerado la información de Kaspersky Lab del 2do. Trimestre 2015, en donde se extrae información de los cuadros con información y de allí se proponemos que existe una correlación entre sus variables de infección como de intentos por los virus informáticos y los países más vulnerables por lo que utilizamos el método ACP

El ACP es un método algebraico/estadístico que trata de sintetizar y dar una estructura a la información contenida en una matriz de datos.

En el ACP, el primer factor o componente sería aquel que explica una mayor parte de la varianza total, el segundo factor sería aquel que explica la mayor parte de la varianza restante, es decir, de la que no explicaba el primero y así sucesivamente. De este modo sería posible obtener tantos componentes como variables originales.

El siguiente paso consiste en la obtención de los valores y vectores propios de la matriz de covarianzas muestral o de la matriz de coeficientes de correlación que se obtienen a partir de la matriz de datos. La elección de una u otra matriz para realizar el ACP es una cuestión controvertida. En este caso vamos a utilizar la matriz de correlaciones.

Se ha identificado a los países más vulnerables con las variables de los virus informáticos atacados o con intentos de infección ya sea móviles o a Pc´s.

País*	usuarios atacados por los troyanos bancarios móviles**	usuarios atacados por lo menos con un intentos de infección	usuarios que han estado bajo mayor riesgo de infectarse mediante Internet*	usuarios en los que los ordenadores de los usuarios han estado bajo mayor riesgo de infección local
Argelia	15525	1902.6669	1096221.626	613896.6169
Bielorrusia	61.179	1641.6365	1254670.298	670259.0588
China	50.9825	3332.2162	1052524.75	640693.6914
Corea del Sur	483.3141	1521.3178	1171456.247	661843.4486
Rusia	177.4191	1735.4443	1481134.107	608360.0312
Ucrania	59.1397	1710.9727	1358402.882	586988.8106
Uzbekistán	73.4148	1745.6408	1432497.584	600498.0796

En la salida de resultados se ha utilizado el software libre R Statistics y con él a través del interface gráfica de Rcommander se ha instalado un “plug in” de Minería de Datos que es el Rattle, herramientas que han ayudado para los cálculos, se observan varias gráficas descriptivas exploratorias donde se presentan varios datos.

El siguiente paso consiste en la obtención de los valores y vectores propios de la matriz de covarianzas muestral o de la matriz de coeficientes de correlación que se obtienen a partir de la matriz de datos.

En este caso vamos a utilizar la matriz de correlaciones.

```
Component loadings:
                                        Comp.1     Comp.2     Comp.3     Comp.4
infectarse.mediante.Internet          0.7114608 -0.1010315  0.07944174 0.6908728
Intentos.de.infeccion                -0.4870948 -0.1285286  0.76852671 0.3944436
PC.con.mayor.riesgo.de.infeccion.local -0.4289048 -0.5402188 -0.58272555 0.4296921
Troyanos.bancarios.m.viles           -0.2694492  0.8254918 -0.25196556 0.4271695

Component variances:
    Comp.1     Comp.2     Comp.3     Comp.4
1.86475295 1.16422068 0.89014606 0.08088031

Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4
Standard deviation     1.3655596 1.0789906 0.9434755 0.28439464
Proportion of Variance 0.4661882 0.2910552 0.2225365 0.02022008
Cumulative Proportion  0.4661882 0.7572434 0.9797799 1.00000000
```

