El papel de la IA en la monitorización de la seguridad en red

Univ. Complutense de Madrid
Mayo 2023

Jesús Esteban Díaz-Verdejo
Departamento de Teoría de la Señal, Telemática y Comunicaciones
E.T.S. Ingenierías Informática y Telecomunicación – Universidad de Granada
C/ Periodista Daniel Saucedo Aranda, s/n - 18071 – Granada (Spain)
Phone: +34-958 242304 - Email: jedv@ugr.es

https://dtstc.ugr.es/neus-cslab

¿Quiénes somos?

NEO TIC154
CYBERSECU

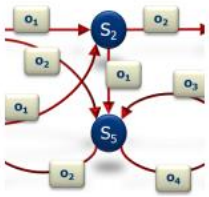
https://dtstc.ugr.es/neus-cslab

APRENDIZAJE AUTOMÁTICO

ANÁLISIS DE GRANDES VOLÚMENES DE DATOS

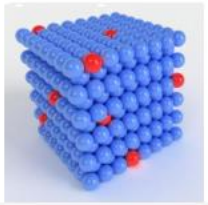
SIMULACIÓN DE REDES

ANÁLISIS DE SERIES TEMPORALES



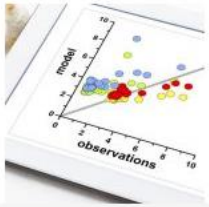
MODELADO DE MARKOV

ANÁLISIS DE RIESGOS



DETECCIÓN DE ANOMALÍAS

ANÁLISIS Y MODELADO DE PROTOCOLOS DE COMUNICACIONES



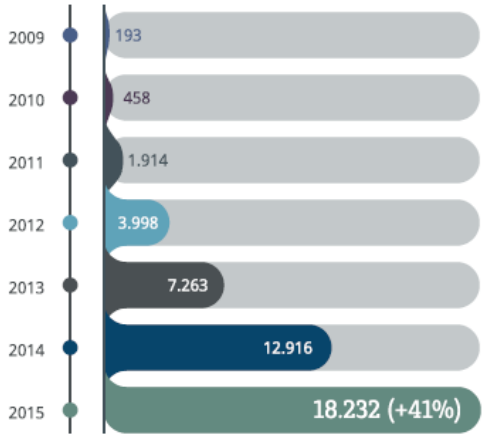
CORRELACIÓN DE EVENTOS

BLOCKCHAIN

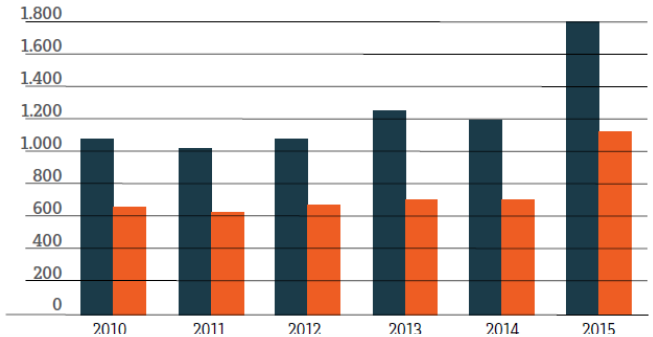
UNIVERSIDAD DE GRANADA

Introducción

- Importancia incuestionable de la seguridad en Internet
  - Aumento del número de incidentes de seguridad
    - Estadísticas de incidentes (Fuente: CCN-CERT):

