Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Telecomunicación

JUEGOS DE GUERRA

IMÁGENES PARA LA BATALLA EN EL CIBERESPACIO

TRABAJO FIN DE MÁSTER

José Luis AZNAR LAHOZ

2017

Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Telecomunicación

Máster Universitario en
Ingeniería de Redes y Servicios Telemáticos

TRABAJO FIN DE MÁSTER

JUEGOS DE GUERRA

IMÁGENES PARA LA BATALLA EN EL CIBERESPACIO

Autor
José Luis AZNAR LAHOZ

Director
PhD. Víctor A. VILLAGRÁ

Departamento de Ingeniería de Sistemas Telemáticos

2017

Resumen

En la era de las tecnologías, el elevado volumen de datos, provenientes de muy diversas fuentes y de muy diferente naturaleza, provoca que su tratamiento de forma manejable para generar información, suponga todo un reto. Los datos deben ser agregados y analizados para que constituyan y proporcionen verdadera información.

Esta información se debe presentar a la persona responsable de la toma de decisión de la empresa, de manera fácil y amigable, para que sea capaz de interpretarla, alcanzando "*conocimiento*", y en base a la "*comprensión*" de su significado, obtener conclusiones que le permitan tomar la decisión más apropiada a las circunstancias del momento, proyectando los eventos y dinámicas actuales a un futuro cercano, anticipando sucesos posteriores y sus implicaciones. Este estado mental es el que se conoce como "*situational awareness*" o conciencia situacional.

La información se puede proporcionar de diversas formas, pero el cerebro humano, como mejor recibe esta información, y la interpreta de una forma sencilla e intuitiva, es utilizando gráficos. La característica principal, que debe mostrar el grafico, es la de sintetizar los elementos útiles de los datos agregados, generando imágenes convencionales que proporcionen información. A raíz de los atentados yihadistas del 11-S, el Departamento de Seguridad Nacional de los Estados Unidos abrió una línea de investigación científica, denominada *Visual Analytics,* con la que, a partir de grandes volúmenes de datos, se ayudara a analizar la información a través de interfaces gráficas, facilitando el combate contra el terrorismo.

Las actividades que suceden en el entorno del ciberespacio, son también susceptibles de representarse a través de gráficos, en consonancia con los procedimientos de Visual Analytics. Mediante este método de exposición, la persona que debe tomar decisiones en este ámbito, obtiene información de la situación actual para alcanzar una conciencia situacional que le permita tomar la mejor decisión, de modo que, provoque cambios en el futuro cercano favorables al negocio propio. Debido a que las operaciones en el entorno del ciberespacio son de muy diversa índole, a la vez que dispares, las técnicas de visualización son muy diferentes para cada uno de los hechos que representan. Técnicas graficas como los símbolos, mapas de colores, diagramas de dispersión, histogramas son sencillas de realizar y de interpretar, pero existen otras más sofisticadas y complejas como los diagramas de coordenadas paralelas, mapas de árbol, gráficos de anillo y flujo de conexiones.

Cada una de estas representaciones graficas tiene la capacidad de simbolizar hechos, acciones, sucesos o circunstancias muy diferentes, desde el flujo de bits de una comunicación, hasta la gestión financiera en materia de seguridad de la información de una empresa, pasando por la arquitectura de sus sistemas de información y telecomunicaciones. Intrínsecamente, cualquiera de estos gráficos presenta características, que en función del tipo y volumen de datos que simbolizan, muestran ciertas ventajas, pero también ofrecen algunos inconvenientes. Dentro del conjunto de inconvenientes, y debido al elevado volumen de datos qué son susceptible de ser representados, se encuentra el problema de distorsionar la imagen, convirtiéndola confusa, llegando a ser caótica e ininteligible, por lo que hay que proporcionar diversas soluciones, resaltando o seleccionando la información requerida.

En algunos casos, diferentes grafismos pueden representar el mismo suceso, por lo que es necesario estudiar y analizar las ventajas e inconvenientes de cada una de ellas y elegir el más adecuado, sin olvidar que la facilidad para su comprensión por la mente humana puede ser un factor determinante para su elección.

Abstract

In the technological age, the large volume of data, coming from many different sources and nature, makes its treatment and analysis a challenge. The data must be aggregated and analyzed to constitute and provide useful information.

This information must be presented to the decision bodies, in an easy and a friendly way, so that it is possible to interpret it, reaching "knowledge". Based on the "understanding" of its meaning, those bodies can draw conclusions that allow them take the most appropriate decision to the circumstances of the moment, projecting the current events and dynamics to a near future, anticipating future events and their implications. This state is known as “situational awareness”.

Information can be provided in different formats, but the best way human brain receive and interprets information is by using graphics. The main characteristic for a graph, is that it can synthesize the most important elements of the aggregated data, generating conventional images that provide useful information. Following the 9/11 attacks, the US Department of Homeland Security began a line of scientific research, called Visual Analytics, which, through large volumes of data, would help to analyze the information by graphic interfaces, providing the fight against terrorism.

Activities that occur in the cyberspace environment are also likely to be represented by graphs, in line with Visual Analytics procedures. Using graphs decision bodies can obtain information of the current situation reaching a situational awareness that facilitates the decision making process for the business benefits. As the operations in the cyberspace environment are very diverse, visualization techniques should be different for each of the type of events they represent. Graphical techniques such as symbols, color maps, scatter plots, histograms are simple to implement and analyze, but there are also more sophisticated and complex ones such as parallel coordinate plots, treemaps, ring graphs and connection river diagrams.

Each of these graphical representations has the capacity to symbolize very different facts, actions, events or circumstances, like communication data flow, company information security financial management, information systems and communications architectures, etc. Intrinsically, any of the types of graphs previously mentioned has characteristics, which, depending on the type and volume of data they

symbolize, show certain advantages, but also offer some drawbacks. Within the set of disadvantages, and due to the large volume of data that can be represented, there is the problem of distorting the image, making it confusing, chaotic and unintelligible, so it is necessary to provide different solutions, highlighting or selecting the more useful information.

In some cases, different graphics may represent the same event, so it is necessary to study and analyze the advantages and disadvantages of each one and choose the most appropriate, keeping in mind that the ease for their understanding by the human mind can be a determining factor for its choice.

Índice general

Índice de figuras

Índice de tablas

Siglas

APT	Advanced Persistent Threat
ASRS	Aviation Safety Reporting System
C&C	Command and Control
CCN-CERT	Centro Criptológico Nacional - Computer Emergency Response Team
CEO	Chief Executive Officer
CIO	Chief Information Officer
COP	Common Operational Picture
CPD	Centro de Proceso de Datos
CPU	Central Processing Unit
DDoS	Distributed Denial of Service
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
DoS	Denial of Service
FTP	File Transfer Protocol
GPS	Global Positioning System
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
I+D	Investigación y Desarrollo
ICANN	Internet Corporation for Assigned Names and Numbers
IDS	Intrusion Detection System
INCIBE	Instituto Nacional de CIBErseguridad
IP	Internet Protocol

IPS	Intrusion Prevention System
ISP	Internet Service Provider
MPG	Miles Per Gallon
NFS	Network File System
NHS	National Health Service
NVAC	National Visualization and Analytics Center
PIB	Producto Interior Bruto
QoE	Quality of Experience
QoS	Quality of Service
RAT	Remote Access Tools
SA	Situational awareness
SMTP	Simple Mail Transfer Protocol
SSH	Secure SHell
SSL	Transport Layer Security
TCP	Transmission Control Protocol
TELNET	TELecommunication NETwork
UDP	User Datagram Protocol

1 Introducción.

La gestión de la información y el conocimiento del negocio es, actualmente, una actividad estratégica para el éxito de cualquier empresa en la consecución de los objetivos planificados. Sin embargo, el análisis de esta información, constituida por datos normalmente heterogéneos y no estructurados, introduce un punto de complejidad que debe ser solventada en aras de tomar la decisión más adecuada para el logro de los mismos.

El analista interpreta los datos de tal forma que obtiene la información que de ellos se puede extraer. Pero no es este actor, dentro de la cadena de decisión, el que evaluará esta información. Es el director del equipo el que recabará de los diferentes analistas, de las diferentes áreas, de los diferentes conocimientos, toda la información que se pueda inferir de los datos, y así, tomar la decisión más oportuna.

De la adecuada interpretación que se realice, condicionada por la correcta representación que se haga a la dirección de esa información, dependerá la bondad de la decisión adoptada para lograr los objetivos de la empresa. Esa información debe ser clara y nítida, además de poderse interpretar de la forma más intuitiva posible.

La persona o personas responsables de adoptar las decisiones estratégicas del negocio, no serán expertos en la materia, más exactamente expresado, en todas las materias involucradas en la consecución de los objetivos marcados. Para adoptarlas, en tiempo real, deberán tener conocimiento de la situación actual y de las posibles evoluciones de las variables en el tiempo para alcanzar situaciones futuras.

En los niveles estratégico y operacional de las operaciones militares, a modo de ejemplo, por ser el ámbito que más profundamente es conocido por el autor, el jefe militar debe coordinar las operaciones en curso. Evidentemente, es imposible que en el moderno entorno operativo multidisciplinar, el jefe militar sea experto en todas las materias que influyen en el normal desarrollo de la acción.

Sin embargo, las decisiones del jefe no se basan directamente en la información, ni tan siquiera del conocimiento que posea sobre la situación. En realidad, las decisiones se basan en la comprensión de la situación, comprensión que es resultado de ciertos procesos humanos y, por ello, intrínsecamente subjetivo [1]. Es evidente, que aun poseyendo la misma información, dos personas distintas pueden tomas decisiones distintas. Aun más, la misma persona puede tomar decisiones distintas en

circunstancias distintas. Ante esta realidad, es de una importancia capital que la información sea presentada al jefe de una forma clara, precisa y concisa, reflejando fielmente los datos agregados.

Vivimos en lo que se ha denominado la "era de la información". En muchos ámbitos, esto ha significado un gran aumento en los sistemas y tecnologías [2]. La globalización de la información ha supuesto que cualquier dato puede ser consultado en cualquier lugar, en cualquier instante, suponiendo un exceso de información. Dentro de nuestras empresas, informes y formularios se han multiplicado y cada aspecto del negocio queda registrado en alguna parte. Tratar toda esta información globalmente de forma manejable es todo un reto. Simplemente hay más información de la que cualquier persona puede manejar.

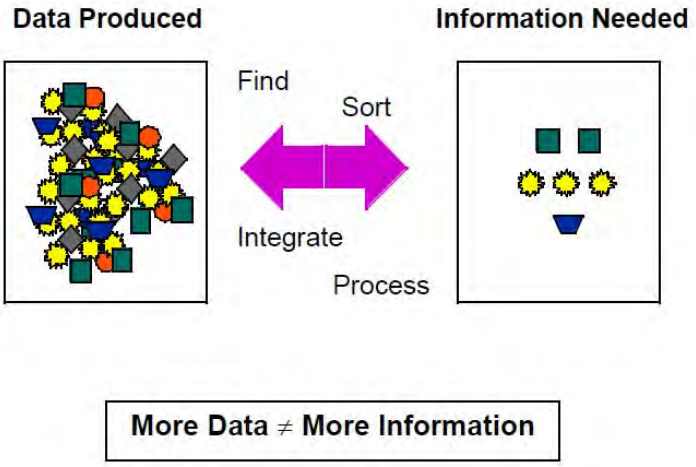


Figura 1. La brecha entre los datos y la información [2].

¿Información?. Atrevernos a denominarla de esa forma quizás sea una temeridad. Se trata de multitud de datos, de muy diversas fuentes y de muy diferente naturaleza, que deberán ser agregados y analizados para que constituyan y proporcionen verdadera información (figura 1).

Por otro lado, no menos importante que "extraer" la información de los datos, es el modo en como esa información es presentada al jefe militar, de tal forma que, de una manera fácil y amigable, sea capaz de interpretarla y en base a las conclusiones que obtenga, tomar la decisión más apropiada a las circunstancias del momento.

Actualmente, los jefes de las fuerzas tienen a su disposición modernos sistemas de telecomunicaciones e información, con los que les resulta posible manejar los grandes volúmenes de información que necesitan para tomar sus decisiones [1].

El proceso de transformación de la información dentro de la jerarquía cognitiva [1], queda representado mediante una pirámide escalonada (figura 2), donde:

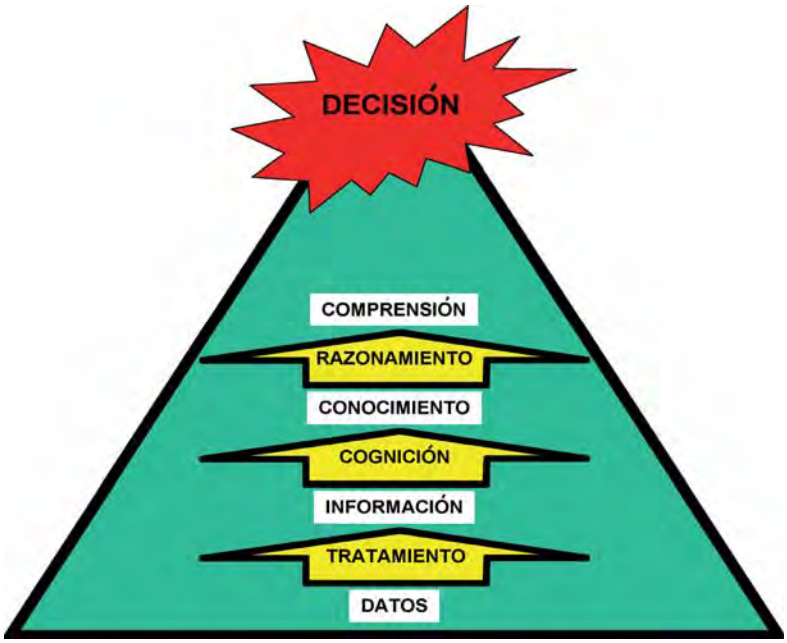


Figura 2. La jerarquía cognitiva [1].

- los *datos*, escalón inferior de la citada jerarquía del conocimiento, son señales o indicaciones aisladas obtenidas del entorno, por ejemplo, por medio de un sensor, que no resultan útiles por sí mismas para ser explotadas, salvo que, a través de un adecuado tratamiento, se les pueda otorgar un significado. Un dato puede ser la coordenada geográfica de una posición de ametralladoras.
- la *información*, constituye el siguiente peldaño, conformada por datos de muy diferente naturaleza que han sido procesados y agregados mediante un proceso de inteligencia que pretende proporcionarle un significado conjunto. El tratamiento de los datos incluye su filtrado, fusión, formateado, organización, correlación, categorización y ordenamiento. Las posiciones (coordenadas geográficas) de diferentes unidades, fotografías aéreas de la zona, conversaciones radio captadas acerca de los movimientos a realizar, pueden conformar diferentes tipos de información de las acciones en curso de ese citado enemigo.
- El *conocimiento* es información que ha sido analizada y evaluada. El análisis aporta significado, la evaluación permite determinar la precisión o certeza de la información, su oportunidad, utilidad e integridad. Así, siguiendo con el ejemplo propuesto, las diferentes informaciones citadas en el apartado anterior, pueden ser integradas con otras informaciones, como

por ejemplo el modo de empleo (doctrina) del enemigo, permitiendo inducir la composición de una fuerza y su disposición para el combate. A partir de este escalón de la jerarquía cognitiva, se obtiene ya un producto utilizable para preparar decisiones, ya que permite deducir interrelaciones y obtener conclusiones.

- Finalmente, en el escalón más elevado de la pirámide cognitiva se encuentra la *comprensión*. Esta se deriva del conocimiento que ha sido sintetizado y sobre el que se han aplicado razonamientos que ayudan a comprender las relaciones intrínsecas de una situación. A partir de aquí es posible saber lo que está sucediendo y por qué, extrapolar significados y hacer inferencias basadas en patrones reconocibles, anticipar consecuencias de las acciones tanto propias como del adversario. La comprensión es una abstracción que, normalmente, se mejora con la colaboración. A partir del conocimiento adquirido mediante las informaciones obtenidas, se pueden inferir las relaciones por las que se ha llegado al momento actual, permitiendo al jefe militar evaluar las diferentes alternativas de actuación mediante un proceso de reflexión, eligiendo la más adecuada.

Todos los procesos referidos conducen al jefe militar, en realidad, a cualquier director, de cualquier organización, que tenga que tomar una decisión, a un estado mental de comprensión de la situación y la proyección de la misma a un estado futuro cercano. Este estado mental es el que se conoce como *"situational awareness"*.

2 Situational awareness en la pirámide del conocimiento.

Conciencia situacional, o en su denominación en ingles, "*situational awareness*", es un estado mental que debe alcanzar aquella persona que toma decisiones en una organización, como consecuencia del entorno en el que se desarrollan las acciones, la evolución de los acontecimientos y otros factores que puedan afectar a su progreso. Intuitivamente se puede percibir la conciencia situacional como "ser conscientes de lo que sucede en nuestro entorno". Mica Endsley, Presidenta de SA Technologies, Inc. y uno de los más afamados diseñadores de sistemas SA y de toma de decisión, define la conciencia situacional de un modo más formal como "*la percepción de los elementos existentes en el entorno en un volumen de tiempo y espacio, la comprensión de su significado, y la proyección de su estatus en el futuro cercano*" [3]. Existen otras definiciones [4] [5], pero esta es la más aceptada por el mayor número de autores [6] [7] [8] dentro de la comunidad científica.

De acuerdo a la teoría de la PhD Endsley, la percepción, la comprensión y la proyección son los tres pilares sobre los que se sustenta el modelo mental que constituye la consciencia situacional en un esquema jerárquico de niveles [9]:

Nivel 1. Percepción: El reconocimiento básico de evidencias es esencial. Sin una correcta captación de la información fundamental, las posibilidades de componer una imagen errónea de la situación se incrementan drásticamente. Investigaciones realizadas [10], manejando informes de la base de datos Aviation Safety Reporting System (ASRS), llevaron a la conclusión que un 76,3% de los errores de conciencia situacional podrían atribuirse a problemas en la percepción de la información básica, independientemente que fueran producidos por fallos del sistema o por dificultades en el proceso cognitivo.

Nivel 2. Comprensión: La conciencia situacional va mas allá de la mera percepción. También engloba cómo los individuos agregan, interpretan, almacenan y retienen la información. Por tanto, incluye algo más que prestar atención o comprender la información, implica también la integración de múltiples elementos de información y la determinación de su relevancia para la consecución de los objetivos. Realizando una analogía con la lectura, implicaría tener un alto nivel de comprensión de lectura, no solamente comprender el significado de las palabras individualmente. Se ha determinado que un 20.3% de errores en conciencia situacional están comprendidos dentro de este nivel 2 [10].

La evaluación de la *"relevancia"* o lo que realmente importa tiene dos componentes: una componente de interpretación subjetiva (conciencia) y otra de significado objetivo o importancia (situación) [8]. Un sujeto en el nivel 2 SA ha sido capaz de descubrir el significado y la importancia operacional de los datos obtenidos del nivel 1.

Nivel 3. Proyección: Este es el nivel más alto de consciencia situacional. Implica la mayor cota de comprensión de la situación y la capacidad de prever sucesos futuros. Esta capacidad de proyectar los eventos y dinámicas actuales para anticipar eventos futuros (y sus implicaciones) permite una toma de decisiones oportuna. En casi todas las materias estudiadas (aeronaves, control de tráfico aéreo, funcionamiento de plantas de energía, mantenimiento, medicina), se ha encontrado que los operadores experimentados confían en las proyecciones futuras en gran medida. Es la impronta que deja un experto en la materia.

Cuando la persona que tiene que tomar la decisión, en función de las metas y objetivos de la empresa, habiendo obtenido los datos correctos, cuantitativa y cualitativamente (nivel 1), alcanzando la comprensión, a través de los mismos, de la situación actual (nivel 2), es capaz de proyectar los valores actuales que definen la situación a un futuro inmediato (nivel 3), habrá adquirido una conciencia situacional completa (figura 3). En estas condiciones estará en situación de adoptar la decisión

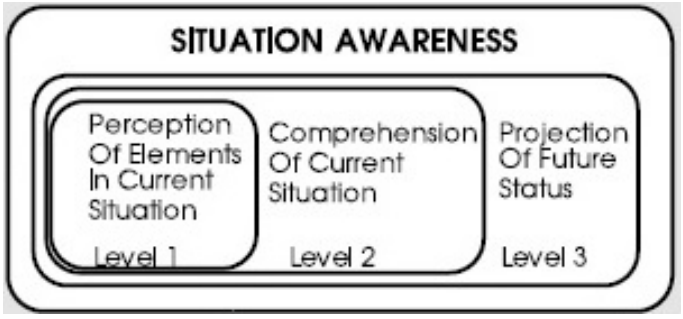


Figura 3. Niveles de Situation Awareness [11].

más correcta. Estas decisiones conllevan la materialización de acciones, provocando una modificación en el estado del entorno, conocida como retroalimentación o "feedback" (figura 4). Esta modificación del entorno desencadena una variación de la conciencia situacional, lo que le confiere un carácter dinámico, provocando que la persona que debe tomar la decisión se vea obligada a revisar sus modelos conceptuales de la situación.

En el modelo propuesto por Endsley (figura 4) la conciencia situacional es un estadio distinto de la toma de decisión y de su propia ejecución [9]. Como se mencionaba en la introducción, la persona que toma la decisión basa la misma en la comprensión que de la situación pueda tener en el momento de adoptarla, no únicamente, de los conocimientos que del negocio posea. Sin embargo, la comprensión

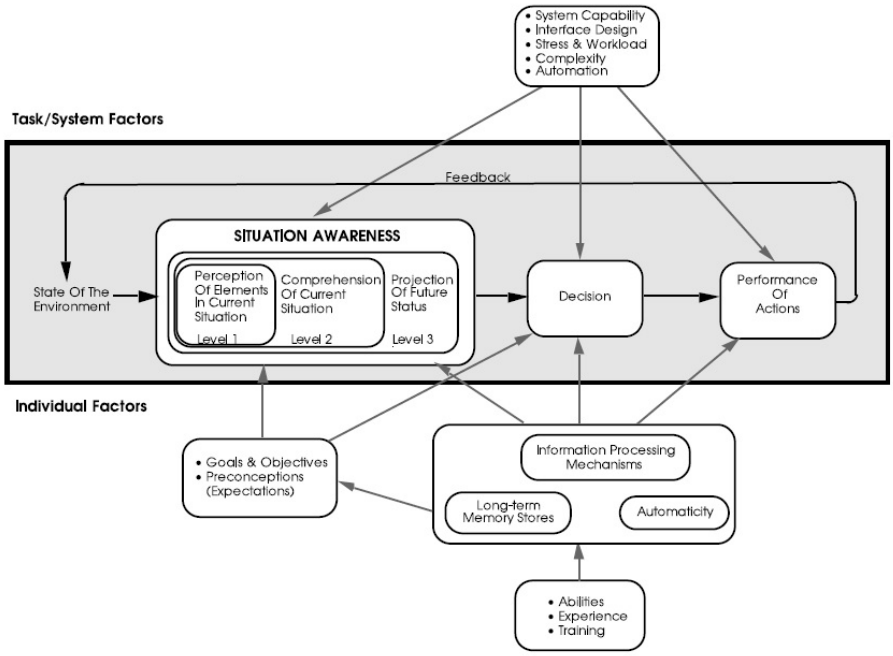


Figura 4. Modelo de "Situational Awareness" para sistemas dinámicos de toma de decisión [12].

de la situación depende del conocimiento que se tenga del entorno y de los patrones de comportamiento que se puedan detectar. Esto es lo que se conoce como experiencia. Esta permite relacionar situaciones actuales con situaciones previas, vividas anteriormente, generando conocimiento actualizado. Sin embargo, si el conocimiento adquirido nuevo se basa en un alto porcentaje en la experiencia, dejando a un lado la situación del momento actual, se corre el riesgo de automatización [11].

La automatización [13] está fuertemente asociada a la memoria a largo plazo, en la cual se guarda el conocimiento y los patrones de comportamiento del entorno. La memoria a corto plazo, por otra parte, es también llamada memoria de trabajo y permite guardar y manipular temporalmente la información captada por los sentidos.

Las personas poseen una capacidad limitada de atención que se pueden asignar a tomar y procesar la información ambiental [14]. También tienen un volumen limitado

de memoria de trabajo de un sistema para el procesamiento y retener la información que se percibe.

El volumen de la memoria de trabajo puede incrementarse a través del entrenamiento, no obstante, está condicionada por el estrés y la sobrecarga cognitiva derivadas de la complejidad y dinamismo del entorno [13].

En entornos dinámicos [14], el desarrollo de la conciencia de situación y el proceso de decisión están restringidos por la limitada capacidad de atención y el volumen de memoria de trabajo de las personas principiantes y de todas aquellas que se encuentran ante situaciones nuevas. Es necesaria la intervención directa, para percibir y procesar el ambiente, y así formar SA, para la selección de acciones y ejecución de respuestas. En entornos complejos y dinámicos, la sobrecarga de información, complejidad de la tarea, y multiplicidad de tareas puede superar rápidamente la limitada capacidad de atención de una persona.

Debido a que la fijación de atención es limitada, dedicar más atención a alguna información puede significar una pérdida de SA en otros elementos [14]. La falta de SA resultante puede dar lugar a malas decisiones que conducen a resultados no deseados.

Con la experiencia, sin embargo, las personas desarrollan mecanismos que pueden superar estas limitaciones [14]. Estos son:

1. generación de expectativas,
2. modelos mentales y esquemas,
3. procesamiento dirigido a objetivos, y
4. automatización.

En la práctica, las personas experimentadas, responsables de la toma de decisiones, son capaces de utilizar almacenes de memoria a largo plazo, muy probablemente en forma de esquemas (prototipos de situación) y modelos mentales, para eludir estos límites de situaciones y entornos. Estos mecanismos proporcionan una guía sobre las características críticas que deben ser atendidas del entorno y para la integración y la comprensión de la información y la proyección en los hechos futuros, ya sea directamente, o a través de prototipos de situación. También facilitan la toma de decisiones sobre la base de información incompleta y en condiciones de incertidumbre.

El desarrollo de los modelos mentales es extremadamente importante para la SA. El uso de modelos mentales proporciona, también, información por defecto para las personas que deben tomar decisiones. Estos valores por defecto (características esperadas de elementos en función de su clasificación) permiten a las personas predecir el comportamiento, aunque la información sea incompleta o incierta.

Pequeños cambios en los valores de incertidumbre pueden producir cambios muy considerables en las conclusiones adoptadas.

Hasta este instante se ha visto como los datos, desde su adquisición en estado crudo, son tratados para generar conocimiento y su posterior comprensión, y cuál es la situación mental a la que se debe llegar, para tomar decisiones que sean las más apropiadas, permitiendo alcanzar los objetivos fijados. Pero ¿cómo es la integración que se produce entre ambos conceptos? ¿qué procesos se generan entre los datos obtenidos del entorno, hasta conseguir una proyección a futuro de la situación actual?

El nivel de percepción, nivel 1, en la teoría de situational awareness, que como ya se ha mencionado, está relacionado con la captación de las evidencias del ámbito de decisión, evidencias que se corresponden con las indicaciones o señales obtenidas del entorno que hemos denominado datos, dentro de la jerarquía cognitiva (figura 5). Ambos son el escalón inferior dentro de sus respectivos campos de investigación, y de la misma forma que los datos no aportan conocimiento y no son explotables por si mismos si no han sido tratados, como se expuso en el apartado 1, la percepción no aporta consciencia plena de la situación, por lo que, la toma de decisiones, basada únicamente en este nivel, no está libre de errores groseros. Sin embargo, en los dos casos, la correcta adquisición de ambas ideas o conceptos permite que los siguientes estadios, en ambas áreas del saber, se completen correctamente.

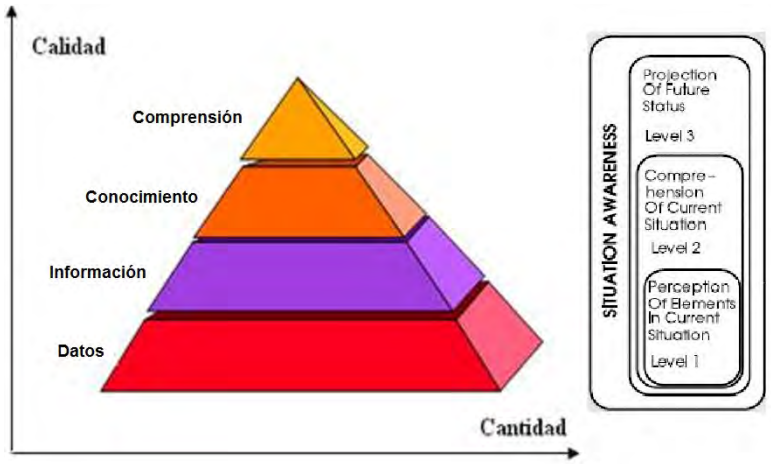


Figura 5. Relación entre la pirámide del conocimiento y Situational Awareness.

El siguiente nivel, nivel 2 de comprensión de la situación actual, de la teoría de situational awareness, abarca los niveles de información y conocimiento de la pirámide cognitiva (figura 5). En ellos los datos son procesados proporcionando un significado,

que analizado y evaluado aporta conocimiento, conocimiento que permite comprender la situación actual, finalidad del nivel 2 de situational awareness. En ambos casos, se alcanza una capacidad cognitiva de la situación actual, mediante la interiorización de la información, transformándola en conocimiento, conocimiento que permite tomar decisiones fundamentadas en la comprensión subjetiva que de la situación tenga la persona que tenga que tomar la decisión.

El nivel de comprensión de la pirámide cognitiva, en el que el conocimiento ha sido extractado, permite alcanzar razonamiento intrínseco de la situación. En este punto, se es capaz de saber que está sucediendo y extrapolar consecuencias de las decisiones adoptadas a futuro, relacionado directamente con el nivel de proyección de situational awareness (figura 5) en el que se pueden anticipar eventos futuros.

La finalidad de este trabajo es la de estudiar las diferentes formas que existen de representación de la actividad en el ciberespacio y relacionarlo con los incidentes de seguridad que ocurren en el mismo, de tal forma que permita presentar, a la persona comprometida con la toma de decisiones, estos incidentes de una forma clara e intuitiva, permitiéndole tener una conciencia de la situación actual, y que le proporcione una proyección, en el futuro, de las decisiones actuales adoptadas, o expresándolo en términos científicos, proporcionar una **Common Operational Picture (COP)** para alcanzar una **Situational Awareness (SA).**

3 Valoración de la necesidad de una representación gráfica.

Desde el inicio del nuevo siglo hasta el momento actual, el aumento de máquinas conectadas a través de las redes ha crecido en forma polinómica. Y en la última década, con el auge de la computación móvil, el crecimiento es exponencial. Cuando el Internet de las cosas sea una realidad el número de equipos conectados a redes será desorbitado. Este es, precisamente, uno de los problemas, entre otros, que IPv6 pretende solucionar, el número de direcciones IP posibles para todos estos equipos, llegando a ser tan elevado como número de estrellas hay en el universo. Con todos estos elementos conectados, el tráfico generado crece proporcionalmente al número de máquinas que lo origina. Evidentemente, esta cantidad de tráfico no se puede analizar simplemente leyendo logs [15]. Se hace necesaria la monitorización automática de las redes en busca de posibles comportamientos anómalos de sus componentes y representar los sucesos de tal forma que permitan alcanzar una *"situational awareness"* en diferentes niveles de decisión.

En este sentido, y antes de relacionar las diferentes técnicas de representación, conviene analizar algunos puntos a considerar en la representación de información y que Kintzel, Fuchs & Mansmann exponen en su artículo "Monitoring Large IP Spaces with ClockView" [15] donde presentan la herramienta ClockView.

Como ya se ha indicado, la finalidad de la monitorización automática es la de detectar patrones anómalos de comportamiento de cada uno de los elementos componentes. Esta detección solo es posible si se dispone de un estado aceptable de referencia. Definir este estado de referencia resulta ser una tarea complicada, que requiere determinarla para cada sistema, por ser innata a él, dependiendo de las características físicas y lógicas del mismo. Para que esta referencia sea eficaz y proporcione información útil en tiempo y forma, deberá entregarse junto con la información de trabajo. De esta forma, el analista podrá determinar, a la simple inspección del gráfico, que el comportamiento del sistema es anómalo. Esto proporcionara una indicación, conducente a la realización de un análisis más profundo, quizá llegando, y siendo apropiado en este instante, a la lectura de logs.

Otro aspecto importante a considerar es la escalabilidad de los elementos a representar. Es necesario que el sistema grafico de representación sea capaz de representar los eventos desde unos pocos ordenadores, en realidad desde una única

máquina, hasta miles de hosts conectados en red y el tráfico que estos generan. Se hará necesario, de alguna forma, agregar los datos de tal forma que facilite una idea general del estado de estos miles de hosts mediante un diagnóstico rápido, preferentemente visual.

Al hilo de esta característica, y mucho más importante que las anteriores, es que el sistema de representación esté capacitado para proporcionar información valida en muy diferentes escalas de decisión. Propondremos una analogía a modo de ejemplo, que proporcione una mayor comprensión del concepto expresado. Si observamos el planeta Tierra desde el espacio, obtenemos una visión global del mismo (figura 6), que permite tener una conciencia de las formas en un modo general, sin poder apreciar el detalle. La percepción que el observador obtiene es de una forma "redonda", pero con un terreno plano en la superficie.

Figura 6. Visión global desde el espacio del planeta Tierra.

A medida que nos acercamos, obtenemos un mayor detalle de los elementos (figura 7), pero perdiendo la perspectiva general. Se empieza a desvanecer la percepción esférica del planeta, pero se aprecia el relieve del terreno. No es que

sean "datos falsos", sino que la percepción de la información proporcionada indica que estas sean las formas del planeta.

Por último, cuando el observador se posiciona sobre el terreno, la percepción que obtiene es que la tierra es plana en su conjunto, con elevadas formas sobre su superficie (figura 8). ¿Cuál de todas estas "informaciones" es falsa? O expresándolo en modo positivo, ¿cuál es la verdadera?

Figura 7. Visión a vista de pájaro del relieve del planeta tierra.

Afortunadamente, ahora sabemos que todas son ciertas, únicamente cambia la forma de presentar la información. A diferentes niveles, la "información es distinta", aunque los datos son siempre los mismos.

Esto nos da idea que una información presentada de forma incorrecta, o para niveles distintos a los que corresponden, puede llevar a la persona hacia una conciencia de la situación errónea, con los consiguientes problemas, errores, que esta conllevaría en la toma de decisión.

Figura 8. Visión sobre el terreno del relieve del planeta tierra.

Continuando con la analogía, si nuestra intención es proporcionar información de la cantidad de hosts que pueden estar comprometidos por una intromisión, quizá no sea lo más adecuado representar cada una de estas máquinas, cuando el número de ellas pueden ser miles

(figura 9). La cantidad de símbolos representados sobre el tablero de gráficos puede impedir que la persona encargada de tomar la decisión no alcance una

Figura 9. Representación de miles de equipos comprometidos.

conciencia situacional adecuada. Parece, en este caso, más razonable jugar con formas geométricas, iconos y colores que indiquen diferentes cantidades en función de códigos acordados (figura 10). Esto puede proporcionar una idea más adecuada de la trascendencia de la situación, sin ofuscar el grafico.

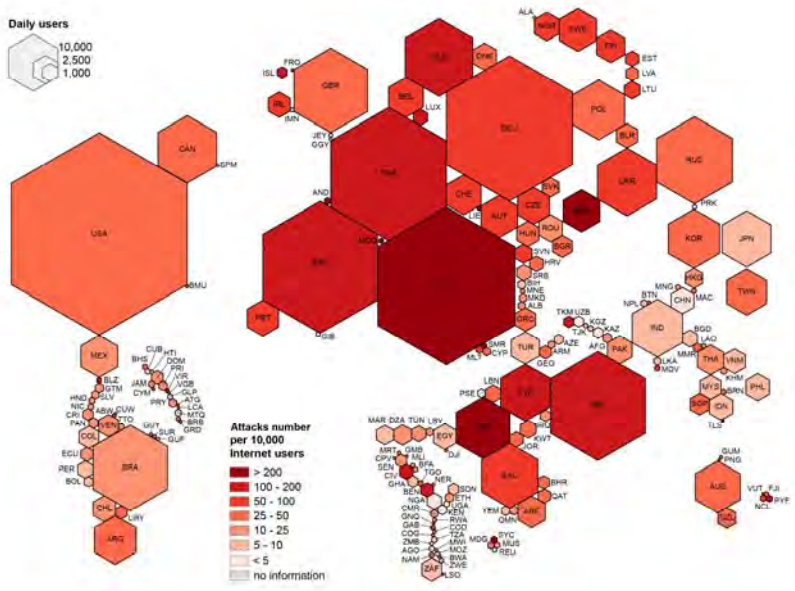


Figura 10. Representación esquemática de miles de equipos comprometidos.

En esta segunda representación (figura 10), se alcanza un mayor conocimiento de las dimensiones de las redes comprometidas (tamaño de la figura) y del grado de uso que de esta se hace (color de la figura). Estos datos no se aprecian en la primera representación (figura 9). De esta sencilla forma, se ha pasado de un conjunto de datos (cada uno de los hosts comprometidos) que no proporcionan información y que no son explotables por sí mismos, a una información sintetizada a través de un proceso de análisis.

Los atentados terroristas del 11 de septiembre de 2001 a las Torres Gemelas, en Nueva York, y al edificio del Pentágono en Washington, pusieron en evidencia que a pesar de la gran capacidad alcanzada en la actualidad para recopilar datos, gracias al avance de las tecnologías de la información destinadas a proporcionar una ventaja técnica que ayude a combatir el terrorismo, la capacidad de analizar la información es muy escasa. La información es masiva, compleja, incompleta e incierta, y abarca todas las formas de datos, idiomas y culturas. Se necesitan tecnologías que apoyen la aplicación del juicio humano para hacer el mejor uso posible de esta información y compartirla con otras personas de modo apropiado para prevenir, disuadir y responder a las amenazas [16].

En 2004, el Centro Nacional de Visualización y Análisis (National Visualization and Analytics Center, NVAC), recibió el encargo del Departamento de Seguridad Nacional de Estados Unidos, de desarrollar un programa de I+D que permitiera abordar la creación de herramientas que aportaran capacidad de análisis visual, suministrando una guía técnica de coordinación entre la investigación estatal y privada. Así nace una nueva disciplina de investigación científica, denominada Visual Analytics, definida por Thomas & Cook como *"la ciencia del razonamiento analítico apoyada por interfaces visuales interactivas"* [16], y ampliada por Keim, D. et al como: "*Visual Analytics combina técnicas de análisis automáticas con visualizaciones interactivas para una comprensión efectiva, razonamiento y toma de decisiones sobre la base de conjuntos de datos muy grandes y complejos*" [17].

Visual Analytics, es una rama científica multidisciplinar, que abarca campos tan dispares como la física, la astronomía, la gestión de emergencias, la biología, la medicina y la gestión empresarial [17], donde se trabaja sobre la base de enormes volúmenes de datos, de naturaleza dinámica y a veces ambiguos, con la finalidad de crear herramientas visuales, capaces de sintetizar la información para obtener conocimiento, detectar lo esperado y descubrir lo inesperado, proporcionar valoraciones oportunas, defendibles y comprensibles y comunicar esas valoraciones con eficacia para la ejecución de acciones [18]. Resulta fácil, por tanto, extender el concepto de Visual Analytics al entorno de las actividades en el ciberespacio,

donde se trabaja con ingentes volúmenes de datos, de características similares a las ya mencionadas, dentro del ámbito de la seguridad.

Sencillamente Visual Analytics se puede contemplar como un enfoque integral (figura 11) en el que se conjugan áreas como [18]:

- la minería de datos y la gestión de datos, debido al vasto volumen, diversidad de tipos y naturaleza de los mismos, necesarios como entrada para ser analizados, datos que requieren ser procesados, mediante la aplicación de algoritmos de análisis automático, generando modelos de aprendizaje y así predecir el comportamiento en base a nuevos datos posteriores.