Los otros elementos importantes en un ACP son los vectores propios asociados a cada valor propio.

```
+   col.quanti.sup="blue", label=c("var", "quanti.sup"), lim.cos2.var=0)

> res$eig
       eigenvalue percentage of variance cumulative percentage of variance
comp 1 1.86475295              46.618824                           46.61882
comp 2 1.16422068              29.105517                           75.72434
comp 3 0.89014606              22.253651                           97.97799
comp 4 0.08088031               2.022008                          100.00000

> res$var
$coord
                                        Dim.1      Dim.2       Dim.3      Dim.4
Troyanos.bancarios.m.viles           0.3679489  0.8906979 -0.23772334 0.1214847
Intentos.de.infeccion                0.6651570 -0.1386812  0.72508613 0.1121776
infectarse.mediante.Internet        -0.9715421 -0.1090121  0.07495133 0.1964805
PC.con.mayor.riesgo.de.infeccion.local 0.5856950 -0.5828910 -0.54978729 0.1222021
```

```
                                        Dim.1      Dim.2       Dim.3      Dim.4
Troyanos.bancarios.m.viles           0.3679489  0.8906979 -0.23772334 0.1214847
Intentos.de.infeccion                0.6651570 -0.1386812  0.72508613 0.1121776
infectarse.mediante.Internet        -0.9715421 -0.1090121  0.07495133 0.1964805
PC.con.mayor.riesgo.de.infeccion.local 0.5856950 -0.5828910 -0.54978729 0.1222021

$cos2
                                        Dim.1      Dim.2       Dim.3      Dim.4
Troyanos.bancarios.m.viles           0.1353864 0.79334269 0.056512384 0.01475854
Intentos.de.infeccion                0.4424338 0.01923247 0.525749901 0.01258382
infectarse.mediante.Internet         0.9438941 0.01188363 0.005617702 0.03860459
PC.con.mayor.riesgo.de.infeccion.local 0.3430387 0.33976190 0.302266069 0.01493336
```

Cada columna representa una combinación lineal (loadings) de las variables originales que proporcionan las componentes principales o factores. Así el primer componente se obtiene con la siguiente combinación:

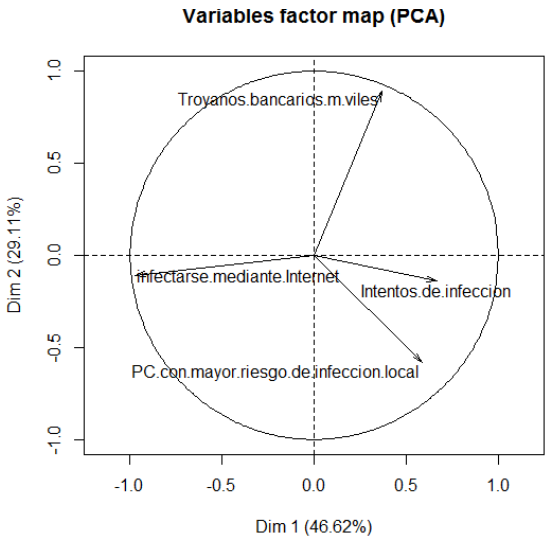
F1 = 0.3679489 troyanos+0.6651570 intentos de infección-0.9715421infectarse mediante internet+0.5856950 PC con mayor riesgo de infección

La determinación del número de factores a retener es, en parte, arbitraria y queda a juicio del investigador. Un criterio es retener los factores con valor propio superior a 1.

La carga del factor es la correlación existente entre una variable original y un factor, obtenido por combinación lineal de las variables originales.

Estas correlaciones se pueden calcular como producto de los coeficientes o loadings y las desviaciones de cada componente:

```
Component loadings:
                                  Comp.1     Comp.2      Comp.3    Comp.4
infectarse.mediante.Internet       0.7114608 -0.1010315  0.07944174 0.6908728
Intentos.de.infeccion             -0.4870948 -0.1285286  0.76852671 0.3944436
PC.con.mayor.riesgo.de.infeccion.local -0.4289048 -0.5402188 -0.58272555 0.4296921
Troyanos.bancarios.m.viles        -0.2694492  0.8254918 -0.25196556 0.4271695

Component variances:
    Comp.1     Comp.2     Comp.3     Comp.4
1.86475295 1.16422068 0.89014606 0.08088031
```

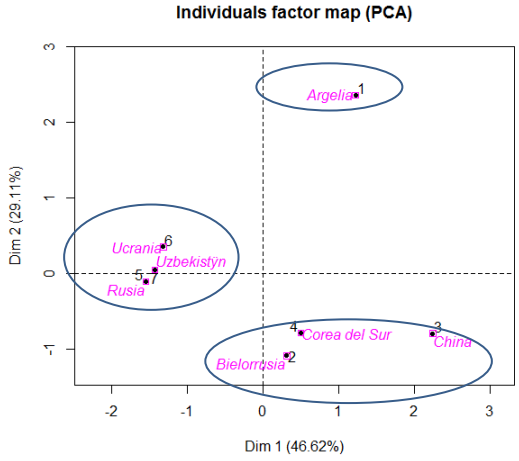


La comunalidad es un valor que se obtiene en el análisis factorial, para cada una de las variables originales, sumando los cuadrados de las correlaciones o cargas de los factores retenidos con la variable para la que se calcula y que expresa la proporción de varianza de la variable extraída o explicada con m factores, donde m es el número de factores retenidos. Si m es igual al número total de variables la comunalidad será igual a 1.