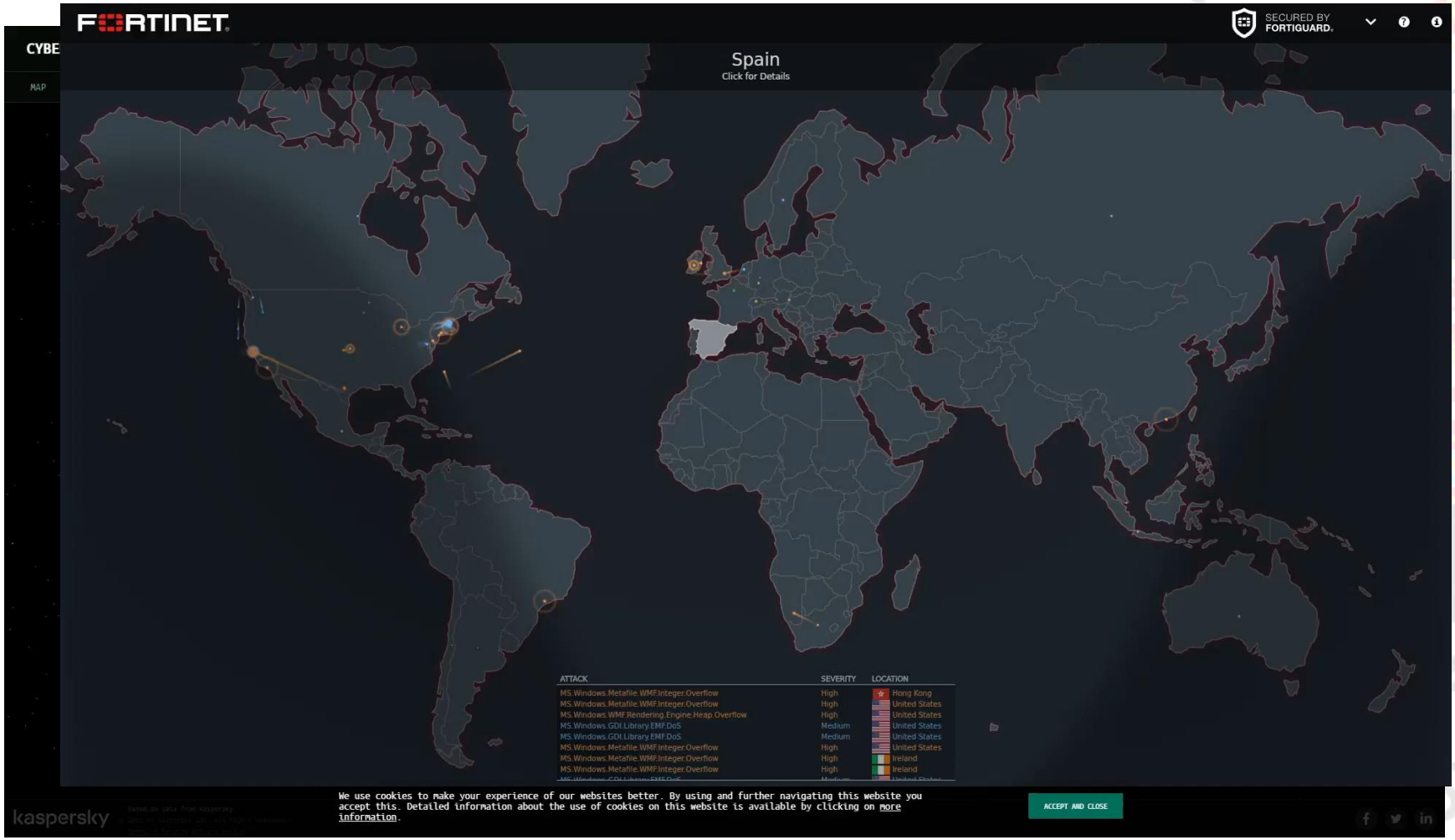


Evolución de los Incidentes gestionados por el CCN-CERT



Introducción

Monitorización de la seguridad en redes



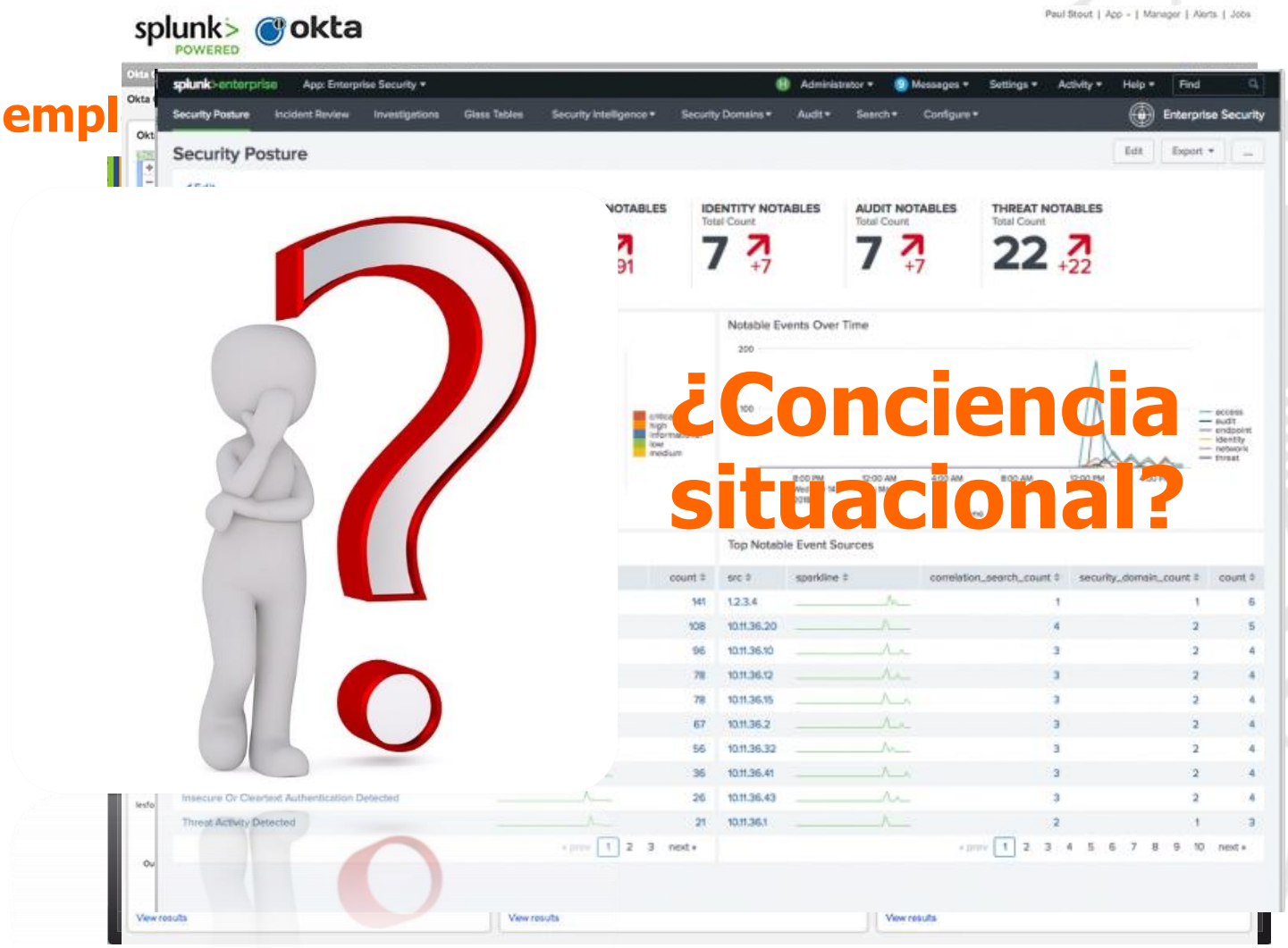
Monitorización de la seguridad en redes

- **Sistemas de monitorización de la seguridad (NSM, SIEM)**
  - **Recolección**, **detección** y **análisis** de los datos de seguridad de la red
  - Objetivo: **Escalado**, declarar la ocurrencia de un incidente para que se pueda activar una **respuesta**

- Attack Sense and Warning (AS&W): detección, correlación, identificación y caracterización de actividad no autorizada e intencional seguida de notificación para la generación de respuesta
  - Doctrina de operaciones de información (IO - DoD) *http://www.au.af.mil/info-ops/doctrine.htm*