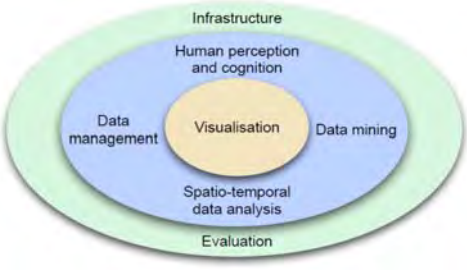


Figura 11. Áreas integradas en Visual Analytics.

- análisis espacio-temporal de datos, dado que habitualmente los datos son incompletos, interpolados, recogidos en momentos diferentes de tiempo, basados en suposiciones a veces contradictorias, referenciados a diferentes escalas espacio-temporales y de una complejidad elevada, son de especial relevancia todos aquellos que hacen indicación al tiempo y el espacio y sirven de base de referencia.
- percepción y conocimiento humano, para lo que es necesario un cuidadoso diseño de las interfaces maquina-humano, aplicando estudios de psicología, sociología, neurociencia y diseño en la elaboración de sistemas de información visuales eficaces, aplicando los criterios de usabilidad, disponibilidad de recursos y desarrollo de nuevos algoritmos.

Estos conceptos facilitan el marco sobre el que desarrollar métodos de visualización de la información, desde una perspectiva científica, para proporcionar una interfaz grafica interactiva entre el ser humano y la maquina, bajo el paraguas de una infraestructura necesaria y suficiente a los servicios de visualización requeridos y una evaluación continua de investigadores y desarrolladores que creen nuevas técnicas, métodos, modelos y teorías, buscando una estandarización que facilite a los usuarios la identificación de los problemas mediante el análisis visual de la situación [17].

De esta forma, el proceso de Visual Analytics combina métodos automáticos y visuales de análisis con la interacción humana en aras de alcanzar conocimiento a partir de los datos. En la figura (figura 12) se expone este proceso, desde la adquisición de los datos hasta la obtención de conocimiento, incluyendo el procedimiento de retroalimentación [17].

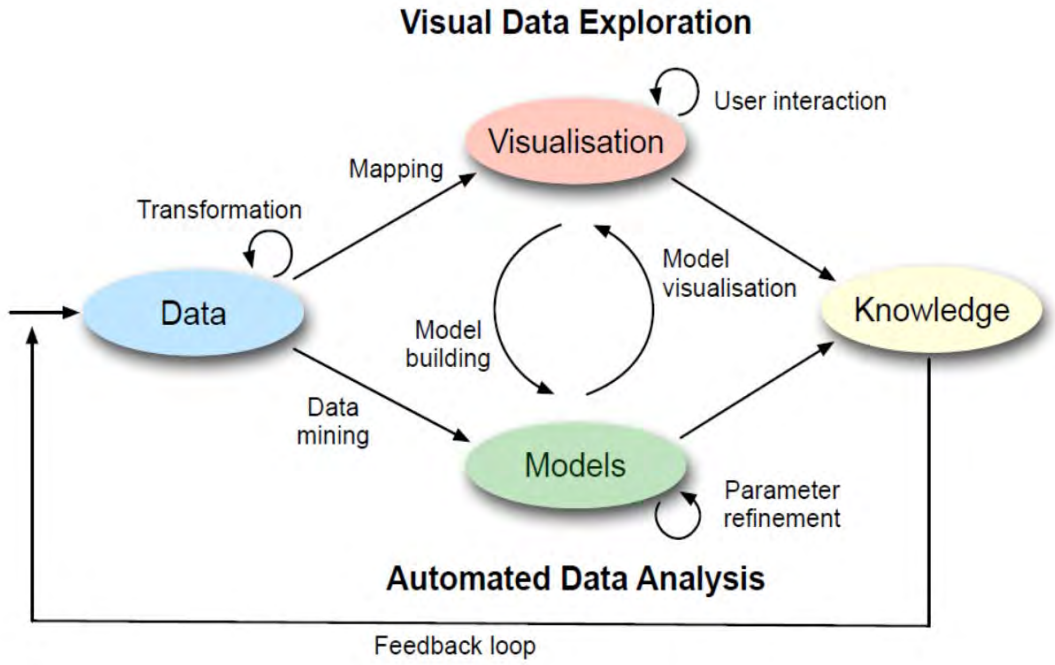


Figura 12. Proceso de Visual Analytics [17].

El primer paso en este proceso, y debido a la heterogeneidad de los datos y sus fuentes, es procesar, normalizar, agregar y transformar estos de forma que se pueda realizar una exploración visual. En el siguiente paso, el analista puede seleccionar la opción de realizar análisis visuales o automatizados. Si la elección es la exploración visual, el usuario puede adaptar el grafico para evaluar los resultados obtenidos pero deberá confirmarlos mediante métodos automáticos. Por el contrario, si la elección es el análisis automatizado, se generará el modelo mediante algoritmos de minería de datos, que puede ser refinado interactuando con los datos, para obtener un modelo visual que el analista nuevamente puede ajustar, modificando parámetros o seleccionando otros algoritmos. Se confirma así, la existencia de un flujo continuo entre la creación del modelo y la visualización del mismo mediante la interacción humana. El resultado final es la obtención de conocimiento a través de un proceso continuo de refinamiento y verificación. Así, se pueden corregir resultados engañosos que pueden ser

descubiertos y descartados, generando un flujo de retroalimentación eliminando los datos erróneos, obviando los elementos superfluos o restringiendo el volumen de los mismos [17].

A menudo se confunde la técnica de Visual Analytics con la de presentación visual de los datos. En el segundo caso, únicamente se detallan los datos en estado crudo, mediante técnicas graficas. Como ha sido expuesto, mediante Visual Analytics, se aplican técnicas automáticas de tratamiento de datos mediante algoritmos de minería de datos, dando lugar a gráficos más elaborados que proporcionan información sintetizada.

En el sencillo ejemplo (figura 10) de representación gráfica de la información sintetizada se puede apreciar cómo se cumple con las condiciones que hasta este momento se han ido exponiendo en este texto. Así, cuando en el apartado 1 se exponía que "*la información debe ser clara y nítida*", se invocaba la necesidad de que esta no sea confusa, proporcionando los elementos necesarios, no más, pero tampoco menos, evitando la distracción de lo superfluo, "*además de poderse interpretar de la forma más intuitiva posible*", siendo necesario, únicamente, unos conocimientos básicos de lo que representa para poder obtener una comprensión de la situación, punto este, entroncado directamente con el hecho, mencionado en el apartado 2, que la "*presentación grafica sea intuitiva, permitiendo tener una conciencia de la situación actual, y que proporcione una proyección, en el futuro, de las decisiones actuales adoptadas*". Por último, debe facilitar, mediante el entrenamiento, alcanzar modelos mentales, que como se expuso en ese mismo apartado 2, "*proporcionan información por defecto para las personas que deben tomar decisiones*". Por último, se aplican técnicas de Visual Analytics, representado relaciones entre tipos de datos y transformándolos a figuras adecuadas a la información transmitida, no así en la figura 9, en la que únicamente se realiza una "*presentación visual*" del dato de equipos comprometidos, un punto por cada equipo, sin ningún tipo de análisis previo.

A continuación, se exponen diferentes técnicas de representación de la información agregada, sus bondades y sus deficiencias, para posteriormente, introducir una valoración de las mismas en la representación de diferentes eventos, y de esta forma elegir la más idónea en cada caso. Podemos adelantar que no existe, como por otra parte es lógico, una única forma de representación, que sea la optima para todos los casos.

4 Técnicas de visualización gráfica.

Las actividades que suceden en el entorno del ciberespacio, como en todos los sucesos de cualquier tarea, son de muy diversa índole, a la vez que dispares. Resulta evidente que las formas de presentar los diferentes hechos, difieren también entre ellas. Cada técnica de visualización utiliza diferentes formas de presentación gráfica de la información. Debe representar las actividades desde muy diferentes taxonomías y todas ellas compatibles. Desde actividades que impliquen a unas pocas maquinas hasta millones de estas, desde un ámbito territorial reducido a un ámbito mundial o quizás superior, sin olvidar que deben representarse actividades, tanto legitimas como ilícitas.

Algunas de las técnicas más usuales y sencillas de visualización, para una representación gráfica de la información son: símbolo o icono (del inglés *glyphs*), mapas de colores, diagramas de dispersión, histogramas, pero existen otras más sofisticadas y complejas como: diagramas de coordenadas paralelas (del inglés *parallel coordinate plots*), mapa de árbol (*treemap*), grafico de anillo y flujo de conexión (connection river). En adelante, aunque utilizaremos ambas denominaciones de cada técnica gráfica, elegiremos preferentemente su denominación en ingles, por ser este el nombre por el que técnicamente se les conoce.

4.1 Glyphs.

Glyph es un pictograma (figura, símbolo o icono) que representa cualquier elemento del sistema que se desee simbolizar. Puede ser una entidad: servidor, router, firewall, unidad de almacenamiento, etc., o un elemento inmaterial: sistema operativo, proceso, flujo de datos, etc.

De esta forma tan sencilla, puede conocerse, mediante una comprobación superficial, una aproximación a la arquitectura de la red (figura 13). Cada elemento se dibuja con un icono distinto, permitiendo así, si están bien elegidos estos, reconocer al elemento que representa, con unos pequeños conocimientos del sistema que simbolizan.

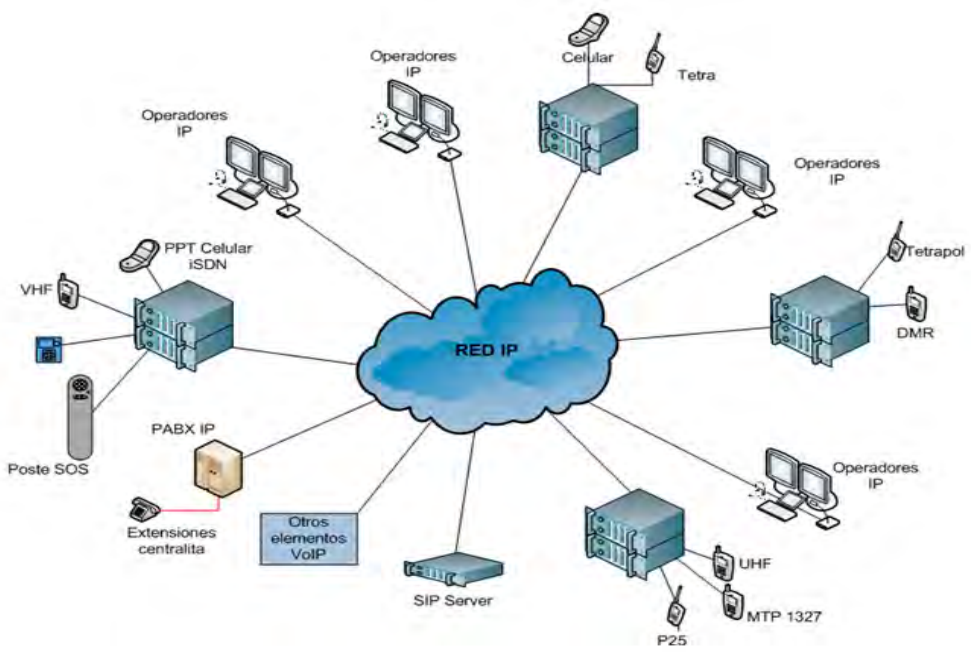
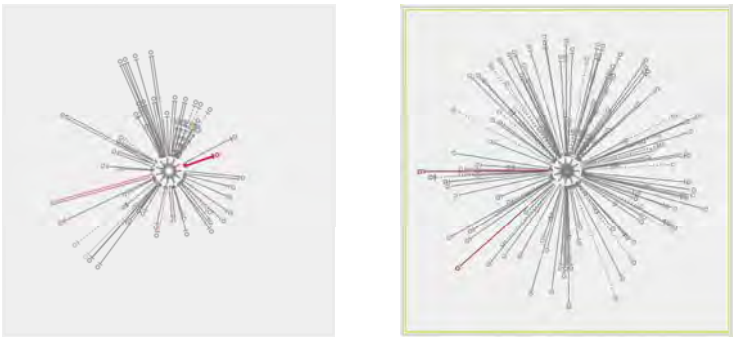


Figura 13. Representación de una red IP mediante glifos o iconos.

En otras ocasiones, los símbolos no son tan reconocibles, dando la información precisa de otro modo. En este caso (figura 14) la información transmitida con otros glifos es bien distinta. En ella se puede apreciar los equipos conectados a un sistema de ficheros mediante el protocolo NFS (Network File System). Esta instantánea (figura 14.a) está tomada al inicio de la mañana, cuando los usuarios realizan la conexión inicial [19]. La segunda instantánea (figura 14.b) está tomada más tarde, pudiéndose apreciar que hay muchas más conexiones. Los nodos en el primer anillo de conexión son hosts de la misma red que el sistema de ficheros. Los nodos en el último anillo, el más exterior, son hosts cuya dirección IP no se ha podido resolver, por lo que son pintados en rojo.



(a) (b)

Figura 14. Representación de una conexión de red mediante glifos [19].

En ambas figuras se puede encontrar mucha más información, como el tramo del día en que se tomaron, posibles direcciones IP de atacantes, etc., solo reconocibles entendiendo los patrones con los que están construidas. En este segundo ejemplo los iconos no resultan tan evidentes como los del primero. Sin embargo, y aunque sean necesarios mayores conocimientos del sistema y de los modelos de construcción, se incrementa la información proporcionada.

Existen muchas técnicas en la utilización de glifos. Un glifo puede ser asignado a entidades multivariables mediante procedimientos de representación de datos multidimensionales, como localización, tamaño, color, forma, opacidad y trasparencia y otras muchas. Generalmente estos atributos, características intrínsecas a la entidad representada, se asignan al glifo antes de ser representado, pudiendo aplicarse procedimientos de comparación entre las mismas [20].

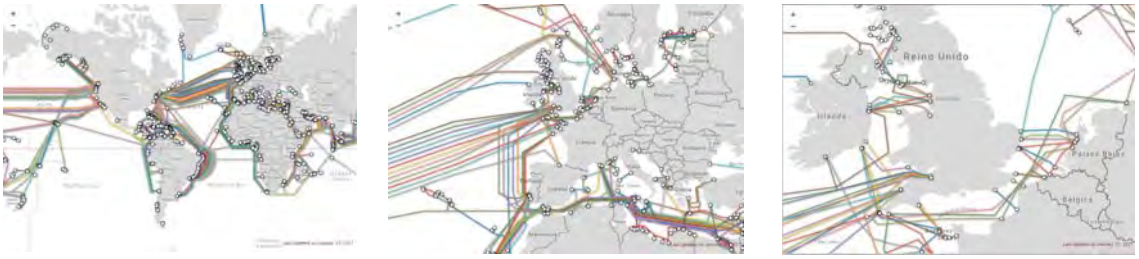
Por último, las conexiones entre equipos o sistemas se pueden representar mediante líneas. En ellas se puede reflejar más información que la simple conexión entre elementos. Utilizando colores, grosor del grafo, tipo de línea y longitud de la misma se puede expresar información añadida, propia de la conexión. Mediante flechas se puede indicar la dirección del tráfico, en el caso que no sea bidireccional.

4.2 Mapa geográfico de situación.

Otro elemento grafico, que permite obtener conciencia situacional del desarrollo de los acontecimientos, son los mapas geográficos. El espectro de ubicaciones a representar abarca, desde la totalidad de la tierra, mediante un mapamundi (tampoco se debe descartar que el limite se pueda situar en el espacio exterior), hasta una pequeña localización, representada por el plano de una ubicación no mayor de unos pocos metros cuadrados, como por ejemplo, un centro de proceso de datos o CPD. En este estudio, únicamente se tendrá en cuenta la localización en grandes ámbitos territoriales.

Muchos son los elementos que se pueden representar como indicadores de incidentes, amén de los elementos constitutivos de las infraestructuras propias, conformando estos últimos, los objetivos a proteger. En primer lugar, se pueden representar las líneas de comunicaciones, soporte de la infraestructura. Para ello,

únicamente trazando los diferentes trayectos, uniendo los nodos, representamos la red de transporte de la información. Dependiendo del ámbito territorial que se esté representando, se alcanzara un mayor nivel de detalle. En la figura (figura 15) se puede observar un mapa del cableado submarino que proporciona las comunicaciones, a través de cable, a todo el planeta [21]. Una de las acciones que se puede realizar es la de zoom, acercando o alejando el foco. Evidentemente, cuando se acerca el objetivo y se



(a) (b) (c)

Figura 15. Diferentes aproximaciones de una red en función del ámbito territorial representado [21].

reduce el plano, se alcanza un mayor nivel de detalle, a costa de perder visión periférica. Así, se aprecia una vista de las líneas mundiales (figura 15.a), que queda reducida a la zona de Europa (figura 15.b), observando un mayor detalle de las líneas, pero perdiendo la visión del resto del mundo. Este mismo criterio es de aplicación para la región del Canal de la Mancha (figura 15.c).

Otros datos que se pueden extraer del grafico, en función de su representación, son: el ancho de banda del cable, el nivel de trafico actual, los puntos o nodos de conexión, estado de servicio actual, longitud, etc., mediante combinación de colores, modos de línea y grados de trasparecía. Para obtener información de otros datos, se pueden extraer mediante cuadro de contesto, posicionando el apuntador (puntero del ratón) sobre el enlace del que se quiere extraer la información.

Al igual, que en la mayoría de los casos estudiados, la excesiva confluencia de líneas resulta farragoso, por lo que una medida a arbitrar puede ser, la de seleccionar dinámicamente la línea, o líneas, sobre la que se realiza el estudio. Eliminando todas las demás o proporcionándoles un cierto nivel de transparencia, permitiría centrar la atención sobre las que se determina el estudio.

Toda esta información, extraída de las comunicaciones propias, es de aplicación acerca de las comunicaciones del posible enemigo, siendo conscientes que difícilmente

se podrá llevar a cabo, por no disponer del conocimiento de sus infraestructuras, como resulta evidente.

Otros elementos, susceptibles de representación gráfica, son aquellos que componen la electrónica de red (figura 16), como routers, firewalls y switches o puntos de acceso, para redes de cable y para interconexionado de equipos wireless, respectivamente. La representación de estos equipos está ampliamente aceptada

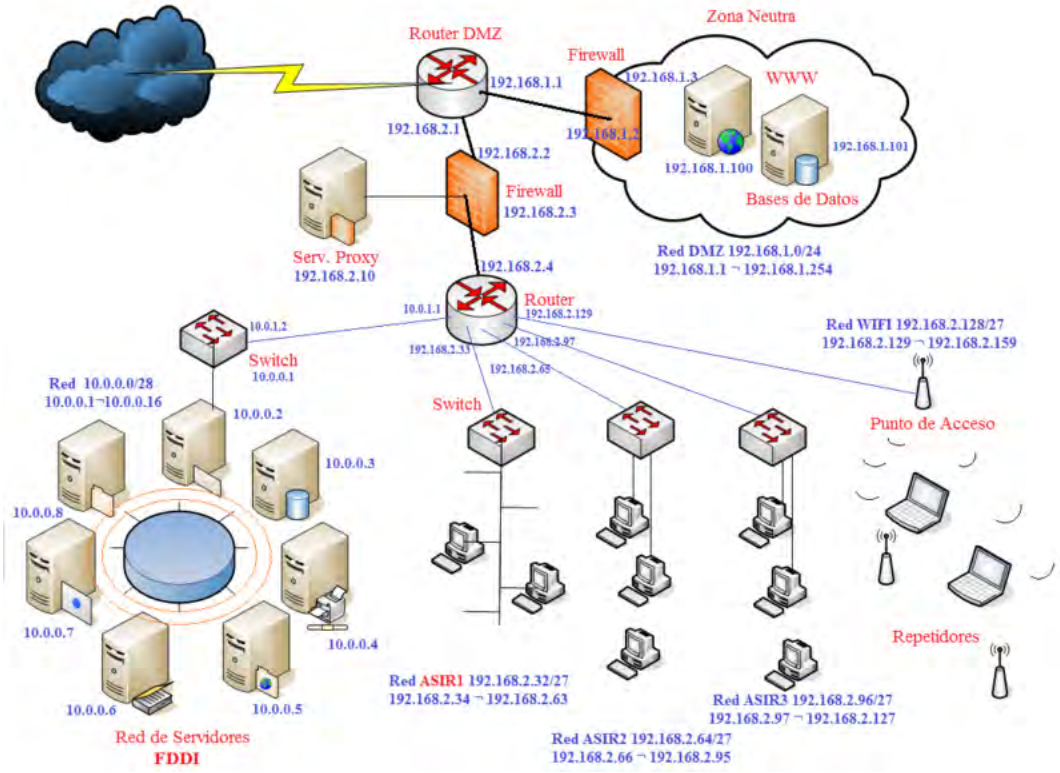


Figura 16. Gráfico de electrónica de red

mediante los símbolos de la figura (figura 17). Tanto si se trata de equipos propios, como si son los de un posible enemigo, el hecho de que cualquiera de estos elementos se encuentre comprometido, supone un grave riesgo para las comunicaciones. Es por este motivo, por el que su representación dentro del mapa o plano es más que necesaria. En este caso resulta conveniente separar las representaciones proporcionando información acerca de aquellos equipos que se tenga capacidad de configuración por ser propios, "*aquellos que no*" por pertenecer a terceros, pero que proporcionan servicio a nuestro propósito y aquellos de los que se tenga información, pero que pertenezcan a un posible enemigo. Fácilmente, pueden diferenciarse unos de otros mediante colores. Como se verá más adelante, es conveniente representar los

elementos propios con un color que denote afabilidad, bien pudiendo ser el verde, mientras que los del adversario deberían estar representados en rojo, que denota animadversión. El otro color, que resta, podría ser perfectamente el azul.

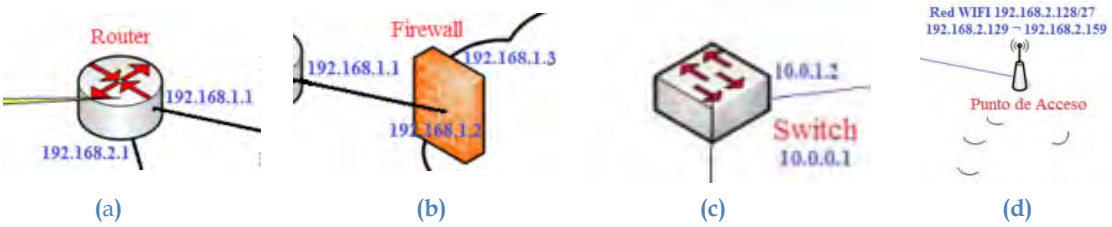


Figura 17. Gráficos de equipos de electrónica de red.

Es necesario, por otra parte, proporcionar información acerca del estado, en los que cada uno de estos equipos, se encuentran: operativo, comprometido, dudoso o bloqueado. El mejor modo para proporcionar esta información es jugando con el tono del color y niveles de trasparencia.

Finalmente, cualquier otra información de relevancia que se pueda proporcionar, como dirección IP, o direcciones para equipos que posean varias, nombre del equipo, localización, marca, modelo, sistema operativo y otras, se facilita a través de un cuadro de dialogo cuando se posiciona el apuntador sobre la representación del equipo.

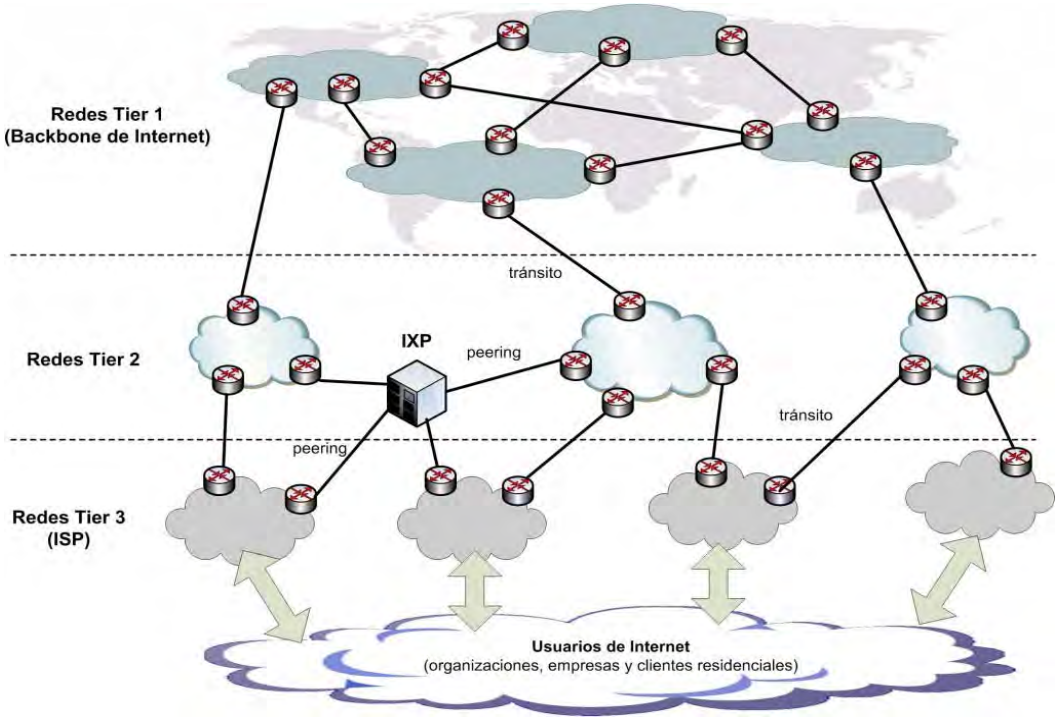


Figura 18. Niveles de ISPs en Internet [22]

Al igual que en todos los casos anteriores, no se pueden representar todos los equipos. Estos deberán aparecer y desaparecer de acuerdo con el orden territorial del mapa que se esté utilizando. Así en el mapa mundial no debe aparecer ningún equipo o de lo contrario estos no podrían diferenciarse unos de otros. Solo cuando el mapa se reduce a unos pocos países, entre 4 y 6 como máximo, pueden aparecer los equipos de ISPs de nivel 1 (tier 1) [23], que proporcionan cobertura internacional (figura 18). A medida que se aumenta el foco y se reduce el campo, deben aparecer los ISPs de nivel 2 (tier 2), con cobertura nacional, pero a su vez desapareciendo los anteriores, o de lo contrario, la imagen se tornará ininteligible. Conforme se desciende en el ámbito territorial aparecen los equipos de operadores locales o tier 3. Igualmente, y con motivo de clarificar el grafico, cuando aparecen nuevos elementos, deben ir desapareciendo los del nivel anterior. Aun así, la gestión del número de equipos del grafico se vuelve inaccesible. Para facilitar la comprensión de los sucesos solo deben aparecer aquellos que el analista elija mediante su selección con el puntero. El resto pueden dibujarse de modo trasparente para que no sean molestos a la vista.

Una vez llegados a este punto, solo quedan por representar los "*equipos de usuario*", en los que distinguiremos por su diferente idiosincrasia tres tipos: servidores, equipos fijos de usuario y dispositivos móviles. Más adelante se tratará que es lo que se va a representar. En este momento únicamente se va a indicar como se va a representar.

Figura 19. Glyphs de posición.

En primer lugar, y antes de representar los distintos equipos, se debe ubicar la posición del mismo. El símbolo grafico más usado para determinar una posición es el glyph de posición (figura 19). En la misma se representan tres marcadores con los mismos códigos de colores utilizados, anteriormente, para los equipos de electrónica de red. Estos se sitúan sobre el plano en la ubicación donde se encuentra el equipo al que se hace referencia (figura 9). Junto al símbolo de localización se adjunta un icono del tipo de equipo estudiado.

El primero de los equipos a representar es el servidor. La representación más comúnmente adoptada es la realizada en la figura (figura 20). Como se puede apreciar, además de representar que el equipo es un servidor, se representa el tipo de servidor que es, mediante un pequeño símbolo que lo identifica. Esta información resulta relevante, tanto para la protección que pueda necesitar, como para las acciones, que

sobre él, se puedan realizar. En la figura (figura 20) se han representado algunos de los tipos de servidores. Existen otros muchos tipos, debiendo elegir los símbolos que lo representen de forma que sean lo más identificativos posibles.

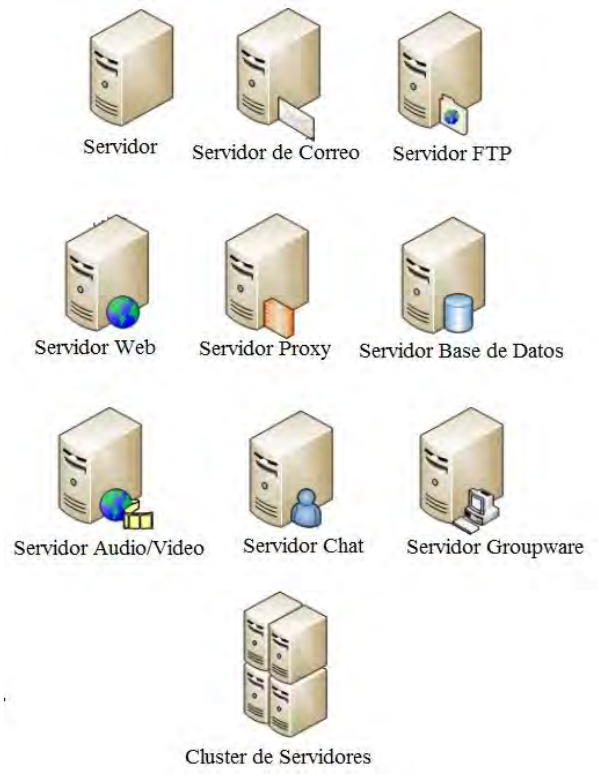


Figura 20. Gráficos de tipos de servidores

Los dos últimos elementos que tienen representación en un mapa geográfico son: los equipos fijos de usuario (figura 21.a) y los dispositivos móviles (figura 21.b).

Deben diferenciarse, porque sus características son diferentes, también lo son sus capacidades, y por ende, los vectores de compromiso de cada uno son muy dispares. El elemento diferenciador entre ellos y, a su vez, distinto al de los servidores, que sirve para discriminar la naturaleza de cada uno de ellos es el sistema operativo. A modo de ejemplo, y utilizando la familia de software de Microsoft, son muy diferentes MS Windows Server, MS Windows y MS Windows Mobile. Tres sistemas operativos, de un mismo fabricante, que diferencian los dispositivos que lo usan. En la figura, se puede apreciar que existe un icono más, la representación de un laptop (figura 21.c). Esta representación se corresponde con la de un equipo bivalente. Aunque su sistema operativo es el de un equipo fijo de usuario, su comportamiento puede ser el de un equipo móvil. Los aspectos que determinan esta situación son, entre otros posibles: que su punto de acceso varía, que se modifica su dirección IP o que traslada su posición a unas nuevas coordenadas geográficas. Si ocurre cualquiera de estas situaciones, u otra que indique que el dispositivo es móvil, su representación debería ser la de la figura 21.c.

(a) (b) (c)

Figura 21. Gráficos de equipos de usuario.

Al igual que los equipos de electrónica de red, en estos últimos equipos también es necesario proporcionar el estado de los mismos. El modo de realizarlo es igual que en el caso expuesto anteriormente. De la misma forma, la información adicional se proporciona a través de cuadros de dialogo, al seleccionar el elemento elegido para estudio.

En el caso de la geolocalización, los equipos propios no presentan ningún problema, ya que su posición es conocida. Los equipos, de un posible enemigo, deberán geoposicionarse por diferentes métodos. Aquellos que estén dotados de tecnología GPS, y sea posible interrogar al terminal, es, sin lugar a dudas, el procedimiento más sencillo y más fiable. De no ser esto posible, y para terminales móviles, la localización se puede realizar por triangulación de antenas, aunque para este método se debe contar con la colaboración de la operadora de telecomunicaciones.

Para el resto de dispositivos, el último recurso disponible es la resolución de las direcciones IP, en función de la asignación de direcciones IP realizada por la Corporación de Internet para la Asignación de Nombres y Números (en inglés: *Internet Corporation for Assigned Names and Numbers*; ICANN) [24]. Dado que esta asignación de números esta realizada territorialmente, puede deducirse una ubicación aproximada en función de su dirección IP. Este método resulta poco fiable, agravado por el hecho que la dirección que, presuntamente, este realizando hechos delictivos puede estar siendo suplantada. Por este motivo, únicamente se puede considerar fiable, con una cierta probabilidad, la penúltima dirección IP obtenida al ejecutar la herramienta "*traceroute*". Dicha fiabilidad es de un 95% a nivel de país, reduciéndose a un 30% a 50% a nivel de ciudad concreta dentro del país [25].

De la misma forma anterior, y para no saturar la imagen, los símbolos deben aparecer y desaparecer en correspondencia con el nivel territorial del mapa.

4.3 Mapas de color.

La paleta/mapa de color es otra herramienta de representación grafica que permite, entre otras funcionalidades, diferenciar elementos similares mediante comparación de los mismos, facilitando la visualización al usuario. PortVis, herramienta para la detección de eventos en los puertos de los sistemas, maneja para visualización, técnicas de representación gráfica en las que se emplea mapas de color.

Como se puede observar (figura 22) en los marcos 1, 2 y 3, se utilizan diferentes colores del mapa de colores para discriminar los distintos elementos representados en cada uno de ellos [26].

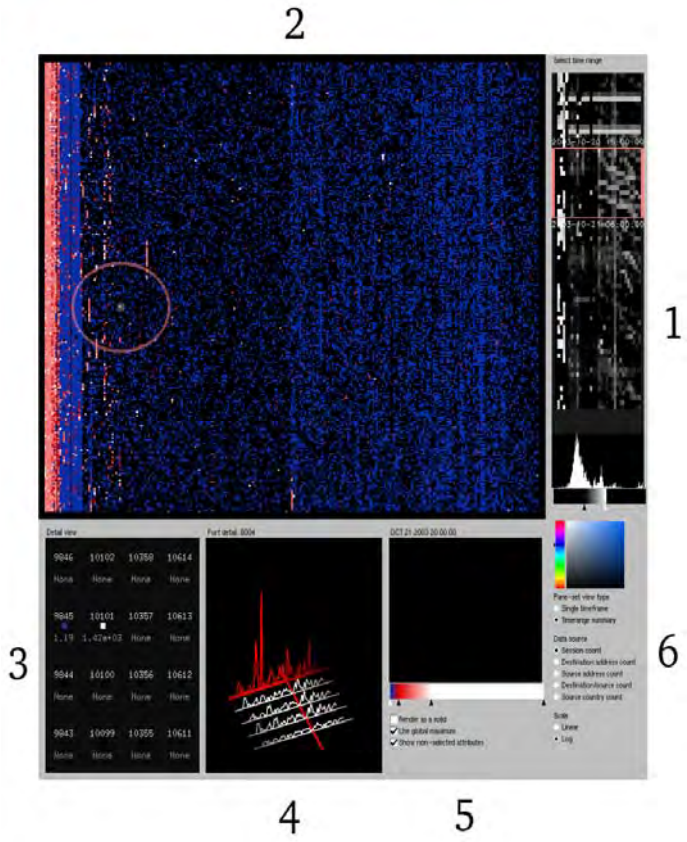


Figura 22. Matriz de 256*256 pixeles (2) para detección de eventos en puertos, mediante códigos de colores [26].

Así, en el marco 1, que es una escala de tiempo (eje vertical), en la que se representa la actividad ocurrida en los puertos de una maquina (eje horizontal) en un rango de 2048 puertos/línea. Cuando ocurre un evento extraordinario, se marcan con colores que permitan diferenciarlos del resto.

En el marco principal de visualización (figura 22), marco 2, se representa la actividad en los puertos del sistema en una unidad de tiempo. Este marco es una matriz de representación de puertos. La dimensión de la misma es de 256 * 256, lo que supone 65.536 puertos. De nuevo se representa la actividad anómala del mismo, mediante la diferenciación por colores. Como se aprecia (figura 22) en el marco 2, el punto rodeado de un circulo, representación del correspondiente puerto, tiene un color diferente, que permite conocer al usuario que la actividad en el desarrollada es, al menos, cuestionable.

Finalmente, en el marco 3 (figura 22) aparece un listado de los puertos. El puerto afectado en el marco 2, aparece en un color distinto en este nuevo marco, para diferenciarlo del resto, y de esta forma discriminarlo inmediatamente de una forma visual.

Habitualmente, se relaciona el uso de mapa de colores a una gradación de valores. Las clasificaciones asociadas se deberán encontrar dentro de un rango de valores de trabajo. Este puede estar definido de muy diversas formas (tabla 1): entre 1 y 10, elevado, alto, medio, bajo y nulo, etc.…

VALORES			
1 ... 10			
SOBRESALIENTE	NOTABLE	APROBADO	SUSPENSO
ALTO	MEDIO	BAJO	
EXCELENTE	BUENO	NORMAL	MALO	DEFICIENTE
ABIERTO	CERRADO		
….			

Tabla 1. Rangos de valores.

Los colores usados habitualmente son: negro, rojo, naranja, amarillo, azul y verde. Los dos primeros, negro y rojo, denotan peligro, especialmente el rojo. El naranja y el amarillo son empleados con un significado de precaución. Advierten de un posible peligro, no confirmado. Por último, los colores azul y verde, indican que el elemento representado está funcionando con normalidad.

4.4 Línea de tiempos o TimeLine.

La línea de tiempos es una herramienta, organizativa principalmente, que permite estructurar de forma cronológica la ocurrencia de sucesos a lo largo de un periodo de tiempo. En el eje de abscisas se representa el tiempo, dentro de un intervalo valido para el estudio que se desee realizar. En el eje de ordenadas se representa la variable, objeto de análisis. De esta forma, el analista adquiere un esquema mental de

la cronografía desarrollada en la red o en un equipo concreto. La organización de la actividad en líneas de tiempo establece un marco de correlación y sincronización de eventos y procesos. Permiten, además, superponer otra información de diferente naturaleza, como por ejemplo una dirección IP, que puntualiza el conocimiento adquirido de la situación. Dado que la ocurrencia de eventos en la red estará almacenada, o así debería ser, la grafica permite avanzar o retroceder en el tiempo, buscando los hitos de una manera sencilla y eficaz, de tal forma que se pueda reconstruir la historia de lo ocurrido.

Figura 23. Cronograma de flujo de tráfico con grandes variaciones.

En la figura (figura 23) se puede observar el cronograma del flujo de tráfico que atraviesa un puerto de un router. Se advierten muy diversas intensidades a intervalos diferentes de tiempos, apreciándose distintos picos de trabajo. Estos picos no tienen que ser, obligatoriamente, eventos maliciosos para el trafico de red, pero si sucesos que requerirán un estudio más profundo.

Por el contrario, en la siguiente figura (figura 24) el tráfico de red no presenta desviaciones importantes, por lo que, a priori, no se detectan anomalías. Esto no significa que el router no esté siendo víctima de algún ataque, pero no podrá ser detectado por este método.

Figura 24. Cronograma de flujo de tráfico regular.

