Traffic Monitoring and Diagnosis with Multivariate Statistical Network Monitoring: A Case Study

José Camacho (josecamacho@ugr.es)

Pedro García-Teodoro (pgteodor@ugr.es)

Gabriel Maciá-Fernández (gmacia@ugr.es)

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Universidad de Granada

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Unión Europea
Fondo Europeo de Desarrollo Regional

Motivation - I

- ✓ Anomaly Detection from Traffic Data
  - ✓ Accuracy & Variate Sources
- ✓ Software for CyberSec:
  - ✓ Pivoting (Specific)
  - ✓ High False Anomalies (Correlation)
- ✓ CyberSec Research ➔ ML
  - ✓ Data Fusion (General)
  - ✓ High Detection but Semantic Gap
- ✓ Multivariate Statistical Network Monitoring
  - ✓ Data Fusion (General)
  - ✓ High Detection & Diagnosis

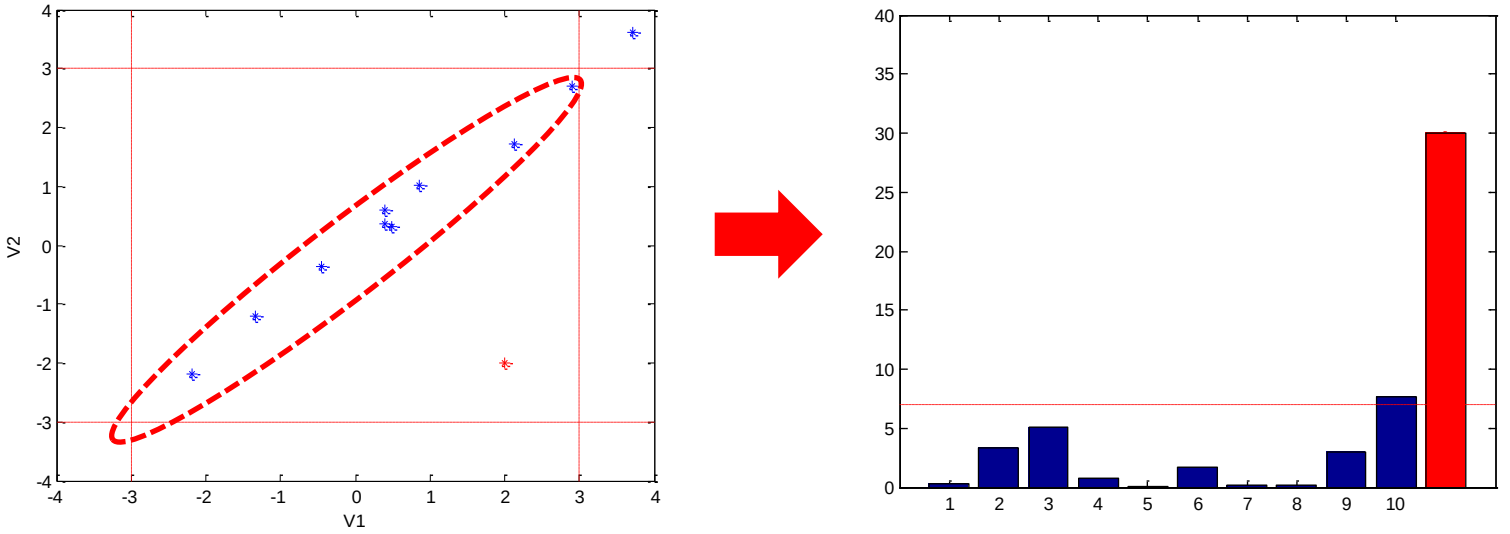
Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Motivation - II

Multivariate approach

In a data set with many measured variables, the interesting information is contained in a (much lower) number of **latent variables**

Multivariate Statistical Control (PCA)



Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Multivariate Network Security Monitoring (MSNM)

COMPUTERS & SECURITY 59 (2016) 118–137

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

ELSEVIER

PCA-based multivariate statistical network monitoring for anomaly detection

José Camacho *, *Alejandro Pérez-Villegas*, *Pedro García-Teodoro*, *Gabriel Maciá-Fernández*

Department of Signal Theory, Telematics and Communications, School of Computer Science and Telecommunications – CITIC, University of Granada, Granada, Spain

2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM Workshop on Security and Privacy in Big Data

Tackling the Big Data 4 Vs for Anomaly Detection

José Camacho, Gabriel Maciá-Fernández, Jesús Díaz-Verdejo and Pedro García-Teodoro
Dpt. of Signal Theory, Telematics and Communications - CITIC, University of Granada
Email:{josecamacho,gmacia,jedv,pgteodor}@ugr.es

Perspective

CHEMOMETRICS

Received: 12 February 2016, Revised: 21 April 2016, Accepted: 25 April 2016, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/cem.2806

Networkmetrics: multivariate big data analysis in the context of the internet

José Camacho[a]*, Roberto Magán-Carrión[a], Pedro García-Teodoro[a] and James J. Treinen[b]

…n all areas of knowledge. In chemistry and related disciplines, the chemometric com-…ffort to understand and solve problems mainly from a multivariate and exploratory …ndeed, of broader applicability, even in areas of knowledge far from chemistry. In …et: the net of devices that allow an interconnected world where all types of data … communication services can be provided. Problems in the Internet or in general in …from chemometric problems. Building on this parallelism, we review four classes of …on, anomaly detection, optimization, and classification. We present an illustrative …ultivariate perspective may lead to significant improvements from state-of-the-art …etter name, we call the approach of treating these problems from that multivari-…letworkmetric problems have their own specificities, mainly, their typical Big Data …ictured data. We argue that multivariate analysis is, indeed, useful to tackle these …n Wiley & Sons, Ltd.