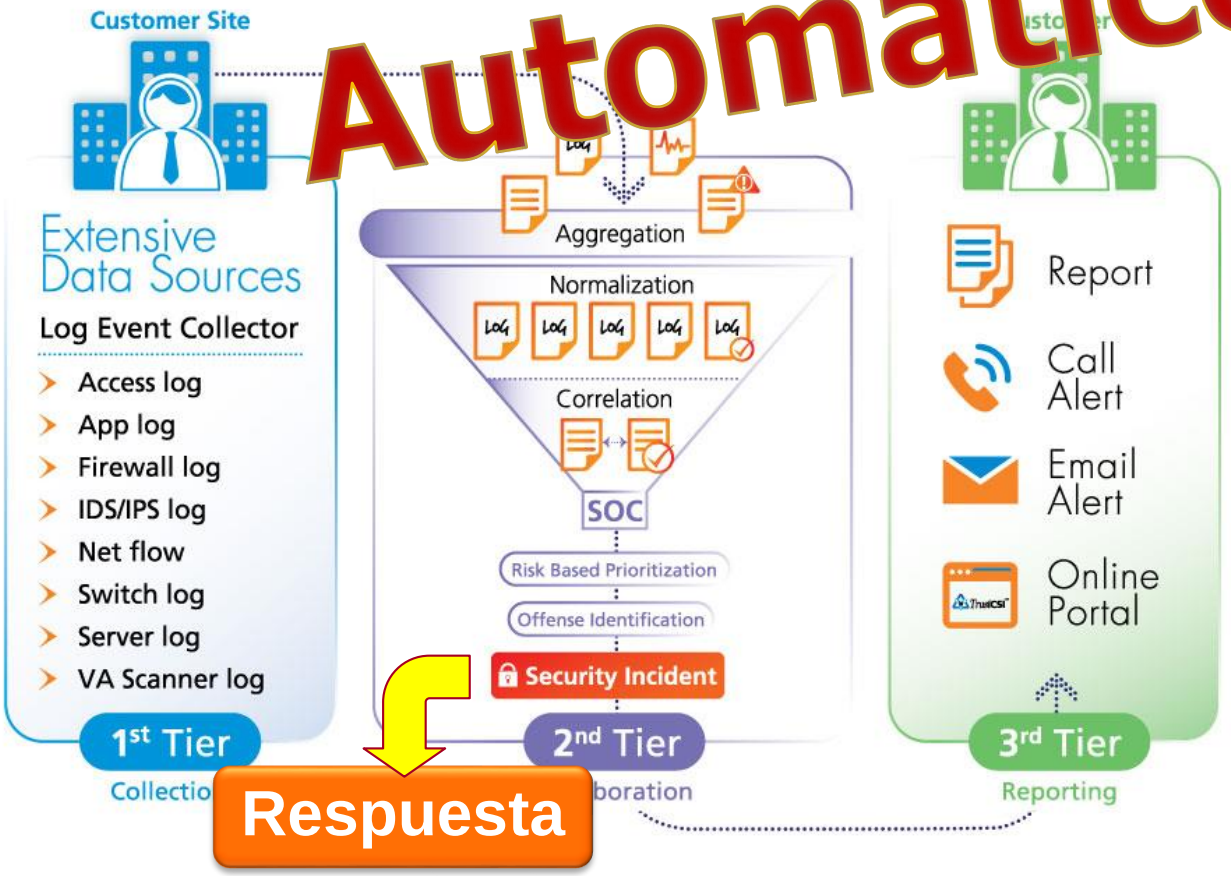
Monitorización de la seguridad en redes

- Ejempl



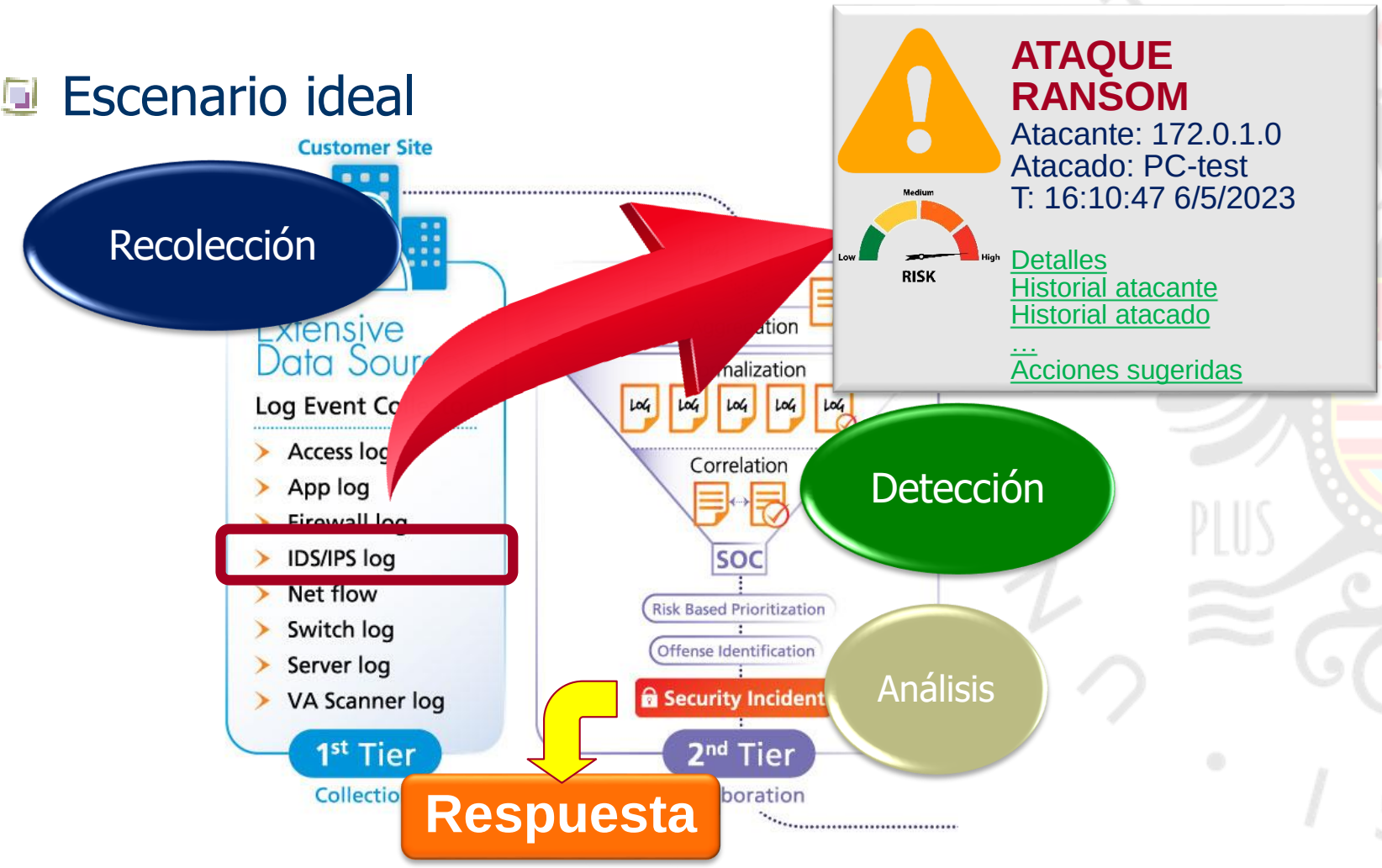
Monitorización de la seguridad en redes

Escenario ideal



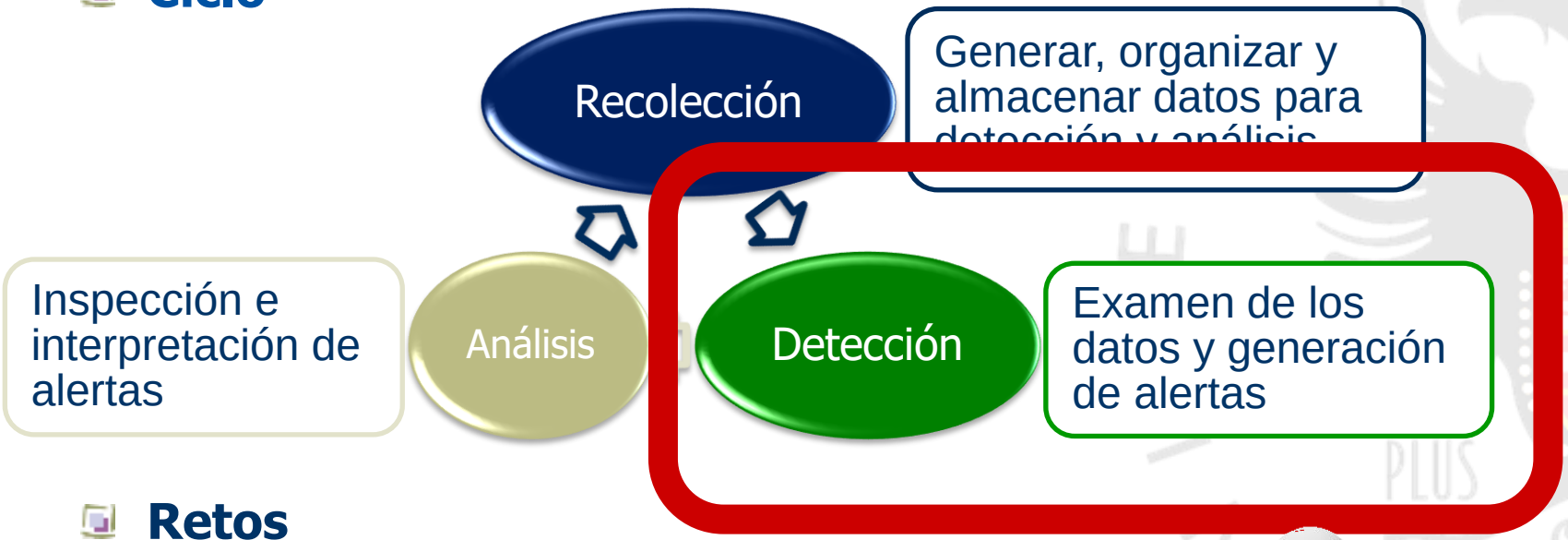
Monitorización de la seguridad en redes

Escenario ideal



Monitorización de la seguridad en redes

Ciclo



Retos

- Tecnología inmadura
  - Terminología / estandarización
- Formación (especialización)
  - Necesidad de profesionales formados (a alto nivel)
- Elevado coste de despliegue y operación
  - Equipamiento / Personal

Sistemas de detección de intrusos

- Metodología básica
  - **Monitorizar eventos y analizarlos mediante los métodos apropiados**
- Elemento (fundamental) de NSM
  - Objetivo: **generar alertas de intrusión**
  - IPS: (y responder)

- Características:
  - Defensa centrada en vulnerabilidades
  - Mayoritariamente basado en firmas

radware
http://www.ndm.net/ips/solutions/radware

CISCO

Check Point SOFTWARE TECHNOLOGIES LTD.
http://www.ndm.net/ips/solutions/check-point

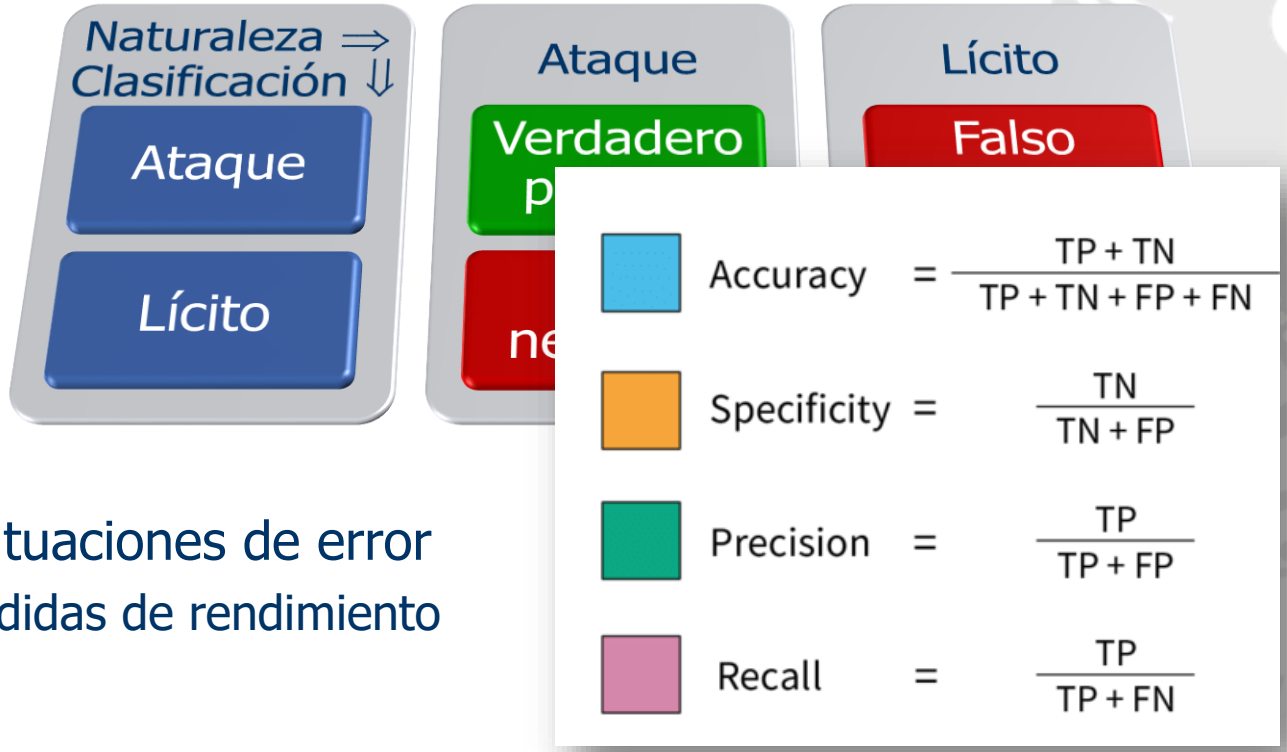
SOURCEfire
http://www.ndm.net/ips/solutions/sourcefire

SNORT™

Medidas de rendimiento

Rendimiento / fiabilidad IDS

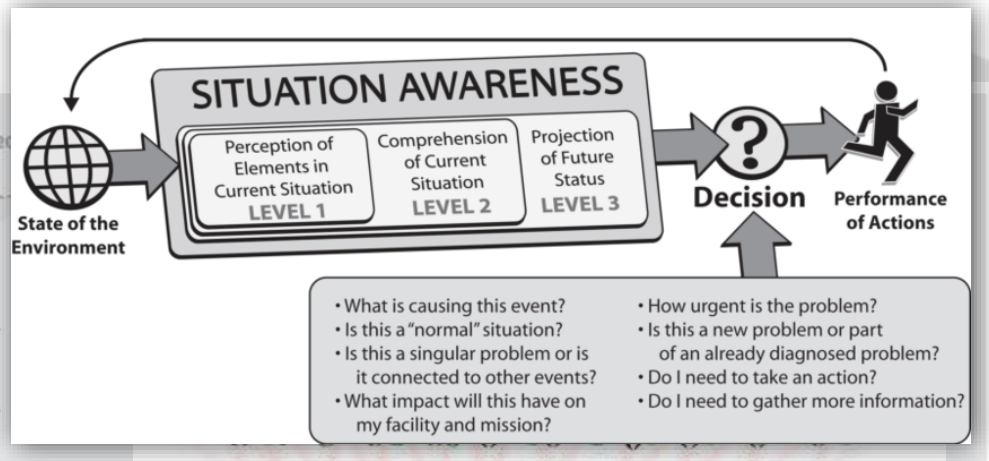
- Clasificaciones posibles:



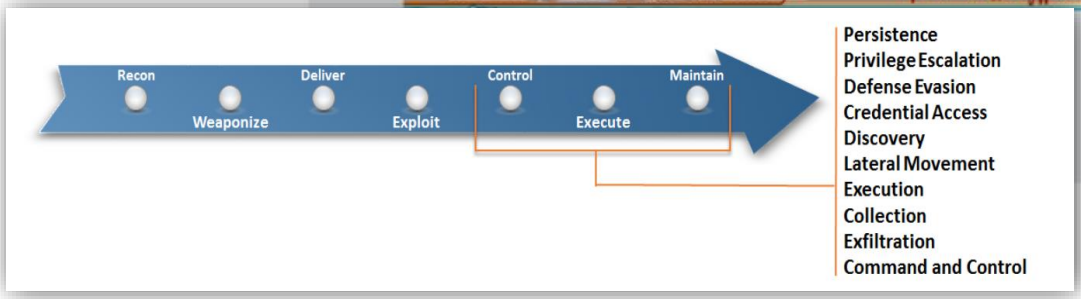
- Dos situaciones de error
  - Medidas de rendimiento

Limitaciones

- Insuficiente:
  - Fal
  - (Fir
  - (Fir
  - Ata
  - AP
  - Op
  - Ata



ETPRO CURRENT_EVENTS Possible Successful Generic Phish (set) Aug 19 2016



Soluciones (posibles)

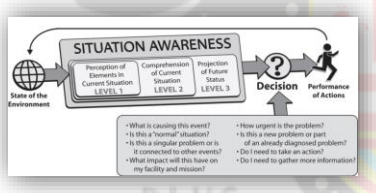
- Técnicas IDS no basadas en firmas (anomalías?)
  - > 300k publicaciones
    - Excelentes resultados (al menos eso afirman)

IA!