Otra funcionalidad, que debe proporcionar la línea de tiempos es la de representar diferentes niveles de detalle. De forma flexible e interactiva el analista tiene la necesidad de expandir o contraer el intervalo de tiempo, objeto del estudio, de tal forma que pueda "precisar" el evento o analizarlo en su contexto histórico. Esto puede materializarse de muy diversas formas, siendo dos de ellas, las más usuales:

1. Representar diferentes líneas de tiempo, con diferentes intervalos.
2. Extender o reducir el intervalo de tiempo.

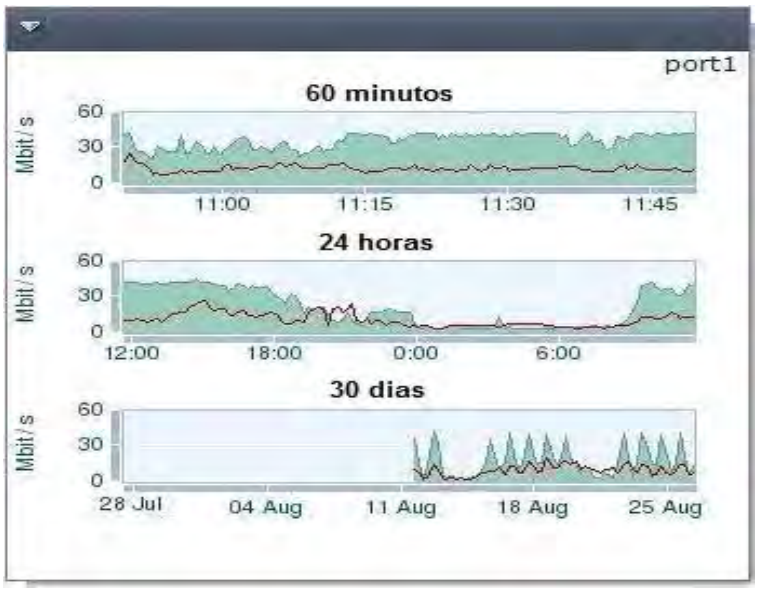
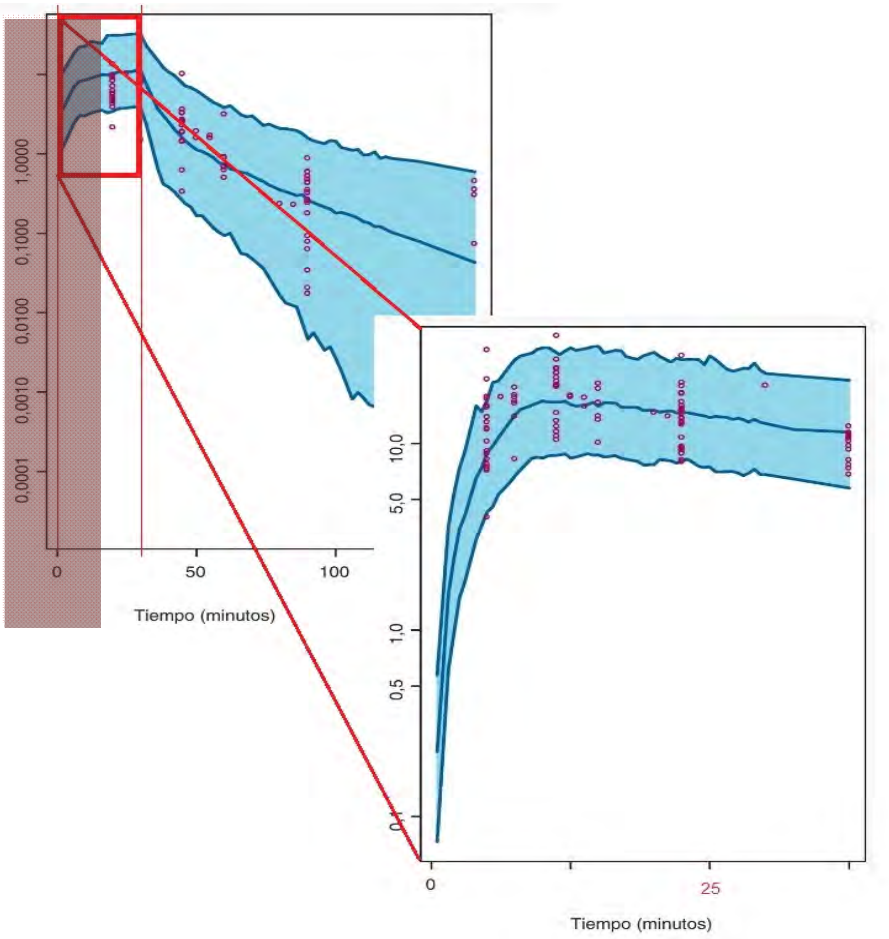


Figura 25. Diferentes intervalos de tiempos de un mismo cronograma.

En la figura (figura 25) está representada la cronología del tráfico de red a través del puerto 1 de un router. Como se puede apreciar, a la simple inspección de la figura, las tres representaciones comprenden el mismo tráfico de red, pero con una sustancial diferencia. El gráfico superior comprende 60 minutos de ese tráfico, incluido en el grafico medio, que representa 24 horas, que a su vez está incluido en el grafico inferior, representando 30 días. De este modo, el analista, de forma visual, puede comprender en una primera aproximación, que es lo acontece en la red y la "historia" que lo provoca. Por contra, al ser los ejes de tiempos fijos, no permite una mayor concreción de los hechos.

La segunda de las formas de representación permite una mayor adaptabilidad a las necesidades de conocimiento del analista. Este, actuando sobre el eje de tiempos, puede modificar el intervalo de tiempo, objeto del estudio, logrando de este modo cambiar la perspectiva temporal de los acontecimientos. Acortando el tiempo de estudio alcanza mayor especificidad, mientras que si lo amplia, consigue una mayor contextualización de lo sucedido. Presenta el inconveniente, frente a la modalidad

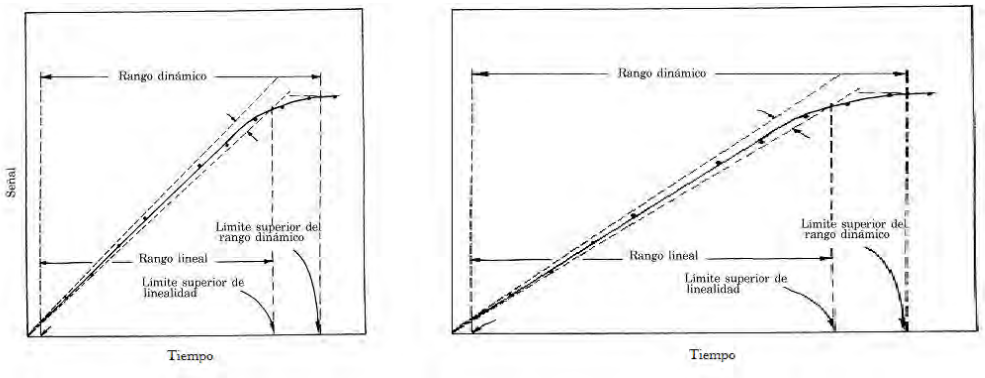


(a) (b)

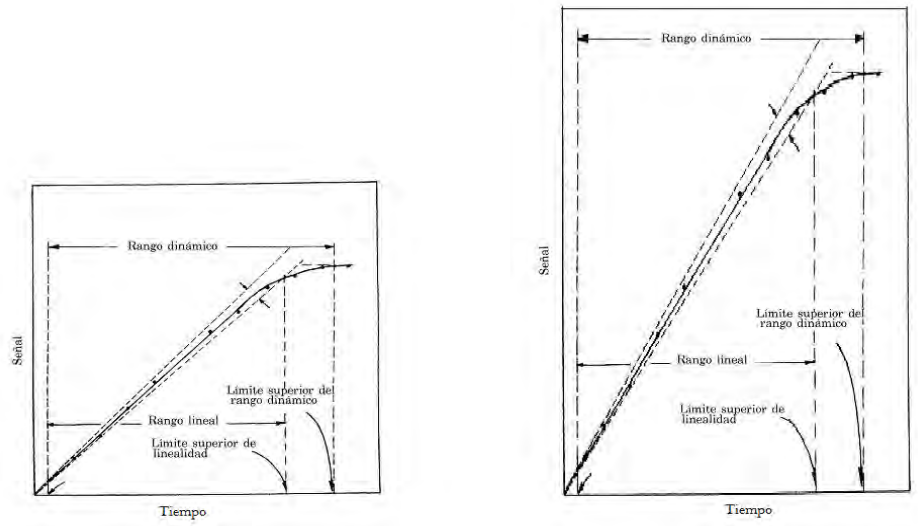
Figura 26. Acotación dinámica del intervalo de tiempo de estudio.

anterior, que no permite la comparación con el grafico del que proviene, ya que este se pierde al generar el nuevo. Una posible solución a este hecho seria que se generara un nuevo marco con la selección de tiempo y el original, de donde nace, no desapareciera, avanzando en los dos el tiempo al unísono.

En la figura (figura 26) se visualizan los valores, y sus medias, que alcanza una variable (eje de ordenadas) a lo largo del tiempo (eje abscisas). En ella se puede comprobar cómo al actuar el analista sobre el eje de tiempos puede alcanzar un mayor nivel de detalle (figura 26.b) u observando menos detalle (figura 26.a) correlaciona los hechos con la "historia" que los ocasiona. También se aprecia como al estrechar el intervalo de tiempo (figura 26.b) aparecen mayor número de valores intermedios que, en un intervalo mayor (figura 26.a), se pierden por simple resolución de la imagen, superponiéndose unos sobre otros.



(a)



(b)

Figura 27. Escalado dinámico del eje de abscisas (a) y del eje de ordenadas (b).

Para obtener mayor nivel de detalle se puede actuar de otro modo sobre el eje de tiempos. Sin modificar el rango de valores estudiados, se puede variar la escala del eje de tiempos (figura 27.a), con lo que el grado de apreciación de los valores individuales aumenta, cuando se ocasionan en intervalos de tiempo pequeños. De igual manera, puede modificarse la escala de los valores de la variable de estudio (figura 27.b), con lo que se mejora el escalado cuando sus valores son muy próximos.

4.5 Diagramas de dispersión.

Los modelos de diagramas de dispersión proporcionan información grafica de funciones estadísticas. Esta es una herramienta grafica que permite representar, mediante un sistema de ejes cartesianos, los valores de dos variables (figura 28) de un conjunto de datos. Los datos se muestran como un conjunto de puntos, donde en el eje de abscisas se distribuyen los valores de la variable independiente y en el eje de ordenadas los valores de la variable dependiente.

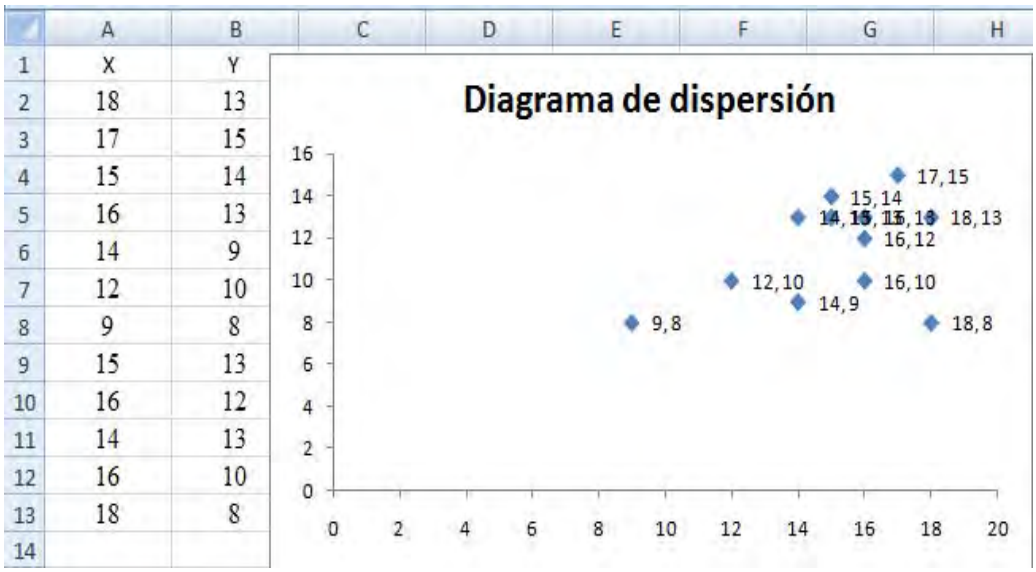


Figura 28. Diagrama de dispersión.

De esta forma se representa la relación entre las dos variables de un conjunto de datos, lo que hace más fácil visualizar e interpretar el vínculo y la tendencia de los datos. Tiene la capacidad de mostrar las relaciones lineales entre variables, así como las no lineales (figura 29). En la figura (figura 29.a) se puede apreciar una correlación lineal positiva, es decir, que existe una correlación, en la que, cuando la variable x aumenta, la variable y también crece. La correlación es negativa (figura 29.b) si al aumentar la

variable x, la variable y disminuye. La relación es no lineal cuando la función que la representa es distinta a una recta (figura 29.d). En este caso, igualmente pueden ser las relaciones positivas (figura 29.d) o negativas (figura 29.e). Además, permite estimar, si este fuera el caso, que las variables no están relacionadas (figura 29.f).

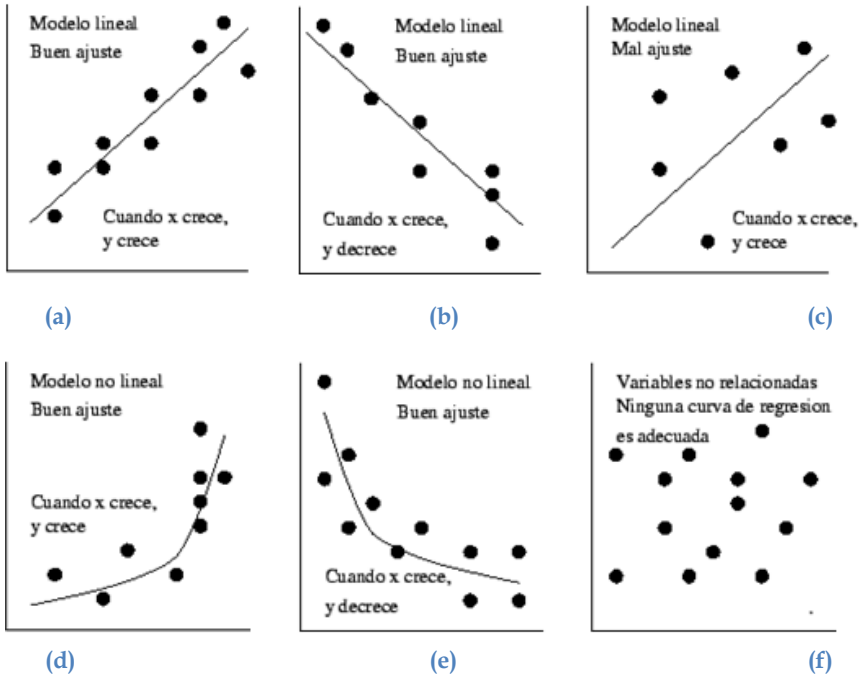


Figura 29. Correlación entre las variables en un diagrama de dispersión.

Existe la posibilidad de representar la relación entre más de dos variables. Esto se consigue mediante el uso de colores y/o figuras. Si se utiliza un color distinto para cada variable, es posible mezclarlas, facilitando el estudio de correlación entre las mismas y las diferencias entre ellas (figura 30).

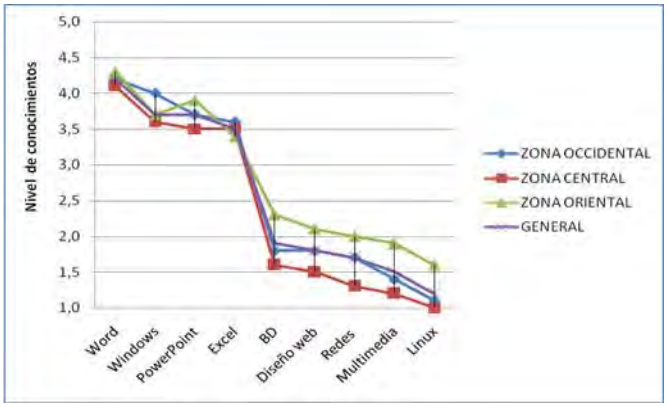


Figura 30. Diagrama de dispersión de varias variables.

Esta técnica, permite representar grandes cantidades de datos en una única pantalla, posibilitando así, contextualizar la información. Esto, que supone una gran virtud, representa también su principal debilidad. Cuando la cantidad de información a representar resulta muy considerable, o las parejas de valores se repiten, en gran medida, es muy probable que se ofusque el grafico (figura 31), por el problema denominado "overplotting" o superposición de datos, en el que diferentes parejas de valores ocupan la misma posición (tienen los mismos valores), no siendo posible discernir los valores individuales de los datos, dificultando el análisis de los mismos [27].

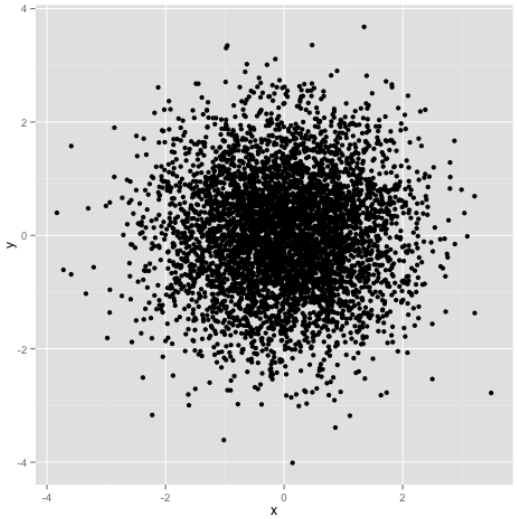


Figura 31. Overplotting [28].

En estos casos, existen diferentes métodos para evitar el problema que provoca el overplotting [28], como la utilización de filtros que, o bien, eliminen la información innecesaria o que permitan visualizar solo aquellos elementos que se requieran. Entre otros, destacan:

- el empleo de jittering (figura 32.a), donde se modifica ligeramente la posición de los puntos, permitiendo de esta forma que no coincidan. Este procedimiento es, únicamente valido, si el número de datos no es muy elevado, y por supuesto, no es muy preciso.
- aplicar transparencia al color de los puntos (figura 32.b). De esta forma se ensombrecen las zonas con mayor densidad de puntos es decir, las zonas más sombreadas son donde coinciden mayor número de puntos.
- desplegar un conjunto de hexágonos sobre el diagrama de dispersión (figura 32.c), para pintar cada una de las celdas de diferente color en función del número de puntos de la zona que abarca. Se hace necesario acompañar el conjunto de una leyenda con el código de colores.

- finalmente, y aunque existen otras técnicas, se pueden utilizar curvas de nivel (figura 32.d), donde estas indican el contorno de las áreas con las diferentes densidades de datos [29].

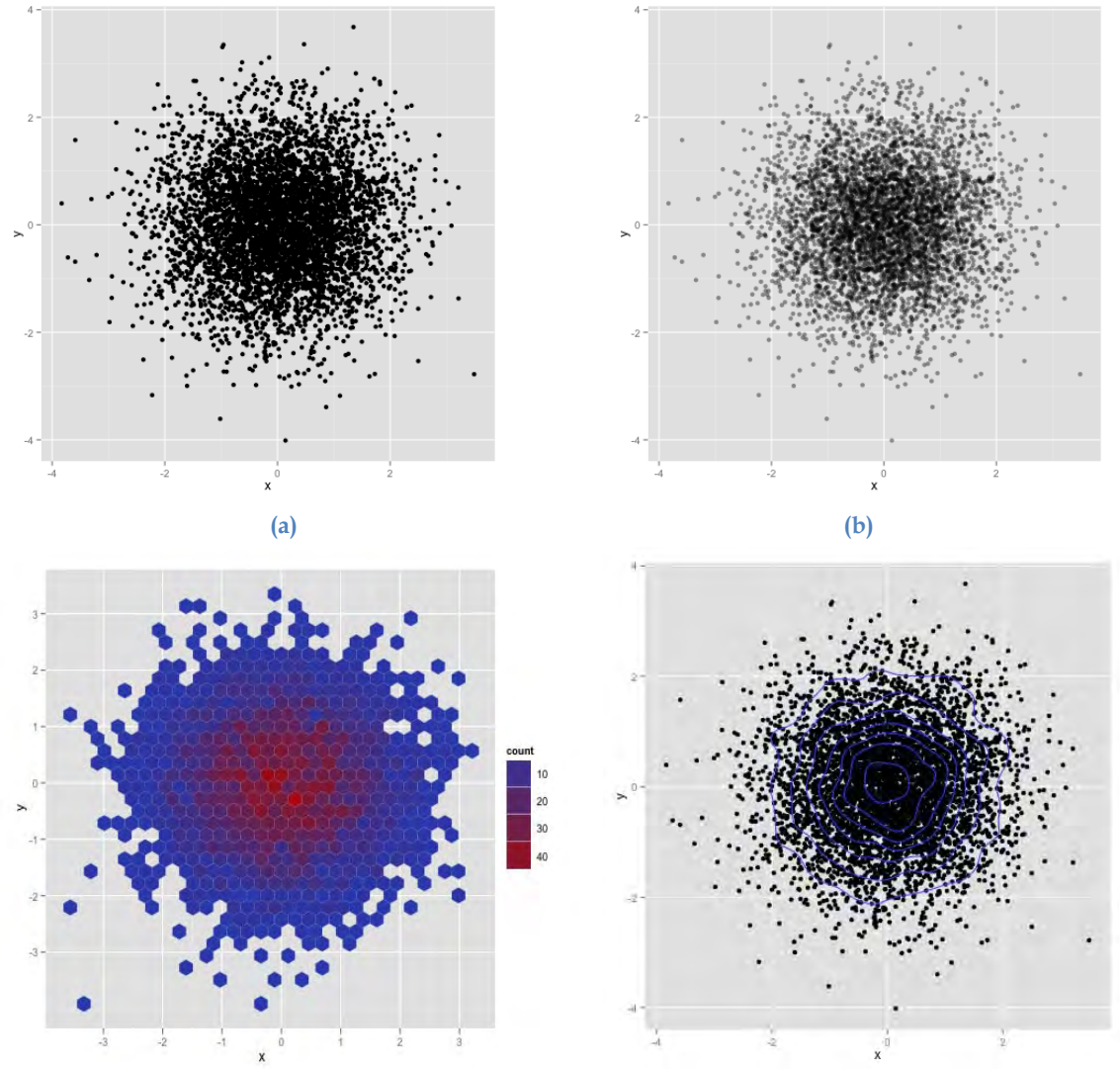


Figura 32. Soluciones al overplotting [28].

En definitiva, mediante cualquiera de las representaciones anteriores, sin olvidar que son diferentes representaciones de un mismo conjunto de datos, el "volumen" de datos, que se están representando en este caso, es el que se puede apreciar en la figura (figura 33), donde existe una gran concentración de datos, puntos con los mismos valores, en el entorno del punto (0,0), siendo más disperso hacia los laterales del grafico.

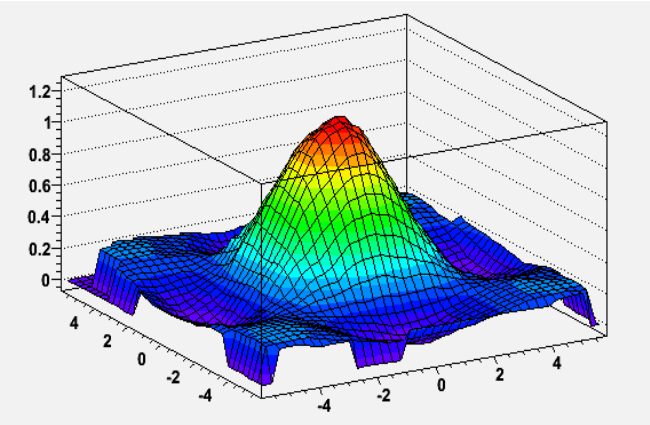


Figura 33. Representación 3D de un diagrama de dispersión que sufre overplotting.

4.6 Diagramas de nodos de enlace.

Los diagramas de nodos de enlace (node-link diagrams) son una técnica de representación grafica, generalmente bidimensional, de una entidad (un ejemplo seria una arquitectura de red) o de suceso (como podría ser el trafico de red). Para representaciones de tres dimensiones, o más, existen otras técnicas que estudiaremos más adelante.

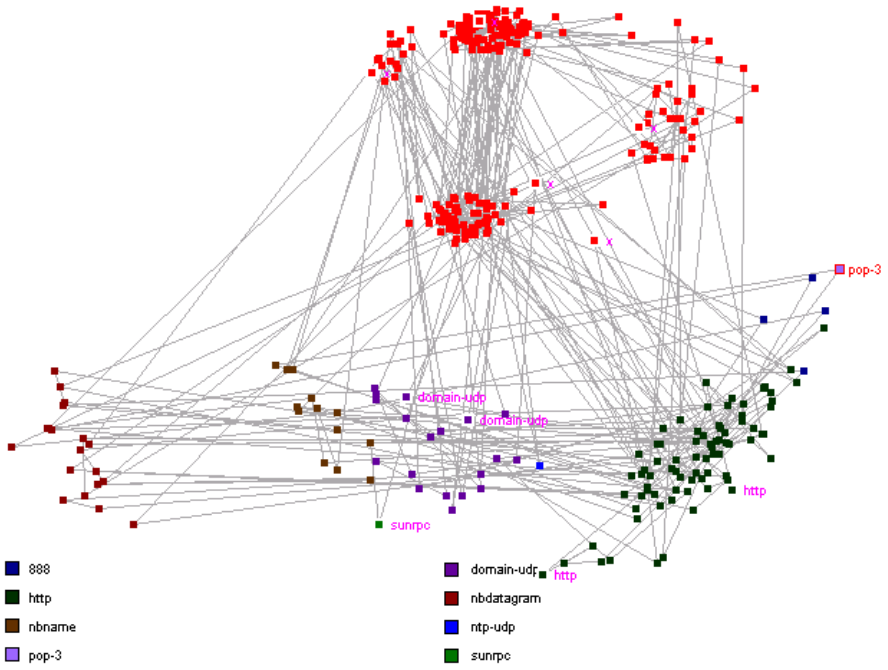


Figura 34. Diagrama de nodos de enlace [20].

Los diagramas de nodos de enlace (figura 34) conjugan iconos y líneas para representar los elementos o nodos que los componen y sus posibles conexiones o interdependencias [20]. Las conexiones de diferentes tipos se pueden representar mediante colores distintos (figura 34). Es posible dotar al grafico de una mayor información mediante la representación diferenciada de los elementos, tamaños diferentes, uso de colores, trasparencias, etc. Aun así, cuando la información representada resulta insuficiente, esta puede ser ampliada mediante ventanas contextuales. Posicionando el señalizador (puntero) sobre el elemento a analizar, puede abrirse una ventana que presente información de más detalle, como nombre, dirección IP, puerto, protocolo, etc. La principal función de un diagrama de nodos de enlace es la de facilitar la visualización de las conexiones entre los diferentes elementos del sistema y sus posibles cambios.

El principal problema al que se enfrenta esta técnica es uno ya conocido, mencionado anteriormente: el grafico, con demasiados nodos y sus respectivos enlaces, presenta un aspecto enmarañado que "impide" alcanzar una comprensión adecuada de la situación (figura 35).

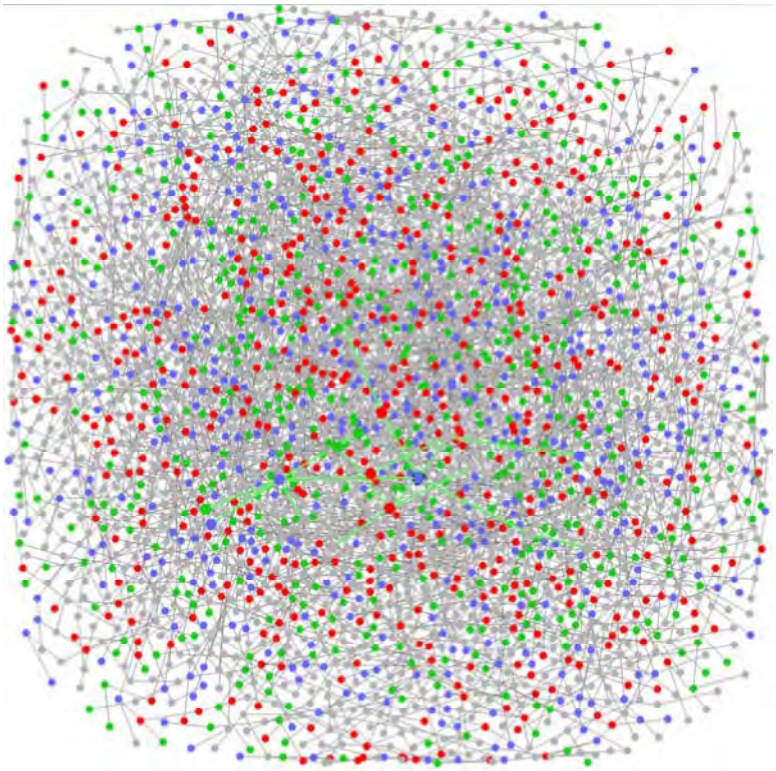


Figura 35. Diagrama de nodos de enlace de 3200 nodos [30].

Para una presentación estática, se estima que se pueden representar hasta 30 nodos, sin que resulte difícil su interpretación. La representación de gráficos con millones de nodos supone un reto insuperable [30]. Sin embargo, para representar

diagramas de tamaño medio, hasta unos pocos miles de nodos, se pueden utilizar técnicas dinámicas como herramientas que resalten algunos de los nodos y sus enlaces, mediante métodos de selección. En la figura (figura 36) se muestra el tráfico de correo de una empresa de tamaño medio de unos 1000 empleados. Cuando se selecciona un nodo o un subconjunto de los mismos, estos y sus enlaces son resaltados, modificando el color, el brillo o el tamaño de ambos. El resto de nodos y sus correspondientes enlaces, se pueden mudar a gris o a colores difuminados. De este modo, se proporciona al usuario una visualización selectiva de los elementos requeridos para el estudio, facilitando su comprensión.

Figura 36. Diagrama de nodos de enlace con 1000 nodos [30].

Existen otros métodos que permiten desplegar diagramas de nodos de enlace y que proporcionan visualidad sobre grafos de tamaño medio. Mediante técnicas de representación en planos hiperbólicos, se pueden desplegar grandes jerarquías de arboles. Estos se proyectan en el plano hiperbólico mapeándolos sobre un círculo, de tal forma que el punto de interés se sitúa en el centro, proporcionándole un mayor espacio relativo que al resto (figura 37). Únicamente se dibujan un determinado nivel de ramas, como contexto, hacia el exterior de la circunferencia, a modo de "ojo de pez", limitación realizada para proporcionar una mayor comprensión del usuario [31].

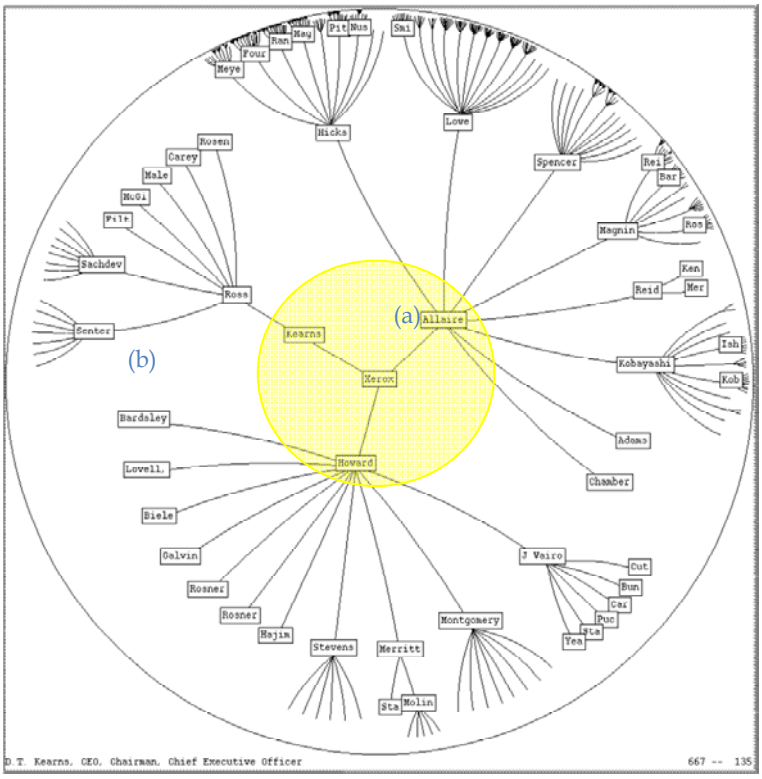
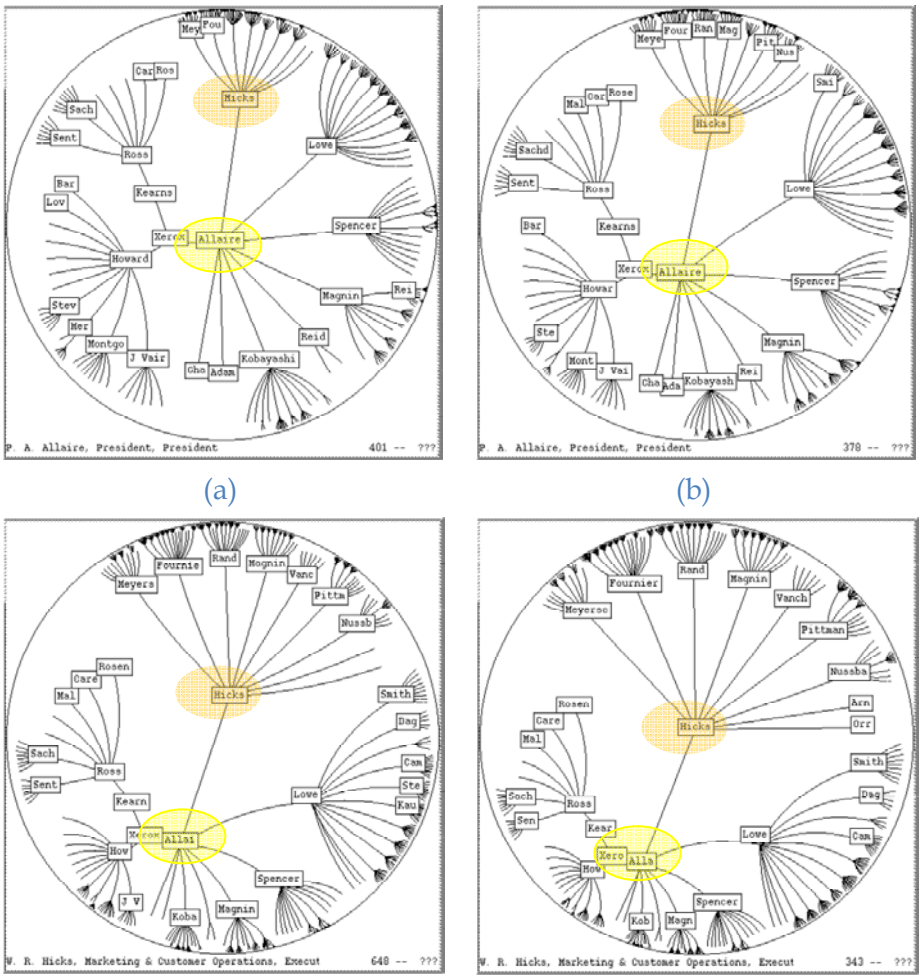


Figura 37. Diagrama de nodo de enlace realizado mediante planos hiperbólicos [31].

De esta forma se pueden representar en una ventana de 600 * 600 píxeles hasta 1000 nodos, 10 veces más que, los aproximadamente 100 posibles, con técnicas tradicionales. La mayor definición proporcionada al punto de interés, permite una navegación mucho más eficaz en las proximidades del foco. El usuario puede navegar por el grafo, cambiando el foco de interés, simplemente señalando con el puntero (ratón) una zona de la representación y arrastrándola. Con este movimiento, el árbol representado en el plano hiperbólico se desplaza, dejando como nuevo foco el punto que al soltar el puntero quede en el centro de la circunferencia (figura 38).

Las transiciones desde un foco a otro, arrastrando con el puntero, deben realizarse suavemente, sin sobresaltos y de forma continua, tal y como se muestra en los marcos a, b, c y d de la figura (figura 38). En ella se aprecia que la transición del foco inicial, pintado en amarillo, al foco final, marcado en naranja, se hace de una forma continua, permitiendo al usuario tener un conocimiento dinámico del grafo.



(a) (b)

(c) (d)

Figura 38. Desplazamiento de nodo de enlace por planos hiperbólicos [31]

4.7 Histogramas.

El histograma es una técnica grafica de representación, en forma de barras, de los valores de una variable, de su frecuencia de aparición. La longitud de cada barra es proporcional a la frecuencia del valor representado (figura 39). En el eje X se representan los diferentes valores que puede tomar la variable. En el eje Y se muestran el número de repeticiones del valor de la

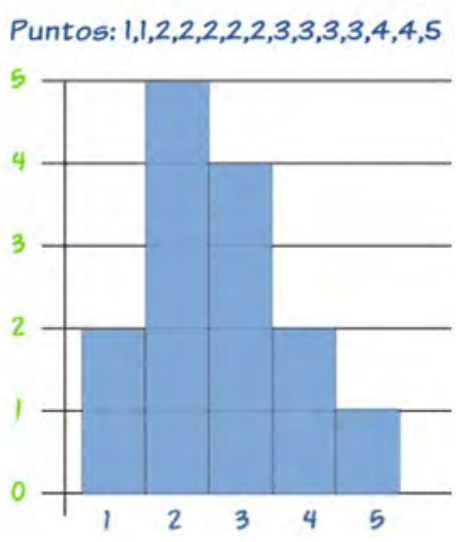


Figura 39. Histograma.

variable. También puede darse como dato del eje de ordenadas, el porcentaje de la frecuencia de repeticiones (valor relativo), en lugar del valor absoluto.

La variable puede tomar valores continuos o discretos. Para el estudio de variables continuas se dividen en intervalos. De esta forma el valor observado de la variable estará comprendido entre los extremos de un intervalo determinado. El número de accesos a un portal web por intervalo de edades seria un ejemplo de un histograma de una variable continua. De esta forma, y mediante una apreciación gráfica, se puede a que sector de población, por edades, está más interesado en el producto.

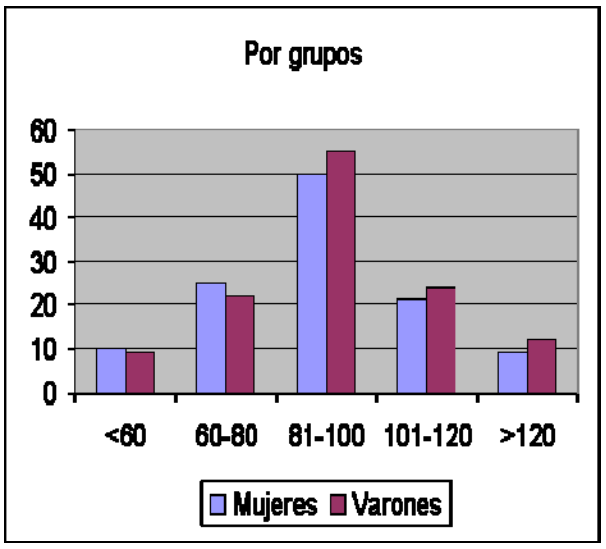


Figura 40. Histograma de dos variables comparables.

Una muestra de representación de frecuencias de variables discretas mediante histogramas, y siguiendo con el ejemplo de accesos a un portal web, seria la medición del número de accesos desde los diferentes departamentos de una empresa. Con este grafico se puede determinar qué departamento tiene más interés por la información ofrecida en el mismo. Además, pueden realizarse histogramas de dos o más variables, que deberán ser comparables. Continuando con el ejemplo propuesto, se puede evaluar la frecuencia de un hecho, pero separado por hombres y mujeres (figura 40). Esto nos permite evaluar el comportamiento de la población ante una acción determinada desglosado por sexos, pudiendo comparar ambas variables de una forma rápida.

De este modo, el histograma permite observar una tendencia o preferencia del espectro de valores de la variable. Así, se constata el comportamiento de la población, evidenciando el grado de uniformidad, o por el contrario, la dispersión de los valores y la heterogeneidad del sistema.

4.8 Gráfico de sectores o diagrama de pastel.

También denominado diagrama circular (en ingles pie chart), permite representar variables de todo tipo, pero se utiliza, principalmente, para representar

variables cualitativas, es decir, que expresan distintas cualidades, características o modalidades. Cada modalidad, que se presenta, indica una categoría o atributo de la variable, y la medición consiste en una clasificación de dichos atributos, de modo cuantitativo.

La representación del grafico se basa en la proporcionalidad entre la frecuencia de cada uno de los atributos y el ángulo central de la circunferencia que le corresponde, abarcando la totalidad de la variable, la totalidad de la circunferencia.