…tworking; networkmetrics; Big Data

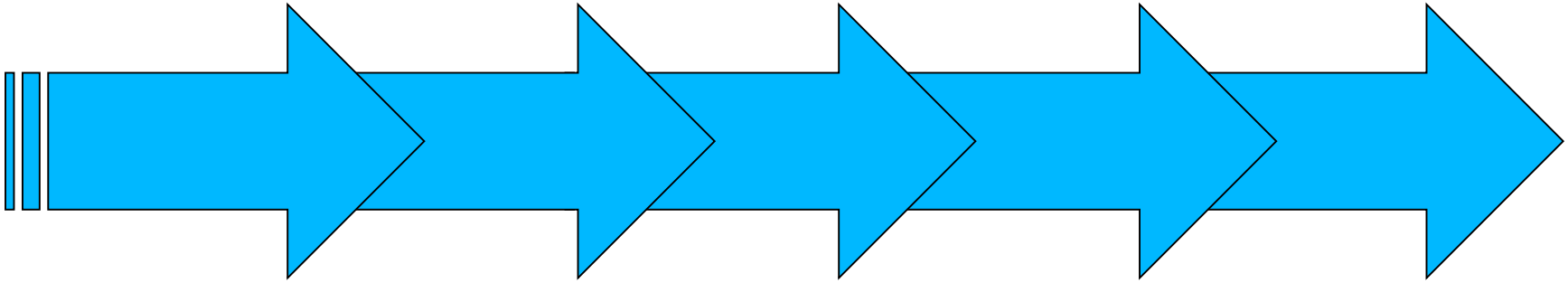
Hierarchical PCA-Based Multivariate Statistical Network Monitoring for Anomaly Detection

Gabriel Maciá-Fernández, José Camacho, Pedro García-Teodoro, Rafael A. Rodríguez-Gómez
Dpt. of Signal Theory, Telematics and Communications
Network Engineering & Security Group, CITIC-UGR
University of Granada - Spain
Email: {gmacia, josecamacho, pgteodor, rodgom}@ugr.es

2016 IEEE International Workshop on Information Forensics and Security (WIFS)

MSNM - II

- ✓ MSNM: 5 steps from the hay to the needle



(1) Parsing

(2) Fusion

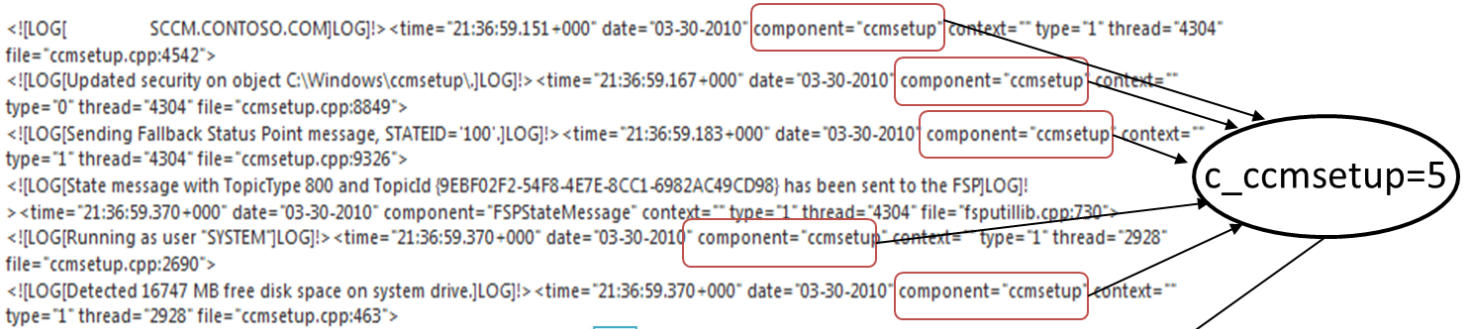
(3) Detection

(4) Diagnosis

(5) De-parsing

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

(1) Parsing: Feature-as-a-counter



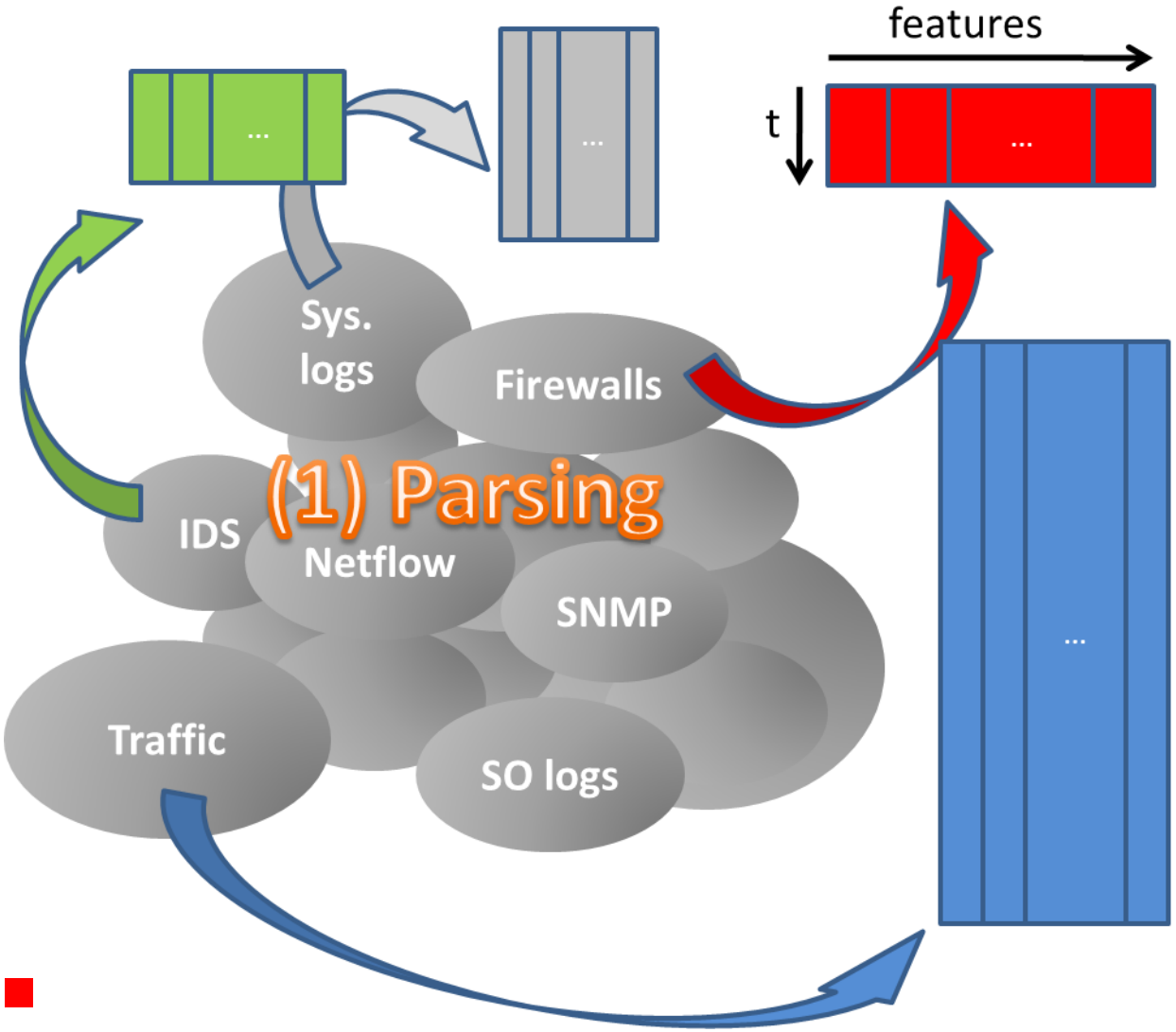
Time	FSPStateMessage	ccmstup	thread_4304
T=20s	1	5	4
T=40s	2	3	3
T=60s	1	3	3
T=80s	1	1	4