Google scholar

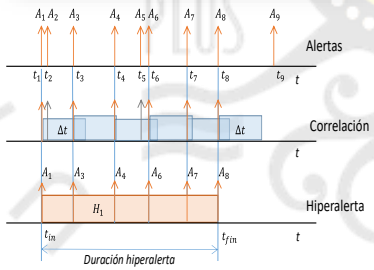
- Contexto (enriquecimiento)
  - Inclusión de información relevante (inferida) en las alertas

IA!



- Modelado de ataques y correlación de eventos

IA!



- **¿Por qué no se despliegan?**

Monitorización de la seguridad

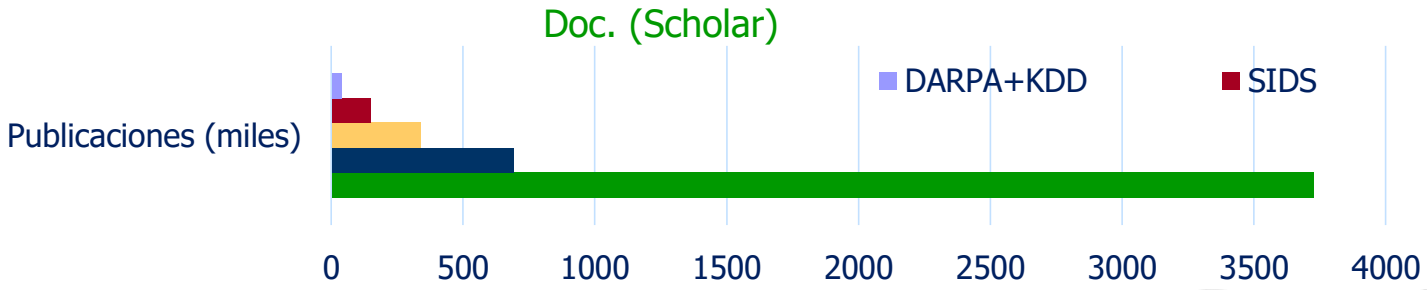
Problema de clasificación de eventos complejos

Desafíos y futuro

- Que **funcione en condiciones reales**
  - Capacidad de generalización

Desarrollo/investigación en IDS

- Problema no resuelto
  - Primeros trabajos: Denning 86
- Gran volumen de actividad investigadora
  - Especialmente en detección de anomalías / híbrida
  - IA (Deep Learning)



Elementos de diseño

- "Ingredientes"

El IDS es específico para cada sistema

¿Qué eventos monitorizar?

Preprocesado

¿Qué técnica/s permite/n detectar qué tipos de ataques?

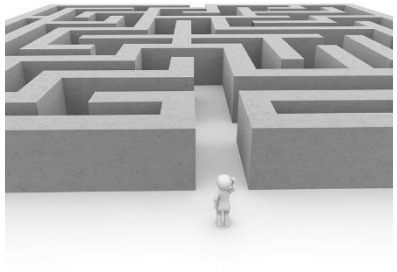
Técnica de detección

¡Datos! (Etiquetados)

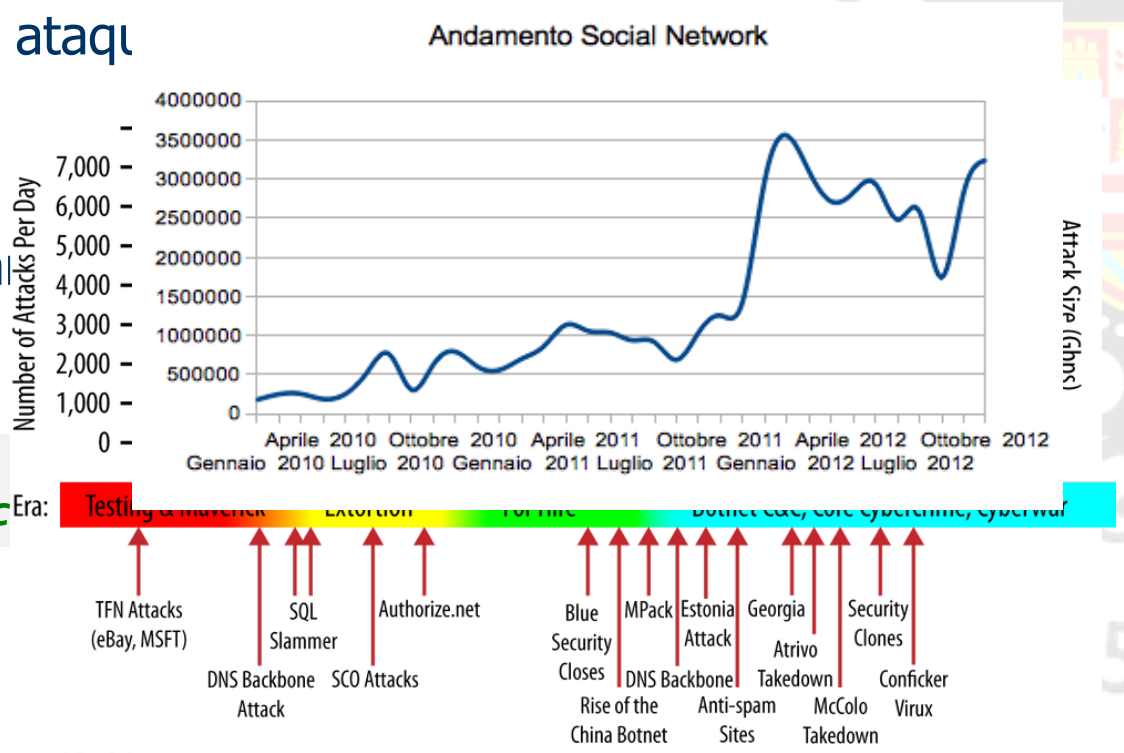
Medidas de rendimiento

Elementos de diseño

- Otros ingredientes
  - Evolución de los sistemas
    - Problema de "data shift"
  - Evolución de los ataq
  - Entorno adversa
    - Evasión

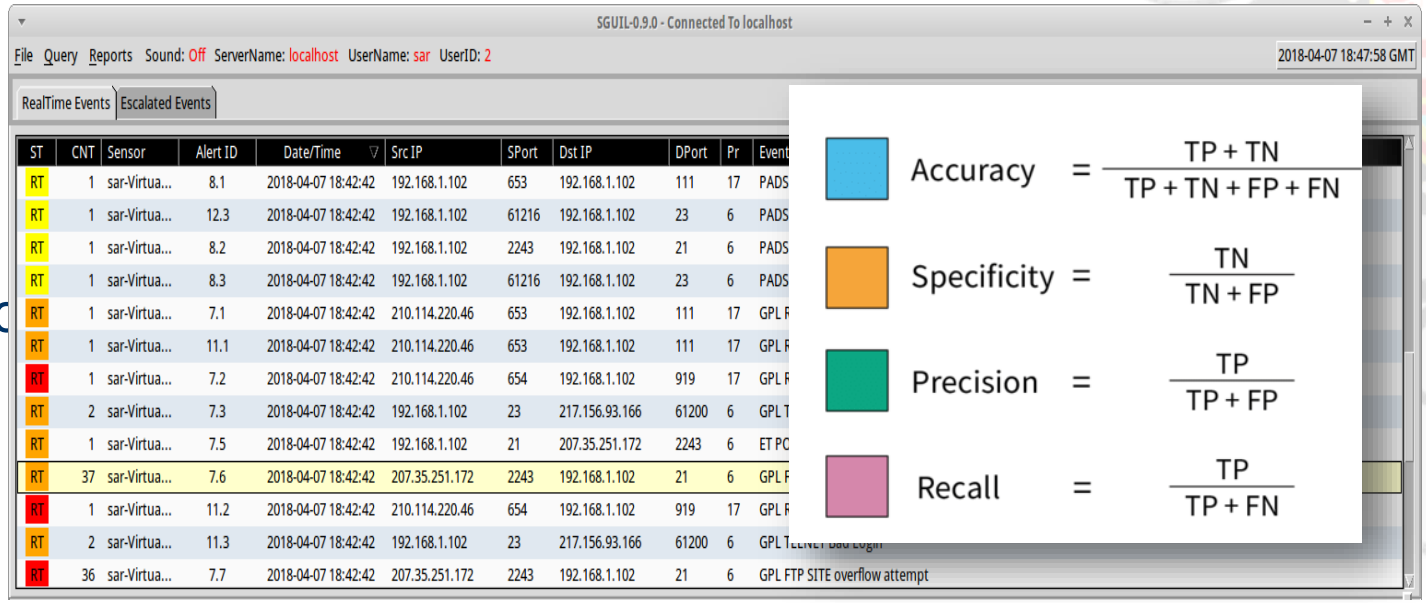


```
GET/
/c+ c
```



Medidas de rendimiento

- Comportamiento deseable:
  - Pocos falsos positivos (idealmente, 0
    - Alto número de alarmas erróneas
      - El IDS se vuelve inútil
  - Ningún falso negativo:
    - Es
  - Pro