Los cosenos son las correlaciones al cuadrado y su acumulación proporciona las comunalidades.

Observamos también que existe 3 clúster

Se identifican los clústeres de los países. Como se puede observar, se marcaron los 3 clústeres o conglomerados. La siguiente tabla presenta la composición de los clústeres mencionados:



Clúster	Individuos
1	Ucrania, Uzbekistan, Rusia
2	Bielorrusia, Corea del Sur, China
3	Argelia

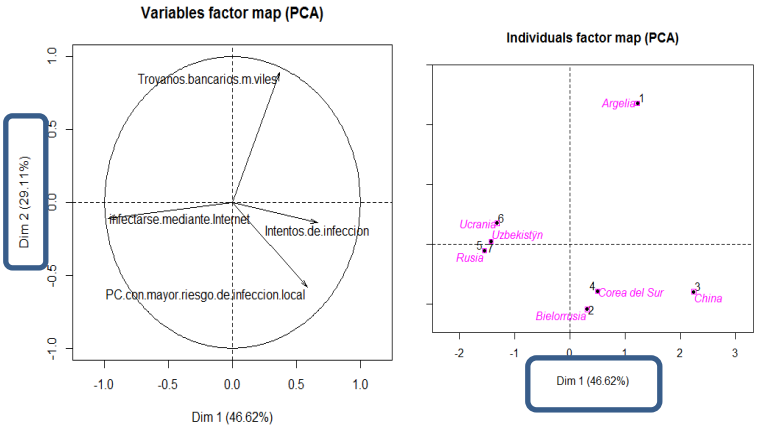
Ahora analizamos las variables del factor map (PCA), para ello utilizamos solo los 2 componentes por contar con buena variación

```
Component loadings:
                                  Comp.1     Comp.2
infectarse.mediante.Internet       0.7114608 -0.1010315
Intentos.de.infeccion             -0.4870948 -0.1285286
PC.con.mayor.riesgo.de.infeccion.local -0.4289048 -0.5402188
Troyanos.bancarios.m.viles        -0.2694492  0.8254918

Component variances:
    Comp.1     Comp.2     Comp.3     Comp.4
1.86475295 1.16422068 0.89014606 0.08088031
```

Y observando el círculo de correlaciones podemos concluir

- Que las “PC con mayor riesgo de infección local” con “intentos de infección” están cercanas entre sí, entonces son positivamente correlacionadas (si estuvieran estas aristas muy cerca significaría que serían fuertemente correlacionadas)
- También observamos que las “PC con mayor riesgo de infección local” con “intentos de infección” y “Troyanos bancarios móviles” es cercano a los 90° entonces no existe ninguna correlación entre ambas variables
- Mientras que las “PC con mayor riesgo de infección local” con “infectarse con internet” están casi opuestas al vértice entonces podemos afirmar que existe una correlación negativa, sin embargo entre las variables de “infectarse con internet” e “intentos de infección” existe una fuerte y negativa correlación.

Ahora aplicamos la Inercia Explicada: Proporción de la cantidad de inercia de cada eje con respecto a la inercia total Indica la importancia de las dimensiones representadas en cada uno de los ejes del mapa



En este caso la Inercia explicada es 46.62%+29.11%= 75.73%, significa que la dispersión de información del estudio se encuentra en 75.73%, es decir lo que se estudia está solo a un 75.73% de toda la información, lo que se está dejando de estudiar es 24.27%.

También se observa que la mejor dimensión es 75.73% que es una asociación entre las variables es altamente significativa.