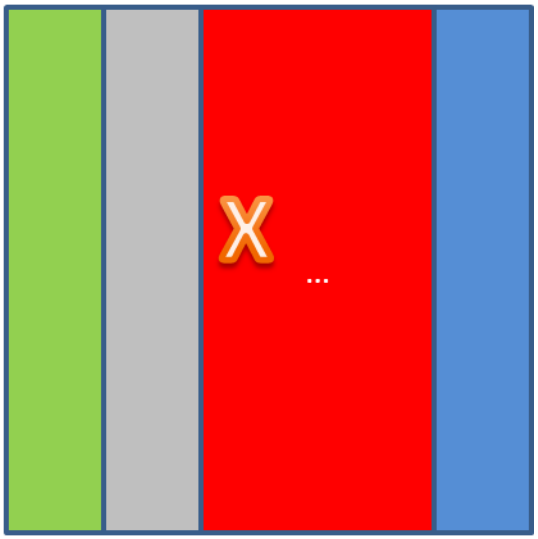
2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM Workshop on Security and Privacy in Big Data

Tackling the Big Data 4 Vs for Anomaly Detection

José Camacho, Gabriel Maciá-Fernández, Jesús Díaz-Verdejo and Pedro García-Teodoro
Dpt. of Signal Theory, Telematics and Communications - CITIC, University of Granada
Email:{josecamacho,gmacia,jedv,pgteodor}@ugr.es

MSNM - III

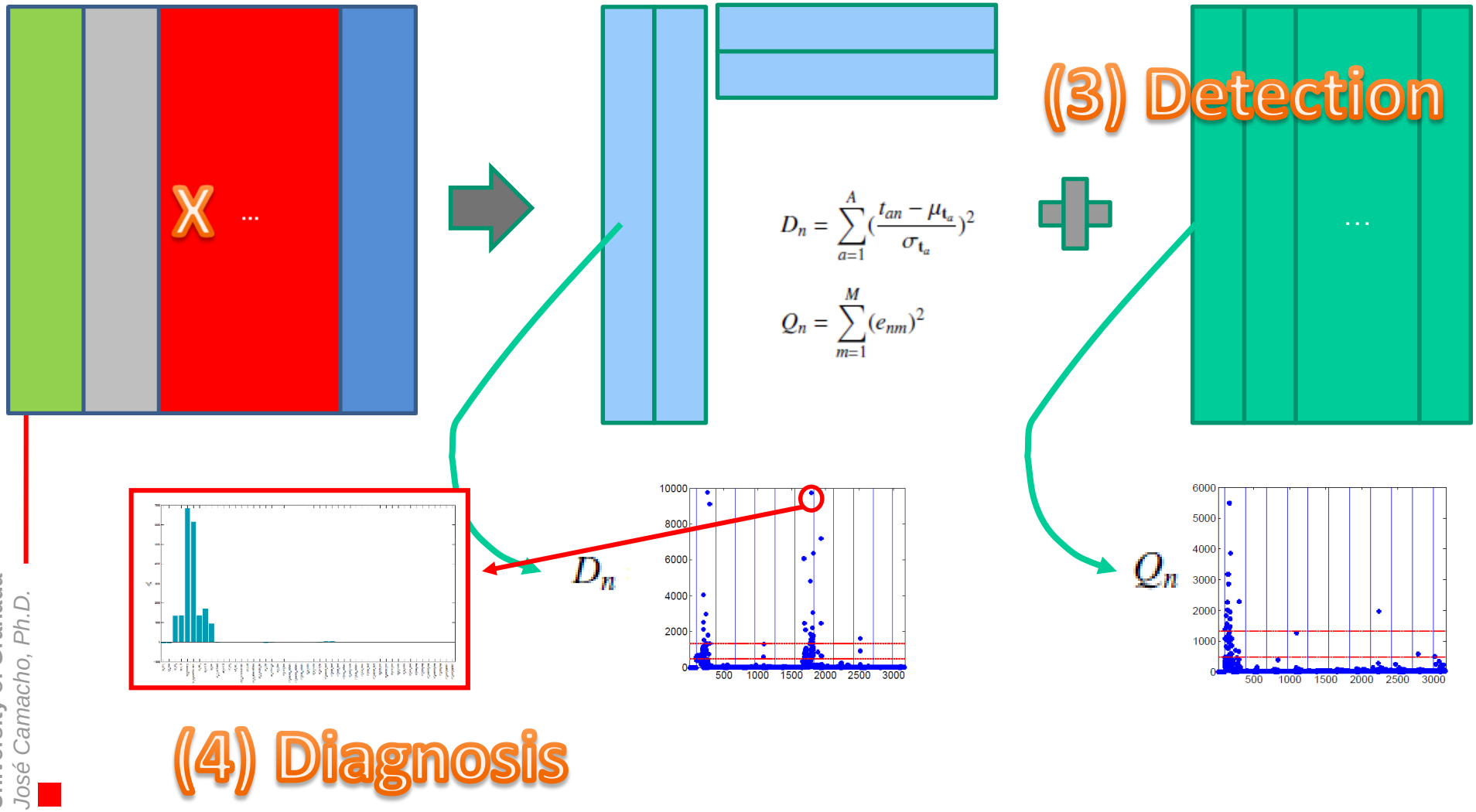




Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Combine any sources: from low level sensors (e.g. netflow) to high level info (e.g. correlation rules at SIEM)