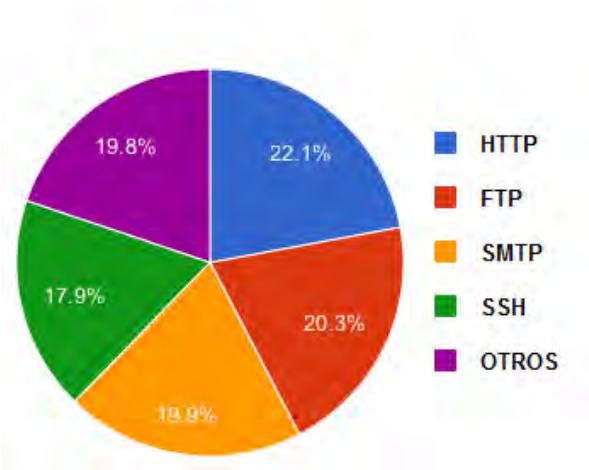


Figura 41. Diagrama circular, representando el tráfico que atraviesa un router.

En la figura (figura 41) se pueden apreciar todos los extremos expuestos para la evaluación de la variable, en este caso: "trafico que atraviesa un router". Como variable cualitativa la podemos dividir en diferentes tipos de tráfico: http, ftp, smtp, ssh y el resto de tráfico que se denomina otros en la figura. Cuantitativamente, se mide la cantidad de tráfico por unidad de tiempo (segundo, hora, día, ...) de cada uno de los diferentes tipos enunciados, que atraviesa los puertos del router. Cada uno de estos volúmenes de tráfico, en relación con el total, deben ser proporcionales al ángulo del sector que los representa, siendo toda la circunferencia la totalidad del tráfico.

El uso de este tipo de gráficos presenta grandes ventajas, pero también grandes inconvenientes. Como principal aspecto favorable, cabe destacar, la facilidad que supone su interpretación. No es necesario explicar a quien lo observa que está contemplando partes de un todo, apreciando, con relativa facilidad, la proporcionalidad entre las partes. Además, como se puede valorar, la figura es muy fácil de construir.

En contra, entre las desventajas descubrimos que el grafico proporciona información acerca de las partes de un todo, pero sin especificar el valor absoluto del todo y los valores parciales de las partes. Además, cuando el número de categorías en las que podemos dividir la variable es muy elevado (figura 42) resulta inviable apreciar los distintos valores o al menos la información es confusa.

Asimismo, resulta difícil comparar los diferentes sectores de un mismo grafico. En la figura (figura 41), es muy complicado, más bien imposible, apreciar la diferencia de tamaño entre los diferentes sectores, de no ser por los guarismos que indican el porcentaje de superficie de cada uno de ellos.

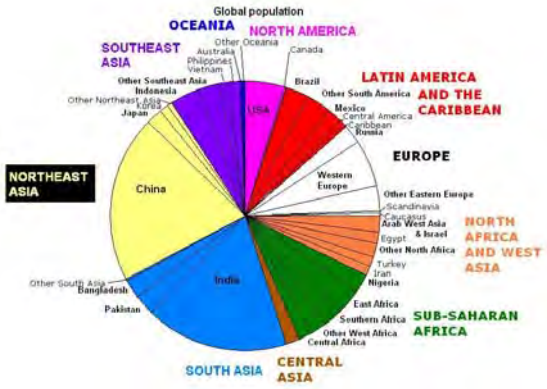


Figura 42. Diagrama circular de una variable con elevado número de categorías.

Por otro lado, se adivina igualmente complejo comparar datos entre diferentes gráficos. Como en el caso anterior, mediante la simple contemplación de los dos gráficos de la figura (figura 43), difícilmente un observador podría afirmar, entre el sector marcado con un 30% del grafico de la izquierda y el marcado con un 32% del de la derecha, cuál de ellos es mayor, sin la ayuda de los guarismos. Más aun, y afectado por sus diferentes posiciones relativas, sería del todo imposible aseverar que el primero de los sectores indicados es mayor que cualquiera de los sectores marcados con un 25%, en el grafico de la derecha.

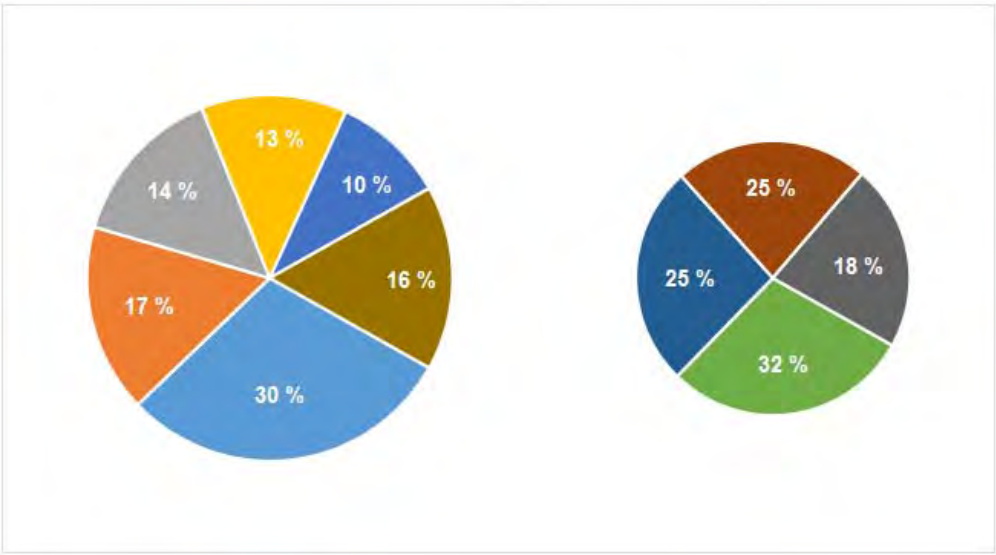


Figura 43. Gráficos circulares con diferentes tamaños de sectores

Todo lo anterior, es debido al inconveniente que presenta este tipo de gráficos para proporcionar valores exactos de los datos representados. Solo resultan útiles, por tanto, cuando el número de categorías es escaso y la diferencia de tamaño entre sectores es elevada.

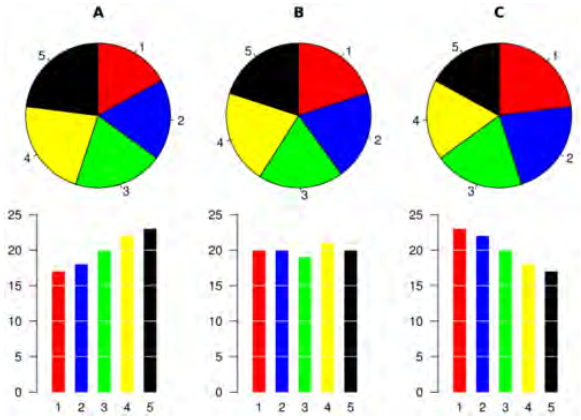


Figura 44. Gráficos circulares y gráficos de barras representando los mismos datos.

Se puede afirmar, que en la mayoría de los casos, la información presentada mediante gráficos circulares se representa, de una forma más adecuada para la comprensión de la mente humana, mediante gráficos de barras, gráficos de cajas o gráficos de puntos. En la figura (figura 44) se aprecia como los gráficos de sectores no son adecuados para apreciar las pequeñas diferencias entre los datos de los diferentes sectores. No así los gráficos de barras, que permiten apreciar estas diferencias con total nitidez. En contra, los primeros dan la sensación de contemplar las partes de un todo, efecto que no proporcionan, en absoluto, los segundos.

4.9 Treemap o mapa de árbol.

En 1990, el PhD Ben Shneiderman diseñó una herramienta para representar gráficamente, de forma interactiva, grandes volúmenes de información estructurada jerárquicamente [32]. Muchas de las estructuras conocidas, y aunque, aparentemente, no lo reflejen, son estructuras jerárquicas. Índices, organizaciones empresariales y familiares, direccionamiento de internet, estructuras de directorio de ficheros, etc., son algunos ejemplos de estructuras jerárquicas cotidianas. Formas tradicionales para simbolizar estas estructuras son la representación mediante esquemas (figura 45) y la representación mediante diagrama en árbol (figura 46).

El método de visualización de la información, propuesto por el profesor Sheneiderman, se denomina treemap o mapa de árbol. La representación de grandes estructuras de directorios de ficheros, almacenadas en los discos duros de los ordenadores, fue la motivación para el desarrollo de esta herramienta. Sheneiderman propone este método aprovechando *"la capacidad humana para reconocer el contenido de una imagen mucho más rápidamente de lo que puede explorar y entender una frase de texto, ya que tiene más perfeccionada la aptitud mental de reconocer la configuración espacial de los elementos de una imagen y observar las relaciones entre estos de modo instantáneo"* [33].

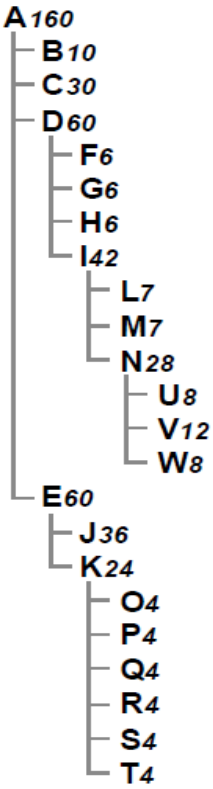


Figura 45. Esquema

Para comprender mejor el planteamiento del PhD Sheneiderman, compararemos los métodos de representación grafica mediante esquemas y diagramas de árbol, frente al método de mapa o treemap, valorando las ventajas que este ultimo proporciona sobre los otros dos. Esta comparación esta realizada sobre la que Shneiderman propuso en su trabajo [32] en 1991. Como se puede apreciar (figura 45), la representación en esquema visualiza una estructura de directorio de ficheros, representada por "A" en la figura, con un volumen total o peso de 160. Esta estructura A, directorio raiz, se subdivide a su vez, en un primer nivel, en cuatro directorios: B, C, D y E, con volumenes de 10, 30, 60 y 60 respectivamente. El resto de la figura es facilmente interpetable, siguiendo las pautas descritas. Como se puede observar, esta es la representacion descriptiva de la estructura de directorio de Windows, Linux y todas sus distribuciones. El diagrama de arbol (figura 46) representa exactamente lo mismo, pero en sentido horizontal.

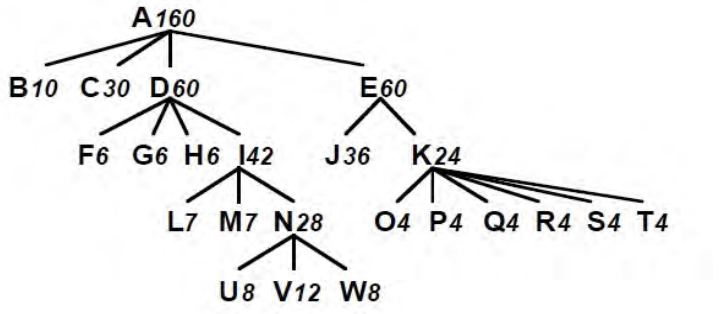


Figura 46. Diagrama de árbol

Aunque estas representaciones son muy practicas para ciertos usos, presentan serios inconvenientes para otros, de los que más adelante se hará empleo para el desarrollo de este trabajo. El primero de estos problemas, y dirigido a la representación

en esquema (figura 45) es la cantidad de líneas necesarias para poder simbolizar todos los ficheros almacenados en un disco duro. Para un conjunto de miles de ellos, la longitud de la representación seria sencillamente intratable. Mucho peor es el caso de la representación en diagrama de árbol (figura 46), ya que se extiende horizontalmente. No habría superficie (papel o pantalla) que los albergara todos, y que permita adquirir una sinopsis de sus dimensiones, relaciones y dependencias. Esto se agrava mucho más, si cabe en ambos casos, utilizando los nombres completos de los ficheros.

Pero mucho más trascendental que lo expuesto hasta el momento, es el hecho que estas organizaciones de la información no proporcionan noción alguna acerca de los nodos. En el ejemplo, el profesor Shneiderman introduce en las dos formas de representación grafica, vistas hasta ahora, un número que representa el volumen del fichero. Este dato adicional, proporciona información acerca de las proporciones entre los nodos, pero es necesario interpretar esta información para crearse una imagen de la misma. Esta misma información, representada mediante la herramienta treemap o mapa de árbol, intenta aportar ese conocimiento de forma grafica, que, en consonancia con lo expresado por el PhD Kamada [33], es mucho más fácilmente inteligible por la mente humana.

Como se entrega este conocimiento, se puede interpretar mediante el grafico de la figura (figura 47). Este es el mapa de árbol que representa la misma información que el esquema (figura 45) y el diagrama de árbol (figura 46), y el modo en el que se construye. El rectángulo inicial representa la totalidad del árbol, con un volumen global de 160, como se ha expresado anteriormente. Este se subdivide en cuatro ramas, B, C, D y E, con diferentes pesos (figura 47.a). Estos cuatro hijos, con sus diferentes pesos, se representan, por tanto, con diferentes superficies, proporcionales al volumen expresado al lado del nombre. La representación de estos nodos se realiza en sentido vertical. Resulta evidente, de este modo, la afirmación formulada anteriormente, en la que se hacía referencia al hecho de ser más fácilmente interpretable por la mente humana las proporciones entre los nodos por medio del grafico que por el número adjunto.

Continuando con el ejemplo, en el siguiente paso avanzaremos por las ramas D y E. El nodo D se subdivide a su vez en otros cuatro nodos, F, G, H e I, de segundo nivel (figura 47.b). Cada uno de estos nodos tiene su propia dimensión: 6, 6, 6 y 42, cuya suma es el volumen total del nodo padre, en este caso, 60. En la figura volvemos a representar los nodos, ahora, dentro de la superficie D, dotándoles de superficies proporcionales a sus valores. El sentido de las superficies es ahora horizontal, para que de esta forma no se confundan con la división anterior. El proceso para el nodo E, con un volumen de 60, lo particionamos en otros dos, J y K con volúmenes de 36 y 24,

respectivamente. De esta forma quedan representados los nodos hasta un segundo nivel. El proceso para la representación de los niveles tercero y cuarto, es exactamente el mismo.

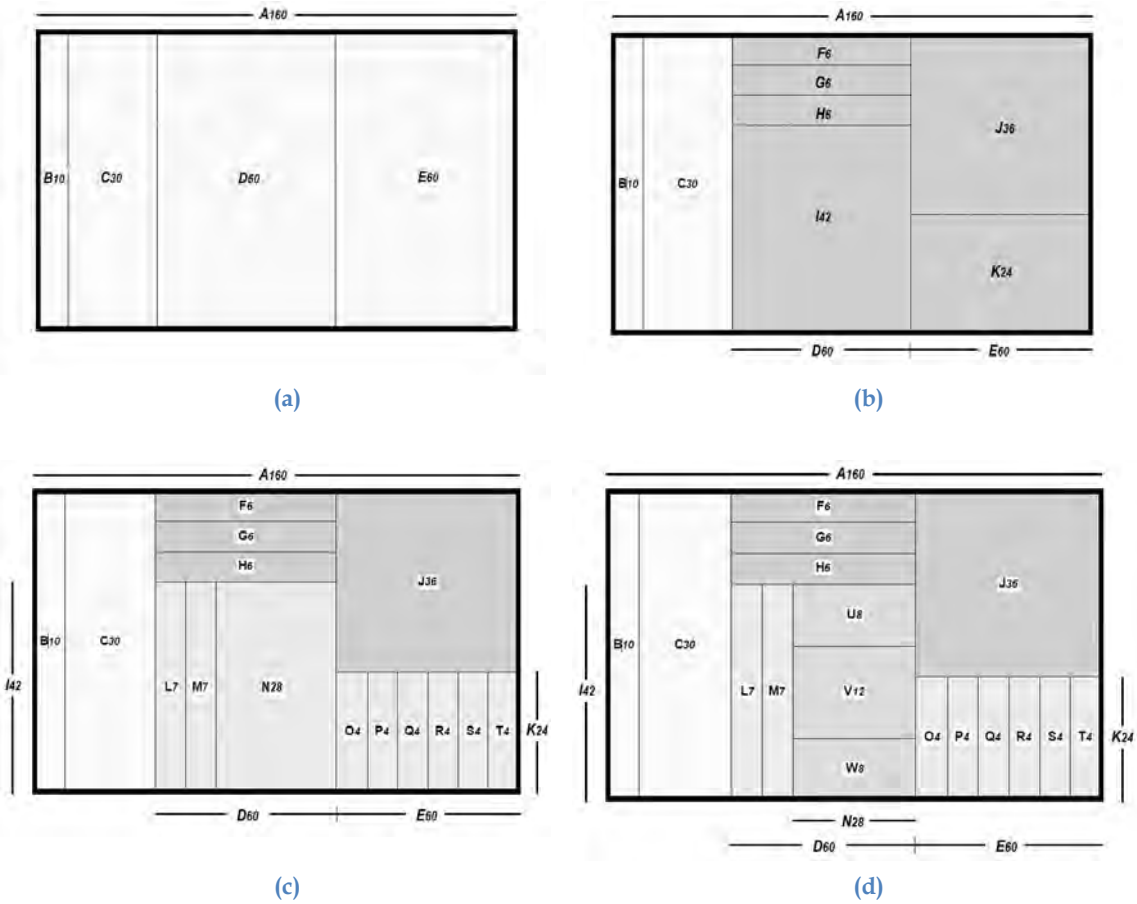


Figura 47. Treemap o mapa de árbol.

Se comprenderán ahora, las afirmaciones realizadas hasta el momento. Además de la información acerca de las relaciones padre-hijo, que las tres representaciones graficas proporcionan, la representación mediante treemap facilita información adicional, una interpretación más fácil y rápida de la misma y algunas ventajas a la hora de representarla. Estas mejoras podemos apreciarlas en la recopilación de los tres tipos de gráficos vistos hasta el momento (figura 48).

Inicialmente, como ventaja más importante, el árbol se representa en un espacio finito (figura 48.c). Una vez dimensionado el árbol, como se puede apreciar en el rectángulo de la figura, este no crece, ni disminuye, sino que continuara particionandose. No así, el esquema (figura 48.a), que crece a lo largo de la figura, llegando a ser intratable, como se mencionó anteriormente, ni el diagrama en árbol

(figura 48.b), que aumenta sus tamaño en sentido horizontal, no pudiendo ser albergado en ninguna superficie bidimensional.

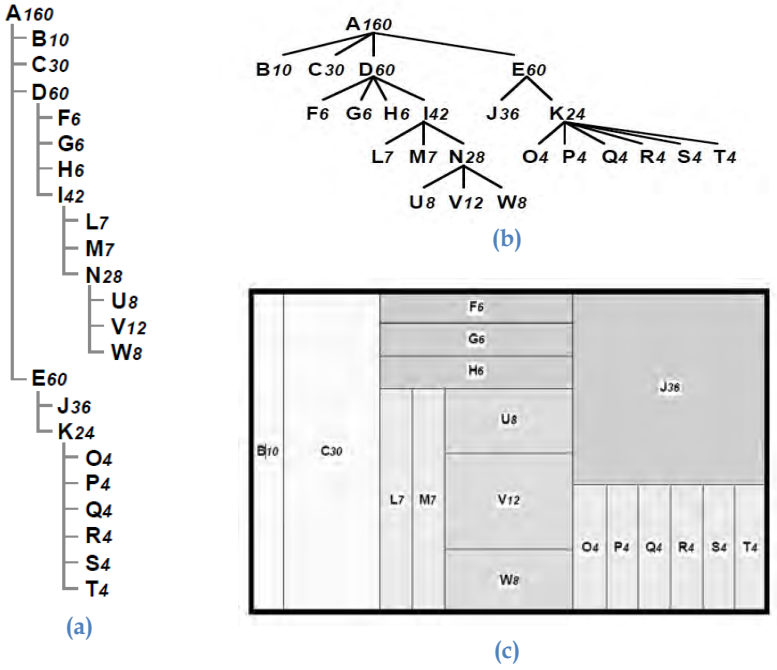


Figura 48. Recopilación de esquema, diagrama de árbol y treemap.

Por otro lado, observando en conjunto las tres representaciones descritas, se advierte que las tres proporcionan información del tamaño de los nodos, y por ende, la relación proporcional de los mismos. Resulta evidente, simplemente inspeccionando los gráficos, que esta es mucho más fácil e intuitiva en el caso de la figura treemap.

Finalmente, se puede añadir información adicional, utilizando ciertos códigos. Siguiendo el ejemplo del profesor Shneiderman [32], se pueden codificar los tipos de ficheros mediante colores. A modo de ejemplo, los ficheros del sistema podrían ser de color verde; los archivos de imagen y video, de color rojo; archivos de texto en color amarillo; los archivos de audio de color azul y el resto de ficheros, que no son parte de estudio, en gris. Así, en la estructura jerárquica de directorios, se pueden apreciar los tipos de archivos que lo componen. Además, en esta representación, se puede analizar donde se encuentran los ficheros de mayor tamaño y de qué tipo son. Esto permite evaluar, de modo global, que tipos de ficheros son los que ocupan el mayor porcentaje de espacio en disco, simplemente observando que color es el predominante. Resultado de este análisis y como ejemplo, una consecuencia de la interpretación de la figura, podría ser, por defecto, comprimir este tipo de ficheros.

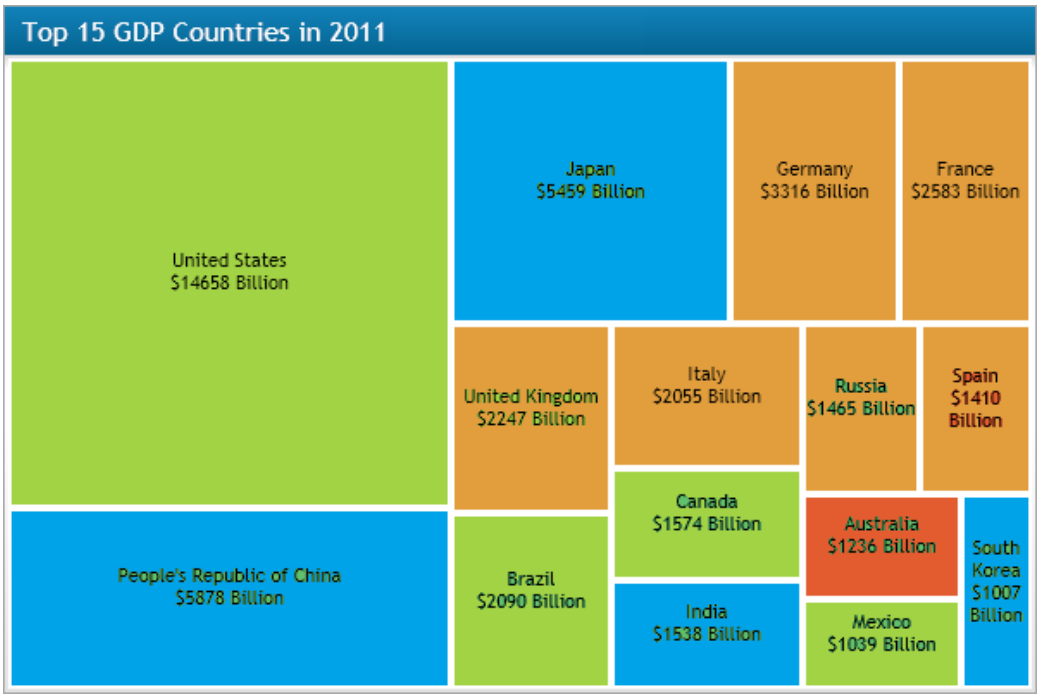


Figura 49. Distribución mediante treemap de los 15 países con mayor PIB en 2011.

Para concluir este apartado, cotejaremos lo expuesto, analizando la distribución por países, de aquellos 15, que poseen mayor PIB (figura 49) del año 2011. Simplemente inspeccionando la figura, evidentemente soslayando el dato numérico, se observa que el país con mayor PIB es Estados Unidos. Por el contrario, el de menor PIB de los 15 es Corea del Sur. Representando los diferentes continentes por colores, advertimos, igualmente, que el continente que suma mayor PIB, es el americano. Oceanía, por el contrario, es el continente con una suma menor de PIB. Resulta de este modo, muy fácil extraer la información y procesarla por la mente humana, para obtener una conciencia situacional del estado del elemento sujeto del estudio.

4.10 Diagrama de coordenadas paralelas *(parallel coordinate plots)*.

La principal virtud de un diagrama de coordenadas paralelas es que permite representar un sistema n-dimensional sobre un plano, esto es, plasmar n dimensiones, únicamente en dos dimensiones (figura 50). Este sistema de representación es el que permite simbolizar más dimensiones de los datos estudiados. La técnica que utiliza es la de que cada dimensión se representa sobre un eje vertical, espaciado entre ellos por una distancia constante. Los valores representados en los ejes

son escalados para que todos tengan la misma altura. De esta forma se aprecia visualmente la dependencia entre variables, además de evidenciar similitudes y/o diferencias y descubrir patrones de comportamiento. Las variables representadas, pueden ser cuantitativas y cualitativas. En el ejemplo (figura 50), se pueden apreciar ambas: la edad seria un modelo de variable cuantitativa, mientras que la raza o el sexo lo son de variables cualitativas.

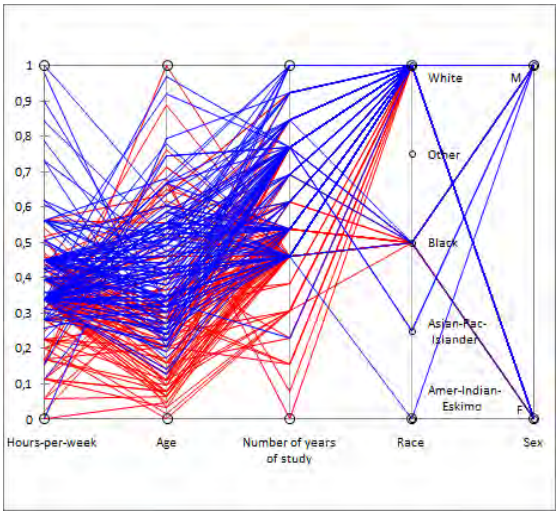


Figura 50. Diagrama de coordenadas paralelas

La representación no da prioridad a una dimensión sobre otra. Representa una combinación de datos simultáneamente, facilitando la apreciación visual de todos ellos, en conjunto. De esta forma el analista no tiene que conocer, a priori, cual es la variable que va a proporcionar la información relevante. Sin embargo, aunque, teóricamente, se pueden representar cualquier número de dimensiones, no es aconsejable representar más de 5 dimensiones [34], por lo que es necesario elegir aquellas consideradas más relevantes y que proporcionan más información del estado del sistema.

	name	economy (mpg)	cylinders	displacement (cc)	power (hp)	weight (lb)	0-60 mph (s)	year
1	AMC Ambassador Brougham	13	8	360	175	3821	11	73
2	AMC Ambassador DPL	15	8	390	190	3850	8.5	70
...	...	...	...	...	...	...	...	...
207	Ford Maverick	24	6	200	81	3012	17.6	76
208	Ford Mustang Boss 302		8	302	140	3353	8	70
209	Ford Mustang Cobra	23.6	4	140		2905	14.3	80
...	...	...	...	...	...	...	...	...
405	Volvo 245	20	4	130	102	3150	15.7	76
406	Volvo 264GL	17	6	163	125	3140	13.6	78
407	Volvo Diesel	30.7	6	145	76	3160	19.6	81

Tabla 2. Datos de 407 modelos de automóviles diseñados entre 1970 y 1982.

Para conocer mejor como proporciona información un grafico de coordenadas paralelas [35], interpretamos un ejemplo con datos relativos a 407 modelos de automóviles (tabla 2) diseñados entre 1970 y 1982 [36]. Los datos estudiados en el ejemplo son: consumo (millas por galón, MPG), numero de cilindros, cilindrada, potencia, peso, aceleración y año de fabricación.

Inicialmente, se percibe a primera vista, la alta densidad de líneas que se le proporcionan al observador para analizar (figura 51). De la forma que se presenta el grafico, parece que se va a obtener muy poca información, debido a lo enmarañado del mismo. Se estima, por tanto, necesario aplicar técnicas que proporcionen claridad al método de presentación de la información. Estas técnicas no son otras que las del filtrado. A continuación se exponen unos casos de éxito en los que el filtrado se realiza por diferentes criterios para obtener diferentes resultados.

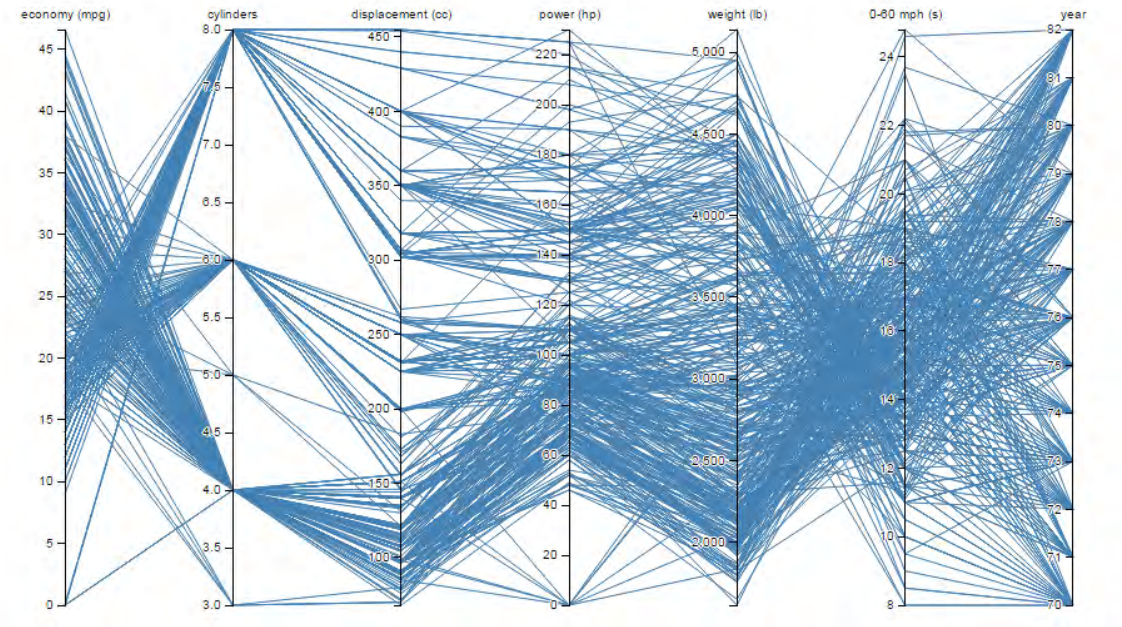


Figura 51. Diagrama de coordenadas paralelas de 407 modelos de automóviles diseñados entre 1970 y 1982 [36].

El primero es, en contra de lo expresado anteriormente, sin ningún tipo de filtro. Se pueden estudiar (figura 51) tendencias globales de los datos analizados. En la figura se aprecia, analizando la segunda columna, *número de cilindros del motor*, que prácticamente la totalidad de los vehículos se agrupan en torno a tres puntos concretos. Estos son: 8, 6 y 4 cilindros. Existen elementos en 3 y 5 cilindros, pero son mínimos. Otra tendencia observable en la misma figura, es la acumulación de concurrencias en torno al intervalo comprendido entre los valores 0 y 150 de la columna *cilindrada*. Por último se aprecia otra acumulación dentro del intervalo comprendido entre 60 y 100, de la columna de *potencia*. Esto significa que la mayor parte de los vehículos estudiados

se encuentran entre los valores mencionados de cada una de las columnas de valores tratadas.

Un segundo tipo de filtrado seria, aquel por el cual, elegimos una única ocurrencia de entre todas las analizadas. En este caso (tabla 3), apreciamos la tendencia de los valores de un vehículo en particular, estudiado dentro del contexto del resto de vehículos (figura 52). De esta forma, analizamos el comportamiento del elemento seleccionado, de los valores de sus variables, comparándolas con la totalidad de valores de las variables de los demás vehículos.

401	Volkswagen Type 3	23	4	97	54	2254	23.5	72
402	Volvo 144EA	19	4	121	112	2868	15.5	73
403	Volvo 145E (Wagon)	18	4	121	112	2933	14.5	72

Tabla 3. Datos del modelo de vehículo seleccionado [36].

Así, observamos que el vehículo seleccionado para estudio, tiene un motor de cuatro cilindros, valor que habíamos determinado, en la figura anterior (figura 51), como uno de los más usuales en el apartado número de cilindros del motor. De igual modo, el dato del vehículo en la columna cilindra es de 121, medida que se encuentra

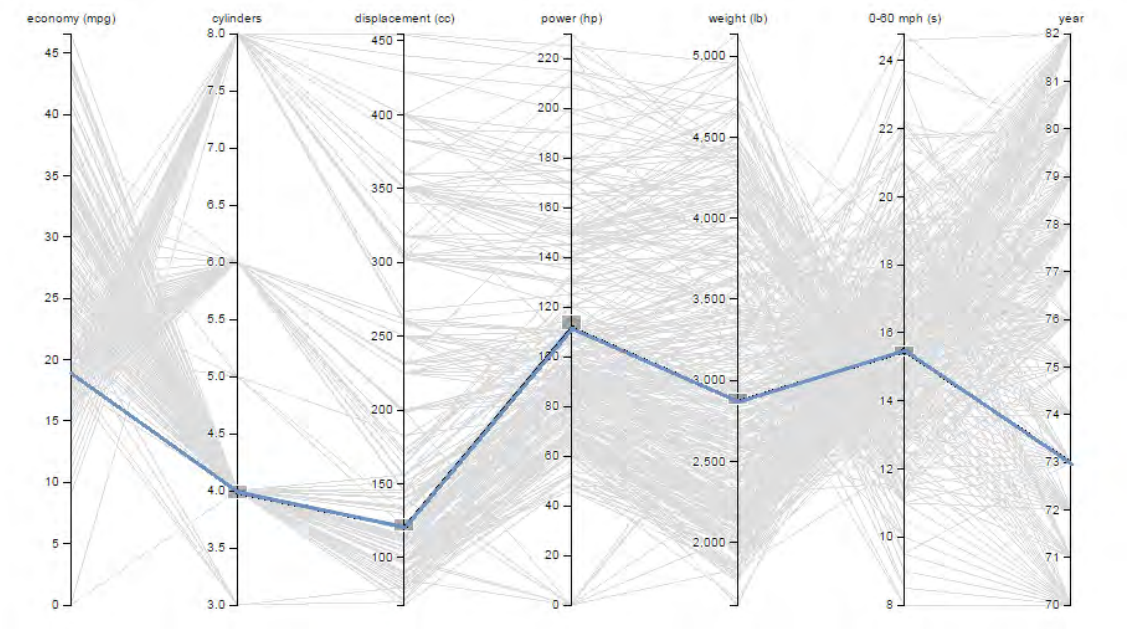


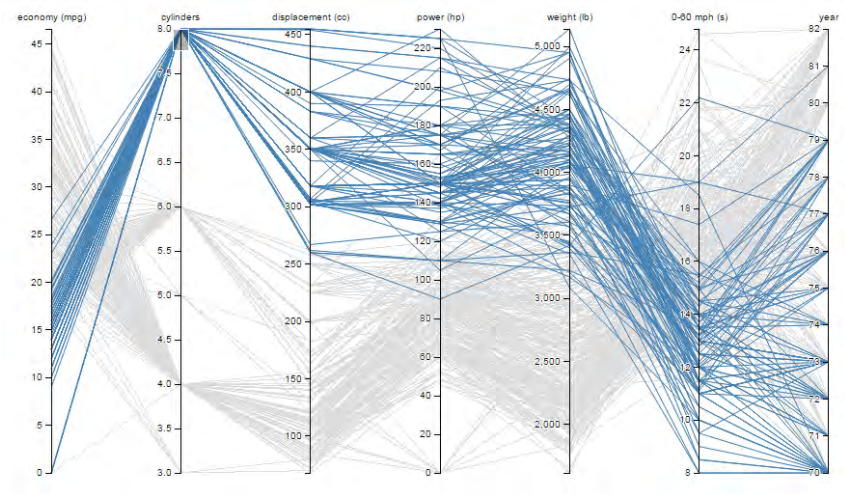
Figura 52. Diagrama de coordenadas paralelas del modelo de vehículo seleccionado [36].

entre los valores 0 y 150, que, igualmente que en el caso del numero de cilindros, son los valores habituales de los vehículos analizados. Por el contrario, si observamos el eje del dato de potencia, el vehículo estudiado presenta una magnitud de 112. Si recordamos, los valores habituales se situaban entre 60 y 100, por lo que este dato se desvía de los considerados frecuentes. De este modo, como se ha podido observar, obtenemos una comparativa rápida de un elemento, con el conjunto de sujetos estudiados.

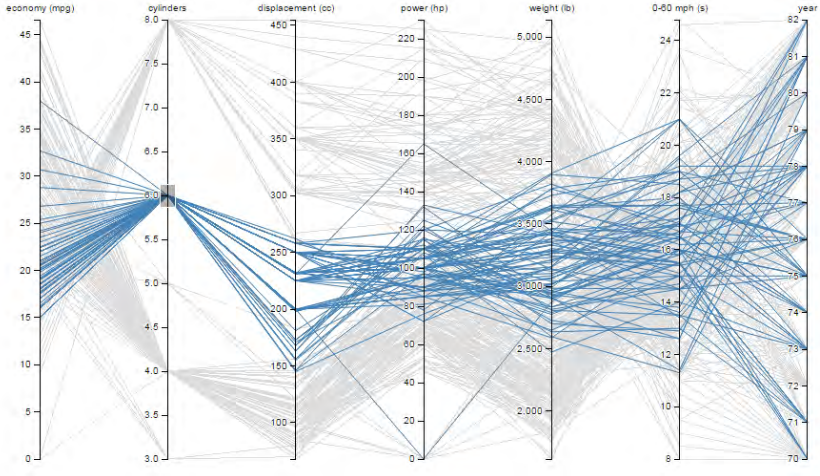
Otro método de trabajo mediante diagramas de coordenadas paralelas consiste en separar los elementos de análisis en base a un criterio, agrupándolos por un perfil concreto. En nuestro caso de trabajo nos volveremos a centrar en el dato *número de cilindros del motor*, representado en el segundo eje de valores, que como se expuso anteriormente, los valores más habituales son 4, 6 y 8. Dividiendo el grafico por estos criterios (figura 53) se puede centrar la atención, más fácilmente, en el comportamiento de estos grupos, de cada uno de ellos por separado. Esto permite comparar sus perfiles, comprobando similitudes de comportamiento de sus elementos y las diferencias con el resto de los grupos. Así, se puede observar que los vehículos con mayor numero de cilindros, 8 en concreto (figura 53.a), tienden a tener mayor cilindrada, mayor potencia y, como consecuencia de todo ello, mayor peso que el resto. Una curiosidad que evidencia la grafica es que, este tipo de coche, era muy habitual hasta el año 79, dejándose de fabricar a partir de entonces. Este dato se evidencia por el simple hecho de filtrar el grupo y solo presentar los datos de los mismos.

Continuando con el estudio del grafico (figura 53.b) se observa que los vehículos con 4 cilindros poseen una cilindrada media, al igual que la potencia y son más ligeros que el grupo estudiado. Por contra, estos modelos de vehículos (figura 53.b) tienen una aceleración muy superior a los de la categoría anterior, estando situados en la parte central de la escala, mientras que los de 8 cilindros, se agrupan en torno al fondo de la misma. En relación con los años de producción, estos vehículos han sido fabricados, por igual, a lo largo de doce años, desde el 70 al 82 objeto de este estudio.