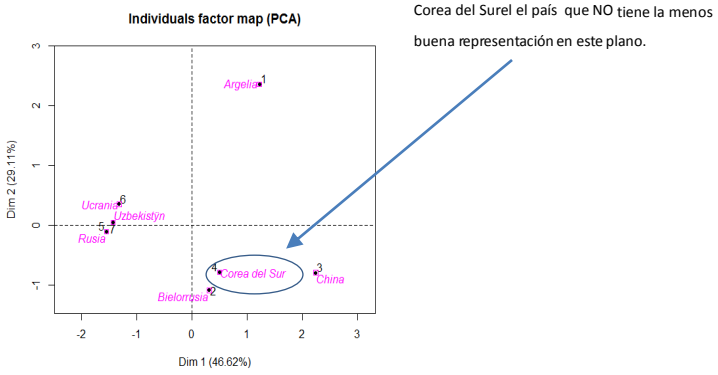
Mientras que para la calidad de representación de cada variable sobre el círculo de correlaciones, será medida con el coseno cuadrado del ángulo de la variable y su proyección, ahora bien entre las variables el coseno es igual a su correlación, por lo que las correlaciones cercanas a 1 son las que impliquen la calidad de la representación de las variables, es decir estarán bien representadas aquellas variables que queden ubicadas cerca de la frontera o borde del círculo de correlaciones

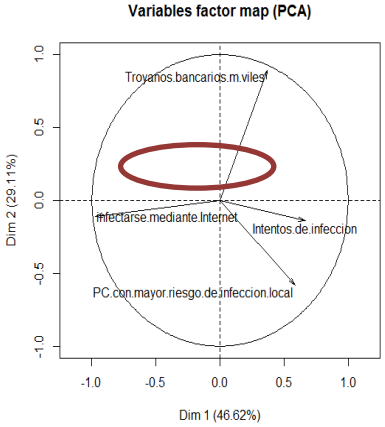
```
$cos2
        Dim.1       Dim.2
1 0.20382702 0.759107676
2 0.03741074 0.438206735
3 0.62259211 0.077120134
4 0.10061089 0.249761165
5 0.90228615 0.004177685
6 0.69827023 0.050726277
7 0.91098851 0.001233952

$contrib
        Dim.1       Dim.2
1 11.4861142 68.51738074
2  0.7730295 14.50323259
3 38.6688388  7.67204758
4  1.9109758  7.59838205
5 18.2612108  0.13542753
6 13.2310453  1.53953516
7 15.6687855  0.03399433
```

Entonces vemos de acuerdo al ángulo del coseno quevariables están cercanas a 1 como que también cuales están alejadas, que serán las que tiene menos representación en éste plano, para ello realizamos el cálculo:

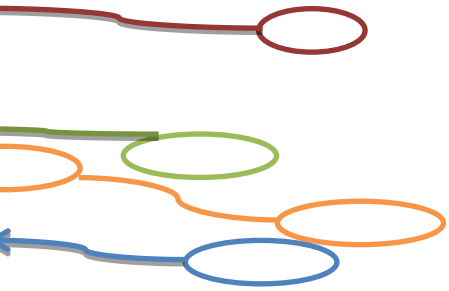
	Dim.1	Dim.2	Suma*100
1	0.2036	0.7591	96.27
2	0.3741	0.4362	81.03
3	0.6226	0.0712	69.38
4	0.1006	0.2498	35.04
5	0.9023	0.0042	90.65
6	0.6963	0.0507	74.7
7	0.9109	0.0012	91.21

La variable 4 que es del país Corea del Sur es la que tiene menos representación en este plano, mientras que los demás países si están bien representado en el plano.





Ahora realizamos la sobre posición de gráficos para relacionar los países con la modalidad de ataque de los virus informáticos



Se observa que el país Argelia es el más atacado por virus troyanos, mientras que los países como Ucrania y Uzbekistán son más propensos a infectarse con virus a través de internet, se nota también que Corea del Sur como China son más atacados por los virus con intentos de infección. Finalmente Bielorrusia es el país donde a través de sus PC´s tienen demasiado riesgo a contagiarse de virus.