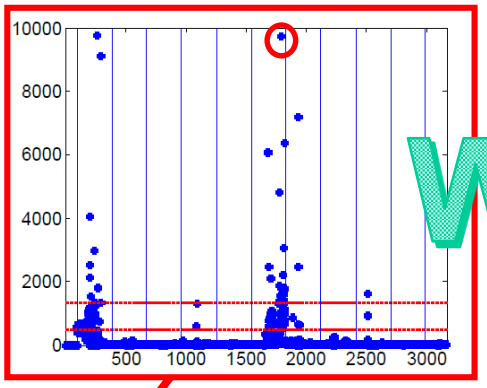
MSNM - VI



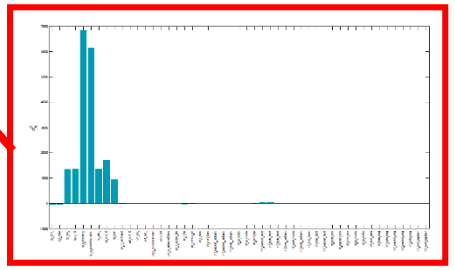
Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

MSNM creates a normality model for detection (which answers when an anomaly takes place) and diagnosis (in which source/s of info it is detected)

MSNM - V



When?



Where?

(5) De-parsing

IDS-03292012-1hr.txt

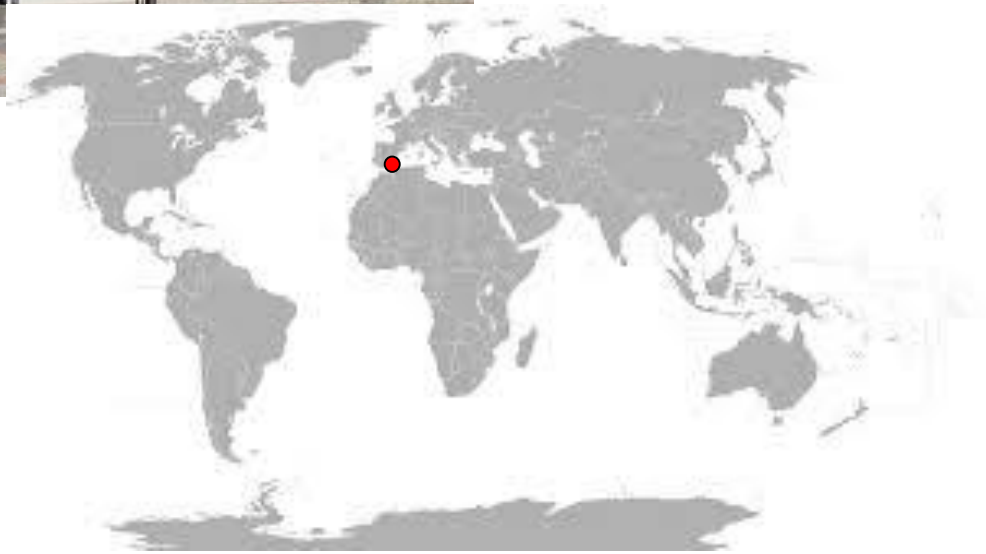
```
[**] [1:2103003:7] GPL NETBIOS SMB-DS Session Setup NTMLSSP unicode asn1 overflow attempt [**]
[Classification: Generic Protocol Command Decode] [Priority: 3]
03/29-14:48:31.019982 172.23.0.216:1251 -> 172.23.0.10:445
TCP TTL:128 TOS:0x0 ID:1696 IpLen:20 DgmLen:1500 DF
***A**** Seq: 0xA9B345B0  Ack: 0x4522D27D  Win: 0xFF3A  TcpLen: 20
[Xref => http://www.microsoft.com/technet/security/bulletin/MS04-007.mspx][Xref =>
http://cgi.nessus.org/plugins/dump.php3?id=12065][Xref =>
http://cgi.nessus.org/plugins/dump.php3?id=12052][Xref =>
http://cve.mitre.org/cgi-bin/cvename.cgi?name=2003-0818][Xref =>
http://www.securityfocus.com/bid/9635][Xref => http://www.securityfocus.com/bid/9633]

[**] [1:2102466:9] GPL NETBIOS SMB-DS IPC$ unicode share access [**]
[Classification: Generic Protocol Command Decode] [Priority: 3]
03/29-14:48:31.024896 172.23.0.216:1251 -> 172.23.0.10:445
TCP TTL:128 TOS:0x0 ID:1698 IpLen:20 DgmLen:138 DF
***AP*** Seq: 0xA9B35020  Ack: 0x4522D402  Win: 0xFDB5  TcpLen: 20

[**] [1:2103003:7] GPL NETBIOS SMB-DS Session Setup NTMLSSP unicode asn1 overflow attempt [**]
[Classification: Generic Protocol Command Decode] [Priority: 3]
03/29-14:48:32.421373 172.23.0.211:1308 -> 172.23.0.10:445
TCP TTL:128 TOS:0x0 ID:1843 IpLen:20 DgmLen:1500 DF
***A**** Seq: 0xA1B4DB42  Ack: 0x5D2556D8  Win: 0xFF3A  TcpLen: 20
[Xref => http://www.microsoft.com/technet/security/bulletin/MS04-007.mspx][Xref =>
http://cgi.nessus.org/plugins/dump.php3?id=12065][Xref =>
http://cgi.nessus.org/plugins/dump.php3?id=12052][Xref =>
http://cve.mitre.org/cgi-bin/cvename.cgi?name=2003-0818][Xref =>
http://www.securityfocus.com/bid/9635][Xref => http://www.securityfocus.com/bid/9633]
```