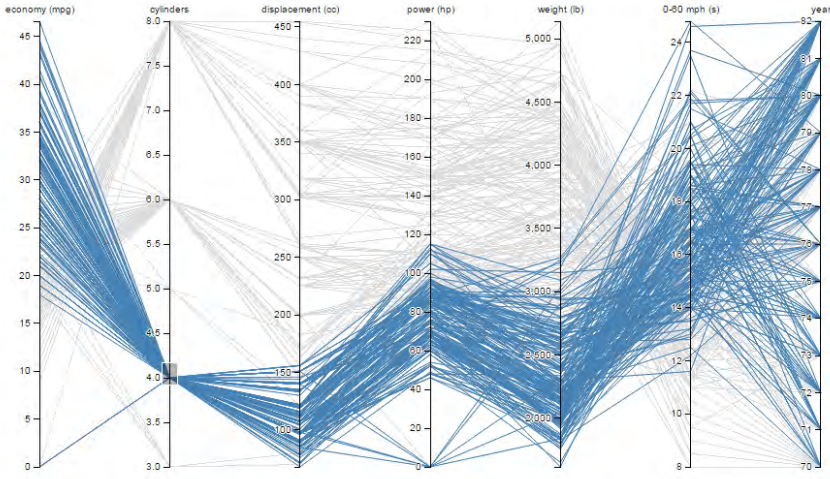
Finalmente, los vehículos diseñados con 4 cilindros, tiene menor potencia que los otros dos grupos, ya estudiados. Son también, por lo general, menos pesados y poseen, como norma, muy superior aceleración. Al igual que los vehículos de 6 cilindros, los años de fabricación son la totalidad de los doce, objeto de este estudio. Como se ha visto, este procedimiento permite conocer comportamientos típicos de grupos de elementos seleccionados mediante el valor específico de una de sus variables.



(a)



(b)



(c)

Figura 53. Selección de trazas de un diagrama de coordenadas paralelas en base a un dato [36].

El último procedimiento de empleo de los ejes de coordenadas paralelas, es realizar una selección multicriterio (figura 54). En el ejemplo se han seleccionado los vehículos con 8 cilindros, una cilindrada entre 300 y 400 cc, con una potencia entre 130 y 230 caballos y un tiempo de aceleración de 0 a 60 mph entre 12 y 25 sg. De esta forma, obtenemos un conjunto de vehículos con un perfil muy específico que los hace muy similares. Los elementos seleccionados mediante este método son altamente comparables, permitiendo hacer una selección, entre ellos, más afín a la búsqueda planteada.

Adicionalmente, buscando una mayor claridad visual de la representación grafica, cualquier herramienta de gráficos de coordenadas paralelas, permite modificar el orden de los ejes e invertir la escala de cualquiera de ellos a voluntad del analista.

En resumen, la técnica de representación mediante gráficos de coordenadas paralelas permite tratar un conjunto de elevado número de datos, proporcionando información relevante en las siguientes líneas:

- permite comparar todos los elementos en conjunto (figura 51),
- proporciona una visión individual frente al conjunto (figura 52),
- discrimina conjuntos por el valor de un dato (figura 53) y
- realiza una selección restringiendo valores de varios datos (figura 54).

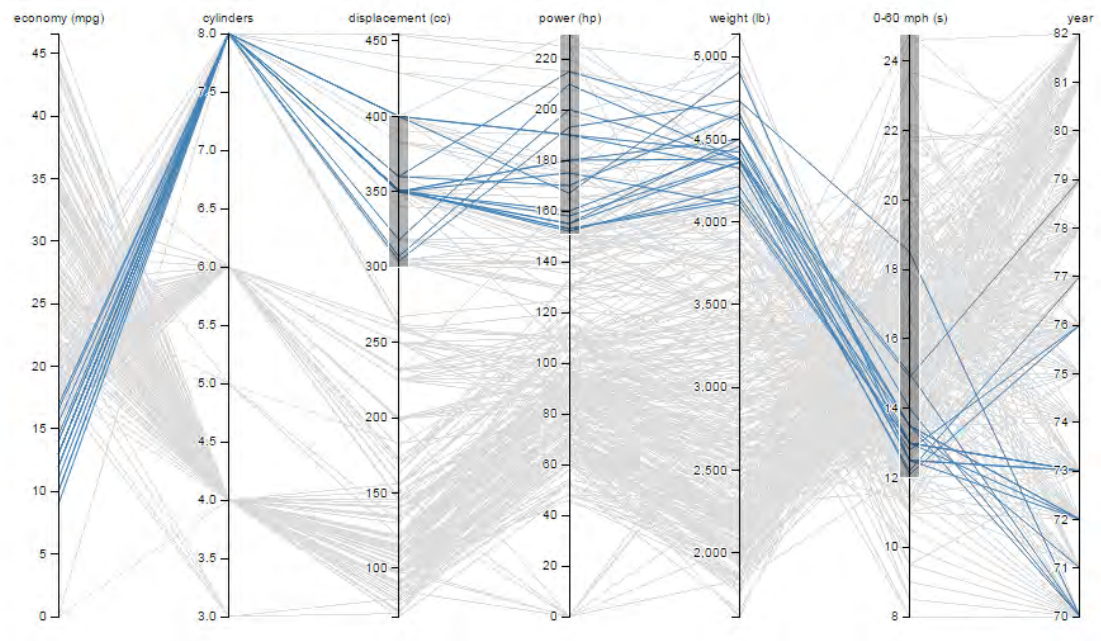


Figura 54. Selección de trazas de un diagrama de coordenadas paralelas en base a una agregación de datos [36].

Sin embargo, y a pesar que esta representación permite, como se expreso al principio de este punto, mostrar varias dimensiones, es necesario simbolizar otra nueva dimensión que proporcione una imagen dinámica a través del ***tiempo*** [34]**.** Este nuevo elemento, el tiempo, proporciona al analista un elemento de comparación, que le permite valorar el comportamiento dinámico del sistema y de esta forma comprender si está sucediendo algo anómalo. Una técnica que proporciona esta visión evolutiva, como veremos más adelante, es mediante una matriz que represente los mismos elementos en diferentes momentos.

4.11 Flujo de conexión.

Esta es una sofisticada técnica grafica, algo complicada inicialmente, pero muy ilustrativa cuando se comprende su significado, del estado de las conexiones en un intervalo de tiempo. El flujo de conexiones, o *connection river* del inglés, utiliza algunas técnicas similares a las de diagrama de coordenadas paralelas (ver apartado 4.10) distribuyendo los datos a lo largo de seis ejes (figura 55), los tres primeros representando datos de los elementos emisores (figura 55.1/2/3) y los tres últimos de los elementos receptores (figura 55.4/5/6) [37].

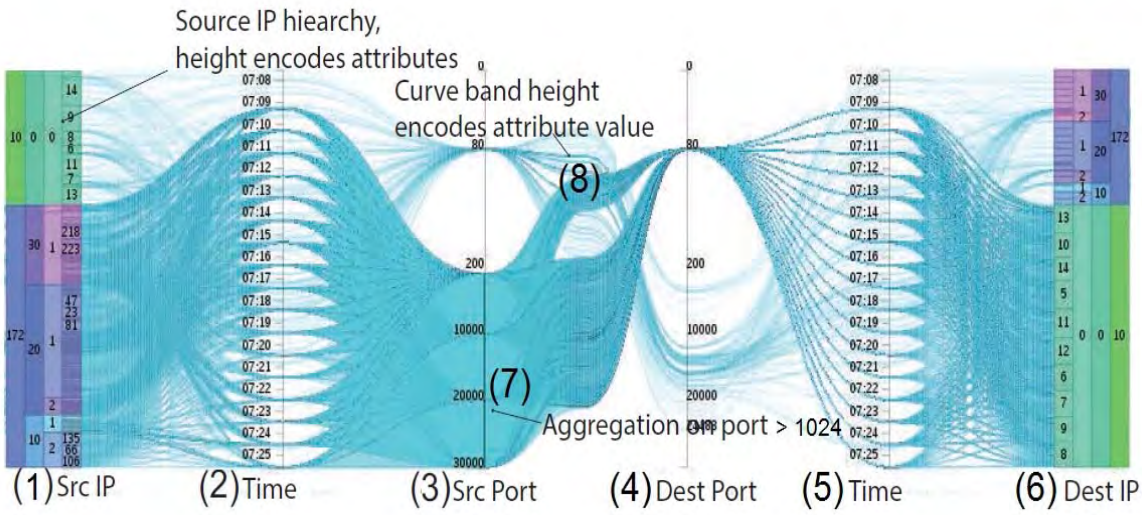


Figura 55. Flujo de conexiones [37].

El primer eje (figura 55.1) representa las direcciones IP de los equipos emisores. Esta representación se confecciona mediante un treemap (ver apartado 4.9), en el que se simbolizan las diferentes redes o equipos (el eje puede representar un único equipo

mediante una única dirección IP) a través de los cuatro bloques de la dirección IP en su formato IPv4. El último eje (figura 55.6) representa exactamente lo mismo, en relación a los equipos receptores.

En el segundo eje (figura 55.2) se incorpora una gradación de tiempos que comprenda el intervalo de tiempo, propósito del estudio. La escala de tiempos es dinámica, es decir, el tiempo se reparte en partes proporcionales a lo largo del eje, por lo que, siendo este de una longitud fija, deberá modificarse cada vez que se elijan tiempos distintos para su estudio. Al igual que en el caso anterior, el penúltimo eje (figura 55.5) refleja el mismo tipo de dato para los equipos receptores. Se significa que no tienen obligación de ser los mismos tiempos, pero debido a la velocidad de las comunicaciones, estos no deben ser muy dispares.

Por último, los dos ejes centrales, tercero y cuarto (figura 55.3 y 4) reflejan los puertos de comunicación. Al igual que los dos casos anteriores, el eje de la izquierda, tercero, son los puertos de transmisión, y el eje de la derecha, cuarto, los puertos de recepción. Una diferencia con los otros cuatro ejes, es que este se divide en dos tramos: uno que comprende los puertos reservados, del 1 al 1024, y otro con el resto, representando estos últimos en un tramo menor (figura 55.7). Los puertos reservados son los más habituales y se representan de forma más nítida. Sin embargo, si es necesario un estudio más detallado del resto de puertos, se pueden acotar estos, disminuyendo el número de los mismos para su estudio.

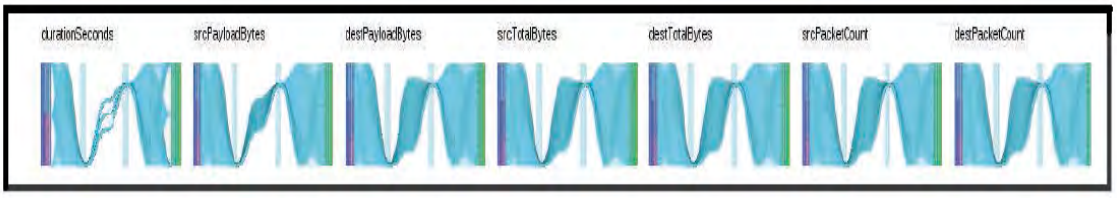


Figura 56. Visualización de diferentes variables en flujo de conexiones

Hasta aquí las similitudes con el diagrama de coordenadas paralelas. El flujo de conexiones posee otras características que lo diferencian y que potencian la representación de determinados comportamientos en las comunicaciones IP. En la figura (figura 55.8) se puede discernir como cada conexión tiene una curva diferente, que alcanza alturas distintas. De esta forma se pueden visualizar y valorar diferentes variables (figura 56) como la duración en segundos de la transmisión, número de bytes, tanto emitidos como recibidos u otras. De este modo se conforman diferentes curvas a diferentes alturas. Al seleccionar alguna de estas presentaciones, esta pasa a ser la presentación principal (figura 55).

Estas diferentes formas de las curvas pueden permitir visualizar comportamientos anómalos del trafico, como extracción de grandes cantidades de bytes mostrando un intenso tráfico mediante una gran altura de la curva, o todo lo contrario, como una denegación de servicio por el escaso tráfico existente, indicado mediante una curva baja. Estos aspectos serán tratados con mayor detalle más adelante.

Al igual que en las graficas estudiadas hasta el momento, al situar el puntero sobre los elementos de la grafica, se puede desplegar un cuadro de contexto con información adicional (figura 57) que ilustre las imágenes, complementándolas, y permita alcanzar un mayor entendimiento de los eventos que están ocurriendo y que son tema de estudio más detallado.

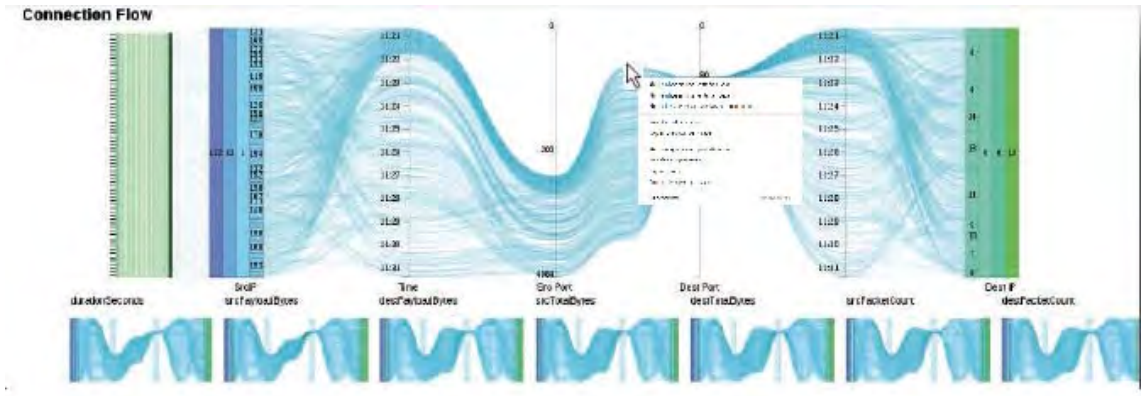


Figura 57. Visualización información adicional mediante cuadro de contexto.

4.12 Gráfico de anillo.

El gráfico de anillo es una herramienta visual que permite representar las comunicaciones entre equipos de una red, y con equipos en el exterior de la misma, en un instante de tiempo [37]. Esto se materializa mediante un conjunto de tres coronas circulares o anillos concéntricos, representando los tres primeros niveles de la jerarquía IP, en su formato IPv4 (figura 58), siendo el anillo más externo el que representa el primer bloque de 8 bits de una dirección IP. El segundo anillo representa al segundo bloque y el tercer anillo, al tercer bloque de la dirección IPv4.

Con la finalidad de una diferenciación más fácil, rápida y eficaz entre las direcciones internas y externas, las primeras se representan con colores que van desde el azul al morado. Análogamente, en las direcciones externas se utilizan colores que van desde el verde al rojo.

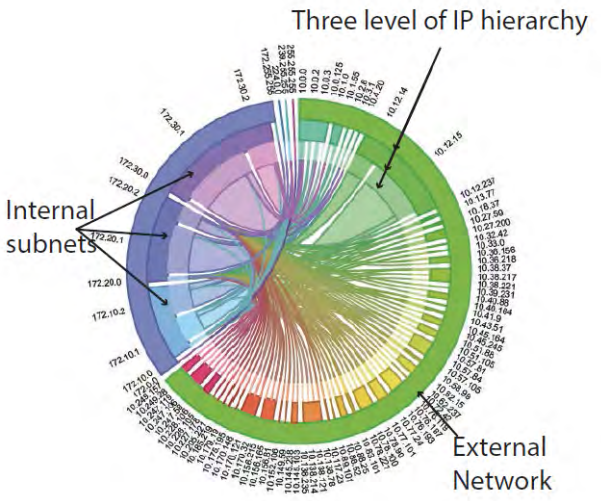


Figura 58. Grafico de anillo

Las direcciones IP representan a equipos tanto internos como externos. La relación de tamaños entre las direcciones internas y externas representa la relación entre las conexiones entrantes y salientes (figura 59). Cuando el número de conexiones externas es muy superior al de internas, predomina el color verde en el grafico (figura 59.a). Por el contrario, si las conexiones entre los equipos de la propia red son las más numerosas (figura 59.b), es el color azul el preponderante.

Las comunicaciones entre los diferentes equipos, ya sean internas o externas, se representan mediante líneas de conexión (figura 60). Estas van dirigidas desde el emisor hacia el receptor. Para identificar gráficamente el sentido de la comunicación, el color de la línea es el del emisor.

Al igual que todos los figuras estudiadas hasta el momento, la saturación de elementos hace que la información pueda quedar oculta entre el maremágnum de líneas y colores. Para facilitar la labor del analista, el grafico de anillo cuenta con la capacidad de seleccionar las direcciones IP a estudiar, o de eliminar las

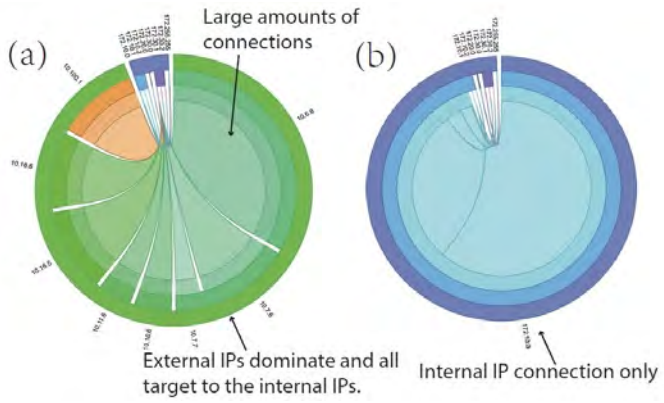


Figura 59. Grafico de anillo con predominancia de conexiones externas (a) e internas (b).

que no sean relevantes, además de permitir las acciones de "*deshacer y rehacer*", de tal forma que se pueda discernir hasta el más mínimo detalle en las conexiones.

Finalmente, y como se ha mencionado anteriormente, este tipo de grafico permite estudiar el comportamiento de las conexiones entre los equipos en un determinado instante de tiempo. Esto significa que no proporciona información a lo largo del tiempo, en un intervalo. Esto, como se verá más adelante, es posible conjugando esta técnica con otras que proporcionen persistencia y permitan un estudio temporal.

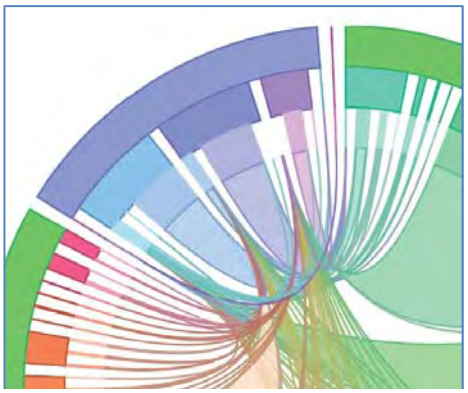
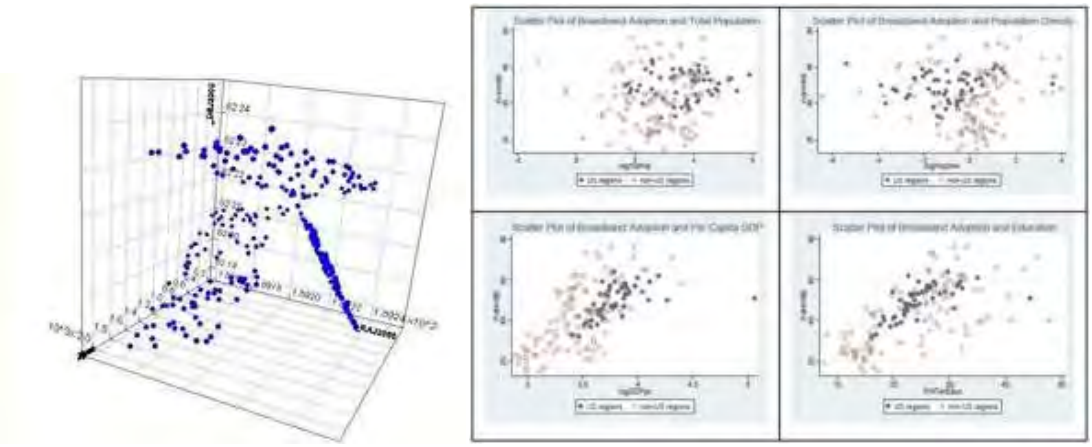


Figura 60. Detalle de conexiones de un gráfico de anillo.

4.13 Grafico de rejilla.

El grafico de rejilla, trellis plot en su denominación inglesa, no es un grafico al uso de los que se han estudiado hasta el momento. Se trata de un conjunto de gráficos, como los estudiados u otros no tratados en este estudio por no ser de interés, todos del mismo tipo, con distintos valores de las variables.

Originariamente, fue utilizado para representar gráficos de dispersión en 3D. Como ya se expuso (ver apartado 4.5), los gráficos de dispersión permiten plasmar los valores de dos variables, una independiente y otra dependiente, mediante ejes de coordenadas cartesianas. Pero, ¿cómo se puede representar esta misma dependencia cuando el número de variables que intervienen es de tres? Esta adversidad puede ser solventada mediante una matriz de diagramas de dispersión bidimensionales [34] que visualicen los diferentes planos que cortan el diagrama de dispersión tridimensional (figura 61). Los planos de corte deben ser equidistantes, para proporcionar una información más completa. En la figura, el diagrama de dispersión tridimensional (figura 61.a) se ha cortado mediante cuatro planos equidistantes (figura 61.b) Evidentemente, se alcanza mayor precisión cuanto mayor es el número de planos con los que cortamos el espacio tridimensional. Por el contrario, un mayor número de planos constituye un inconveniente de implementación y una mayor dificultad de comprensión. Por tanto, el número de planos a representar, debe ser un compromiso entre economía y eficacia. Ambos puntos, facilidad de comprensión y determinación del número de planos a proyectar, serán tratados más adelante.



(a) (b)

Figura 61. Representación de un gráfico de rejilla de cuatro cortes (b) de un diagrama de dispersión tridimensional (a).

El grafico de rejilla permite, también, representar un conjunto de datos que conformen un espacio continuo de valores. Al igual que en el caso anterior, el espacio tridimensional es atravesado transversalmente por planos paralelos equidistantes (figura 62). La sección cortada en cada uno de los planos se dibuja en el mismo. Lógicamente, el conjunto de todos estos planos representados en una matriz, conforma la figura del grafico de rejilla.

La finalidad buscada es la de permitir al analista comparar las diferentes figuras, estudiando similitudes que proporcionen patrones de comportamiento, y que permitan inferir conocimiento acerca de presuntas irregularidades, que bajo un análisis más detallado, conduzcan a juicios definitivos.

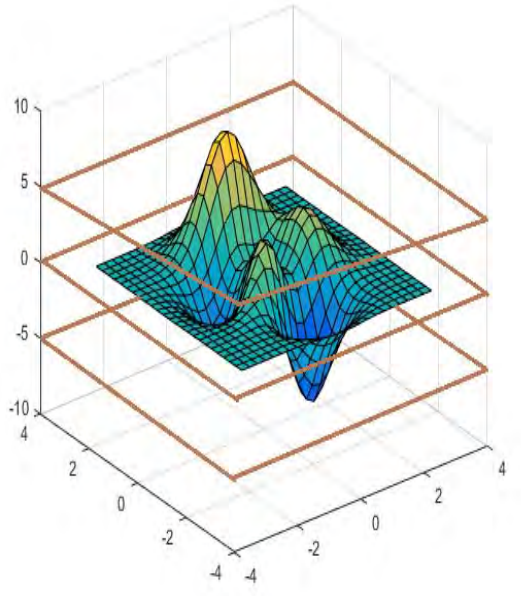


Figura 62. Seccionado de un espacio continúo de valores

Como ya se ha indicado, este tipo de grafico nació para representar los valores de variables, ya sea mediante gráficos de dispersión o de valores continuos, en tres dimensiones. Pero cabe una

tercera posibilidad, y es la de representar cualquier tipo de grafico y sus variaciones a lo largo del tiempo, manejando el tiempo como "tercera dimensión". En la figura (figura 63) está representada una matriz de 8*8 gráficos de coordenadas paralelas.

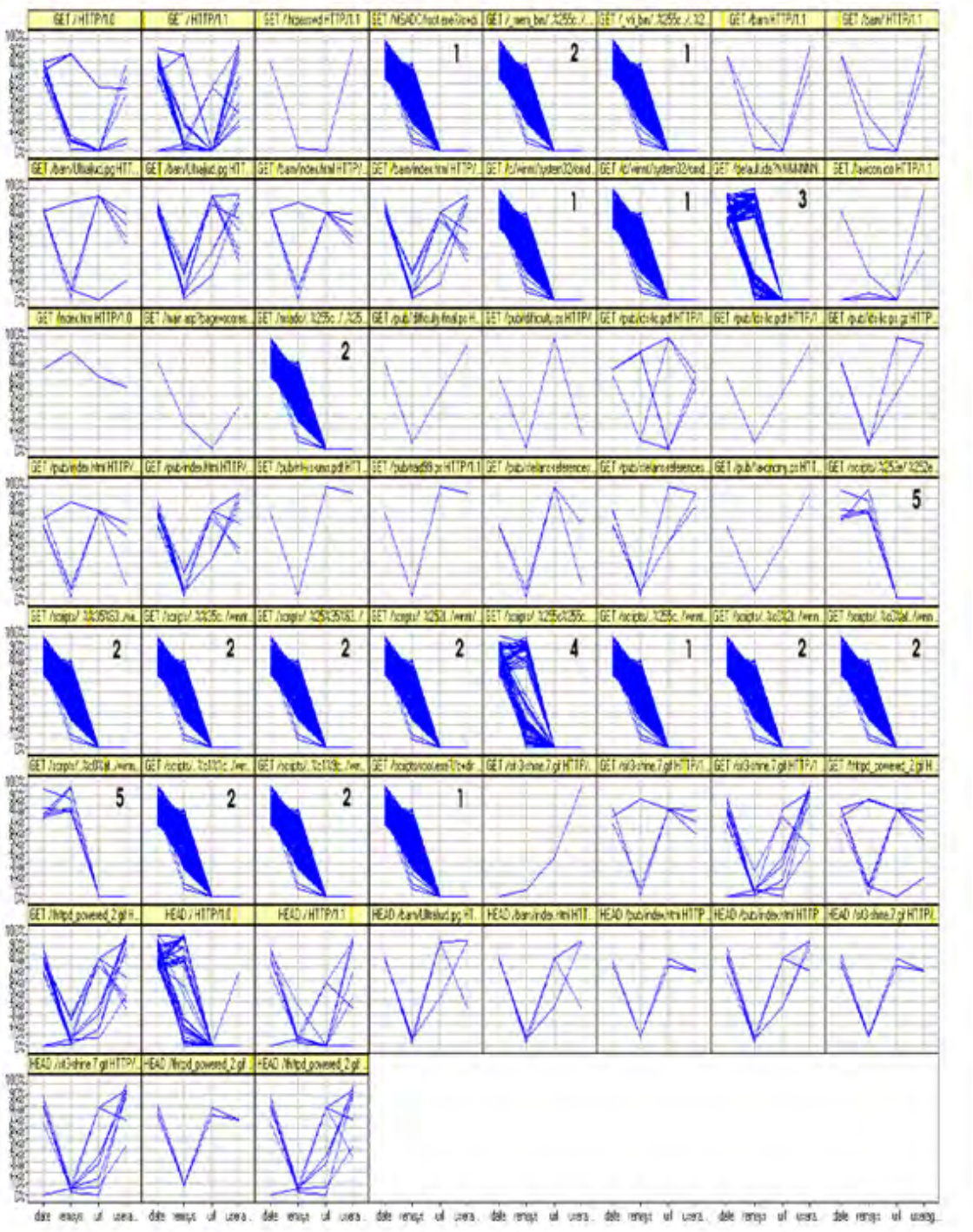


Figura 63. Trellis plot de 8*8 gráficos de coordenadas paralelas [34].

Estas representan diferentes situaciones del sistema estudiado a lo largo del tiempo. Como se puede apreciar en la figura los gráficos representados en cada una de las

celdas de la matriz no son gráficos de dispersión, ni una representación de un espacio continuo. Y aunque en este caso se trata de diagramas de coordenadas paralelas, podría ser cualquier otro tipo de gráfico que experimentara variaciones a lo largo del tiempo, como treemaps, gráficos de anillos u otros no examinados en este capítulo. De esta forma se puede comparar el estado de un sistema, con otro estado de ese mismo sistema, transcurrido un cierto intervalo de tiempo. Para que este tratamiento sea eficaz en el estudio de sistemas de telecomunicaciones, y por norma general, el intervalo de tiempo entre uno y otro debe ser pequeño, no superior a unos pocos minutos, como máximo. Resulta obvio, por tanto, que entre un gráfico y otro existe un espacio continuo de estados del sistema que no van a ser estudiados. Si algún evento ocurriera durante esos intervalos de tiempos, lo que evidentemente será lo más habitual, pasara inadvertido para el analista, no siendo consciente de lo ocurrido. Este hándicap, desaconsejaría, en una primera instancia, utilizar este método. Como se propondrá más adelante, esto puede ser resuelto mediante algunas modificaciones de fácil implementación. Imaginemos por un instante, que en lugar de ser cada celda una fotografía estática de la situación del sistema en ese instante de tiempo, esta fuera en realidad un recorrido animado de esas situaciones en el tiempo que va desde que se toma esa fotografía inicial, hasta el momento de inicio de la celda siguiente. El analista tendría una visión general de estados, por los que el sistema habría ido pasando. Es fácil deducir, que si es necesario visualizar varias celdas al mismo tiempo, estas deberán reducir su número, o de lo contrario, el cerebro del analista no será capaz de procesar toda la información ofrecida.

5 Correlación de grafismos con incidentes.

12.00 am. El comandante militar del Mando de Ciberdefensa, responsable de las políticas de Seguridad de la Información en los Sistemas de Información y Telecomunicaciones del Ministerio de Defensa, entra en la sala. Todos los oficiales, suboficiales y personal civil le reciben en pie. Como todas las mañanas, van a exponerse al General, y a todos los presentes, los incidentes de seguridad que han sucedido en los sistemas de información propios de las Fuerzas Armadas en las últimas 24 horas. Avanzan los minutos y en las pantallas aparecen los informes de los sucesos, mientras el oficial encargado explica lo ocurrido desde la reunión del día anterior. Uno tras otro se identifican los episodios acontecidos en los sistemas de información: hoy el protagonista ha sido APT 28, la semana pasada el problema lo protagonizó un caso de phishing, por cierto, con algún éxito, y anteriormente se había producido un caso de ransomware, del tipo cryptoLocker, que afecto a un miembro de las Fuerzas Armadas en su domicilio particular y fue investigado por si había trasladado el software malicioso a su ordenador de trabajo, con resultado negativo.

El autor ha de reconocer en este punto que, en un momento dado, su mente deja de prestar atención al relato de los sucesos y fija su interés en la elaboración del presente trabajo, perdiendo momentáneamente el discurrir de los acontecimientos, imaginando como esa información que está siendo trasladada, en su mayor parte en texto por las pantallas y comentada mediante la palabra por el analista correspondiente, podría ser desarrollada mediante gráficos, mucho más fácilmente reconocibles e interpretables por la mente humana y que permitieran al Comandante "*planear y ejecutar las acciones relativas a la ciberdefensa en las redes y sistemas de información y telecomunicaciones del Ministerio de Defensa u otras que pudiera tener encomendadas, así como contribuir a la respuesta adecuada en el ciberespacio ante amenazas o agresiones que puedan afectar a la Defensa Nacional*" [38] (trasládese esta misma aseveración, en los términos que correspondan, a Director Gerente en cualquier otra empresa). Este planteamiento, no es otra cosa distinta a que, como ya se expuso al principio, en el apartado 2, le permita alcanzar una "«*situational awareness*», *ese estado mental que debe alcanzar aquella persona que toma decisiones en una organización, como consecuencia del entorno en el que se desarrollan las acciones, la evolución de los acontecimientos y otros factores que puedan afectar a su progreso*".

Además, y como ya se expresó en el apartado 3, "*el sistema de representación debe estar capacitado para proporcionar información valida en muy diferentes escalas de decisión*". No es atribución única del Comandante tomar las decisiones. Si es cierto que, las que tomen estos actores serán las de más alto nivel, pero la vida moderna está plagada de

decisiones a muy diferentes niveles, basadas en la situación actual de los acontecimientos.

Es por esto, que no solo centramos la atención en presentar todos estos datos al "responsable máximo de la empresa", que en definitiva, como ya se podrá imaginar, van a ser representaciones estadísticas en su mayoría, sino, además, como representar diferentes casos a técnicos y analistas: exfiltraciones de información, denegaciones de servicio y otros casos diferentes. Evidentemente, al ser incidentes de muy diferente naturaleza, las técnicas de representación van a ser muy diferentes.

En este último punto de este trabajo se intentará relacionar la información de la que se dispone en cada uno de los casos con las diferentes herramientas visuales que hasta este momento han sido presentadas. El orden en que se van a presentar las diferentes soluciones es de menor a mayor responsabilidad en la decisión. Se comenzara por las soluciones más técnicas, para terminar con las de alta dirección. En la medida que sean fáciles de implementar y mas intuitiva su compresión harán que su eficacia sea mayor.

5.1 Alguien está llamando a la puerta.

Normalmente, aunque no es obligatorio, la materialización de un ataque empezara por un escaneo de puertos. Se denomina *"puerto"* a una interfaz a través de la cual los diferentes tipos de datos se pueden enviar y recibir. Dicha interfaz puede ser de tipo físico, o puede ser a nivel de software, en cuyo caso, se usa el término "*puerto lógico*". Se designa puerto lógico a una zona, o localización, de la memoria de un ordenador que se asocia con un puerto físico o con un canal de comunicación, y que proporciona un espacio para el almacenamiento temporal de la información que se va a transferir entre la localización de memoria y el canal de comunicación. En el ámbito de Internet, un puerto es el valor que se usa, en el modelo de la capa de transporte, para distinguir entre las múltiples aplicaciones que se pueden conectar [39].

La expresión "escaneo de puertos" se refiere a la realización de esta acción sobre puertos lógicos. Esta consiste en analizar, normalmente por medio de un programa (uno de los más usados actualmente es Nmap), el estado de los puertos de una maquina víctima, comprobando si están abiertos, cerrados o protegidos de alguna forma. De este modo, dado que como ya se ha comentado, un servicio está asociado a un puerto, se comprueba que servicios ofrece la máquina. Con esta técnica se puede investigar qué sistema operativo gobierna la máquina, y las versiones de las aplicaciones que proporcionan los servicios [40]. Con todos estos datos se saben las vulnerabilidades conocidas (y hasta las "no conocidas") a las que está sujeta la víctima.

De las dos definiciones anteriores, se deduce que para visualizar gráficamente que "alguien está llamando a la puerta de nuestras maquinas", comprobando los puertos lógicos de las mismas para conocer su estado, se necesita un tipo de grafico que, lógicamente, represente estos puertos. Tres son los gráficos presentados que muestran esta característica: la matriz de puertos mediante mapa de colores, el diagrama de coordenadas paralelas y el diagrama de flujo de conexiones. Del análisis de cada uno de ellos, de sus ventajas e inconvenientes, se puede deducir, en una primera aproximación, cuál de ellos es el que presenta al analista el ataque de la forma más conveniente.

Recordemos que la matriz de puertos (figura 22) representa a los puertos lógicos de una maquina mediante pixeles. Así, de un total de 65.536 puertos (256 * 256), se establece una "vigilancia" sobre todos ellos. Cuando algún puerto sea accedido desde el exterior, este cambia su color. Si la maquina está sufriendo un escaneo de puertos, este será indiscriminado, es decir, se tocaran todos los puertos, o al menos los más habituales, del 1 al 1024. El reflejo que esta circunstancia tiene sobre la matriz es que uno a uno se irán "iluminando" consecutivamente todos los puntos, representación de los puertos, y probablemente, casi con total seguridad, desde el 1, en incremento de uno, hasta el último. El analista puede apreciar esta circunstancia y decidir si ha sido un acceso autorizado o por el contrario está sufriendo un ataque. Si se ha accedido a todos los puertos, no cabe ninguna duda que es un ataque. Si el acceso ha sido a un único puerto, lo más probable es que sea un acceso legítimo, pero no se puede descartar que se trate de un escaneo del puerto. Analizado el comportamiento del gráfico, ha de tenerse en cuenta diversas circunstancias. La primera es que el grafico representa un instante de tiempo, motivo por el cual cuando un puerto se "ilumine", inmediatamente se "apagara", lo que proporciona un escaso tiempo para la observación, pudiendo pasar inadvertido. Si se accede a todos los puertos, será decididamente más fácil observar el suceso, pero sigue existiendo el inconveniente, para este tipo de gráfico, que no ofrece persistencia en el tiempo. La segunda, y no menos importante, es que este grafico pertenece a una máquina, lo que implica que se deberán construir tantos gráficos como maquinas deban ser vigiladas.

Antes de pasar a evaluar las otras dos formas graficas para representar un escaneo de puertos, debemos tener presente, de entre las diferentes taxonomías de este ataque, aquella que se origina en función de a que sistemas o puertos concretos va dirigido. Se denomina escaneo horizontal [41] cuando el atacante realiza una exploración de un único puerto en varias maquinas. Este tipo de exploración resulta útil cuando se descubre una nueva vulnerabilidad de una aplicación (recuérdese que un puerto está asociado a un servicio) y se buscan todas aquellas maquinas que tienen abierto el puerto para posteriormente lanzar el ataque dirigido a esa vulnerabilidad. Si el ataque se dirige a una única maquina, pero explorando todos los puertos de la misma para comprobar su estado, el escaneo se denomina vertical [41]. Este tipo de escaneo se realiza cuando el ataque se desea dirigir contra una maquina en concreto y se busca una posible vulnerabilidad en cualquiera de los servicios que ofrece.

Repasando la representación que proporciona la matriz de puertos, se comprende fácilmente que cuando se visualiza un escaneo vertical resulta relativamente fácil apreciar el suceso. Como ya se ha explicado, se van a ir "iluminando", uno a uno, los diferentes pixeles que representan los puertos en una única máquina. ¿Pero como se ve un ataque horizontal?. En este caso, cambiara el color del pixel que representa el puerto ¡en cada una de las matrices con las que representemos a cada una de las maquinas!. Individualmente, para cada matriz únicamente se "ilumina" un punto. Tal y como se comento anteriormente, esta circunstancia resulta relativamente fácil que pase inadvertida para el analista. Los dos hándicaps expuestos limitan enormemente la resolución que ofrece este grafico acerca del ataque por escaneo de puertos. Analizaremos, ahora, las otras dos figuras propuestas, buscando solventar los inconvenientes mostrados.

En primer lugar, estudiaremos el grafico de flujo de conexiones. Recapitulando los atributos expuestos en el apartado 4.11, recordamos que el grafico se componía de seis ejes, los tres primeros dedicados al emisor y los tres últimos al receptor. Los datos representados para unos y otros son los mismos: en un eje las direcciones IP de las maquinas implicadas, otro eje de tiempos y en el último los puertos lógicos numerados.

En el grafico se representa el flujo desde la maquina origen, en un momento determinado y por un puerto de la misma. En el segundo tramo del gráfico, el flujo se asocia a puerto de entrada, en un momento del tiempo y a una maquina concreta (figura 55). Al ser uno de los ejes el asociado al tiempo, se pueden representar varios eventos durante un intervalo de tiempo. De esta forma un escaneo vertical queda representado (figura 64) mediante flujos asociados, uno a uno, a cada puerto de la maquina víctima.

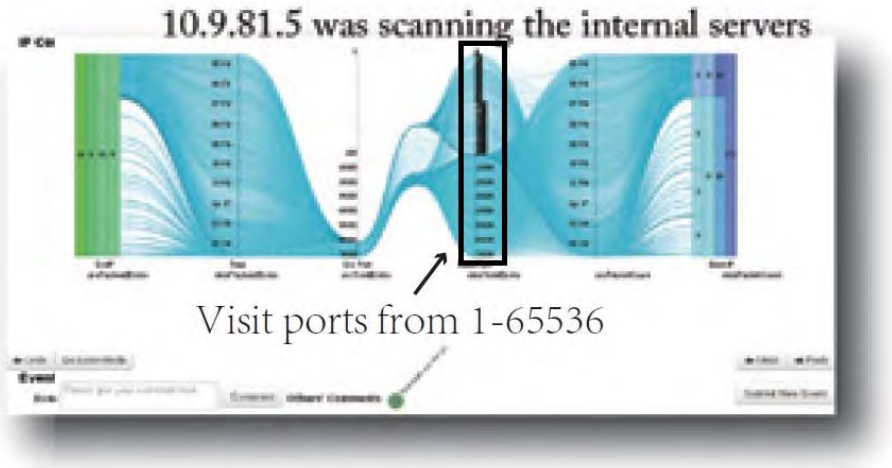


Figura 64. Escaneo vertical de una maquina mediante el grafico connection river.