Conclusiones:

- Se ha identificado a los países más vulnerables con las variables de los virus informáticos atacados o con intentos de infección ya sea por móviles o a Pc´s, información tomada de Kaspersky Lab del 2do Trimestre del año 2015:

País*	usuarios atacados por los troyanos bancarios móviles**	usuarios atacados por lo menos con un intentos de infección	usuarios que han estado bajo mayor riesgo de infectarse mediante Internet*	usuarios en los que los ordenadores de los usuarios han estado bajo mayor riesgo de infección local
Argelia	15525	1902.6669	1096221.626	613896.6169
Bielorrusia	61.179	1641.6365	1254670.298	670259.0588
China	50.9825	3332.2162	1052524.75	640693.6914
Corea del Sur	483.3141	1521.3178	1171456.247	661843.4486
Rusia	177.4191	1735.4443	1481134.107	608360.0312
Ucrania	59.1397	1710.9727	1358402.882	586988.8106
Uzbekistán	73.4148	1745.6408	1432497.584	600498.0796

- A través del método ACP observamos que existe 3 clúster o conglomerados.

Clúster	Países
1	Ucrania, Uzbekistan, Rusia
2	Bielorrusia, Corea del Sur, China
3	Argelia

- Las “PC con mayor riesgo de infección local” con “intentos de infección” están cercanas entre sí, entonces son positivamente correlacionadas (si estuvieran estas aristas muy cerca significaría que serían fuertemente correlacionadas)
- También observamos que las “PC con mayor riesgo de infección local” con “intentos de infección” y “Troyanos bancarios móviles” es cercano a los 90° entonces no existe ninguna correlación entre ambas variables
- Mientras que las “PC con mayor riesgo de infección local” con “infectarse con internet” están casi opuestas al vértice entonces podemos afirmar que existe una correlación negativa, sin embargo entre las variables de “infectarse con internet” e “intentos de infección” existe una fuerte y negativa correlación.
- Al ser la inercia explicada es 46.62%+29.11%= 75.73%, significa que la dispersión de información del estudio se encuentra en 75.73%, es decir lo que se estudia está solo a un 75.73% de toda la información, lo que se está dejando de estudiar es 24.27% pero es significativo por su concentración.
- De los 7 países estudiados, la variable 4 que es del país Corea del Sur es la que tiene menos representación según el plano, mientras que los demás países si están bien representado en el plano
- Se observa que el país Argelia es el más atacado por virus troyanos, mientras que los países como Ucrania y Uzbekistán son más propensos a infectarse con virus a través de internet, se nota también que Corea del Sur como China son más atacados por los virus con intentos de infección. Finalmente Bielorrusia es el país donde a través de sus PC´s tienen demasiado riesgo a contagiarse de virus

Definiciones

- Un **exploit** habla de un programa o código que se aprovecha de un agujero de
- Seguridad (vulnerabilidad) en una aplicación o sistema, de forma que un atacante podría usarla en su beneficio.
- El **ACP** es un método algebraico/estadístico que trata de sintetizar y dar una estructura a la información contenida en una matriz de datos.
- **R** es un lenguaje y un entorno diseñado para análisis estadístico. Fue desarrollado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993. R se distribuye bajo la licencia GNU GPL y está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.
- **Rattle** (del inglés, R Analytical Tool to Learn Easily) es un paquete de software libre desarrollado por Togaware que proporciona una interfaz gráfica basada en Gnome para la **minería de datos** desarrollada en el lenguaje de programación R.

Bibliografía y referencias:

http://support.kaspersky.com/mx/614

http://www.convosenlaweb.gob.ar/media/1204493/guiadeamenazas.pdf

http://recursostic.educacion.es/observatorio/web/en/listado-monograficos?start=2

[1] Jolliffe, I.T. (2002). Principal Component Analysis, (2nd edition) Springer-Verlag, New York. edition) Springer-Verlag, New York..

[2] J.M. Marín and M.T. Rodríguez-Bernal "Multiple hypothesis testing and clustering with mixtures of non-

central t-distributions applied in microarray data analysis". Computational Statistics and Data Analysis (2012).

[3] Santesmases, Miguel. Diseño y análisis de encuestas en investigación social y de mercados, Ed. Pirámide, Madrid, 2009.

[4] Peña, D. (2002). Análisis de Datos Multivariantes, Mc-Graw Hill, Madrid.