For Help, press F1 | 30 | 1 | Read | Ovr | Block | Sync | Rec | Caps

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

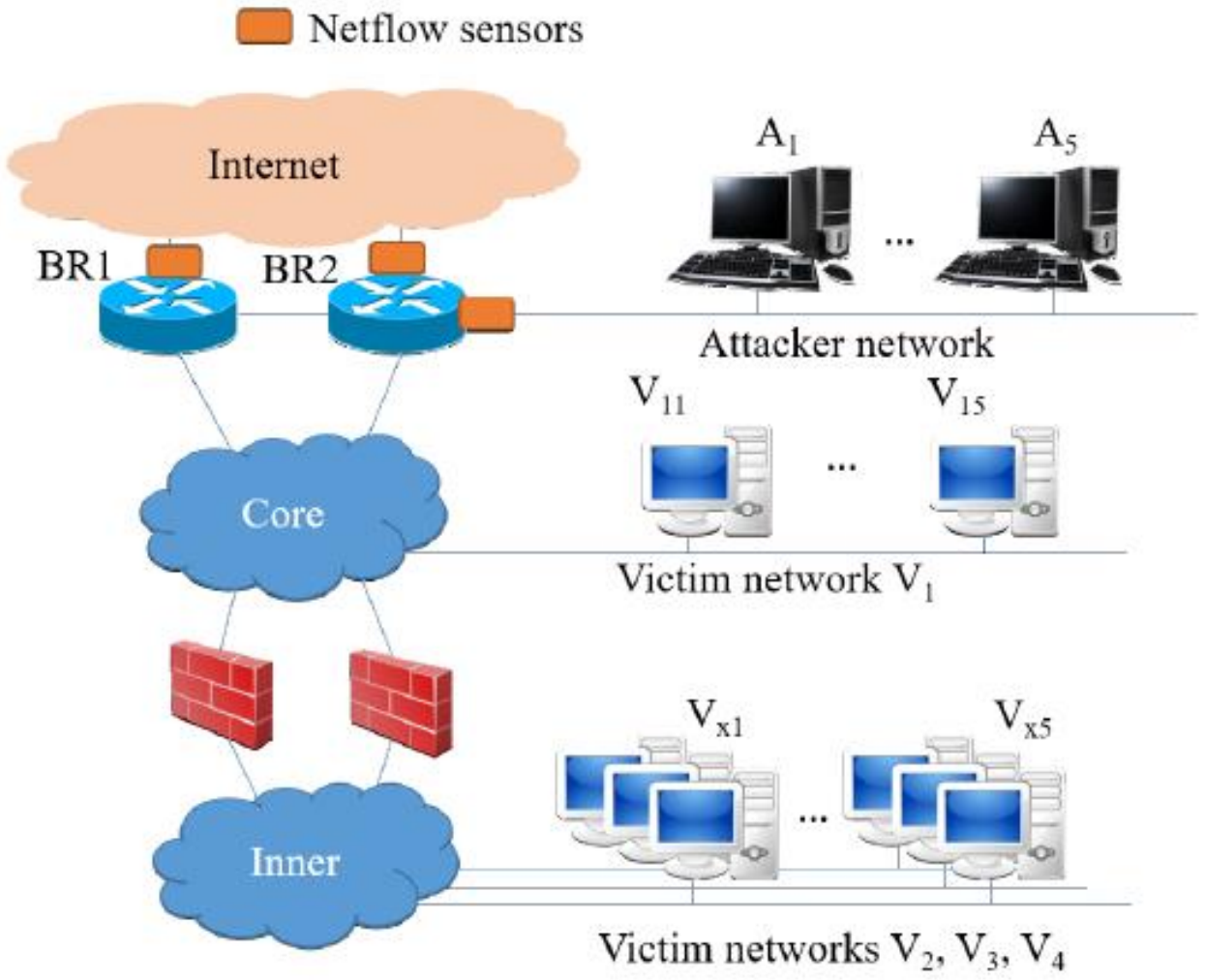
Logs with detailed info of the anomaly are manually identified from the information provided by MSNM

Case Study - I



Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

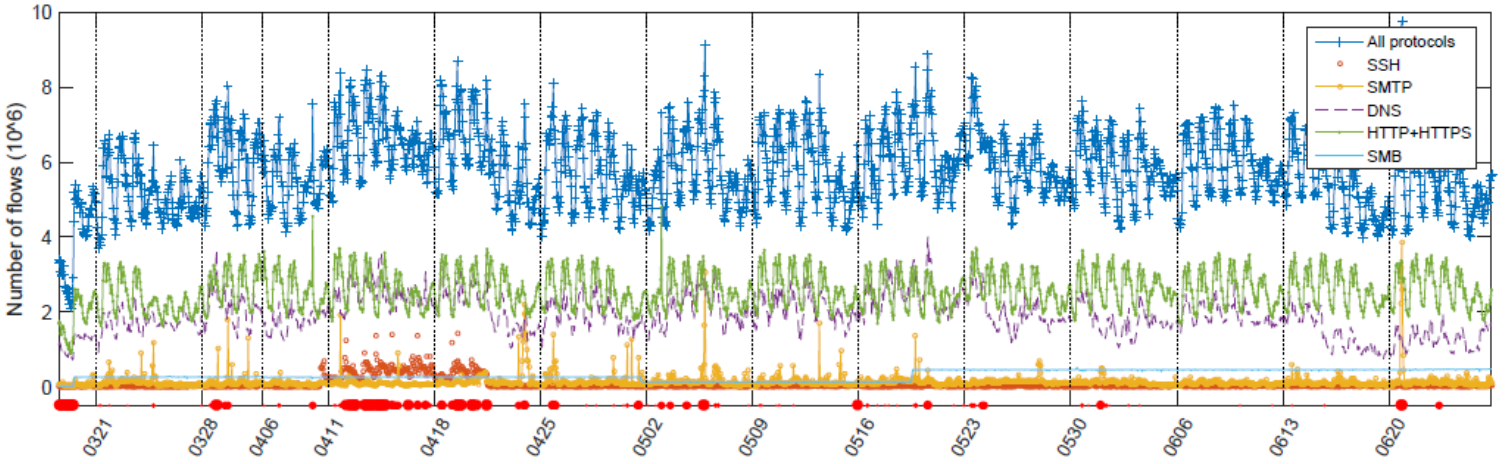
Case Study - II



- ✓ Synthetic attacks: DoS, scan, exfiltration (3 types)