Si el ataque es un escaneo horizontal, su representación consistirá en una serie de flujos asociados a un puerto de todas las maquinas, o al menos un rango de ellas, de la red objetivo. En la figura (figura 65), el atacante está intentando localizar un servidor

web, por lo que los flujos están asociados al puerto 80 de todas las máquinas de la red víctima.

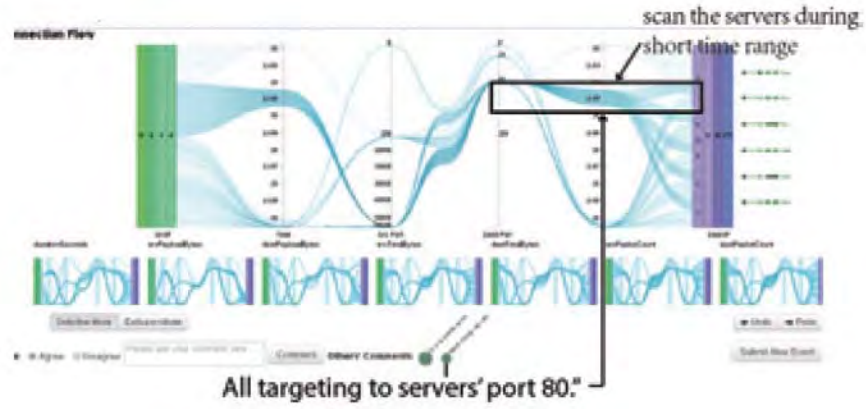


Figura 65. Escaneo horizontal de varias maquinas mediante el grafico connection river.

Por último, únicamente resta describir la representación mediante grafico de coordenadas paralelas. Recordaremos (apartado 4.10) que permitía representar varias dimensiones, mediante, al igual que el grafico anterior, ejes paralelos para cada una de las dimensiones que se quieren representar, reflejando en estos el rango de valores que toman las variables, pudiendo ser cuantitativas o cualitativas. Para representar un escaneo de puertos es necesario modelar, al menos, los puertos, las maquinas propias (sus direcciones IP) y las maquinas atacantes. Si se desea que no sea únicamente el reflejo de una situación instantánea, sino que permita trazar un seguimiento en un intervalo de tiempo, ese tiempo debera conformar otro eje. Como se observa, son los mismos datos del grafico connection river, lo que implica que para este caso de estudio son graficos muy similares.

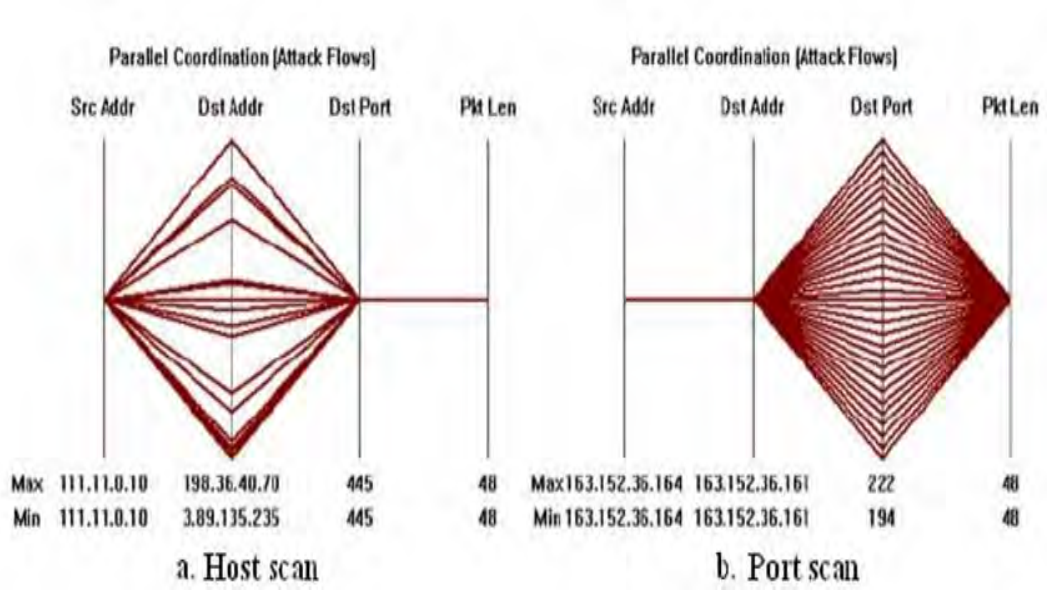


Figura 66. Escaneo horizontal (a) de varias máquinas y escaneo vertical (b) de una maquina mediante el grafico parallel coordinate.

En la figura (figura 66) se puede observar una representación de los casos estudiados en los ejemplos anteriores. En la figura 66.a se reproduce un escaneo horizontal. En ella se aprecia como desde una dirección fuente se alcanza varias direcciones destino en un único puerto, 445, de todas ellas. En la figura 66.b, lo que se observa es como una maquina origen solo realiza la llamada contra una única maquina destino, llamando a los puertos desde el 194 al 222. Se trata por tanto de un escaneo vertical.

Analizando todos los datos expuestos hasta el momento, de los tres gráficos estudiados para el caso, primeramente se debe descartar la utilización de la matriz de puertos. Como ya se ha expuesto no proporciona información de tiempos, por lo que no se pueden analizar los sucesos a lo largo de un intervalo de tiempo. Además, para comprobar un escaneo horizontal, obliga a representar tantas matrices como maquinas se quieran vigilar.

Por otro lado, ambos gráficos restantes, connection river y parallel coordinate, son muy similares como se planteó anteriormente. De hecho el uso de uno preferentemente sobre el otro, únicamente se plantea en base a la "comodidad" de su empleo. Si se necesita implementar, exclusivamente, los ejes de direcciones IP, tiempo y puertos de las maquinas origen y destino, el grafico más idóneo es connection river.

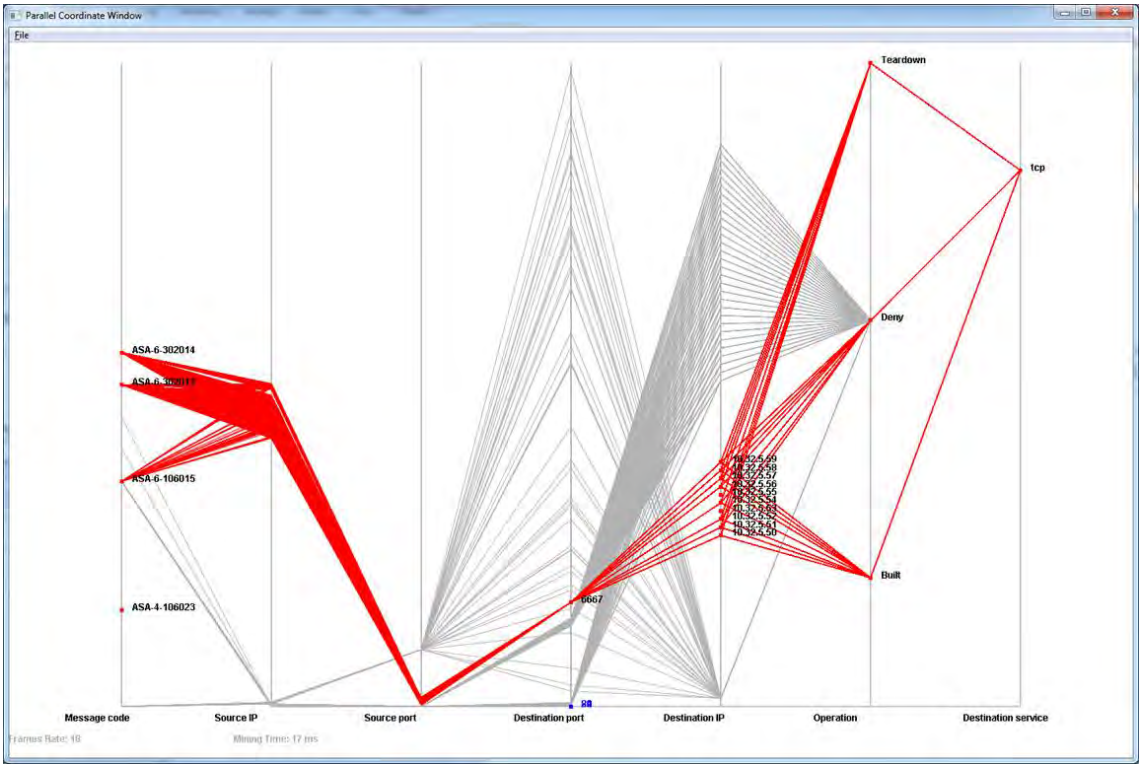


Figura 67. Grafico parallel coordinate con variables añadidas.

Si, por el contrario, es necesario representar más variables, por otros motivos ajenos a la vigilancia de escaneos de puertos, será obligatorio usar el grafico de coordenadas paralelas (figura 67), ya que el anterior no las implementa.

5.2 ¿No hay acceso a los servicios del sistema?.

Un ataque de denegación de servicio se caracteriza por impedir que usuarios legítimos puedan acceder a los servicios autorizados [42]. Un ataque de denegación de servicio distribuido posee además la característica que es protagonizado por numerosas maquinas realizándolo simultáneamente desde muy diferentes ubicaciones, normalmente, contra el servicio comprometido.

Existen muy diversos métodos de materializar este tipo de ataque, buscando siempre agotar los recursos finitos de los diferentes elementos que intervienen en la comunicación. Entre otros muchos procedimientos, y sin ser exhaustivo, se encuentran los siguientes [42]:

- enviar a la victima un flujo continuo de paquetes que consuman un recurso clave, un ejemplo podría ser la memoria, que impida a la maquina ofrecer el servicio con normalidad o, aun peor, no poder ofrecerlo.
- otro modo de ataque consiste en transmitir el atacante paquetes modificados de tal forma que alteren el protocolo de comunicación, provocando en la victima un bloqueo o incluso reiniciarse.
- modificando el software de maquinas legitimas dentro de la red atacada, de tal forma que, como cliente legitimo, consuma recursos de la misma y no permita alcanzar al resto de clientes los servicios que se proporcionan.
- suplantando la identidad de la víctima, solicitando diferentes servicios legítimos en su nombre. Cuando los diferentes servidores requeridos responden, consumen los recursos de la víctima.
- consumiendo el ancho de banda de la red legítima [43], introduciendo "ruido" en la misma, coordinando acciones de numerosas maquinas maliciosas.
- y otros muchos métodos descubiertos en la actualidad y otros que todavía están por descubrir.

¿Por qué se producen estas modalidades de ataque? ¿Qué permite que tengan éxito?. Se puede asegurar que estas son las principales causas [42]:

1. Con la paulatina mejora de las comunicaciones, a lo largo del tiempo, se ha ido trasladando la responsabilidad, en el transporte de paquetes IP, de la correcta entrega de los mismos, a los host finales, emisor y receptor. Son estos host finales, siguiendo el paradigma del mejor-esfuerzo ("best-effort"), los que garantizan la calidad de servicio («QoS» *Quality of*

Service), o todavía más exigente, asegurando la calidad de experiencia («QoE» *Quality of Experience*). Los elementos intermedios de red, "únicamente" tienen el cometido de reenviar los paquetes que reciben hacia su destino. Si en este modelo de comunicación, uno de los dos implicados, emisor o receptor, actúan maliciosamente, puede infligir un grave daño a su homónimo.

2. Los recursos de todos los elementos que intervienen en la comunicación son limitados. Esto implica que cualquiera de ellos puede ser agotado mediante la técnica adecuada.
3. Anonimato. Es muy fácil suplantar otra identidad con la que poder atacar a la víctima, evitando así, ser identificado como posible usuario malintencionado.
4. La administración en Internet corresponde al ISP de turno por el que pasa nuestra comunicación. No hay una estandarización, por lo que la política de seguridad depende, en gran medida, de quien provee el servicio.

Y lo peor de todo esto, es que desde hace mucho tiempo ¡existen herramientas capaces de automatizar el escaneo, la explotación, la implementación y la propagación de malware para la ejecución de ataques DoS! [44].

Recapitulando, cuando se ataca a una maquina mediante técnicas de denegación de servicio, esta agresión tiene, entre otras características, las siguientes:

- se pueden utilizar numerosas maquinas infectadas (zombies) o únicamente la propia maquina del usuario. Esto último, por el argumento de la anomización será muy poco probable que se realice directamente, lo más lógico será hacerse con el control de otra máquina y que esta sea la que lo efectué.
- puede estar siendo realizada contra la propia víctima o contra otras muchas que le proporcionen respuestas legitimas, al unísono;
- puede realizarse mediante herramientas automáticas o por el contrario de forma manual;

Cabe preguntarse, entonces, como identificar el proceso de denegación, si es tan dispar en su ejecución y variado en los recursos afectados. La respuesta es obvia, serán necesarias diferentes técnicas de identificación (y por tanto diferentes representaciones graficas) para diferentes procesos. Además, se deberá diferenciar si lo que se está observando, y a su vez representando, son técnicas de monitorización o técnicas forenses, es decir, se producen los incidentes en tiempo real durante la vigilancia de los sistemas o por el contrario, en el proceso de investigación a posteriori, se reconstruyen los hechos acaecidos en busca del incidente.

Comenzaremos por las primeras, valorando la información para poder apreciar incidentes que ocurran durante la monitorización. En este apartado, encontramos los ataques mediante los que se desbordan elementos físicos de la maquina víctima. De estos elementos, los más comunes son la memoria y el procesador. Si cualquiera de

estos dos elementos están ocupados en su totalidad (la memoria cargada con datos "basura" y la CPU consumiendo capacidad de proceso inútilmente), será imposible ofrecer los servicios requeridos.

Una forma de representarlo podría ser mediante un diagrama de sectores (figura 68), en el que el sector azul muestra la capacidad de proceso de la CPU en uso y el sector rojo, al contrario, la que no está siendo utilizada. Para la representación de la ocupación de memoria se utilizaría exactamente el mismo grafico. Cuando la capacidad de proceso estuviera siendo colapsada, el sector azul se agrandaría, llegando a ser totalmente azul si el procesador estuviera ocupado en todo momento. Es evidente que no podría responder a otras peticiones, si además se pueden "enmascarar" como peticiones de procesos del kernel del sistema operativo.

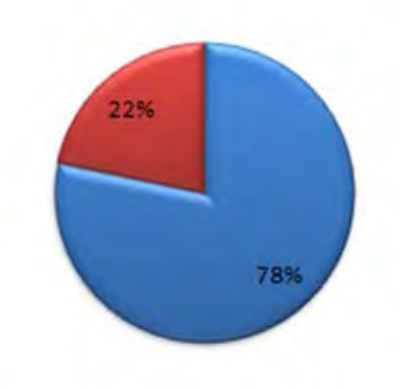
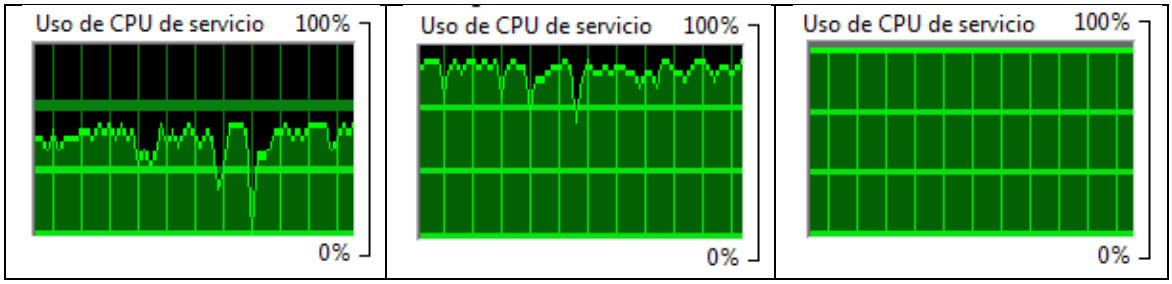


Figura 68. Diagrama de sectores del uso de CPU.

Aun pudiendo ser útil esta figura, adolece de un componente que es muy útil en este tipo de incidentes, y es el tiempo. La figura, aunque puede variar con el tiempo, únicamente representa la situación del procesador en un instante determinado. Este hándicap podría solventarse con una complicada técnica (trellis plot), que se verá más adelante en otro caso, pero existe otra figura que nos proporciona información a lo largo del tiempo. Esta no es otra que la línea de tiempos (timeline, ver apartado 4.4).



(a) (b) (c)

Figura 69. TimeLine de uso de CPU.

Como ya se explicó, esta grafica sitúa el tiempo en el eje de abscisas, y en el caso que nos ocupa, el porcentaje de uso de procesador en el eje de ordenadas (figura 69). Nuevamente se hace hincapié que el caso de ocupación de memoria se trataría de igual forma. En cada instante, cada punto del eje de abscisas, la grafica representa el porcentaje de capacidad de proceso en uso. De esta forma, se puede apreciar (figura 69.a) que el uso es adecuado. De producirse un incidente que afecte a la

capacidad de proceso en uso (figura 69.b) el porcentaje pasaría a ser mucho más elevado, pudiendo llegar a ser del 100% (figura 69.c). Es de advertir que si esta circunstancia es continua, probablemente, se deba a un mal dimensionamiento de la capacidad de proceso del sistema (de la memoria, en el otro caso) y no a un posible ataque.

De esta forma se obtiene, para cada instante de tiempo, el mismo resultado que con el uso del grafico de sectores, pero con el valor añadido de poder comparar con los instantes anteriores. Esto hace, de esta grafica, un elemento mucho más útil de visualización, siendo preferible su empleo.

En el caso de desbordamiento por inundación del ancho de banda de red, mediante técnicas de generación de ruido, por ejemplo, el grafico es el mismo, pero presentando una importante salvedad. Debido a que el ancho de banda depende, en gran medida, del modelo de servicio contratado con la operadora y que la ocupación es muy fluctuante, incluso a lo largo del día, es necesario un elemento más dentro del grafico. Este no es otro que una media de ocupación de acuerdo a un patrón de cada momento. Es importante tener presente que la ocupación del ancho de banda de la red es muy dependiente del tramo horario en el que se está trabajando (figura 70). Este no

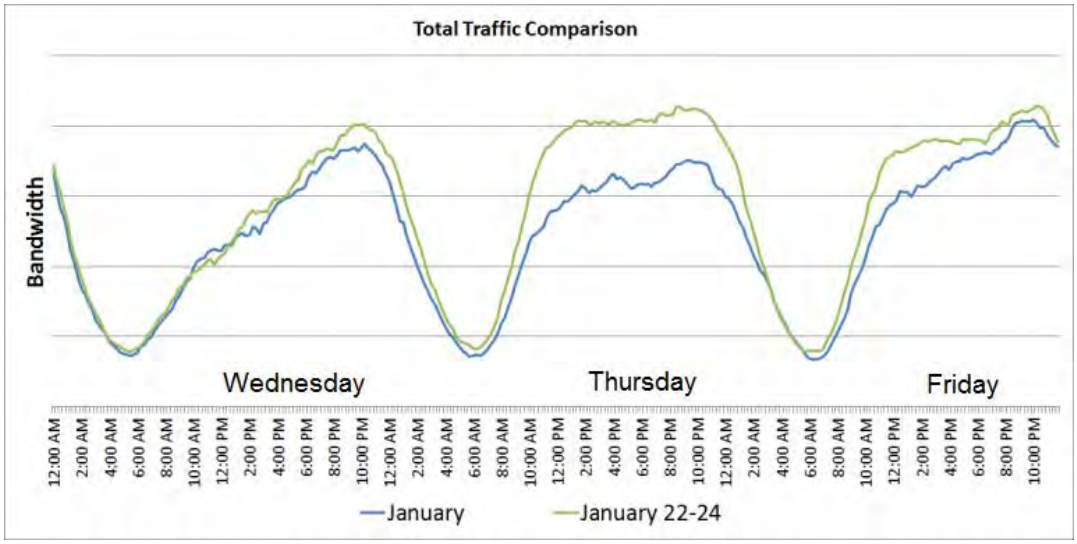


Figura 70. TimeLine de trafico de red.

es el mismo en horario de oficina que fuera de este y mucho mas por la noche, a no ser que nuestra empresa este repartida a lo largo de todo el mundo, en cuyo caso, el horario de oficina se prolonga las 24 horas del día. También es habitual que el trafico dependa del día de la semana, porque tampoco se genera el mismo trafico de lunes a viernes que los fines de semana, igualmente incluso por la generación de eventos sociales, y valga como ejemplo, aunque resulte extraño, al día siguiente a un importante encuentro de futbol.

Para poder establecer un nivel de referencia, que indique que el tráfico generado en el momento actual es correcto deben tenerse en cuenta todos estos factores. Este nivel de referencia debe calcularse con los datos de un pasado cercano, en el ejemplo (figura 70) se puede apreciar que los datos evaluados para obtener una media son los del mes en curso. Este puede considerarse un buen intervalo de tiempo para este objetivo, ya que ampliar el periodo de tiempo de estudio, en lugar de proporcionar un grafico más exacto, probablemente desvirtuaría el resultado. En diferentes épocas del año se realizan actividades diferentes y no proporcionan una mejor información.

Con el nivel de referencia establecido, permite que el analista tenga una cota con la que establecer una comparación. Como ya se ha determinado, esta es distinta para cada red. De esta forma, si los niveles no se separan en exceso de los de la referencia, cabe pensar que el flujo a través de la red es normal. Por el contrario, si se separan en exceso, no se puede afirmar que se está produciendo un incidente, pero supone una alerta y la consiguiente activación de un análisis más exhaustivo. Si finalmente el positivo resultara cierto, habría que eliminar los datos obtenidos del nivel de referencia para siguientes estudios, como resulta obvio.

El ultimo modo de denegación de servicio, a analizar, desde el punto de vista de la monitorización, es aquel en el que un gran número de maquinas atacantes realizan peticiones simultaneas a una misma máquina víctima, lo que se conoce como denegación de servicio distribuida (DDoS, del ingles "*Distributed Denial of Service*"). Ya sea por el número de maquinas que al unisonó realizan la correspondiente llamada, o porque el protocolo de aplicación por el que se va a realizar la comunicación haya sido "debidamente" manipulado, como por ejemplo el ataque TCP/SYN rompiendo el proceso de establecimiento de conexión TCP de 3 vías [42], la principal característica que lo identifica es, normalmente, que las comunicaciones mediante ese protocolo empiezan a ser mucho más numerosas que por el resto de protocolos. Es también importante resaltar el hecho que cuando el ataque se realiza por este método, se utiliza un único protocolo, no siendo habitual, aunque posible, hacerlo mediante dos o más protocolos.

Son estas dos particularidades las que proporcionan el medio para representar un ataque de este tipo. Para ello contamos con dos tipos útiles de figuras: el diagrama de sectores y el mapa de árbol o treemap. Analizando la información a modelar y como la presenta una u otra representación se puede seleccionar la más adecuada.

El diagrama de sectores posee la capacidad de representar las partes de un todo (ver apartado 4.8). De esta forma, cuando el porcentaje de uso de uno de los protocolos aumenta considerablemente, hace que el resto disminuya en la misma proporción (además de disminuir porcentualmente respecto del que se está incrementado, también disminuye en valor absoluto, únicamente por capacidad de la maquina). En la figura (figura 71) se representan, a modo de ejemplo, el tráfico actual de entrada que está

gestionando una maquina, categorizado en los diferentes protocolos que soporta: el protocolo web HTTP; el protocolo de correo SMTP; el protocolo de configuración dinámica de host DHCP; el protocolo intérprete de órdenes seguro SSH y el protocolo de resolución de nombres de dominio DNS. En un momento dado, cientos o miles de maquinas distribuidas por todo el mundo, y que previamente han sido capturadas (este ataque no funciona poniéndose de acuerdo todos los usuarios para hacerlo al mismo tiempo, se necesita una sincronización automática) realizan peticiones HTTP contra la máquina, es decir, reclaman una página del servidor web que tendrá instalado. Instantáneamente, la victima intentará responder a todas esas peticiones legítimas, por lo que el trabajo que realiza en ese protocolo aumenta, disminuyendo porcentualmente en el resto, por un lado, además de perder capacidad de gestión, en valor absoluto por otro lado, en estos otros protocolos, como ya se ha indicado.

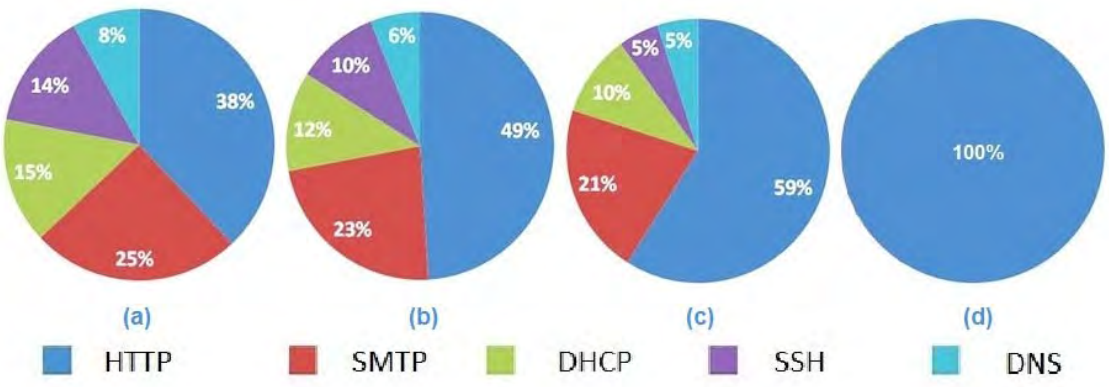


Figura 71. Representación gráfica por diagramas de sectores de un ataque DDoS a un servidor web.

En la figura (figura 71.a) se representa el funcionamiento correcto del sistema, con porcentajes iniciales apropiados al mismo. Cuando se inicia el ataque, la maquina víctima se ve obligada a responder a todas las peticiones web que le llegan, por lo que los porcentajes de gestión del flujo HTTP crecen (figura 71.b y c). Si el ataque persiste, la maquina llegara a gestionar únicamente este protocolo (figura 71.d), viéndose saturada y no pudiendo responder a todas las peticiones, en realidad, lo más viable es que no esté respondiéndose a ninguna.

Una vez desarrollado como se puede visualizar un ataque DDoS mediante un diagrama de sectores, comprender como la materialización grafica mediante un mapa de árbol o treemap, es relativamente sencillo. La construcción es prácticamente igual, con un sutil matiz diferenciador. Como ya se menciono, un grafico treemap ofrece representar una estructura jerárquica. Dado que el protocolo de red IP se divide en la capa de transporte en los protocolos TCP y UDP, orientado y no orientado a la conexión respectivamente, se puede dividir el treemap en estas dos jerarquías. El resto es exactamente igual que el diagrama de sectores, analizado anteriormente.

Figura 72. Representación gráfica mediante treemap de un ataque DDoS a un servidor web.

En la figura (figura 72) se analizan los mismos datos del estudio anterior. De esta forma (figura 72.a), nuevamente, observamos el estado inicial de la maquina víctima, con idénticos porcentajes de flujo que se consideraron apropiados en el ejemplo. Cuando se inicia el ataque se aprecia (figura 72.b) además de la disminución del flujo de todos los demás protocolos, que la jerarquía UDP, parte derecha de la figura 72.a/b/c, también disminuye. Por contra, lógicamente, la jerarquía TCP aumenta (parte izquierda de la figura 72.a/b/c/d).

Llegados a este punto, en ambas representaciones, el ataque DDoS se habría realizado con éxito. Siguiendo la evolución del ataque a través de los gráficos, el analista habría podido tener conciencia de lo que estaba ocurriendo, y tomar una decisión para contrarrestarlo. Evidentemente, para tener posibilidad de tomar medidas de recuperación en tiempo real, y dado que el tiempo de materialización del ataque es muy reducido, se requiere una vigilancia de las redes 24/7.

¿Qué grafico resulta más adecuado? En este caso, ambos son igualmente útiles, aportando algo más de información el esquema de árbol, aunque no resulta excesivamente relevante. Recordemos, como se indicaba en el apartado 3, que la presentación grafica debe ser lo más intuitiva posible, de tal forma que con unas pequeñas indicaciones, sea fácil su comprensión. Este puede ser motivo suficiente para elegir la representación mediante diagrama de sectores como medio grafico, para mostrar este tipo de eventos, sobre el grafico treemap.

Hasta este punto se ha presentado gráficamente una denegación de servicio de muy diferentes características, pero siempre desde el lado de la monitorización, es decir, presentando cómo evoluciona el elemento que se satura hasta que se colapsa, dando a la persona que toma la decisión la oportunidad de adoptar alguna medida. Pero, ¿si el objetivo no es tomar medidas que palien el ataque? ¿si el propósito es analizar offline, mediante un estudio forense, un ataque que ya se ha producido, y de esta forma conocer como, cuando y, si es posible, por quien ha tenido lugar el ataque? De esta reconstrucción es posible evaluar las consecuencias y elaborar conocimiento para prevenir ataques futuros.

Para realizar una evaluación forense debe representarse la consecuencia de un ataque de este tipo, que no es otra, que no se ofrece servicio durante algún tiempo. Son estas variables las que se deben representar: el servicio (la ausencia de él) y el tiempo. Recordando que el servicio se asocia a un puerto lógico y la entrega del mismo es un flujo de datos a través del tiempo, se advierte que para poder realizar la representación son necesarias, al menos, tres dimensiones. Para materializar este hecho, múltiples dimensiones, disponemos de dos figuras ya estudiadas: el flujo de conexiones y el grafico de coordenadas paralelas.

Estudiaremos en primer lugar la representación de una denegación de servicio mediante un grafico de flujo de conexiones. Este grafico contiene los elementos mínimos que se han determinado necesarios para modelizar un ataque de este tipo: los servicios asociados a los puertos; las maquinas, tanto las atacantes como las victimas y el tiempo; además del flujo de datos que transita de una maquina a otra a través de los puertos en diferentes instantes de un intervalo de tiempo. La "presencia" de un ataque por denegación de servicio es, como se indico anteriormente, la "ausencia" de ese flujo de datos, debido a que la maquina víctima se encuentra saturada y no es capaz de responder a las peticiones que se le formulan, tanto las licitas como las ilícitas.

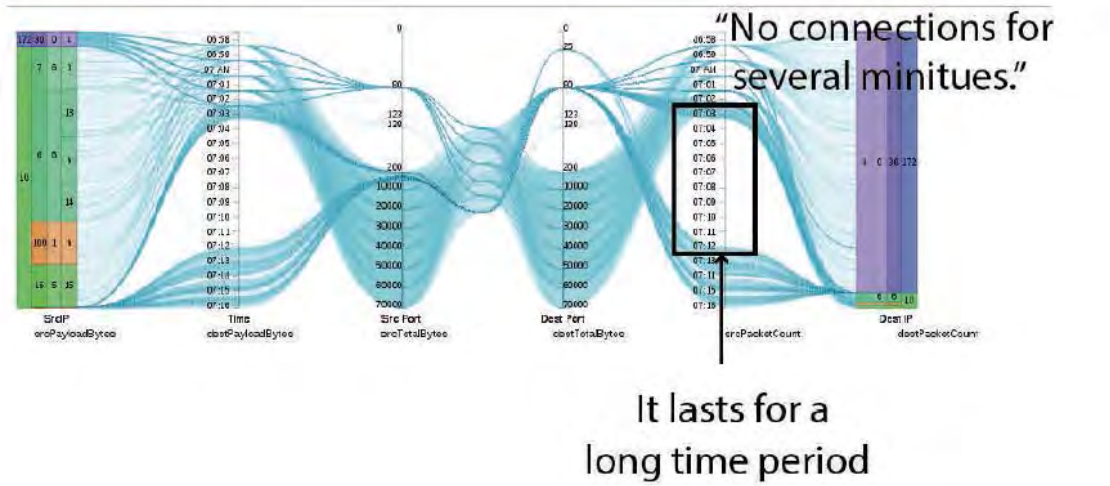


Figura 73. Representación grafica mediante connection river de un ataque DDoS a un servidor web.

En la figura (figura 73) se aprecia el flujo de información desde maquinas origen, treemap de direcciones IP de la izquierda, hacia maquinas destino, treemap de direcciones IP de la derecha. Estos flujos están asociados a servicios, ejes de puertos, tanto de salida como de entrada. Restan, por tanto, describir otros dos ejes: los de tiempos. Estos son los que nos proporcionan la visualización del ataque. En el eje de la derecha, maquinas destino, se ha recuadrado el tiempo en el que el servidor web, entrada por puerto 80, no es capaz de responder a las solicitudes recibidas. La grafica permite tomar consciencia que se ha producido un ataque por denegación de servicio. Esta técnica solo es útil a posteriori, mediante una reconstrucción forense, ya que se

obtiene la citada consciencia cuando ha tenido lugar, ya ha pasado, un intervalo de tiempo en el que no se obtiene respuesta.

Queda ahora solventar el problema de reconstruir los sucesos a lo largo de intervalos de tiempo más grandes, como por ejemplo un día completo. Observando la figura (figura 73) apreciamos que cada corte del eje de tiempos representa 1 minuto, por lo que el intervalo representado en el eje no comprende mas allá de 20 minutos, y aun así, la figura resulta farragosa. ¿Cómo se puede lograr aumentar el intervalo sin recortar el tamaño de cada corte?. La respuesta es sencilla: haciendo que el eje de tiempos sea dinámico, es decir, que el tiempo avance a lo largo del eje. Evidentemente implica que ambos ejes de tiempo deben tener su avance coordinado. Esto hará que los flujos varíen a medida que se van leyendo los logs para generar el grafico. No es necesario que el tiempo avance con el tiempo natural, pudiendo progresar de forma acelerada o frenada, dependiendo si se desea obviar un periodo, acelerándolo, o por el contrario observar con más detalle y pausadamente, frenándolo. Además para poder localizar con más precisión el evento debe ser posible retroceder en el tiempo a voluntad del analista que lo controla.

Antes de terminar con el apartado de ataque por denegación de servicio, cabe especificar que muy probablemente existirá un indicador de planificación del ataque. Este no es otro que un escaneo de puertos. En primer lugar es muy posible que se realice un escaneo horizontal, buscando la maquina donde reside el servicio que se quiere denegar. Una vez localizada la maquina, y siempre que el ataque no sea por saturación del servicio, realizando llamadas al mismo desde múltiples maquinas, en cuyo caso no resulta necesario, se practicaría un escaneo vertical, en busca de alguna vulnerabilidad conocida a la que sea susceptible. Para ambos casos, se ha estudiado sendas figuras que delatarían el hecho.

5.3 ¡Datos a la fuga!.

Se denomina fuga de información al incidente que pone en poder de una persona ajena a la organización, información propia de la misma, de forma intencionada o no, y que sólo debería estar disponible para los individuos que la organización designe "tengan necesidad de conocer", sean integrantes de esta o personal externo [45]. La fuga de información supone, por tanto, la pérdida de la confidencialidad, de forma que, información que a priori no debería ser conocida más que en el ámbito de una organización, área o actividad, termina siendo visible o accesible para otros [46].

¿Cómo se materializa un ataque para extraer información sensible de los sistemas de información propios de la empresa? La respuesta no es única, pudiendo realizarse, sencillamente, mediante un insider que robe la información en un USB o simplemente desde una conexión exterior a una maquina de la red (telnet o ssh),

tomando el control de la misma (este método es tan burdo que no se puede considerar un método, por lo que no es usado). El modelo de exfiltración que se va a "dibujar" es el originado por una Amenaza Avanzada Persistente (APT, de su denominación en ingles Advanced Persistent Threat).

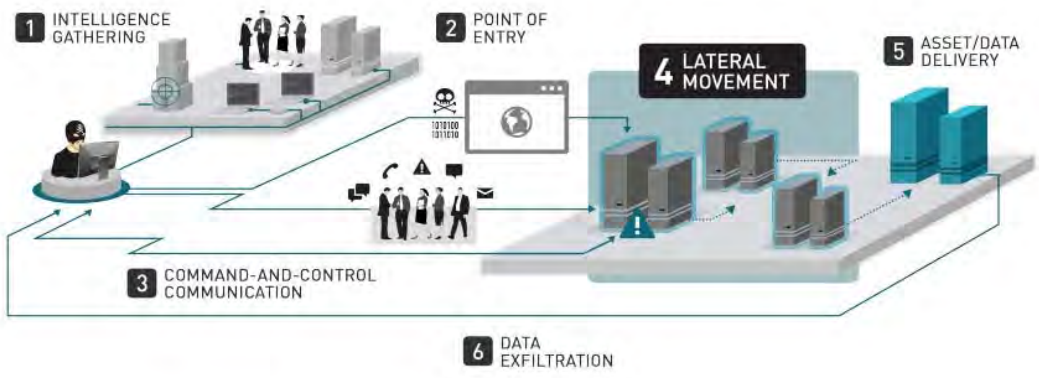


Figura 74. Fases de un Advanced Persistent Threat (APT) [47]

El patrón de un APT, con diversas variantes, se inicia con la búsqueda de un punto débil del sistema de información, que proporcione un factor inicial de ataque. Pueden ser varios, pero normalmente será un usuario legítimo del sistema que habitualmente no cumple los protocolos de seguridad, por desconocimiento o por desidia (figura 74.2). Dos son entonces los principales vectores de ataque: un correo malicioso enviado por un usuario "legitimo" del que se confía y que no levanta sospechas o mediante navegación por páginas webs infectadas (pueden ser varios los métodos de infección) [48]. Existen más métodos como la infección mediante la conexión de un dispositivo externo USB, con el malware incluido, a alguna de las maquinas de la red o la descarga de software "pirata" de paginas diseñadas para tal fin, dirigidas a usuarios a los que no seduce la idea de pagar por la adquisición de software legal, y que por supuesto, se encuentra infectado [49]. Por cualquiera de los métodos, si el usuario es engañado, infectara su máquina con un pequeño código malicioso (figura 75). Este pequeño código hará que la maquina intente conectarse con otro equipo exterior (aunque hay otros, este es el método más habitual) que le entregara el APT completo.

El APT quedará latente, intentando camuflar su actividad para pasar inadvertido y recopilar la información ordenada para la que ha sido creado. Cabe destacar que el periodo de latencia puede ser de varios años. El malware denominado APT28, creado por activistas rusos para recabar información con fines políticos, fue descubierto en octubre de 2014, por la empresa Trend Micro, más de un lustro después de que comenzara su actividad registrada [50], evidenciado por los indicios encontrados, pero podría ser anterior.

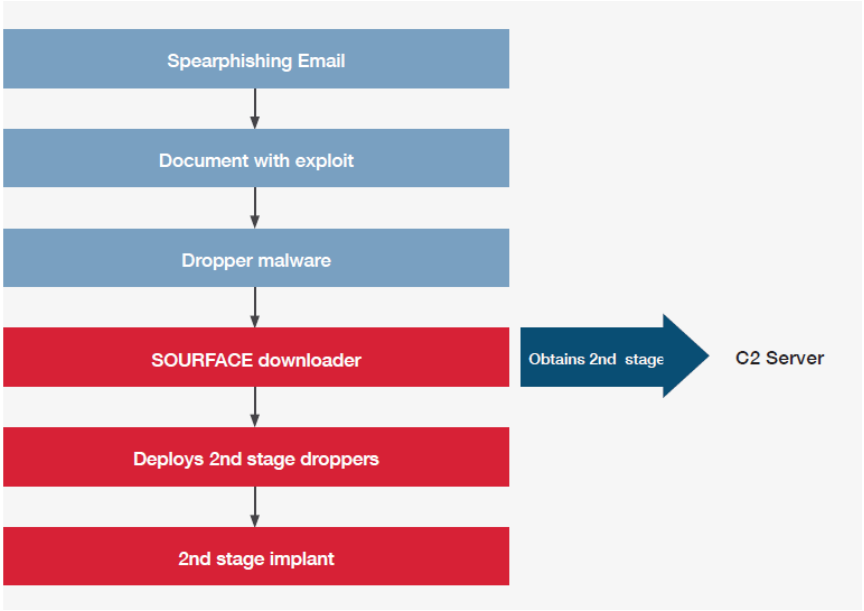


Figura 75. Esquema de actuación de Advanced Persistent Threat (APT) [50]

Durante todo este tiempo que no es descubierto, el malware recopila información y la exfiltra al exterior. También se puede hacer de diversas formas, pero siempre buscando el enmascaramiento de la actividad. Se logra, y no hay otra forma, mediante el mimetismo del tráfico ilegitimo con el tráfico legítimo. El tráfico de exfiltración de información hacia el exterior tiene que ser lo más parecido posible al tráfico ordinario.