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

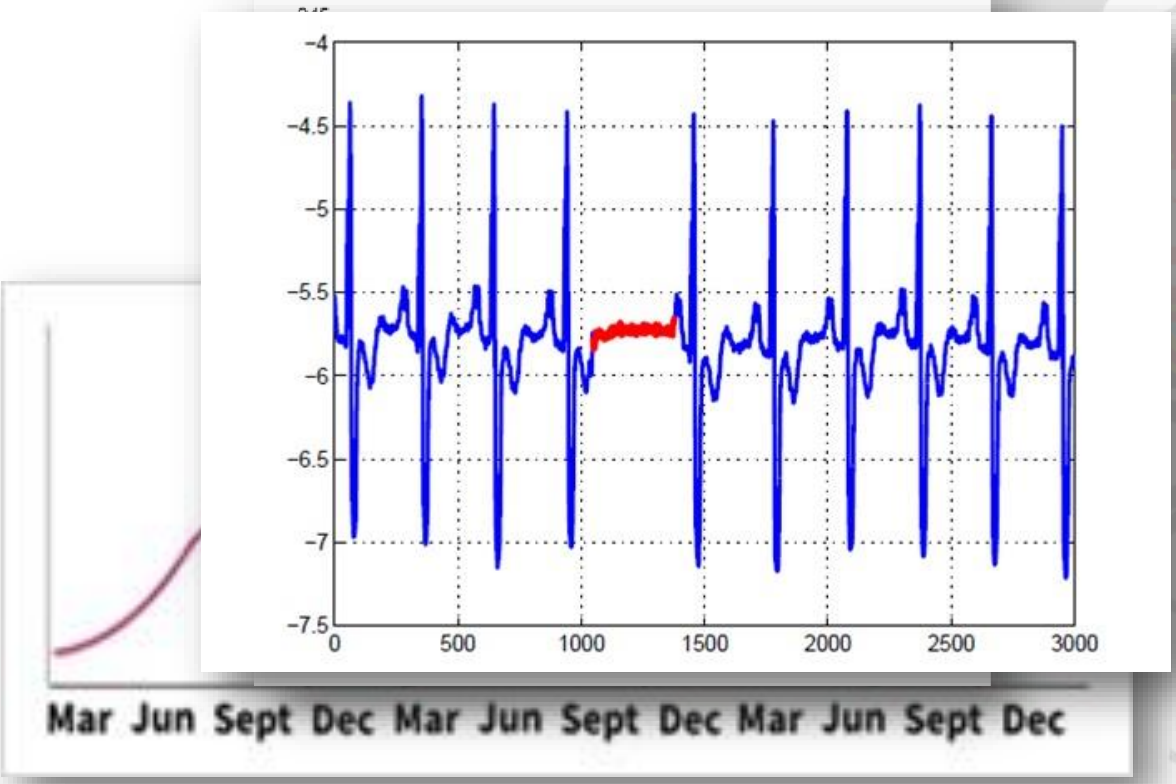
$$\text{Recall} = \frac{TP}{TP + FN}$$

Técnicas de detección

VARUN CHANDOLA, ARINDAM BANERJEE, and VIPIN KUMAR
Anomaly Detection: A Survey,
ACM Computing Surveys, Vol. 41, No. 3, Article 15, 2009.

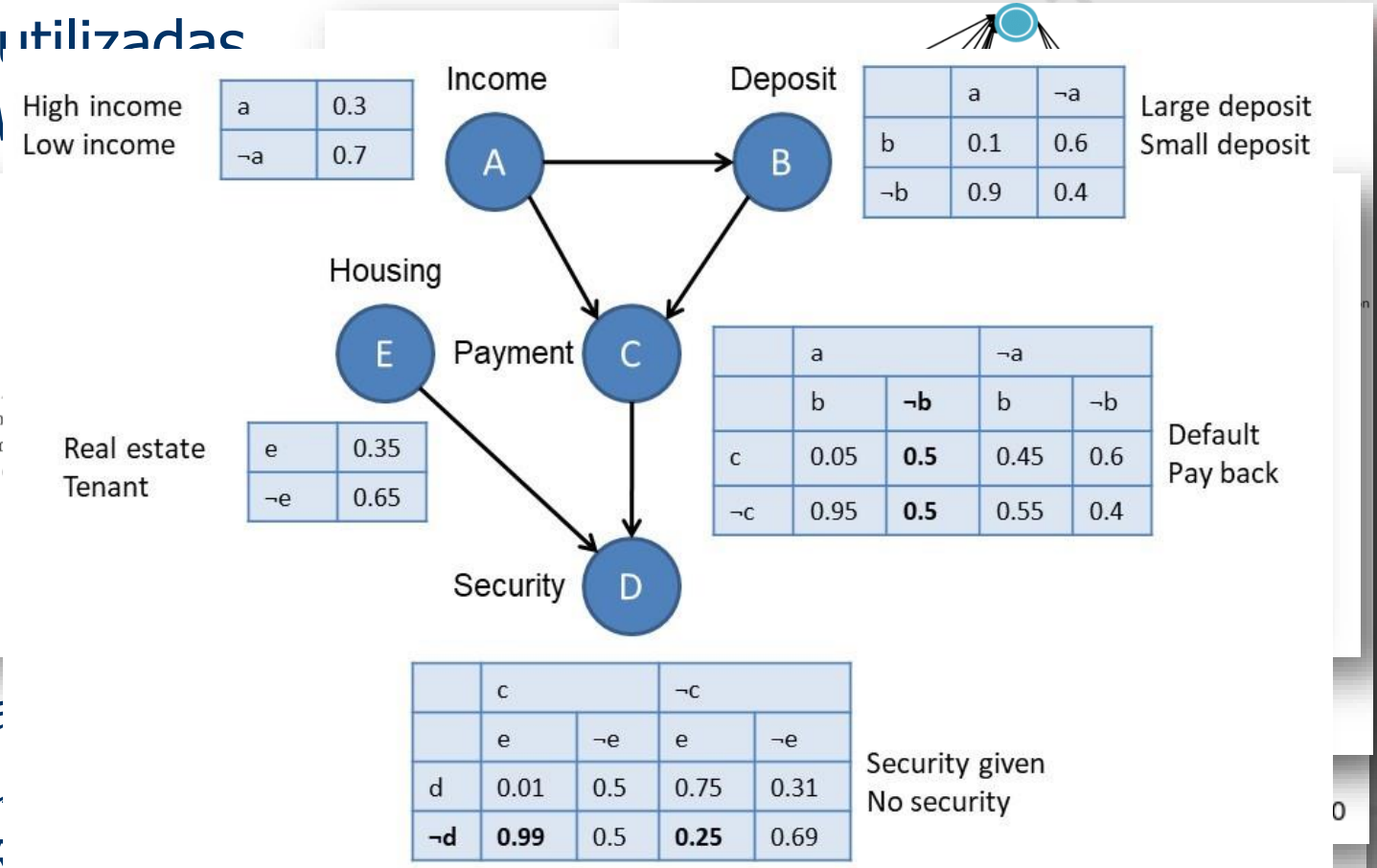
Tipos de anomalías

- Puntuales
- Colectivas
- Contextuales



Técnicas de detección

- Más utilizadas
  - *kN*
  - *G*
  - *A*
  - *A*
  - *S*
  - *A*
  - *A*
  - *A*
  - *Ga*
  - *Au*
  - *AR…*
  - *Redes bayesianas*



a	0.3
¬a	0.7

	a	¬a
b	0.1	0.6
¬b	0.9	0.4

	e	0.35
¬e	0.65	

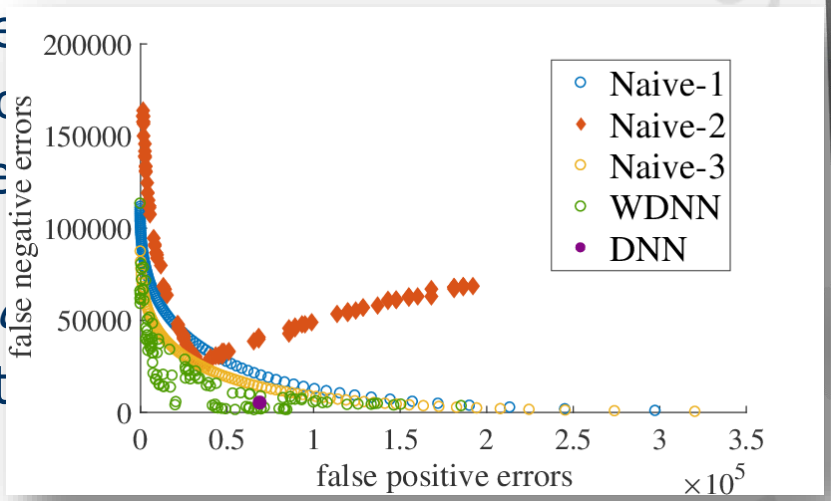
	a		¬a	
	b	¬b	b	¬b
c	0.05	**0.5**	0.45	0.6
¬c	0.95	**0.5**	0.55	0.4

	c		¬c	
	e	¬e	e	¬e
d	0.01	0.5	0.75	0.31
¬d	**0.99**	0.5	**0.25**	0.69

Deficiencias típicas

Estado del arte

- Análisis deficie
  - No es un pro
- Aplicación "cie
campos)
  - P.e. *machin*
- Es fácil subest



- Múltiples contribuciones se limitan a mostrar que su propuesta mejora los resultados de otra sobre el mismo conjunto de datos
  - Falta de generalidad
  - Sobreajustes (sobre-entrenamiento)

Deficiencias típicas

- **Big data / deep lear**
  - Usabilidad / resultados
    - Gartner

Ausencia de datos ade

Falta de conocimiento

Resolución del proble

Falta de valor adiciona

Falta de ética

85 87

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Opinion piece

Check for updates

Cite this article: Succi S, Coveney PV. 2019 Big data: the end of the scientific method? *Phil. Trans. R. Soc. A* **377**: 20180145. http://dx.doi.org/10.1098/rsta.2018.0145

Accepted: 25 July 2018

One contribution of 11 to a theme issue 'Multiscale modelling, simulation and computing: from the desktop to the exascale'.