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

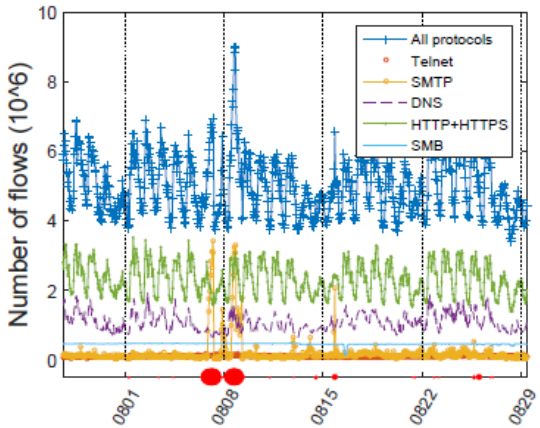
Case Study - III



(a) Calibration Set

TABLE IV
FEATURES OF THE CALIBRATION AND THE TEST SETS.

Feature	Calibration	Test
Capture start	10:52h 03/18/2016	13:43h 07/27/2016
Capture end	18:27h 06/26/2016	09:27h 08/29/2016
Attacks start	N/A	00:00h 07/28/2016
Attacks end	N/A	12:00h 08/09/2016
Number of files	17	6
Size (compressed)	181GB	55GB
# Connections	≈ 13,000M	≈ 3,900M



(b) Test Set

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Case Study - V

- ✓ Features (138, no Fusion)

Variable	#features → values
Source IP	2 → *public, private*
Destination IP	2 → *public, private*
Source port	50 → *specific services, Other*
Destination port	50 → *specific services, Other*
Protocol	5 → *TCP, UDP, ICMP, IGMP, Other*
Flags	6 → *A, S, F, R, P, U*
ToS	3 → *0, 192, Other*
# Packets in	5 → *very low, low, medium, high, very high*
# Packets out	5 → *very low, low, medium, high, very high*
# Bytes in	5 → *very low, low, medium, high, very high*
# Bytes out	5 → *very low, low, medium, high, very high*

2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM Workshop on Security and Privacy in Big Data

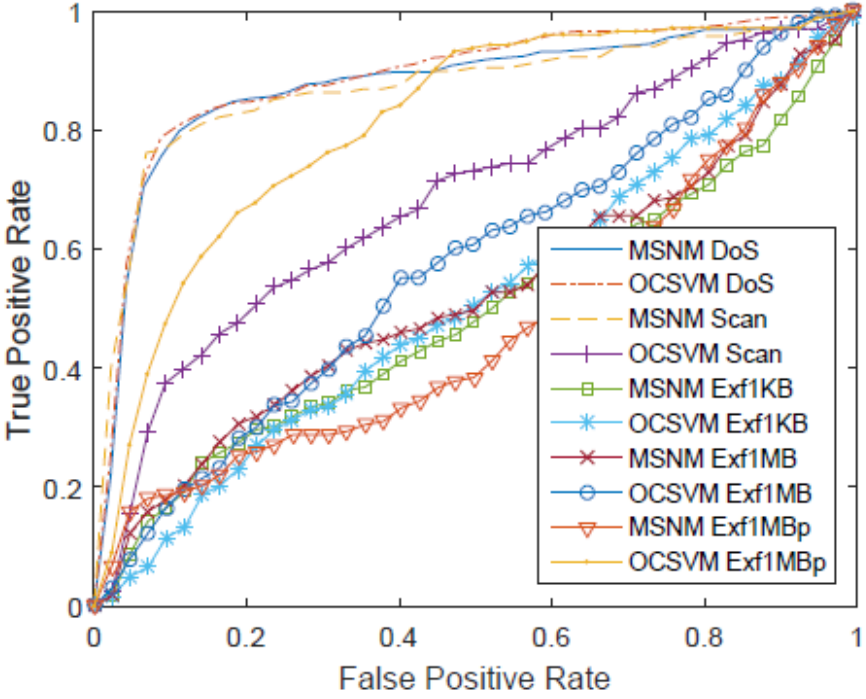
Tackling the Big Data 4 Vs for Anomaly Detection

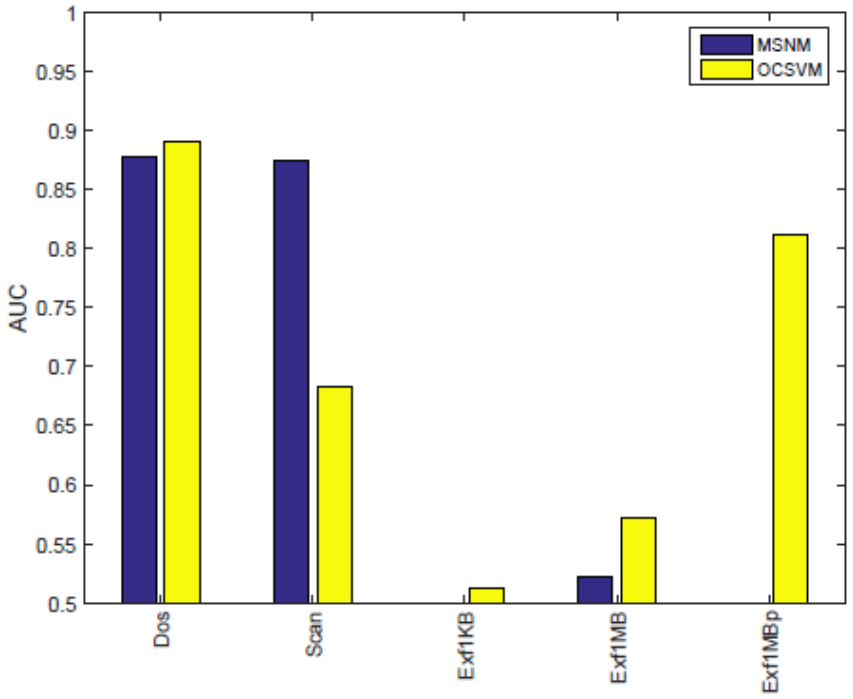
José Camacho, Gabriel Maciá-Fernández, Jesús Díaz-Verdejo and Pedro García-Teodoro
Dpt. of Signal Theory, Telematics and Communications - CITIC, University of Granada
Email:{josecamacho,gmacia,jedv,pgteodor}@ugr.es

- ✓ One-class SVM
  - ✓ B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “New Support Vector Algorithms,” Neural computation, vol. 12, no. 5, pp.1207–1245, 2000.
  - ✓ B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution,” Neural Computation, vol. 13, no. 7, pp. 1443–1471, 2001.