Para lograr este mimetismo el atacante tiene que realizar muy diversas acciones en aras de la discreción y la ocultación de los movimientos que efectúa. Tiene que utilizar muy pocos recursos del sistema: tiempo de procesador, memoria, ancho de banda, espacio "continuo" de almacenamiento y tiempo "continuado" de transmisión. Si usa de manera exhaustiva cualquiera de estos elementos, delatará su presencia. La táctica para realizarlo se formula a continuación, pero no es obligatorio realizar todos los procedimientos que se exponen, sabiendo que no desarrollarlos aumenta el nivel de exposición y la posibilidad de ser descubierto.

En un primer momento, partiendo de la situación en que la maquina objetivo ya ha sido infectada, el malware intentara conectarse con una maquina exterior, como se enuncio anteriormente. Este equipo exterior puede formar parte del sistema Command & Control (C&C), o sencillamente ser un repositorio desde donde descargar los diferentes componentes del malware, como herramientas de acceso remoto (RAT, Remote Access Tools) [51], herramientas de descubrimiento de conexiones activas y herramientas de escaneo de puertos [47], entre otras. Ambas opciones son válidas. Si se trata de la primera, el atacante corre el riesgo de exponer un equipo o equipos con mayor utilidad posterior. Si la opción es la segunda el equipo es fácilmente sacrificable una vez realizada la descarga.

El sistema de C&C, constituido por un único equipo o un conjunto de equipos, permite al atacante mantener un canal de información con el sistema comprometido, por el que emitir comandos y verificar el estado del mismo (figura 74.3). Dado que los sistemas de monitorización, los firewalls, los IDS y los IPS son cada vez más capaces de identificar y descubrir tráfico malicioso y anómalo, la comunicación del sistema comprometido con el atacante, con su C&C, se realiza mediante técnicas de ofuscación y enmascaramiento para ocultar el control y el tráfico de red de control [51].

¿Por qué se busca siempre que sea el usuario desde el interior del sistema el que arranque las acciones que comprometen el sistema? ¿Qué motiva que la dirección del vector de ataque sea desde el interior hacia el exterior, y no al contrario como parece más lógico?. La respuesta ya se ha esbozado. Los sistemas modernos, soporte de monitorización, firewalls, IDS e IPS, son cada vez más capaces de determinar si el tráfico es malicioso o no [51]. Para evitar que los sistemas de detección, cualquiera de ellos, interpreten que la comunicación desde el atacante hacia el sistema comprometido es maliciosa y la bloquee, esta se inicia desde un usuario legitimo, y que mas legitimo, ¿qué un usuario registrado de nuestra propia organización?. He aquí la razón: resulta difícil determinar que una petición de un usuario propio no es legítima, en realidad, y muy seguramente nuestros sistemas de defensa perimetral no la bloquearan, al menos en un primer momento, hasta que sea descubierto el engaño.

Una vez determinado el sentido más propicio del ataque y su motivación, los siguientes pasos a desarrollar tienen como propósito recabar la información y almacenarla para poderla enviar hacia el atacante en el momento más adecuado. La siguiente acción es establecer el canal de comunicación con el C&C (hasta ahora solo había descargado el malware, y probablemente no sería con los equipos C&C) (figura 74.3). Simultáneamente, desarrolla la fase denominada "movimiento lateral" (figura 74.4) por la que el atacante realiza un reconocimiento del sistema, intentando obtener inteligencia; roba credenciales; eleva privilegios y se infiltra en otras maquinas [47]. Contrariamente a lo que pueda parecer esta es la fase más crítica. Es esta fase intermedia, la que ocupa más tiempo desarrollarla al atacante y donde es más vulnerable [52].

Desde este momento el malware trabajará reconociendo el sistema, buscando aquellos equipos encargados de las transacciones críticas o el almacenamiento de datos sensibles que puedan ser capturados. Para estas tareas, al igual que otras posteriores, la necesidad de disimular al máximo el trabajo realizado conlleva la creación de distintos tipos de maquinas: unas que dictan las órdenes y otras que las ejecutan. De esta forma, si una maquina ejecutora es descubierta, es posible variar la orientación de las acciones y continuar su desempeño sin perder la estructura creada, y que como se recordara, se ha invertido una gran cantidad de esfuerzo y tiempo en construir.

Una vez que el atacante dispone del mapa de la red y conoce los equipos, sus capacidades y vulnerabilidades, hace acopio de las maquinas que puede capturar y distribuirá las ordenes para realizar el ataque. Al igual que el resto de actividades, las

comunicaciones entre el sistema y el C&C también deben ser sigilosas. ¿Cómo realizarlas para conseguirlo?. Tres son los modos más empleados [51]:

- mediante el uso de cuentas de correo web. Cuando el malware se conecta a servicios conocidos como Gmail o Yahoo! Mail, la sesión está protegida por cifrado SSL y, por lo tanto, el software de monitorización de la red no podrá determinar si el tráfico es malintencionado o no. Los atacantes utilizan estas cuentas de correo web para enviar comandos a hosts comprometidos, actualizarlos con herramientas o componentes de malware adicionales y en el exfiltrado de datos.
- utilizando sitios web legítimos comprometidos. Esto permite a los atacantes, incluso, si la comunicación de la red se detecta como anómala, después de una inspección adicional el sitio web se determinará que es legítimo. Un atacante simplemente incrusta comandos dentro de las etiquetas de comentario HTML en páginas web de sitios web legítimos pero comprometidos. El malware simplemente visita estas páginas y extrae y decodifica los comandos. Además, los atacantes están haciendo uso de certificados SSL robados en un intento de hacer que su tráfico de red parezca legítimo.
- alojando componentes adicionales de malware en los servicios de almacenamiento en la nube. El uso de tales servicios proporciona a los atacantes una infraestructura de Command & Control que no puede ser fácilmente detectada como maliciosa. Al igual que en los otros dos casos anteriores el atacante hace uso de la comunicación cifrada, incluso creando túneles que dificulten la monitorización.

Mapeada la red y recibidas las ordenes desde el C&C, el malware recopilará la información y la almacenará. Para ello seguirá la misma táctica que en las fases anteriores, buscando la discreción para prolongar su permanencia en el tiempo. Unos equipos transmitirán las órdenes de recopilación de la información, otros la ejecutarán y unos terceros la almacenarán. Para un mayor disimulo, los equipos que ejecutan la orden de búsqueda de la información rotan para no ser siempre el mismo. Igualmente, no se recoge la información en un único equipo, en su sistema de almacenamiento. Si se recogieran en el mismo varios cientos de GB seria fácilmente detectable. En este punto se ha de tener en cuenta que la información buscada no está analizada, se analiza posteriormente cuando el atacante la recibe, por lo que los equipos atacantes deberán recopilar tanta información como les sea posible, siempre y cuando haya indicios de que "puede ser útil". Esto implica que el volumen de información es muy considerable, obligando a fraccionarlo "para disimularlo".

Una vez que el malware tiene recopilada información deberá empezar a transmitirla al atacante, prosiguiendo, paralelamente, con la captura de mayores cantidades de información. Emitirá una señal de control al C&C indicando que se encuentra en disposición de transmitir los almacenes de datos disponibles. Se continúan con las mismas técnicas buscando la ocultación mediante el camuflaje y el

enmascaramiento. Los métodos son los mismos que para la transmisión de órdenes entre el C&C y el sistema [51]:

- usando cuentas de correo web.
- empleando servicios de almacenamiento en la nube.
- manipulando sitios de transmisión de datos, tales como FTP y HTTP. Para mayor seguridad se hace uso del cifrado y de la compresión de los ficheros. También puede tunelizarse la transmisión.
- manejando la red de anonimato TOR.

Al igual que los datos no deben almacenarse juntos "para no levantar sospechas", tampoco deben transmitirse en un único proceso de comunicación. No deben usarse largos periodos de tiempo para transmitir, o de lo contrario, quedara al descubierto la presencia del malware. Del mismo modo, no debe transmitirse siempre desde el mismo equipo, ni tampoco, al mismo equipo receptor, buscando dificultar la detección de la comunicación por parte de los medios de monitorización y equipos de defensa perimetral del sistema víctima.

Siguiendo estas tácticas y técnicas, el atacante puede permanecer residente en el interior del sistema durante un largo periodo de tiempo. Estudiando y analizando los usos y costumbres de los usuarios y administradores; se diseña y desarrolla el malware personalizado para cada organización; siendo capaz, de esta forma, de evadir las medidas de seguridad y los sistemas de vigilancia de red; aprovechando las aplicaciones y el entorno de trabajo; aprendiendo del comportamiento de los usuarios mientras reside y se mueve por la red y busca datos susceptibles de poder ser aprovechados en beneficio de los atacantes [53].

Conociendo como funciona, la pregunta es ¿qué debe buscarse que proporcione indicios, la certeza es un concepto imposible sin un análisis detallado, que se está realizando un ataque persistente, y por ende, una exfiltración de datos?. A lo largo de la exposición anterior, se ha ido planteando como el "disimulo" es la característica principal de este tipo de ataque. Se realiza la comunicación y la trasferencia de datos mediante protocolos comunes; con tamaños de ficheros frecuentes y tiempos de transmisión ordinarios. Pero hay algo que no se puede disimular, y son las conexiones con otros equipos para completar la transmisión de la información. De la misma forma que el malware hace inteligencia con los patrones de comportamiento del sistema y de sus usuarios, la defensa está obligada a estudiar el comportamiento del malware en busca de debilidades, y esta es encontrar un patrón de conexiones que le delate. Esto, como se comprenderá no es inmediato, requiere un tiempo en el que el malware se encuentra instalado en el sistema, consumando su función.

De las figuras estudiadas en el capitulo anterior, la que proporciona una información más completa de las conexiones entre equipos, ya sean de la propia red de empresa o exteriores a la misma es el grafico de anillo. Como se expuso en el

apartado 4.12, en el mismo grafico (figura 58) se dedica una parte del mismo a los equipos propios, internos de nuestro sistema colores (colores desde el azul al morado) y otra parte a los equipos externos (colores desde el verde al rojo). Ambos espacios son variables en función de la proporción de equipos internos/externos que se representen (figura 59), siendo las direcciones IP de cada máquina el identificador que lo representa en el gráfico. Cuando entre dos equipos se establece un canal de comunicación, una línea une los dos equipos (figura 60) dirigida desde el emisor hacia el receptor, del color del emisor. El concepto de conexión debe entenderse como una idea más amplia de la que se entiende en telecomunicaciones, desde el punto de vista de los protocolos de comunicaciones. En una comunicación no orientada a la conexión se establece una transmisión de información sin establecimiento de la conexión. Dado que los identificadores de emisor y receptor, sus direcciones IP en una comunicación UDP, se encuentran en las cabeceras de los paquetes, se entiende que hay una conexión entre ambos equipos, representada en el grafico de anillo por una línea que los une, que une a sus identificadores, a sus direcciones IP.

De esta forma tan sencilla podemos representar todas las conexiones dentro de la red corporativa. Esto que puede parecer una ventaja, al igual que en todos los casos anteriores, si se representan todas las conexiones, puede suponer un inconveniente (¡lo es de hecho!), originando probablemente, que se sature el gráfico. Para solucionar este hándicap, no existe otra solución, aportada en casos anteriores, que la de reducir el número de equipos que se visualizan. En este caso no resulta interesante eliminar equipos, ya que todos son susceptibles de ser víctima del ataque. La medida más aconsejable a adoptar es la de "segmentar la red". La referencia al concepto de "segmentar la red" no es el de segmentación física, ni tampoco el de segmentación lógica mediante vLAN [54], que también se puede aplicar y es además una de las soluciones a ataques persistentes [55]. El objeto que se quiere segmentar es el grafico. Dado que la profusión de líneas que impedirían una visión nítida de la información y que, a priori, no se pueden descartar heurísticamente equipos, la solución se centra en la clasificación de los equipos por categorías. Las categorías pueden ser muy dispares:

- redes y subredes, de esta forma se crearía un anillo diferente para cada red. Así se reduce el número de direcciones IP que gestiona cada anillo, pero hace necesario tantos anillos como subredes dispongamos. Destaca como ventaja que presenta la información de un entorno de red completo, por lo que en el caso de advertir algún problema, se puede aislar rápidamente y con facilidad la zona hasta solventarlo.
- por servicios prestados, creando un anillo por cada servicio, y generando otro(s) para las estaciones de trabajo. Presenta como ventaja la visualización de vulnerabilidades soportadas por algún servicio determinado, pero tiene como inconvenientes que no se relaciona con la ubicación en la que se encuentra y que las estaciones de trabajo se visualizan aparte.

- por criticidad de los equipos. Esto implica la realización de una catalogación previa de los equipos. Presenta la ventaja que el riesgo sobre el equipo es directamente proporcional a su criticidad, pero padece de los mismos inconvenientes que la clasificación anterior.
- otras.

Dado que este tipo de ataque es personalizado, la defensa tiene que construirse también ad hoc, vinculada a la estructura de la organización y a la información con la que trabaja, por lo que no se puede proponer alguna de las clasificaciones anteriores mejor que las otras. La mejor será la que facilite una defensa más optima de la propia organización, fruto de la experiencia, y muy probablemente, si el sistema de información de la empresa es profuso, sea una combinación de varias.

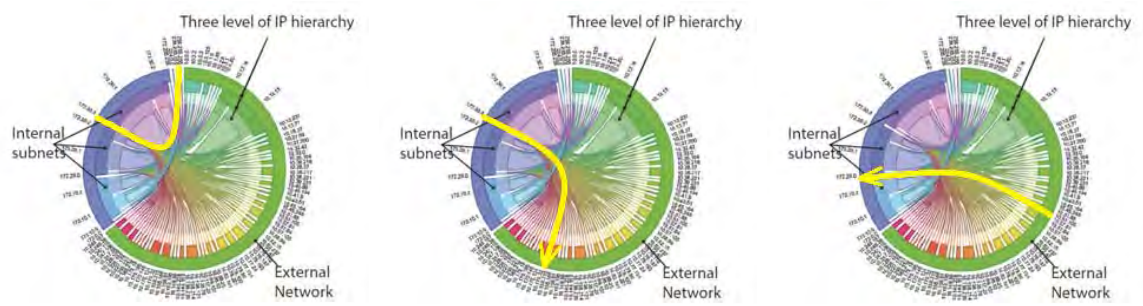


Figura 76. Gráficos de anillo de conexiones interna, externa saliente y externa entrante.

Así, pueden visualizarse todas las comunicaciones de la red corporativa, ya sean conexiones internas, entre dos equipos de la empresa (figura 76.a), o externas, entre un equipo de la propio de la empresa y otro que no pertenece a la red propia, ya sea en dirección saliente (figura 76.b) como entrante (figura 76.c). Es el momento de buscar patrones de conducta que resulten anómalos y que proporcionen indicios que se está produciendo un ataque. Conductas que resultan anómalas y que pueden ser consideradas como indicio, entre otras, pueden ser:

- conexión con una dirección IP exterior desconocida que se repite varias veces a lo largo del día, de la semana o del año. Evidentemente requiere que exista una base de datos o repositorio con las direcciones IP a donde nuestras maquinas se conectan. También debe evaluarse como sospechoso si la conexión al equipo exterior, aun no siendo siempre el mismo, pertenece a la misma red.
- conexiones con direcciones IP externas a horas extrañas, intentando eludir la vigilancia de las redes. Con los modernos sistemas de monitorización, firewalls, IDS e IPS en activación 24/7, pierde sentido, pero si las redes solo tienen vigilancia de 8/5, tiene muchas probabilidades de permanecer oculto más tiempo. Si la actividad de los sistemas disminuye drásticamente fuera de los horarios de oficina, el atacante tiene que perfilar muy bien el

flujo de transmisión para que no se produzcan grandes picos sospechosos que facilitan su detección.

- conexiones con direcciones IP externas cuando baja el caudal de trafico del sistema. Es una variante de la opción anterior que en lugar de vigilar el horario, vigila el trafico. Si consigue equilibrar el flujo de información, a través de la red, cuando este disminuye, para que se mantenga en unos niveles medios, dificulta su descubrimiento. Esto obliga a vigilar no únicamente las conexiones, sino también el tráfico y el tipo de este.
- conexiones cifradas y/o tunelizadas. Las primeras son muy habituales ya que muchos servicios son ya hoy, por defecto, cifrados, por lo que marcar estas comunicaciones podría ser incluso contraproducente. Muy distinto caso es el de la tunelización, por lo que se podría marcar en el grafico para que resaltara. Podría ser tan sencillo como marcar con un "*" las direcciones IP que establecen una conexión tunelizada. Que la conexión sea tunelizada no es coyunturalmente un incidente malicioso, pero si digno de investigación.
- conexiones entre estaciones de trabajo (no servidores) de la red corporativa. Anteriormente se ha visto como entre estos equipos se emiten órdenes para ejecutar acciones. Cuando las conexiones entre estaciones de trabajo se hacen frecuentes, suceso que aunque parezca muy normal, no lo es tanto, ya que lo habitual es comunicarse con servidores, puede ser un indicio que está ocurriendo algo anómalo, digno de indagación.
- conexiones con servidores de la red corporativa, ya sean desde estaciones de trabajo o desde otros servidores. El primer caso, conexiones establecidas desde estaciones de trabajo, es el modo normal de trabajo, por lo que hay que buscar circunstancias que salgan de esa normalidad. Algunos ejemplos pueden ser trafico HTTP fuera de horario de oficina, cuando nadie navega por páginas web. Trafico SMTP a esas mismas horas, etc. De igual forma que en casos anteriores esto no es más que indicio, debiendo ser analizado. Mas sospechoso resulta establecer trafico SMTP con un servidor web, pero habiéndose diseñado el malware a medida de la red víctima, esta será una circunstancia que difícilmente tendrá lugar. Al hilo de esta afirmación y continuando con el segundo caso, conexiones entre servidores, será muy difícil que dos servidores con distinto protocolo de aplicación y que en principio no tengan que establecer conexión, lo hagan. Pero no debe descartarse. Son los errores los que hacen que el engaño quede en evidencia y sea detectable.
- cualquier otra conexión, que por su naturaleza, pueda resultar extraña y despierte inquietud al analista que la está observando.

A lo largo de todas las circunstancias anteriores, que por su naturaleza puedan resultar anómalas, se aprecia que en casi todas ellas uno de los puntos de interés que apoya el análisis para detectar si se trata de un ataque son los protocolos de aplicación con los que se está trabajando y el caudal de trafico que en cada momento se gestiona. El grafico de anillo no ofrece en ningún momento información de estos dos aspectos, por lo que se hace necesario complementarlo con otra figura que si la proporcione. Estas, como ya se estudio, pueden ser dos: el flujo de conexiones y el diagrama de coordenadas paralelas. Al igual que en el análisis realizado en el apartado 5.1, si no son necesarios más datos que los que aporta el flujo de conexiones, y este es el caso, este es preferible sobre el diagrama de conexiones paralelas.

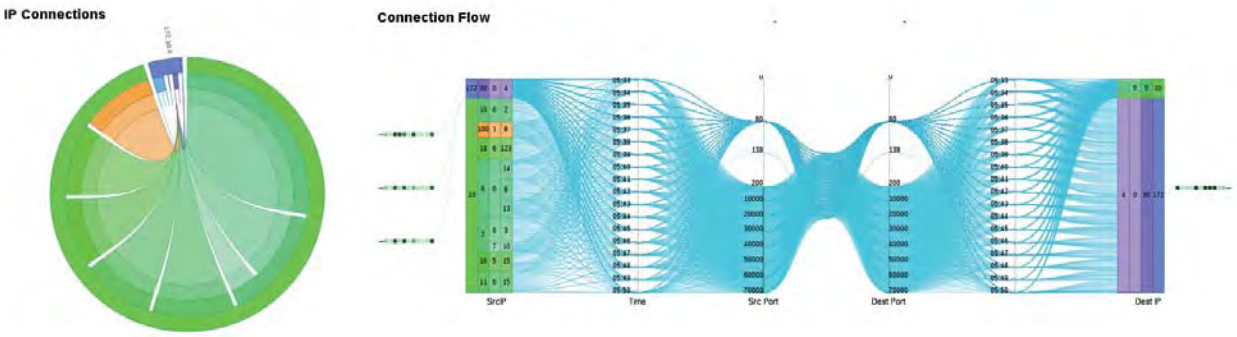


Figura 77. Gráfico de anillo complementado con un gráfico de flujo de conexiones.

Con esta nueva figura, constituida por un grafico de anillo y un flujo de conexiones (figura 77), se pueden vigilar las conexiones internas, conexiones externas, salientes y entrantes, protocolos de aplicación con los que se realizan (recordemos que va ligado al puerto lógico) y el caudal del trafico que se genera en cada momento. Complementado con una base de datos donde almacenar las conexiones proporcionando memoria al sistema, estamos en disposición de generar inteligencia.

Aun así, y para facilitar la labor del analista que le lleve a tener una fotografía mental de la realidad (situational awareness), falta por representar un dato que, hasta el momento no se ha tenido en cuenta. Cuando se intenta establecer una conexión con un equipo con una dirección IP (o también con un dominio) incluido en una lista negra, donde se registran las direcciones IP de hosts maliciosos que generan spam de forma voluntaria o involuntaria [56], o empleados de repositorios de malware o destinados como Command & Control de botnes; y que se utilizan comúnmente para proteger los sistemas informáticos contra amenazas de malware [57]. Si los equipos de defensa perimetral corporativa están bien configurados y actualizados, estos deberán bloquear la conexión, con lo que esta no existe. ¿Debe esta "no conexión" aparecer en el grafico estudiado? La respuesta es, sin ninguna duda, ¡si!. Esto obliga a modificar

parcialmente la figura, generando un espacio en el anillo para ello. El mejor color, que genera conciencia de lo que está ocurriendo y que permite saber que es una conexión que en realidad no ha tenido lugar es el negro (figura 78).

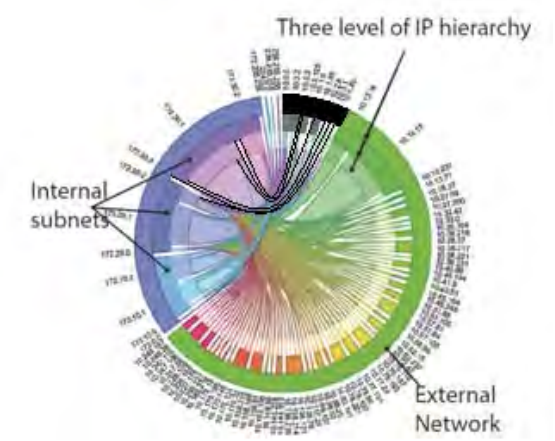


Figura 78. Grafico de anillo con conexiones a listas negras

Hasta aquí se ha descrito el grafico propuesto para una visualización de un ataque persistente con exfiltración de datos. Esta representación resulta útil en el proceso de monitorización, vigilando las redes "online". Pero, ¿como se puede hacer una reconstrucción forense de este mismo ataque? ¿Cómo encontrar el momento en el que hay un indicio o una evidencia que esto está ocurriendo dentro del sistema de información corporativo, si no es conocido el instante de tiempo en el que ocurrió? Esto lleva al analista a jugar con el tiempo para realizar la búsqueda.

En este nuevo cometido es necesario realizar una reconstrucción a lo largo de un periodo de tiempo, de algo que ya ha ocurrido. Leyendo los "logs" del periodo de tiempo a reconstruir se puede realizar esta tarea. Para facilitar una más rápida reconstrucción es por lo que se propone construir esta nueva figura. Mediante el uso del grafico de rejilla o del ingles "trellis plot", se puede construir un marco de celdas donde situar diferentes gráficos de anillo, representado la misma red en diferentes instantes de tiempo (figura 79).

Para que la percepción sea eficaz los intervalos de tiempos entre cada uno de los gráficos debe ser constante, de esta forma el analista conoce el tiempo que le falta a un grafico para convertirse en el siguiente. Dado que hay que vigilar todos los gráficos al mismo tiempo no pueden ser un número elevado. Para un grafico de rejilla cuadrado, un número congruente seria 9, pudiendo llegarse hasta 16. Un grafico de 25 celdas resulta excesivo, perdiendo el analista capacidad de fijación, pudiendo no considerar relevantes detalles que, verdaderamente, fueran significativos para la resolución del incidente. El grafico puede no ser cuadrado, por lo que también se pueden considerar como validos el numero de 6, 8 y 12 celdas de rejilla. Nuevamente, 18 celdas se consideran excesivas. Probablemente, una persona con entrenamiento en el tiempo sea capaz de aumentar el número de celdas a vigilar, pudiendo apreciar los detalles de cada una de ellas.

Otro aspecto a sopesar es el tiempo de evaluación. Este no puede ser desmesurado. Se considera que un analista experimentado, con los números expresados de celdas, no es capaz de concentrar toda la atención, vigilando la actividad

de todas concurrentemente, más allá de un minuto, aproximadamente, lo que supone para un grafico de rejilla de 16 un total de 16 minutos. Si no se ha observado nada anómalo, se volvería a empezar en el minuto siguiente donde se acaba la presentación actual. En el ejemplo expresado seria en el minuto 17.

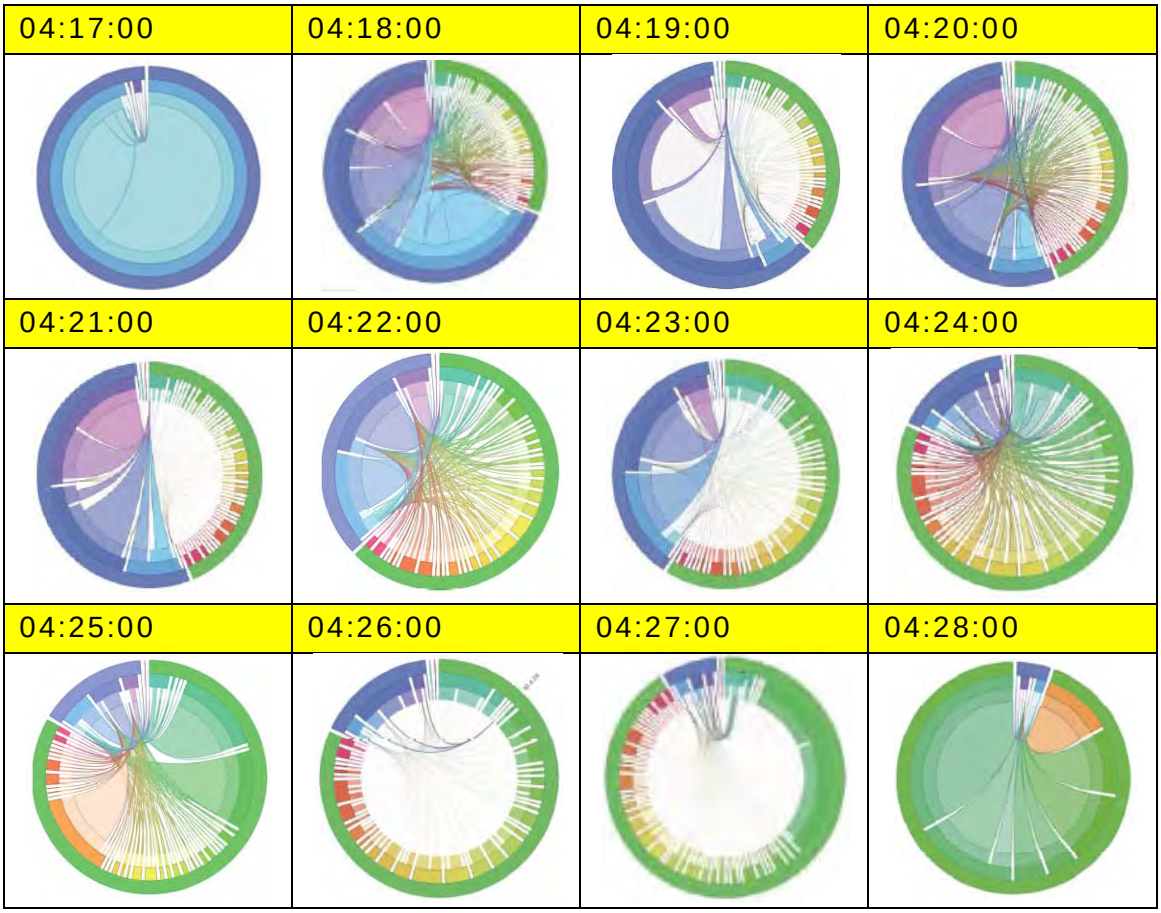


Figura 79. Trellis plot de 4*3 de gráficos de anillo.

Construido el grafico, el paso siguiente es que empiece a "correr". Para realizar esta función cada uno de los gráficos que componen la rejilla debe avanzar en el tiempo hasta alcanzar el instante reflejado en la casilla siguiente, momento en que el grafico se detiene, y se vuelve a componer en otro intervalo de tiempo. Para una mayor facilidad en el avance del grafico se refleja la hora local de la red estudiada. Es importante que sea la hora de trabajo de la red, y no la hora en que se está estudiando, para que proporcione información del momento en que se produce el posible evento. El avance del reloj no debe ser lineal con el tiempo, pudiendo manipular la velocidad del mismo, acelerándola o ralentizándola, avanzando o retrocediendo a voluntad, mediante mandos de avance, pausa, stop, avance rápido y retroceso rápido (figura 80).

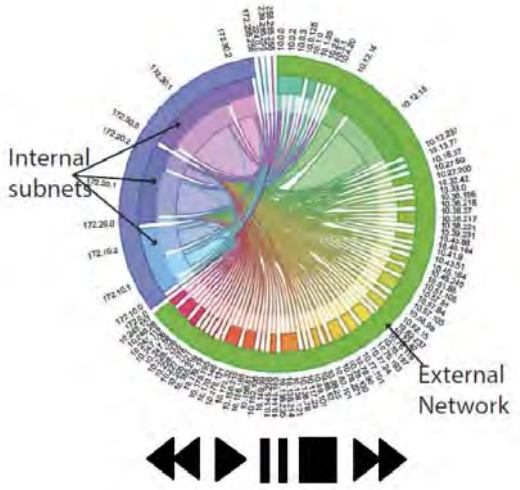


Figura 80. Gráfico de anillo con mandos de avance.

Cuando el analista encuentra un posible incidente, el cuadro seleccionado se debe poder apartar para estudiarlo individualmente, de forma exhaustiva, prestándole toda la atención. En este momento, cuando solo hay un grafico, se acompaña de un grafico de flujo de conexiones como ya se había expresado, facilitando información adicional del flujo de transmisión.

5.4 For Your Eyes Only.

Siguiendo el guion inicial, únicamente resta ya, presentar la información mediante gráficos a la alta dirección. Evidentemente, y como ya se ha reflejado anteriormente, no se trata de crear conciencia de un incidente: como se inicia, como se desarrolla y como se concluye; pudiendo realizarse la presentación de estas tareas, en momentos puntuales, y de forma somera, aclarando y puntualizando aquellos detalles sobre los que se debe fijar especial atención para una mejor comprensión.

A la alta dirección deben presentarse datos globales que permitan alcanzar una conciencia situacional del entorno de trabajo corporativo, acerca de los vulnerabilidades, riesgos y ataques a los que está sujeta la empresa. Hoy, viernes 12 de mayo, casualmente, se ha realizado un enorme ciberataque a nivel global que ha golpeado sistemas informáticos en decenas de países. Los análisis del Instituto Nacional de Ciberseguridad (INCIBE) de España han determinado que el software malicioso es un WanaCrypt0r, conocido ransomware, que ha afectado, entre otros, a los equipos de la sede de Telefónica en Madrid, al sistema de salud británico o el ministerio del Interior ruso [58]. Sirve de poco que a los presidentes de cualquiera de estas empresas se les explique con gráficos cómo evoluciona el virus cifrando los elementos de almacenamiento de la información en un equipo, pero puede resultar fundamental iniciar la exposición de los responsables de seguridad con gráficos que permitan conocer donde se ha detectado el virus a nivel global, mediante un mapa (figura 81) y paralelamente otro que visualice las sedes de la propia empresa donde ha sido descubierto.

Figura 81. Mapa creado por Malware Tech con los sitios en los que se ha detectado el ransomware [58].

Si al presidente de la empresa se le ofrece un listado con todos los lugares donde el virus ha tenido alguna incidencia (figura 82), se le está ofreciendo, exactamente, la misma información que en el grafico anterior, pero resulta evidente que la persona no alcanza el mismo grado de conciencia con ambas informaciones. Se deja a elección del lector decidir cuál es la que le proporciona una mejor apreciación de la situación.

Clasificación	Ciudad	Índice
1	Sorrento (Italia)	84,71%
2	Dresde (Alemania)	84,13%
3	Gdańsk (Polonia)	84,01%
4	Dubrovnik (Croacia)*	83,71%
5	Siem Reap (Camboya)*	83,35%
6	Cracovia (Polonia)*	83,25%
7	Budapest (Hungría)*	83,20%
8	Chicago (IL, EEUU)	82,98%
9	Varsovia (Polonia)*	82,95%
10	Split (Croacia)*	82,86%
11	Brujas (Bélgica)*	82,62%
12	Washington D.C. (EEUU)	82,61%
13	Hanói (Vietnam)	82,44%
14	Verona (Italia)*	82,32%
15	Marrakech (Marruecos)*	82,18%
16	Bucarest (Rumanía)	82,07%
17	Kioto (Japón)*	82,02%
18	Praga (República Checa)*	82,01%
19	Santiago (Chile)	81,92%
20	Riga (Letonia)*	81,92%
21	Venecia (Italia)*	81,89%
22	Lisboa (Portugal)*	81,68%
23	Viena (Austria)*	81,67%
24	Berlín (Alemania)*	81,86%
25	Oporto (Portugal)*	81,58%
26	Florencia (Italia)*	81,50%
27	Tokyo (Japón)*	81,46%
28	Nueva Orleans (LA, EEUU)	81,41%
29	**Valencia (España)***	**81,08%**
30	Dallas (TX, EEUU)	81,02%
31	San Petersburgo (Rusia)*	81,01%
32	Buenos Aires (Argentina)*	81,00%
33	**Madrid (España)***	**80,91%**
34	Estambul (Turkía)*	80,77%
35	**Granada (España)***	**80,54%**
36	Atenas (Grecia)*	80,51%
37	**Barcelona (España)***	**80,49%**
38	**Sevilla (España)***	**80,47%**
39	Edimburgo (Reino Unido)*	80,45%
40	Atlanta (GA, EEUU)	80,39%
41	Melbourne (Australia)*	80,33%
42	Montreal (Canadá)	80,28%
43	Estocolmo (Suecia)*	80,20%
44	Dubai (Emiratos Árabes Unidos)	80,16%
45	Palermo (Italia)	80,09%
46	Osaka (Japón)	79,96%
47	Hamburgo (Alemania)	79,95%
48	Brisbane (Australia)	79,92%
49	Surat Thani (Tailandia)	79,91%
50	Kiev (Ucrania)*	79,90%

Figura 82. Listado del grafico anterior.

El primer elemento, imprescindible para tomar consciencia de la situación, sin tener en cuenta ningún incidente de seguridad, es un mapa de los elementos de red como activos de la infraestructura propia. Como ya se expuso en el apartado 4.2, los elementos deben aparecer y desaparecer en función del nivel de detalle que se estén presentando. A nivel mundial, únicamente deberían representarse las líneas de comunicaciones internacionales y los nodos principales, o de otro modo, la figura se transformara en ininteligible (figura 9). En este punto son de aplicación todas las indicaciones expuestas en el apartado 4.2 acerca de la representación de más datos con funciones graficas. A medida que la localización representada es más reducida deben ir aumentando los elementos. Se pueden mostrar diferentes equipos hardware, eligiéndolos por distintos criterios: únicamente routers, routers y firewalls, firewalls por criterios de familias, todos los servidores, servidores por sistema operativo, servidores por servicio prestado, terminales de usuario, solo algunas por criterios dispares y todas las combinaciones imaginables.

Siguiendo, como hilo conductor de la exposición, con el caso del ciberataque de las últimas horas, se proponen diferentes presentaciones graficas a la alta dirección, en función de cómo haya sido damnificada. El ciberataque ha afectado a un centenar de países, y en España lo han sufrido alrededor de una decena de grandes empresas de servicios (de las que se tenga noticia), siendo la compañía más afectada Telefónica. Varios centenares de ordenadores de su sede central del Distrito C de Madrid se vieron infectados [59]. De acuerdo con la información proporcionada por el CCN-CERT, se ha identificado el ataque de ransomware como una variante de WannaCry, que infecta la máquina, cifrando todos sus archivos, explotando una vulnerabilidad de Windows utilizando EternalBlue/DoublePulsar, y pudiendo trasladarse al resto de sistemas Windows conectados en esa misma red que no estén debidamente actualizados. Esto permite la ejecución de comandos de forma remota a través de Samba (SMB) [60]. Los sistemas Windows afectados son: Microsoft Windows Vista SP2, Windows Server 2008 SP2 y R2 SP1 Windows 7, Windows 8.1, Windows RT 8.1, Windows Server 2012 y R2, Windows 10 y Windows Server 2016.

Una vez publicada la noticia y conocida la amplitud del ataque, cualquier empresa que no se ha visto afectada debería preocuparse por la situación. Que no haya sido víctima de la agresión puede ser solo cuestión de tiempo, y en esta ocasión muy escaso. Un CEO (del inglés Chief Executive Officer) preocupado por lo que pudiera ocurrir, debería inmediatamente pedir información a su CIO (del inglés Chief Information Officer) de los riesgos a los que está sujeto. Una vez instruido en el incidente y sus peculiaridades, la primera inquietud debería ser, saber si en la empresa se dispone de equipos con esas características. Nuevamente, darle al CEO un listado con todos los equipos que las cumple, no es la solución. Para informarle adecuadamente, se puede desplegar un mapa con todas las sedes de la empresa, dibujando mediante glyphs adecuados, como se expuso en el apartado 4.1, aquellos equipos que soportan los sistemas operativos mencionados. Seguidamente resaltar aquellos que no se hayan parcheado siguiendo las indicaciones publicadas por el fabricante y que bloquean el ataque. Si el número de estos equipos es muy elevado se puede recurrir a presentarlos mediante formas, colores y números que proporcionen la información de modo apropiado y claro (figura 10). Cualquiera de los dos métodos que sea elegido, da una importante medida de las vulnerabilidades a las que está expuesto el sistema de información de la empresa.

Dado que el problema de este ciberataque, uno de muchos, es su propagación entre equipos vulnerables, el siguiente paso, debería ser conocer como está interconectado el sistema de información. Esto se consigue mediante diagramas de nodos de enlace. El uso del plural y no del singular se debe a que, si existen varios dominios de información, lo que se refleja en la creación de redes diferenciadas, es decir que no existe conexión entre ellas, por tratar asuntos distintos, o tener distinta calificación de seguridad y no poder interconectarse, deberán dibujarse tantos diagramas como redes. Evidentemente, se superpondrán, siendo el mejor método para discriminarlas el uso de colores que las diferencie. Aun así, probablemente, el número

de líneas se hará excesivo, debiendo poder seleccionar uno u otro, para delimitarlas con mayor claridad.