Subject Areas:
computer modelling and simulation, computational physics, artificial intelligence

Keywords:
Big data, multiscale modelling, simulation, artificial intelligence

Author for correspondence:
Peter V. Coveney
e-mail: p.v.coveney@ucl.ac.uk

Big data: the end of the scientific method?

Sauro Succi[1,2] and Peter V. Coveney[3,4]

[1]Center for Life Nano Sciences at La Sapienza, Istituto Italiano di Tecnologia, viale R. Margherita, 265, 00161, Roma, Italy
[2]Institute for Applied Computational Science, J. Paulson School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, USA
[3]Centre for Computational Science, Department of Chemistry, University College London, London, UK
[4]Yale University, New Haven, USA

PVC, 0000-0002-8787-7256

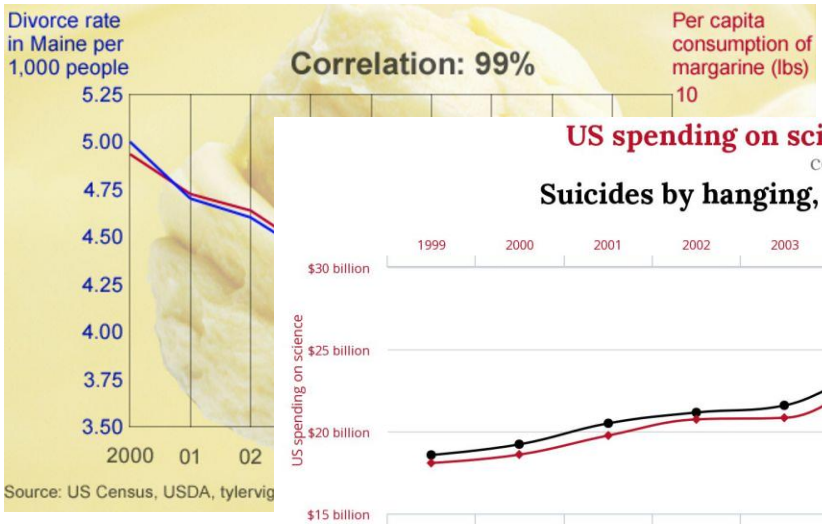
For it is not the abundance of knowledge, but the interior feeling and taste of things, which is accustomed to satisfy the desire of the soul. (Saint Ignatius of Loyola).

We argue that the boldest claims of big data (BD) are in need of revision and toning-down, in view of a few basic lessons learned from the science of complex systems. We point out that, once the most extravagant claims of BD are properly discarded, a synergistic merging of BD with big theory offers considerable potential to spawn a new scientific paradigm capable of overcoming some of the major barriers confronted by the modern scientific method originating with Galileo. These obstacles are due to the presence of nonlinearity, non-locality and hyperdimensions which one encounters frequently in multi-scale modelling of complex systems.

Deficiencias típicas

- **Big Data**
  - Correlaci
    - Correlaciones espúreas

José Camacho, Gabriel Maciá-Fernández, Jesús Díaz-Verdejo, Pedro García-Teodoro
Tackling the Big Data 4 Vs for Anomaly Detection
Proc. 2014 IEEE INFOCOM, pp. 506-511.

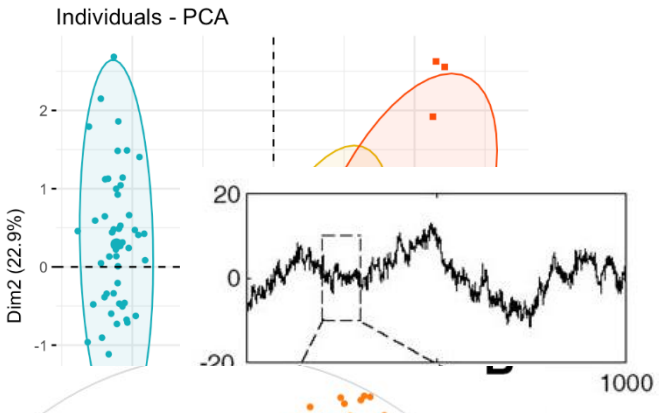


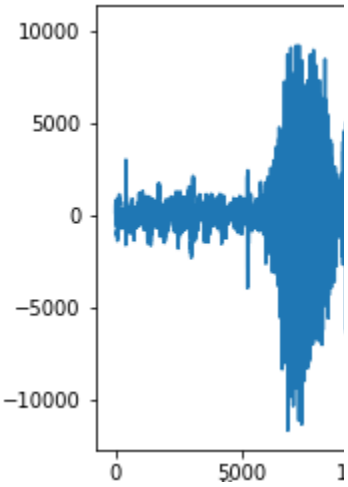
- No todos los datos son útiles

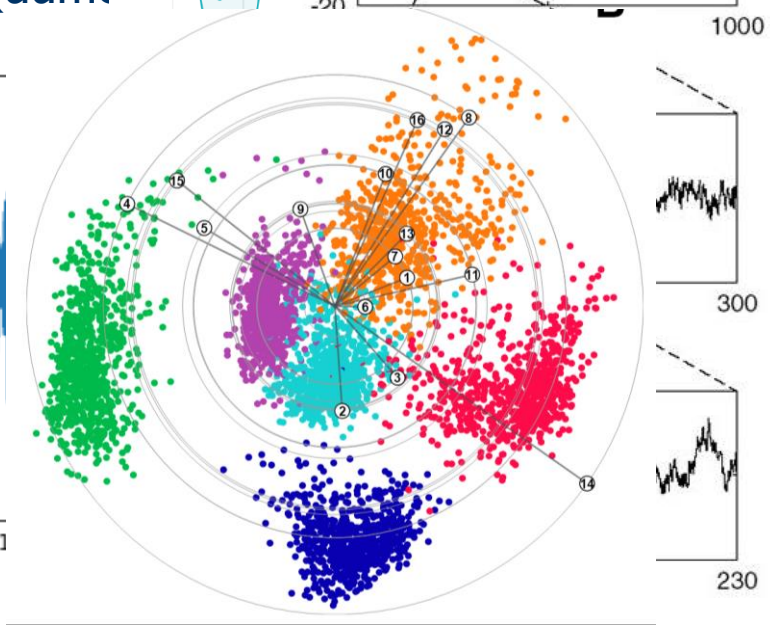
Deficiencias típicas

Preprocesado

- Datos en bruto
  - Aproximación típica Big Data / minería datos
  - Más datos ≠ mejores resultados (aume SNR)
- Parametrizacione
  - Más parámetros
  - Información disc asociada a mayo
- Granularidad (te
  - Hay evolución te mágnitudes
  - Tráfico autosimilar