Results - I

✓ Synthetic attacks

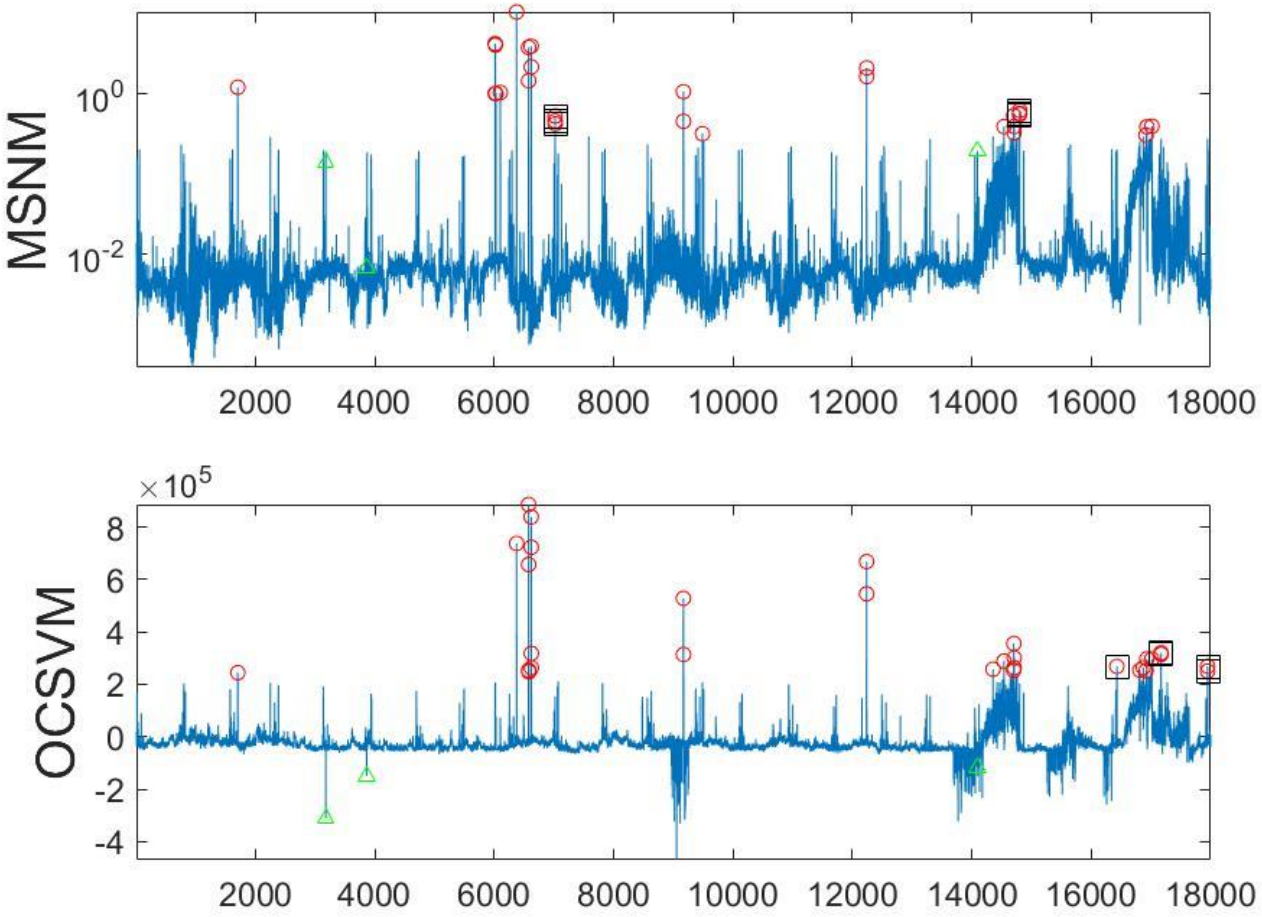




Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Results - II

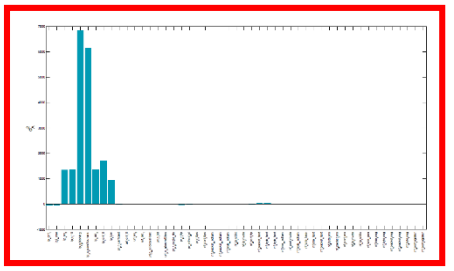
✓ Real attacks: Detection scores



- 20 Top detections
- Top & Synth.
- Least agreement

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Results - III



- ✓ Real attacks: Diagnosis

Time interval 20160806t2039-20160807t0559
'dport_smtp' 'in_nbytes_low' 'tcpflags_SYN' 'protocol_tcp' 'tcpflags_PSH' 'in_npackets_low' '—protocol_udp' '—dport_dns' 'tcpflags_RST' 'srctos_zero'

↓

Many SMTP short connections

Advantage over ML

Diagnosis is useful, but too complicate!