En los nodos especiales, aquellos que puedan, por diferentes motivos de criticidad, ser considerados de singular relevancia, se puede particularizar su presentación mediante la técnica de diagrama de nodo de enlace realizado con planos hiperbólicos. Como se recordara del apartado 4.6, esta técnica permite centrar el foco en el nodo que se está analizando, distorsionando el resto, siendo esta distorsión mayor cuanto mayor es la distancia a este nodo (figura 37). De esta forma, se comprueban de manera más eficiente las amenazas que puedan provenir de otros nodos laterales y, sobre que otros nodos, puede ser más pernicioso el que se está estudiando, en caso de estar comprometido. Este análisis, permite estudiar cómo podría realizarse una posible desconexión de alguna de las partes, de tal forma que pudiera permanecer aislada mientras es tratada y poder continuar operando con el resto del sistema, si esto fuera posible.

Hasta aquí, esta información podría crear una imagen mental de la situación de los sistemas de información de la empresa al CEO de la misma, desde la posición de no haber sido víctima del ataque. Ahora, y desde otro punto de vista, ¿cómo se podría informar gráficamente al CEO de otra empresa, pero que hubiera sido víctima del ataque?. En primer lugar, al igual que en el caso anterior, es necesaria la información acerca de la naturaleza del ataque y que sistemas son los afectados (el CEO no siempre posee conocimientos de telecomunicaciones y como se pueden comprometer los sistemas). De nuevo la figura más explicativa vuelve a ser un mapa con las sedes de la empresa, pero esta vez señalando todas aquellas que se han visto afectadas. La elección de colores más lógica seria rojo, para aquellas que se han visto afectadas, y verde para el resto. Transmitir información de relevancia y dimensiones de la sede se puede realizar a través de intensidades del color y tamaño del glyph. Para comunicar más datos, no se encuentra otra opción, en detrimento de ofuscar la figura si no se realiza, que presentar estos mediante un cuadro de dialogo al posicionar el cursor sobre la sede.

Para adquirir una mayor conciencia de la situación, la realización de un cuadro con el número de equipos distribuidos por sistemas operativos, es un dato muy relevante. Este dato es también interesante dividirlo por sedes, pero realizarlo sobre el mapa provocaría un exceso de información que impediría apreciarla. Es conveniente, por tanto, que este dato aparezca dentro del cuadro de dialogo que se indico anteriormente, y que aparece al posicionar el cursor sobre la sede. Esto supone que la información que se proporciona mediante cuadros de dialogo, no es fija. La configuración de estos cuadros de dialogo debe proporcionar cierta flexibilidad, que permita modificarla en función de los eventos que se traten, y el asesoramiento concreto que se desee transmitir a la dirección.

Como en el caso de no haber sido afectado por el ataque, resulta también interesante, como medida preventiva frente a sedes, o elementos de sedes, que todavía no lo han sufrido, conocer por parte de la dirección las posibilidades de desconexión, si es que existieran, y la capacidad de continuidad de negocio con estas sedes, o quizás parte de ellas. Ante la envergadura de la acción, esta correspondería a la dirección. Nuevamente, la presentación mediante diagramas de nodo de enlace realizado con planos hiperbólicos es una buena solución, resaltando aquellos departamentos que fueran susceptibles de poder ejecutar esta medida.

Finalmente, y para cerrar la información proporcionada al CEO en este caso concreto, se puede visualizar cómo y cuando se realiza la conexión con el dominio malicioso que actuaba como elemento de control de la propagación de la infección. En el caso de no haber sido damnificado, seria ver como no se conecta. Como se recordara la figura que nos proporciona esta información es el grafico de anillo (figura 58), válido en ambos casos.

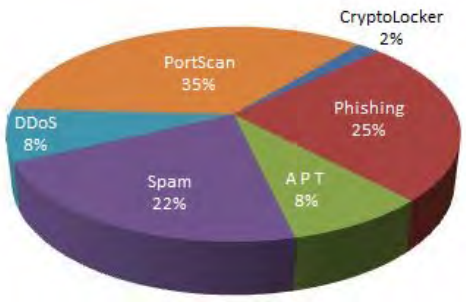


Figura 83. Grafico de sectores de ataques sufridos.

Volviendo a la información genérica, necesaria para gestionar el normal funcionamiento de los sistemas de información, y dejando de lado el incidente tratado, es indispensable generar otros elementos gráficos que proporcionen consciencia de la situación actual y que permitan obtener una visión de futuro. En buena parte, esto puede conseguirse presentado el todo compuesto por sus partes. La figura que representa esta especificación es el grafico de sectores (figura 83). Pero, ¿que todo puede representar el grafico de sectores? El todo se refiere a cualquier variable que pueda ser descompuesta cualitativamente, principalmente, y variables cuantitativas, en menor medida. Un ejemplo podría ser el de la figura (figura 83) en el que se representa el caso de tipos y porcentajes de ataques sufridos por la empresa en el último año, proporcionando a la dirección un conocimiento que le permita entender que inversiones, en materia de seguridad, son mas optimas y rentables.

También se puede, siendo esto menos habitual, representar variables cuantitativas mediante esta técnica, teniendo siempre como condición que se represente la totalidad de los elementos que la componen. Una muestra de esta cualidad podría ser la representación, en números totales, de los sistemas operativosque gobiernan las maquinas de la empresa. En la figura (figura 84) se representan los números de equipos con el sistema operativo Microsoft Windows, en

sus diferentes versiones, instalados en los terminales de usuario de una gran empresa, de acuerdo a los números expresados en el grafico. En el mismo se aprecia que más de un 50% de las maquinas están soportadas por la versión Windows XP. ¡Si esta imagen se la hubieran presentado al responsable del departamento de tecnología del sistema de salud público británico (NHS Digital)!. Sabiendo que Windows XP dejo de estar soportado por Microsoft hace ya algo más de dos años, y por tanto, no existiendo desde entonces actualizaciones de seguridad, es fácil deducir que existe un problema serio de vulnerabilidad del sistema en la empresa. Debido al alto coste que suponía renovar las licencias del sistema operativo, el sistema de salud británico (National Health Service, NHS) decidió "ahorrar" en esa partida presupuestaria [61]. Hoy el ahorro supone un gasto superior, además de una grave extorsión, en los dos sentidos. Adquiriendo conciencia de la situación, el paso lógico es invertir, paulatinamente, en licencias del software que actualicen el parque y que eviten el problema antes que suceda. Un grafico como este (figura 84), y conociendo los riesgos a los que se enfrenta, puede ser muy convincente. En este punto, cabe la satisfacción de resaltar que los sistemas de información de nuestras Fuerzas Armadas, resistieron este ataque global de un modo sobresaliente, gracias a las políticas de actualización, promulgadas por el CIO del Ministerio y con el esfuerzo de cientos de profesionales por desarrollarlas, no registrándose ningún incidente, que pusiera en riesgo la Defensa Nacional.

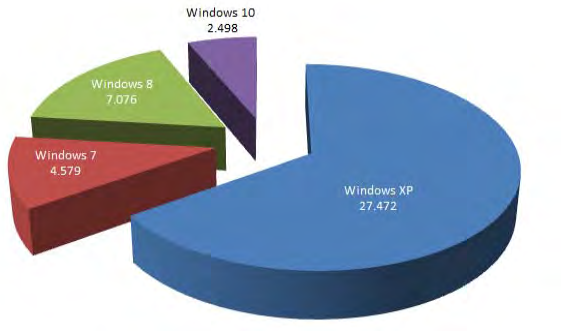


Figura 84. Gráfico de sectores de S.O. de usuario de una gran empresa.

Sin embargo, como ya fue expuesto en el apartado 4.8, el grafico de sectores no permite apreciar con nitidez suficiente la diferencia de valores que sean relativamente próximos. Si el grafico de sectores se acompaña de un histograma o de un gráfico de barras (figura 85), dependiendo del tipo de variable, este procedimiento queda clarificado. Es cierto que no son necesarios ambos gráficos juntos. Cada uno de forma independiente, proporciona la información necesaria. Se pretende, además, que de esta forma cree una idea mental al analista que la estudia. El grafico de sectores genera la idea de examinar la totalidad de los elementos de la variable, pero genera imprecisión en la cuantificación de los valores tomados. Al contrario, el grafico de barras es muy preciso en proporcionar visualmente el valor, pero no proporciona una sensación de observar todos los valores que la variable puede tomar. Este es el valor añadido que se obtiene al trabajar con ambos gráficos.

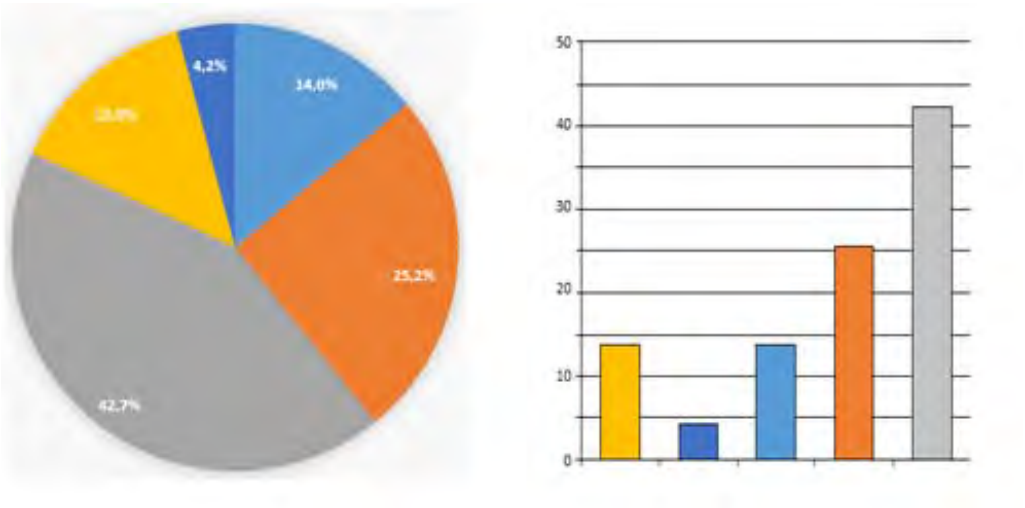


Figura 85. Gráfico de sectores y grafico de barras anexado.

Finalmente, y aunque no esté directamente relacionado con la presentación grafica de incidentes en el ciberespacio, es necesario convencer y concienciar a la alta dirección que la inversión en seguridad minimiza el riesgo y produce beneficios. Aunque esta afirmación parece evidente y cualquier individuo, a priori, está plenamente de acuerdo con ella, requiere una amplia justificación ante la alta dirección. El gasto para disminuir los riesgos de los sistemas de información frente a las amenazas que los acechan, es normalmente muy elevado y ante la dirección, en apariencia, no ofrece un retorno adecuado a la inversión. Se hace necesario convencer al equipo directivo que la financiación en seguridad es productiva, mucho más que reparar los daños. Esto último, a día de hoy, hay muchas empresas que ya lo han comprendido, pero han tenido que pagar un alto coste, tanto económico, como en imagen.

Un grafico adecuado para confeccionar una imagen de resultados positivos, siempre que lo sean, es el gráfico de dispersión, buscando aportar sensaciones de objetivos alcanzados. A modo de ejemplo, en la figura (figura 86) se puede observar como la inversión económica que, a lo largo de diez años, realiza la compañía en diferentes aspectos de seguridad (hardware, software, formación, concienciación…) motiva una disminución paulatina en los diferentes tipos de incidentes en los que se ve involucrada.

En la figura se visualizan dos aspectos positivos: el primero, como ya se ha indicado, una tendencia positiva que hace disminuir la cantidad de incidentes, con el paso de los años, debido a las medidas adoptadas para paliar los riesgos y a la experiencia adquirida por todos los actores; y el segundo, cuando se produce una acentuada inflexión que altera la tendencia positiva (ver figura 86, línea "insider abuse", entre los años 2014 y 2015) corroborar que ante una grave incidencia, se han tomado medidas que han sido correctas y han logrado restablecer el rédito anterior.

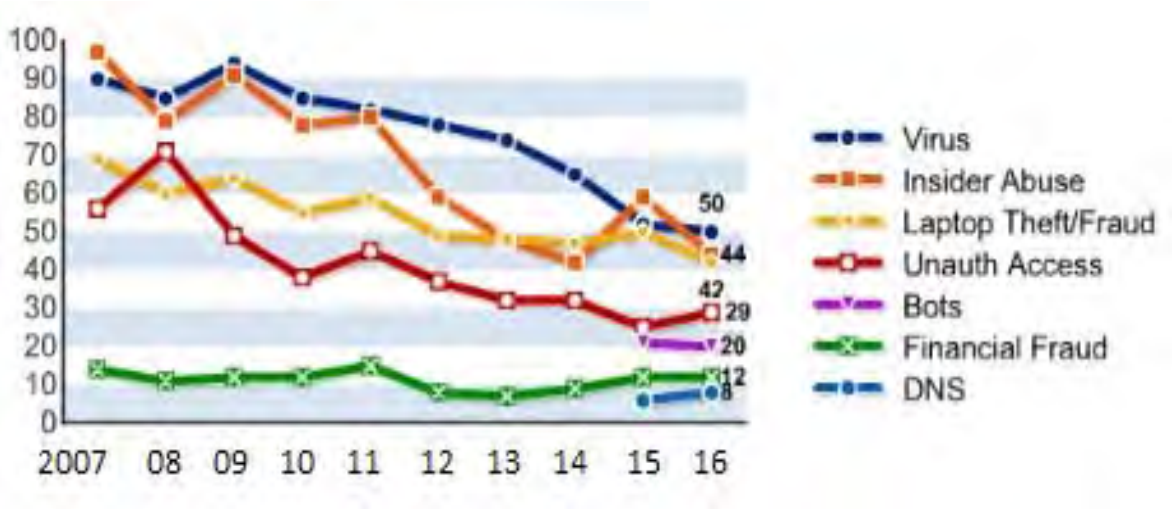


Figura 86. Evolución de incidentes en un diagrama de dispersión.

Este tipo de información crea en el responsable de las decisiones económicas la impresión de realizar inversiones correctas, y representarlas en este tipo de figuras ayuda en gran medida a crear esa imagen mental de una financiación adecuada. Como ya se he expuesto, resulta muy difícil justificar un gasto que, aparentemente, no tiene retorno. ¡Este es el retorno!

En otro ejemplo (figura 87), se observa como la inversión en medidas de seguridad hace disminuir el riesgo, siendo los usuarios más capaces de discernir cuando son o pueden ser víctimas de un ataque, consiguiendo ellos mismos, o por los medios dispuestos de la empresa, evitarlos. Evidentemente, el grafico ratifica el acierto de la inversión ante la dirección de la empresa, pudiendo conseguir que continúe en el tiempo. En este caso (figura 87), la relación entre las dos variables (miles de dólares e incidentes de seguridad) es no lineal negativa (apartado 4.5, figura 29.e), donde la función que las vincula no es una recta y el aumento de la inversión conlleva una disminución de los incidentes detectados.

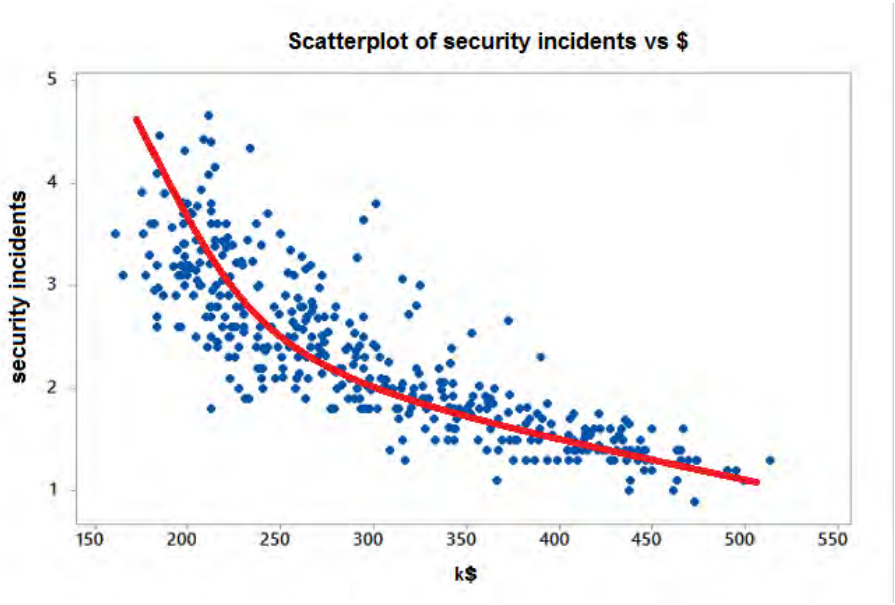


Figura 87. Evolución de la relación entre gasto en seguridad e incidentes detectados por empleado.

En el siguiente caso (figura 88), donde se presenta la relación entre el tiempo en meses que los empleados de una sede no reciben formación y/o concienciación en seguridad y el número de incidentes que la sede sufre, la relación es prácticamente lineal y positiva, es decir, al aumentar el tiempo que no se recibe formación aumenta el riesgo, y por tanto, el número de incidentes. Este grafico, resultaría útil si el objetivo es obtener financiación para realizar cursos o campañas de concienciación entre los empleados, si se observa que se está ocasionando algún menoscabo, en algún sentido, en este apartado.

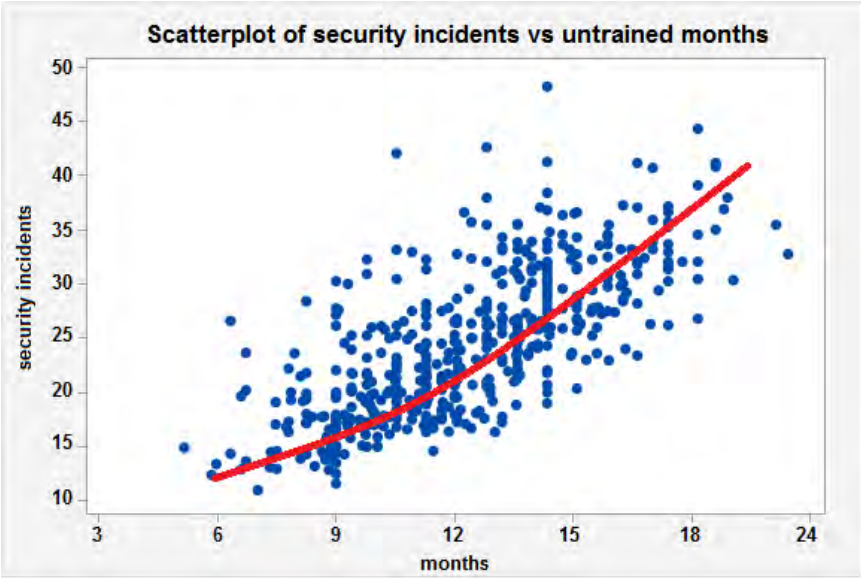


Figura 88. Evolución de la relación entre tiempo en meses sin formación e incidentes detectados por sedes.

Como se puede adivinar, ambas figuras son útiles a los fines que se pretenden alcanzar: convencer, en cada caso, que la inversión en la materia genera rentabilidad a la empresa. Presentarlas a la dirección puede ser una buena opción para ayudarles a tomar conciencia positiva de la labor realizada. Pero no todas las acciones tienen porque dar resultados positivos. Figuras como la siguiente, la información que proporciona no resulta conveniente presentarla ante la dirección, ya que el resultado que se observa no es positivo hacia los intereses de la empresa, pero si puede proporcionar al analista que lo estudia información sobre un problema que sería conveniente solucionar.

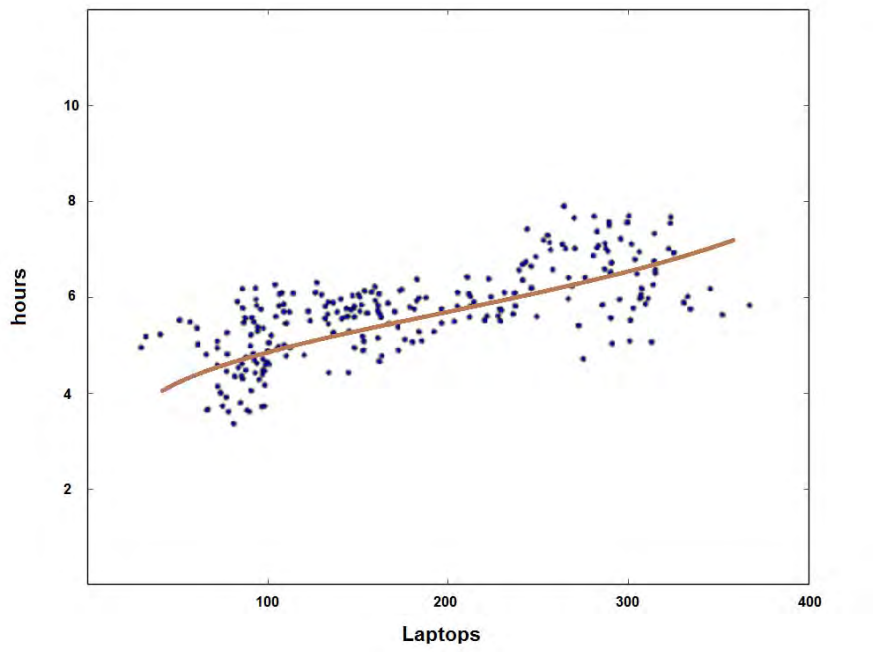


Figura 89. Evolución de la relación entre portátiles de empresa y horas semanales de navegación web.

En la figura (figura 89), se presenta la relación que existe entre el número de portátiles (no se tienen en cuenta los equipos de sobremesa) que la empresa proporciona a sus empleados, que por las tareas que desarrollan se determina que lo necesitan, y el número de horas semanales, dentro del horario laboral, que cada uno de ellos utiliza para navegar en internet. Se aprecia como con el aumento de ordenadores portátiles se experimente un aumento en las horas de navegación por internet que cada empleado dedica, un aumento prácticamente lineal y positivo, siendo esta tendencia negativa para los intereses de la empresa. Esto significa que la empresa tendrá que restringir el uso de portátiles, imponiendo condiciones más restrictivas, o desarrollar políticas que eviten que sus empleados no aprovechen adecuadamente el tiempo de trabajo. Estas conclusiones se obtienen habiendo "comprendido" la información aportada mediante el grafico, alcanzado una "situational awareness" que permita inferir comportamientos futuros.

6 Conclusiones

La toma de decisión se basa actualmente en la "comprensión" que, la persona o personas responsables de tomarla, adquieran de la situación. Habiendo analizado la información para adquirir "conocimiento", permite proyectar su estatus en el futuro cercano [3], proceso mental que se conoce como "situational awareness" o conciencia situacional.

Cuando el conocimiento se adquiere desde enormes volúmenes de datos, siendo estos complejos, ambiguos, obtenidos de muy diversas fuentes y de naturaleza heterogénea, se hace necesario el tratamiento automatizado de los mismos. Los datos, tratados mediante técnicas de Visual Analytics [16], son agregados y sintetizados, y con la ayuda de herramientas de análisis automáticas, transformados en información efectiva a través de interfaces graficas interactivas, siendo este, el modo grafico, el más sencillo y natural de asimilación por el cerebro humano, alcanzando esa situational awareness, que permita realizar la toma de decisión.

Las actividades que suceden en el ciberespacio son, también, susceptibles de representarse de modo grafico. Ante la diversidad de sucesos y situaciones que pueden ocurrir en este ámbito, es necesario seleccionar los gráficos que mejor puedan simbolizarlos. Cada suceso o situación debe exponerse con el grafico que mejor reconstruya la esencia de la información, presentando lo fundamental e ignorando lo superfluo. Por otro lado, no se debe olvidar que existen diferentes jerarquías de decisión, desde un nivel más básico, puramente táctico o técnico, hasta otros de rango operacional y, en lo más alto de la escala, de condición estratégica. Para cada nivel, es necesaria distinta información, presentada mediante imágenes distintas.

Normalmente, cualquier ataque en el ciberespacio se inicia con un escaneo de puertos, donde se buscan maquinas que presenten vulnerabilidades. Para explorar un servicio concreto se realiza un escaneo horizontal [41], donde se llama a un puerto determinado (servicio) de todas las maquinas de la red. Si la búsqueda se realiza sobre todos los puertos de una maquina concreta, el escaneo es vertical [41] y se está buscando posibles vulnerabilidades de la misma. Ya sea el escaneo vertical u horizontal, existen dos herramientas graficas: connection river [37] y parallel coordinate [35], que proporcionan a los responsables técnicos consciencia de lo que está sucediendo. Ambas permiten visionar el flujo de bits durante un periodo de tiempo, variable fundamental para percibir el ataque. Connection river facilita la representación de: la maquina, el instante de tiempo y el puerto, tanto del atacante como de la víctima. Si no son necesarios más datos esta herramienta es adecuada para representar el escaneo. Si se requieren más variables, se hace necesario el grafico parallel coordinate.

En un ataque por denegación de servicio, DoS o DDoS [42], se impide que usuarios legítimos accedan a servicios autorizados. Se materializa mediante el agotamiento de algún recurso del sistema (memoria, capacidad de proceso, ancho de banda, …) impidiendo que este, el sistema atacado, responda a la petición de servicios. La representación del ataque se efectúa mediante gráficos que representan el recurso y el uso que se hace de él. En los modelos de agotamiento de memoria, capacidad de proceso y anchos de banda, el grafico que más fielmente refleja esta situación es la línea de tiempos. En ella se relaciona el porcentaje de ocupación del recurso con el tiempo. Esto permite establecer una relación visual de la evolución de uso en un intervalo de tiempo, y "predecir" lo que va a ocurrir en un futuro cercano. En el caso de la representación del empleo del ancho de banda, es necesario establecer un límite dinámico del consumo habitual que del recurso se hace en diferentes instantes de tiempo. Para representar el agotamiento del servicio, el grafico más representativo es el diagrama de sectores, en el que se representan los diferentes servicios. Si se realiza un ataque, se observará que uno de ellos prevalece sobre el resto, no pudiendo responder al resto, ni al atacado. Si el objetivo es realizar una reconstrucción a posteriori del ataque sufrido (análisis forense), se puede visualizar muy eficazmente mediante un grafico connection river [37], en el que se visualiza que no existe flujo de información durante un intervalo de tiempo.

Los ataques de fuga de información [45], ponen está en poder de personas ajenas a la organización. El modelo de ataque de APT [47] queda residente en el sistema, exfiltrando la información hacia el exterior durante largos periodos de tiempo. Para no ser descubierto tiene que transferir la información de modo que se "camufle" con el normal funcionamiento del sistema. Los gráficos tienen que representar aquellos detalles que no sigan esta norma. El anillo de conexiones [37] es un grafico que representa las comunicaciones entre equipos. Para que sea eficaz debe representar todas las comunicaciones, incluso, las que no se efectúan, por ser interrumpidas por los equipos de defensa perimetral, permitiendo conocer con que equipos prohibidos intenta conectarse. Pequeñas anomalías como horarios extraños de conexiones, número de conexiones con equipos desconocidos, tamaño del flujo de información transferida, cifrado o tunelización de la comunicación u otras que puedan resultar sospechosas pueden delatar el ataque. Distintas combinaciones del anillo de conexiones pueden proporcionar la información necesitada. El grafico connection river [37] puede proporcionar información adicional que facilite la investigación.

En todos los casos la detección del ataque se basa en la búsqueda de patrones anómalos al normal funcionamiento de los sistemas que se están vigilando. Cualquier actividad que se separe del normal discurrir del sistema, resulta sospechosa y es susceptible de ser investigada. Utilizando sistemas inteligentes que tengan "capacidad de aprender" se logrará que puedan identificar estas anomalías, para posteriormente, mostrarlas de modo automático.

Por último, y aunque no se trata de figuras que muestran ataques, se hace necesario concienciar a la alta dirección de la empresa que la inversión en seguridad es una inversión rentable. Con tal fin se usan principalmente dos modelos de figuras: mapas y planos, dónde se representan las estructuras del sistema, por un lado, y los elementos comprometidos por otro; y un segundo modelo como son figuras estadísticas, de diversos tipos, que denoten una tendencia positiva de resultados, fruto de la inversión.

Bibliografía

[1] Mando de Adiestramiento y Doctrina, «Sistemas de Telecomunicaciones e Informacion (CIS),» de *OR3-501 ORIENTACIONES*, Madrid, Ejercito de Tierra. Ministerio de Defensa, Noviembre 2007.

[2] Endsley, M.R., «Designing for Situation Awareness in Complex System,» de *Proceedings of the Second Intenational Workshop on Symbiosis of Humans, Artifacts and Environment*, Kyoto, Japan, 2001.

[3] Endsley, M.R., «Design and evaluation for situation awareness enhancement,» de *Proceedings of the Human Factors Society 32nd Annual Meeting*, Santa Monica, CA, Human Factors Society, 1988, pp. 97-101.

[4] Bedny, G. & Meister, D., «Theory of activity and situation awareness,» *International Journal of Cognitive Ergonomics,* vol. 3, nº 1, p. 63–72, 1999.

[5] Smith, K. & Hancock, P., «Situation Awareness Is Adaptive, Externally Directed Consciousness,» *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, nº 1, pp. 137-148 , March 1995.

[6] Wickens, C., «Situation Awareness: Review of Mica Endsley's 1995. Articles on Situation Awareness Theory and Measurement,» *Human Factors and Ergonomics Society,* vol. 50, nº 3, p. 397–403, June 2008.

[7] Stanton, N. et al, «Situational Awareness and Safety,» *Safety Science,* vol. 39, pp. 189-204., 2001.

[8] Flach, J., «Situation awareness: Proceed with caution,» *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 37, nº 1, pp. 149-157, March 1995.

[9] Endsley, M.R., «Theorical Underpinnings of Situation Awareness: a Critical Review,» de *Situation Awareness Analysis and Measurement*, Mahwah, New Jersey, Lawrence Erlbaum Associates, Inc, 2000, pp. 1-24.

[10] Jones D. & Endsley M.R., «Sources of situation awareness errors in aviation.,» *Aviation, Space, and Environmental Medicine,* vol. 67, nº 6, pp. 507-512, June 1996.

[11] Herrera, A., «Modelo de Awareness Basado en Topologias de Interaccion para Espacios Virtuales de Trabajo Colaborativo,» *Revista Latinoamericana de Ingenieria del Software,* vol. 2, nº 4, pp. 219-261, 2014.

[12] Endsley, M.R., «Toward a Theory of Situation Awareness in Dynamic Systems,» *Human Factors and Ergonomics Society,* vol. 37, nº 1, pp. 32-64, March 1995.

[13] Figueroa, J., «Integracion de los requerimientos de conciencia situacional y grupal al desarrollo de sistemas colaborativos y dinamicos usando un enfoque basado en modelos,» Granada, Departamento de Lenguajes y Sistemas Informaticos, Universidad de Granada, Septiembre 2012.

[14] Endsley, M.R. & Jones, W., «Situation Awareness Information Dominance & Information Warfare,» Texas, United States Air Force Armstrong Laboratory, February 1997.

[15] Kintzel, C. et al, «Monitoring Large IP Spaces with ClockView,» Pittsburg, PA, USA, VizSec '11, July 2011.

[16] Thomas, J. & Cook, K., «Illuminating the Path. The Research and Development Agenda for Visual Analytics,» National Visualization and Analytics Center, 2005.

[17] Keim, D. et al, «Mastering the Information Age. Solving Problems with Visual Analytics,» Goslar, Germany, Eurographics Association, September 2010.

[18] Keim, D. et al, «Visual Analytics: Definition, Process, and Challenges,» March 2008.

[19] Erbacher, R., «Visual Behavior Characterization for Intrusion Detection in Large,» de *International Conference on Visualizaation*, Marbella, Spain, September 3-5, 2001.

[20] Kasemsri, R., «A Survey, Taxonomy and Analysis of Network Security Visualization Techniques,» Atlanta, Georgia State University, 2006.

[21] «HUAWEI MARINE NETWORKS,» Huawei Marine Networks Co., Limited, 27 February 2017. [En línea]. Available: http://www.submarinecablemap.com/. [Último acceso: 03 March 2017].

[22] «Redes Telemáticas. Plataforma para la difusión de conocimientos dentro del ámbito de las redes informáticas, redes de datos e Internet,» 24 September 2012.

[En línea]. Available: http://redestelematicas.com/arquitectura-de-internet/. [Último acceso: 05 March 2017].

[23] Labovitz, C. et al, «The Impact of Internet Policy and Topology on Delayed Routing Convergence,» de *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 22-26 April 2001.

[24] Internet Corporation For Assigned Names and Numbers, ICANN, 2016. [En línea]. Available: https://www.icann.org/es. [Último acceso: 05 March 2017].

[25] «Microsiervos,» IPligence Community Edition, 07 January 2007. [En línea]. Available: http://www.microsiervos.com/archivo/internet/ipligence-geolocalizacion.html. [Último acceso: 05 March 2017].

[26] McPherson, J., et al, «PortVis: A Tool for PortBased Detection of Security Events,» de *VizSEC/DMSEC'04*, Whashington, DC. USA., October 2004.

[27] Few, S., «Perceptual Edge,» Visual Business Intelligence Newsletter, September/October 2008. [En línea]. Available: https://www.perceptualedge.com/articles/visual_business_intelligence/over-plotting_in_graphs.pdf. [Último acceso: 01 September 2016].

[28] Hauck, T., «High Density Scatter Plots. Various ways to represent lots of points in a scatter plot.,» Lambda Omega Lambda, 01 May 2013. [En línea]. Available: http://blog.trenthauck.com/posts/high-density-scatter-plots/. [Último acceso: 19 September 2016].

[29] Turner, S., «Fix Overplotting with Colored Contour Lines,» Getting Genetics Done, 6 July 2012. [En línea]. Available: http://www.gettinggeneticsdone.com/2012/07/fix-overplotting-with-colored-contour.html. [Último acceso: 14 September 2016].

[30] Ware, C. & Bobrow, R., «Supporting Visual Queries on Medium Sized Node-Link Diagrams,» *Information Visualization,* vol. 4, nº 1, pp. 49 - 58, February, 2005.

[31] Lamping, J. & Rao, R., «Laying out and Visualizing Large Trees Using a Hyperbolic Space,» de *Proceedings of the ACM Symposium on User Interface Software and Technology*, Palo Alto, CA 94304, November 1994.

[32] Johnson, B. & Shneiderman, B., «Treemaps: a space-filling approach to the

visualization of hierarchical information structures,» de *2nd International IEEE Visualization Conference*, San Diego, USA, October 1991.

[33] Kamada, T., «On visualization of abstract objects and relations,» Tokyo, JAPAN, University of Tokyo, Department of Information Science, December 1988.

[34] Axelsson, S., «Visualisation for Intrusion Detection Hooking the Worm,» de *The proceedings of the 8th European Symposium on Research in Computer Security (ESORICS 2003),*, Gjøvik, Norway, 13–15 October 2003.

[35] Few, S., «Multivariate Analysis Using Parallel Coordinates,» Perceptual edge, September, 2006.

[36] Davies, J., «Jason Davies's Block 1341281,» bl.ocks.org, 5 October 2015. [En línea]. Available: http://bl.ocks.org/jasondavies/1341281. [Último acceso: 15 February 2016].

[37] Chen, S. et al, «OCEANS - Online Collaborative Explorative Analysis on Network Security,» de *VizSec '14*, Paris, France, 10 November 2014.

[38] «Boletin Oficial del Ministerio de Defensa,» de *Disposiciones Generaes. Organizacion*, Madrid, Imprenta del Ministerio de Defensa, 18 January 2016, pp. 982-1022.

[39] «Wikipedia,» Wikimedia Foundation, Inc., 20 March 2017. [En línea]. Available: https://es.wikipedia.org/wiki/Puerto_(informática). [Último acceso: 09 April 2017].

[40] «Wikipedia,» Wikimedia Foundation, Inc., 12 November 2015. [En línea]. Available: https://es.wikipedia.org/wiki/Escáner_de_puertos. [Último acceso: 09 April 2017].

[41] Bailey C. et al, «Detection and Characterization of Port Scan Attacks,» University of California, San Diego, Department of Computer Science & Engineering.

[42] Mirkovic, J. & Reiher, P., «A Taxonomy of DDoS Attack and DDoS Defense Mechanisms,» Los Angeles, Computer Science Department UCLA, 2004.

[43] «Software Engineering Institute blog,» Software Engineering Institute. Carnegie Mellon University, 1997. [En línea]. Available: https://www.cert.org/information-for/denial_of_service.cfm?.

[Último acceso: 26 March 2017].

[44] Houle, K. & Weaver, G., «Trends in Denial of Service Attack Technology,» Pittsburgh, Pensilvania, Carnegie Mellon University, October 2001.

[45] Cabarique, W. et al, «Factores y causas de la fuga de información sensibles en el sector empresarial,» de *Cuaderno Activa*, vol. 7, Ciudad de Medellín, Departamento de Antioquia, República de Colombia, Fundación Universitaria Maria Cano, Universidad de Antioquia, October 2015, pp. 67-73.

[46] Instituto Nacional de Tecnologias de la Comunicación, «GESTIÓN DE FUGA DE INFORMACIÓN,» León, Spain, May 2012.

[47] Trend Micro, «LATERAL MOVEMENT: How Do Threat Actors Move Deeper Into Your Network?,» Cupertino, CALIFORNIA, Trend Micro, Inc, 2013.

[48] Sood, A. & Enbody, R., «Targeted Cyberattacks: A Superset of Advanced Persistent Threats,» de *Cyberwarfare*, Michigan, IEEE Computer and Reliability Societies, January/February 2013, pp. 54-61.

[49] Sevillano, F., «MOOC Ciberseguridad: Ataques y contramedidas,» Universidad Rey Juan Carlos, 29 February 2016. [En línea]. Available: https://www.youtube.com/watch?v=v02jU53ooaM. [Último acceso: 25 April 2017].

[50] FireEye, «APT28: A WINDOW INTO RUSSIA'S CYBER ESPIONAGE OPERATIONS?,» Milpitas, California, United States, FireEye, Inc, October 2014.

[51] Villeneuve, N., «Trends in Targeted Attacks,» Cupertino, CALIFORNIA, TREND MICRO INC., October 2011.

[52] Smokescreen, «THE TOP 20 LATERAL MOVEMENT TACTICS,» Smokescreen Technologies Pvt. Ltd., August 2016.

[53] Trend Micro, «APT C&C Communication. Superior Detection with Trend Micro Custom Defense,» Cupertino, CALIFORNIA, TREND MICRO INC., 2013.

[54] Allied Telesis, «VLANs (Virtual LANs),» Bothell. Washington. United States, Allied Telesis, Inc., 2008.

[55] Miller, R., «Amenazas Persistentes Avanzadas. La defensa desde adentro hacia

afuera,» Nueva York. Estados Unidos, CA Technologies, July 2012.

[56] Wikipedia, 3 January 2017. [En línea]. Available: https://es.wikipedia.org/wiki/Lista_negra. [Último acceso: 5 May 2017].

[57] Kührer, M. et al, «Paint It Black: Evaluating the Effectiveness of Malware Blacklists,» Grant 01BY1110, MoBE, Horst Görtz Institute for IT-Security. Ruhr-University Bochum. Germany, 2014.

[58] Oliveira, J. & Jiménez R., «EL PAIS. Tecnologia,» 12 May 2017. [En línea]. Available: http://tecnologia.elpais.com/tecnologia/2017/05/12/actualidad/1494586960_025438.html. [Último acceso: 12 May 2017].

[59] «EL PAIS. Tecnología,» 13 May 2017. [En línea]. Available: http://tecnologia.elpais.com/tecnologia/2017/05/13/actualidad/1494668788_755982.html. [Último acceso: 13 May 2017].

[60] Centro Criptológico Nacional, «CCN-CERT. Alerts and vulnerabilities,» 13 May 2017. [En línea]. Available: https://www.ccn-cert.cni.es/en/updated-security/ccn-cert-statements/4464-ataque-masivo-de-ransomware-que-afecta-a-un-elevado-numero-de-organizaciones-espanolas.html. [Último acceso: 13 May 2017].

[61] The Guardian, «Cyber-attacks highlight growing vulnerability of us all,» theguardian.com, 14 May 2017. [En línea]. Available: https://www.theguardian.com/technology/2017/may/14/cyber-attacks-highlight-growing-vulnerability-of-us-all. [Último acceso: 20 May 2017].