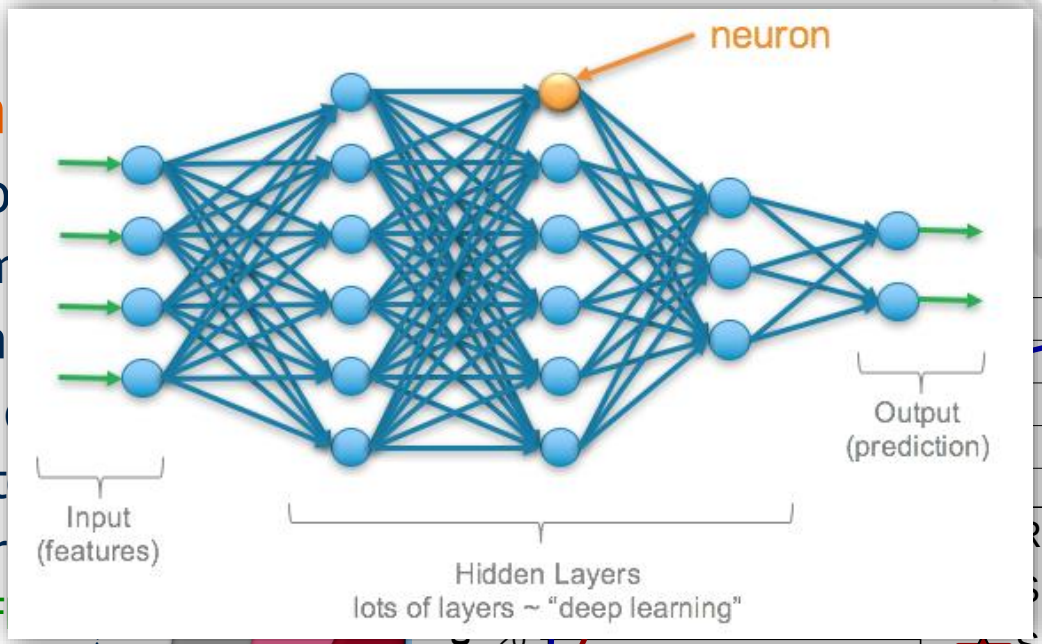


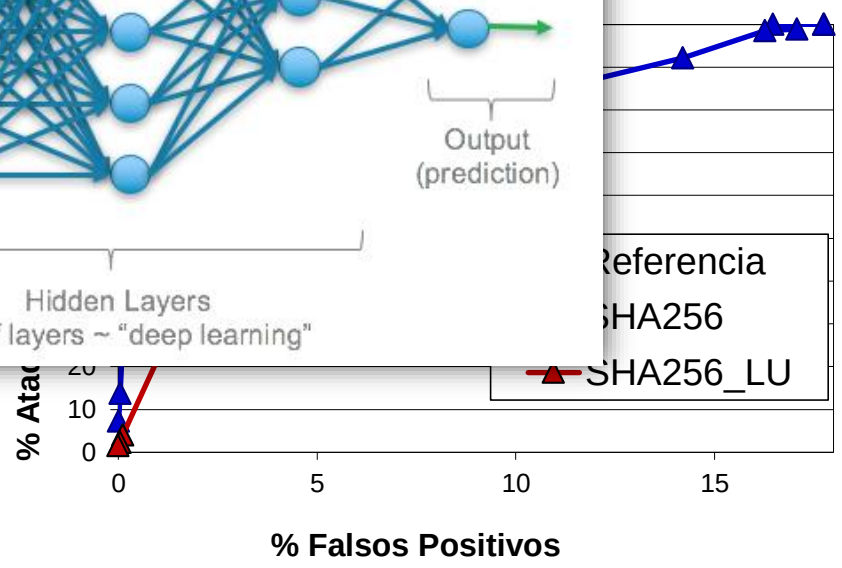




Deficiencias típicas

- **Entrena**
  - Particio
    - Mism
  - Medida
    - Uso
    - Punt
    - Sobr
      - F
  - Ajuste d
    - Núme
  - Capacid



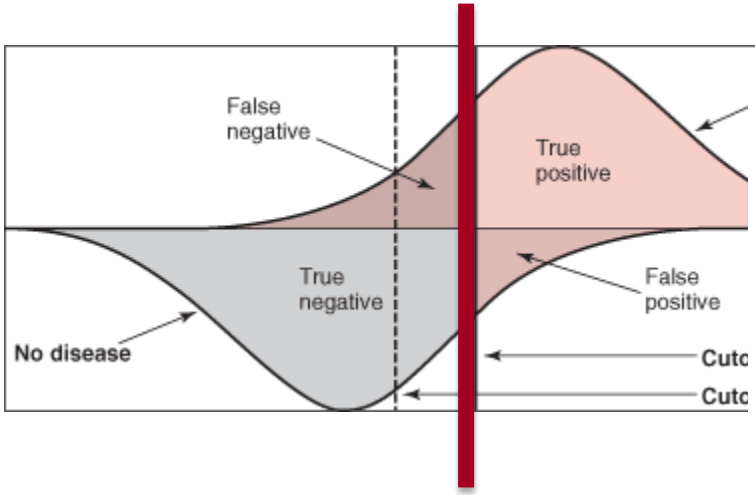


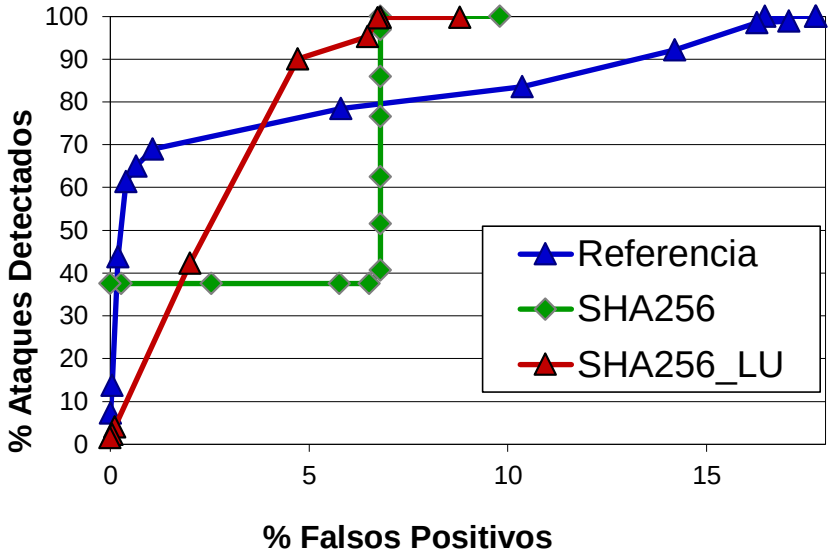
The Base-Rat… Detection, S. …

M. Bermúdez-Edo, R. Salazar-Hernández, J. Díaz-Verdejo, and P. García-Teodoro
Proposals on Assessment Environments for Anomaly-Based Network Intrusion Detection Systems
Lect. Notes in Computer Science, 4347 / 210 – 221 / Springer-Verlag / 2006.

Punto de operación

- Selección inaceduada del punto de operación
  - No basada en curvas ROC





Deficiencias típicas

Black box

Jesús Díaz-Verdejo

Apr 25, 17:32

-**2**: (reject)

El presente tra
de tráfico a pa
parámetros de
presentada re
más allá de in

La propuesta,
existen numer
propuesta. Est
aplicación.

Se omiten det
posible detecc
"ciegas" al pro
(netflow en es
expectativa de
ámbito de la s
dependiente d
generalización

Por otra parte
no proporcion
exploratorio d
pretende dete

Por otra parte
detectar en es

Por último, au
el uso de herramientas de generación de tráfico netflow para tráfico "normal" y de ataques sin la consideración de un modelo o caracterización estadística del tráfico real.

Detección de ataques sobre tráf
mediante análisis de imágenes en r
convolucionales

Resumen—Actualmente, el interés por la detección de ataques en la red está creciendo, pero debido a las cuantías tan numerosas de paquetes que se mueven en el mundo por segundo es inasequible el análisis del contenido de cada uno de ellos para el rastreo de diversos tipos de ataques. En este artículo de investigación en desarrollo, se presenta una solución para la detección de este tipo de eventos en la red, mediante únicamente, flujos de Netflow, protocolo estándar de los dispositivos de red. La investigaciones en esta línea en su mayoría se han basado en algunas técnicas de machine/deep learning pero pocas de ellas han hecho uso de redes convolucionales para ello. Por lo tanto, la detección que presentamos en este trabajo se lleva a cabo haciendo uso de una red residual la cual es nutrida por imágenes (2D) creadas a partir de flujos Netflow (1D).

Index Terms—Netflow, detección de ataques, red convolucional.

Tipo de contribución: *Investigación en desarrollo*

I. Introducción

La tecnología NetFlow desarrollada por Cisco nació para poder recolectar información sobre tráfico IP de una forma simple, pudiendo hacer así el seguimiento de los flujos. Este protocolo de red, que nos permite obtener distintas características de los paquetes que pasan por los dispositivos de red, se ha convertido en un estándar en la industria. Este protocolo dispone de varias versiones pero las versiones estándares más utilizadas son la versión 5 [1] y la 9 [2], que cuentan con características como la IP de origen, la IP destino o los puertos tanto origen como destino, entre otros.

Hoy en día, debido a que es imposible analizar de forma exhaustiva cada uno de los paquetes que pasan por las redes, el interés por la detección de ataques en base a los flujos de NetFlow está creciendo de una forma estrepitosa. Esto complica la detección de posibles ataques dado que no se hace uso del contenido del paquete en el análisis sino simplemente de su flujo y de las escasa información con la que este cuenta. Para ello, se está haciendo uso de técnicas de inteligencia artificial como por ejemplo KNN o SVM [3]-[7]. Muchas de estas técnicas pueden trabajar de forma conjunta

en una herramienta m
MoEv [8]. A1 unísono
también algunos artíc
convolucionales para l
cuales han ayudado a
con esa línea de inve
describirá el proceso
detección de diversos
CNN. Las redes conv
una diversidad muy g
caso, se hará uso de
problema de la detec
flujos.

El presente docum
en la que se explica
desde la creación de l
los resultados obtenid
así como una segund
describen posibles lín

II

En esta sección se
proyecto para la dete
Como se ha ido des
primeramente se gener
de los flujos de Netflo
creadas son introducid
mos de entrenar. Por
sus respectivas subsec
y los resultados obten

II-A. Fuente de dato

En este apartado
datos utilizadas en la
soluciones en caso d
datos se están utilizan
sistema pero la expec

MDPI

Article

Intrusion detection based on gray-level co-occurrence matrix and 2D Dispersion Entropy

1
Version April 20, 2021 submitted to Appl. Sci.

Abstract: The Intrusion Detection System (IDS) is an important tool to mitigate cybersecurity threats in an Information and Communication Technology (ICT) infrastructure. The function of the IDS is to detect an intrusion to an ICT system or network so that adequate countermeasures can be adopted. Desirable features of IDS are computing efficiency and high intrusion detection accuracy. This paper proposes a new anomaly detection algorithm for IDS, where a machine learning algorithm is applied to detect deviations from legitimate traffic, which may indicate an intrusion. To improve computing efficiency, a sliding window approach is applied where the analysis is applied on large sequences of network flows statistics. This paper proposes a novel approach based on the transformation of the network flows statistics to gray images on which Gray level Co-occurrence Matrix (GLCM) are applied together with an entropy measure recently proposed in literature: the 2D Dispersion Entropy. This approach is applied to the recently public IDS data set CIC-IDS2017. The results show that the proposed approach is competitive in comparison to other approaches proposed in literature on the same data set. The approach is applied to two attacks of the CIC-IDS2017 data set: DDoS and Port Scan achieving respectively an Error Rate of 0.0016 and 0.0048.