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Results - IV

✓ Real attacks: De-parsing example

Time interval 20160806t2039-20160807t0559

'dport_smtp' 'in_nbytes_low' 'tcpflags_SYN' 'protocol_tcp' 'tcpflags_PSH' 'in_npackets_low' '—protocol_udp' '—dport_dns' 'tcpflags_RST' 'srctos_zero'

```
MSNM rocks :-) $ nfdump -R $inputs -t 2016/08/06.20:39:00-2016/08/07.06:00:00
 'dst port = 25 or (bytes > 150 and bytes < 1001) ... '   -w $output
```

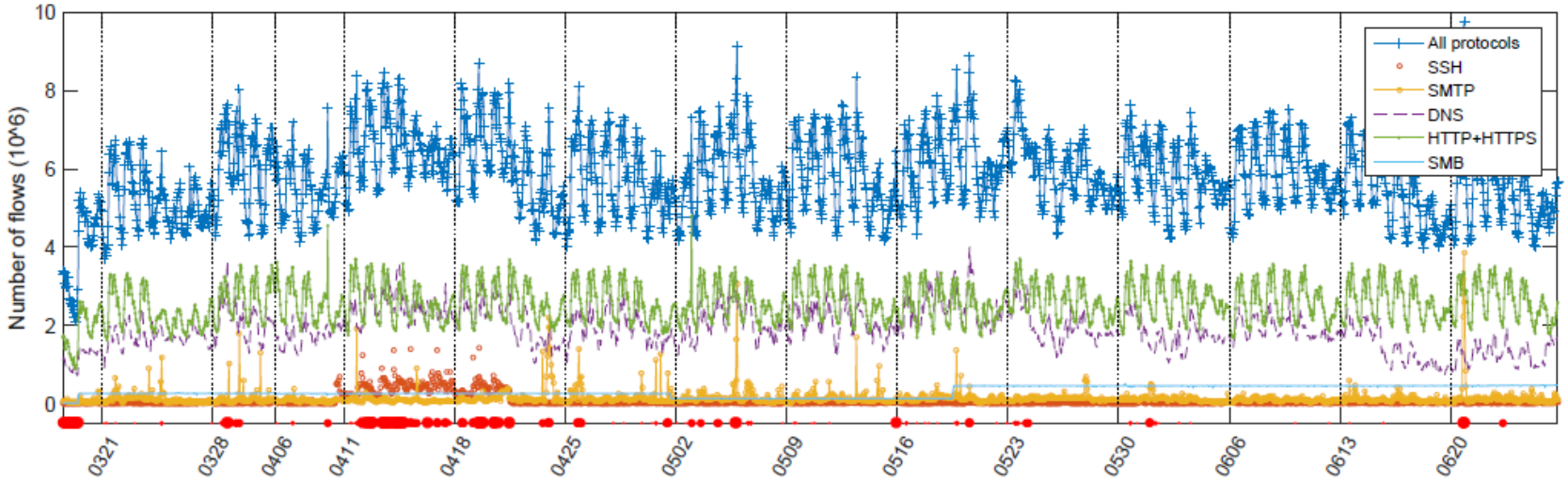
SPAM campaing

```
MSNM rocks :-) $ nfdump -r $output -s dstport:p/packets -n 5
Top 5 Dst Port ordered by packets:
Date first seen          Duration Proto          Dst Port    Flows(%)      Packets(%)
2016-08-06 20:39:01.448 33657.436 GRE                 0    16537( 0.1)  170.8 M(31.0)
2016-08-06 20:40:21.164 33579.712 TCP                25   11.5 M(38.2)   72.0 M(13.1)
2016-08-06 20:39:02.756 33658.236 TCP               443    1.4 M( 4.6)   45.6 M( 8.3)
2016-08-06 20:39:08.644 33652.352 TCP                80    1.5 M( 4.9)   45.1 M( 8.2)
2016-08-06 21:00:20.876 27042.808 TCP                20     109( 0.0)   25.9 M( 4.7)
```

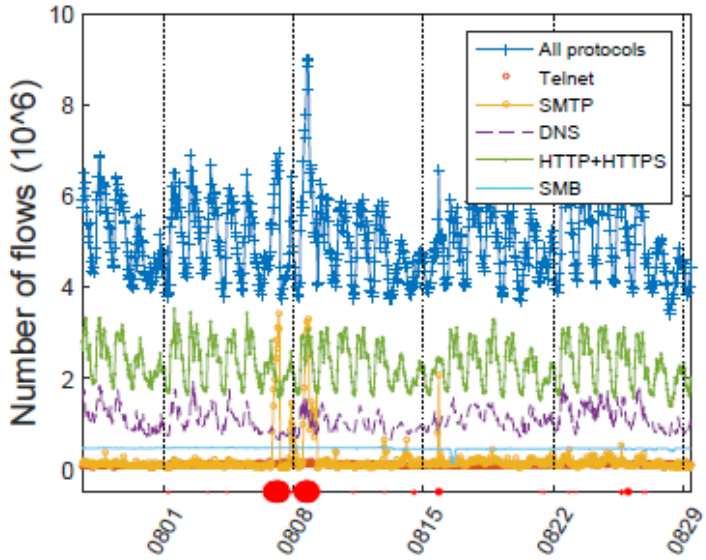
Normal

```
Top 5 Dst Port ordered by packets:
Date first seen          Duration Proto          Dst Port    Flows(%)      Packets(%)
2016-08-20 11:07:01.452   895.896 GRE                 0     378( 0.1)    2.0 M(25.6)
2016-08-20 11:07:27.128   869.912 TCP                80   32569( 9.5)    1.0 M(12.8)
2016-08-20 11:07:27.116   870.680 TCP               443   26252( 7.6)   633500( 7.9)
2016-08-20 11:11:27.008   583.996 TCP             26242       4( 0.0)   590509( 7.4)
2016-08-20 11:07:01.000   888.752 TCP                22     898( 0.3)   505988( 6.3)
```

Results - V



(a) Calibration Set



(b) Test Set

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Conclusion- I

- ✓ MSNM ~ OCSVM in detection.
- ✓ Good detectionperformance when including real attacks.
- ✓ MSNM has diagnosis support ➔ reduces the time from detection to response.
- ✓ Unlike other methods, MSNM takes advantage of a large number of features ➔ improves Diagnosis
- ✓ Future work:
  - ✓ Other sources (Host), MSNM-SIEM, Privacy & Scalibility, …

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.

Traffic Monitoring and Diagnosis with Multivariate Statistical Network Monitoring: A Case Study

José Camacho (josecamacho@ugr.es)

Pedro García-Teodoro (pgteodor@ugr.es)

Gabriel Maciá-Fernández (gmacia@ugr.es)

This work is partly supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds through project TIN2014-60346-R

Network Engineering and Security Group
University of Granada
José Camacho, Ph.D.