Keywords: Intrusion detection systems; security; machine learning; communication

1. Introduction

Our society is becoming increasingly dependent on the internet and communication services but the risk of cybersecurity threats has also increased. Intrusion Detection System (IDS) can be a powerful tool to mitigate cybersecurity attacks. Research in IDS is more than 20 years old and various types of IDS have been proposed in literature: signature-based IDS, which focuses on the recognition of traffic patterns associated to a threat, anomaly-based IDS which detects deviations from a model of legitimate traffic and often relies on machine learning or reputation-based IDS based on the calculation of reputation scores [1]. Requirements or preferred features of IDS have been already defined in literature [1],[2] and they can be summarized in: a) fast detection of the attack, b) high detection accuracy and c) low computing complexity of the detection algorithm to support the capability to analyze a large amount of traffic due to the high throughput of the current networks. The successful fulfillment of these three main requirements can be challenging because there are trade-offs between them. For example, algorithms, which are able to obtain high detection accuracy, may require considerable computing resources or they may not be able to achieve a fast detection. The advantage of anomaly-based IDS, in comparison to signature-based IDS, is to potential detect new attacks which have not been recorded before and where the corresponding signature has not been created yet. On the other side, the detection of anomalies in high throughput traffic would benefit from dimensionality reduction while preserving an high detection accuracy. To achieve this goal, anomaly-based IDS have

iva a la detección de anomalías en los flujos
ración de "imágenes" a partir de los
iales convolucionales. La descripción
os preliminares, no se aportan resultados

carece de elementos novedosos por cuanto
lerando incluso el uso de CNN como en la
limitadas y algunas fuera del ámbito de

e, sobre su justificación en relación a la
del elevado conjunto de aproximaciones
se basa en considerar una parametrización
a técnica de clasificación (CNN) con la
a aproximación carece de utilidad en el
ultados y, adicionalmente, es altamente
s. Es difícil evaluar la capacidad de

nente relevante. El uso de más parámetros
o"), por lo que se echa en falta un análisis
en relación a su justificación y a lo que se

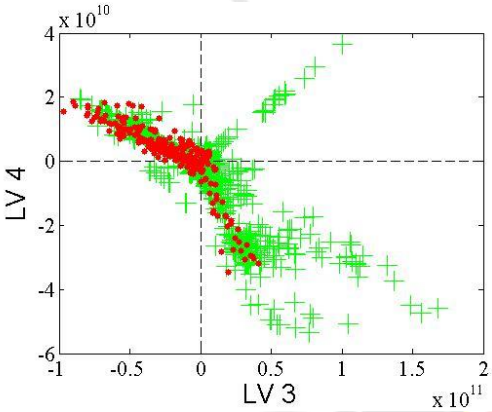
imágenes. ¿Qué es lo que se pretende
le características?

presentativos del tráfico real, lo que invalida

Deficiencias típicas

Datasets

- Falta de representatividad / escala
  - Datos propios
  - Simulaciones
- Obsolescencia
  - Evolución de los ataques en últimos años
  - Evolución de los sistemas en los últimos años
- Ocultismo
  - No se prop
    datasets
  - Presencia
  - Artefactos
  - **Ground-**
- Tamaño
  - **Entrenamiento insuficiente**



NASA95: /htbin/wais.pl?wnet

BIBLIO:
/imce?app=ckeditor%7Csendto%40ckeditor_imceSend
To%7C&CKEditor=edit-body-und-0-

Rafael Estepa Alonso, Jesús Díaz-Verdejo, Antonio Estepa Alonso, Germán Madinabeitia
How much training data is enough?. A case study for HTTP anomaly-based intrusion detection
IEEE Access, 8:44410-44425, 2020.

DARPA / KDD99

- **DARPA'98 (y 99)**
  - Datos de Lincoln Labs (1998 y 1999)
  - Trazas de actividad, incluyendo ataques, de una hipotética base aérea
  - Una de las pocas bases de datos etiquetadas disponibles
- Problemas:
  - Sintética
    - Múltiples artefactos, p.e. TTL
  - Vieja
  - Sobreutilizada
- Se sigue utilizando todavía

- **KDD'99**
  - Versión parametrizada de DARPA'98
  - Propuesta para una competición de ML (KDD)
- Hereda los problemas de DARPA'98
- Parametrización inadecuada ("ciega")
- Aún más utilizada que DARPA'98

Testing Intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory, John McHugh, ACM Transactions on Information and System Security 3(4), 2000

Desafíos y futuro

- Que **funcione en condiciones reales**
  - Capacidad de generalización
  - Capacidad de adaptación
- **Retos técnicos** más relevantes
  - Elevado número de datos
  - Alta dimensionalidad de los datos
  - Naturaleza temporal
    - Los datos próximos en el tiempo suelen estar correlacionados
  - Distribución descompensada
    - La aguja en el pajar
  - Preprocesado de los datos

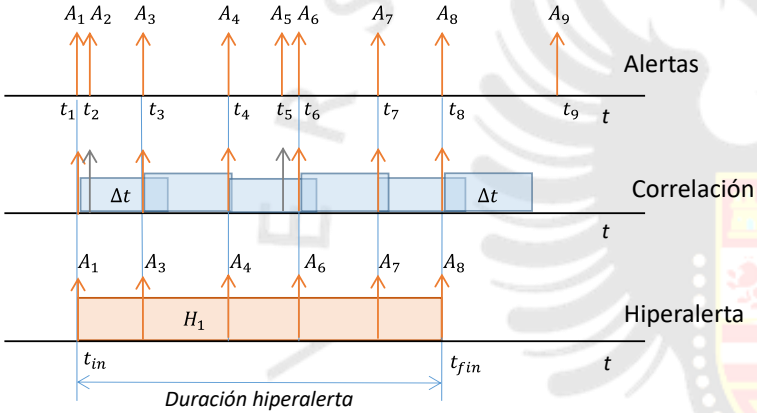
Desafíos y futuro

Problemas abiertos

- Baja tasa de eficiencia (problema de altas tasas de falsos positivos)
  - Se necesitan aproximaciones más estructuradas
  - Técnicas más eficientes
- Bajo "throughput" y alto coste
- Inexistencia de metodologías de evaluación válidas
- Bajo número de mecanismos de respuesta
  - Se requieren mecanismos más eficientes y más robustos
- Análisis de datos cifrados

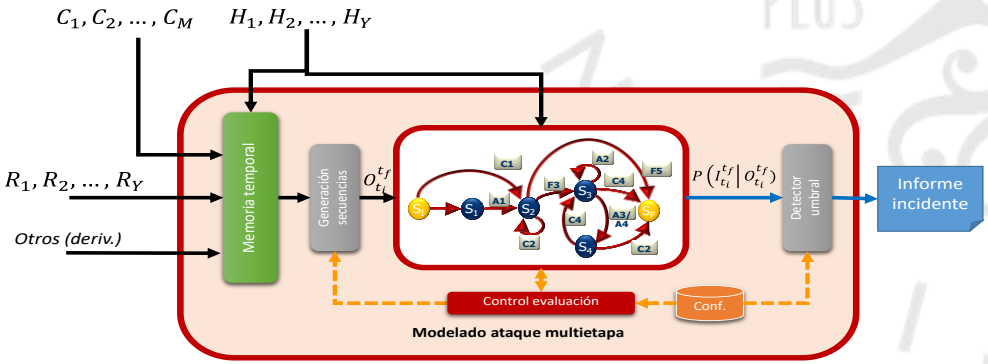
Desafíos y futuro

- Evolución hacia sistemas híbridos distribuidos
- **Posprocesado**
  - Técnicas de correlación de eventos



- Modelado de (fases de) ataques
- Sensores adicionales

ATT&CK™
Adversarial Tactics, Techniques & Common Knowledge



Más información

- ***R. Sommer, V. Paxson;***
  Outside the Closed World: On Using Machine Learning For Network Intrusion Detection
  Proc. IEEE Symp. On Security and Privacy, 305-316, 2010.
- ***P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, E. Vazquez;***
  Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges
  Computers & Security, 28:18-28, 2009.
- ***M. Tavallaee, N. Stakhanova, A. Ghorbani,***
  Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods
  IEEE Transactions on Systems, Man, And Cybernetics —Part C, 40(5):516-524, 2010.
- ***VARUN CHANDOLA, ARINDAM BANERJEE, VIPIN KUMAR,***
  Anomaly Detection: A Survey
  ACM Computing Surveys, Vol. 41, No. 3, Article 15, 2009.
- ***Zuech, R.; Khoshgoftaar, T.; Wald, R.***
  Intrusion detection and big heterogeneous data: a survey
  Journal of Big Data, 2:3 (2015).

El papel de la IA en la monitorización de la seguridad en red

Jesús Esteban Díaz-Verdejo
Departamento de Teoría de la Señal, Telemática y Comunicaciones
E.T.S. Ingenierías Informática y Telecomunicación – Universidad de Granada
C/ Periodista Daniel Saucedo Aranda, s/n - 18071 – Granada (Spain)
Phone: +34-958 242304 - Email: jedv@ugr.es

https://dtstc.ugr.es/neus-cslab

Esquema