# SIP-based VoIP Traffic Behavior Profiling and Its Applications

Juan R. Cayon Alcalde
GRSST -- Automatica y Computacion -- UPNA

# Autores

- Hun Jeong Kang, Zhi-Li Zhang (Universidad de Minesota)
- Supranayama Ranjan, Antonio Nucci (Narus Inc.):
  - *”Narus is working closely with prominent academics to address the issues of providing real-time traffic insight to profitably manage, secure and deliver services over IP and the threat of increasingly advanced cyber-attacks. ”*

# Objetivos

- Entender las características del tráfico SIP
- Identificación de anomalías
- Diagnosis de fallos y detección de ataques

# Metodología

- Análisis de tráfico REAL: trazas capturadas en un proveedor de servicio VoIP (3 trazas de unos 40m capturadas a distintas horas de un mismo día)

- Multilevel Profiling:
  - Máquina servidor
  - Entidades lógicas del servidor (registar, call-proxy, ...)
  - Usuarios individuales

# Identificación de Servidores

- Típica cabecera de un paquete con un método INVITE

```
U 2008/09/30 09:19:05.221536 198.33.1.12:5070 -> 198.33.1.20:5060

INVITE sip:0034123456789@voztele.com:5070 SIP/2.0.

Record-Route:
   <sip:198.33.1.12:5070;lr=on;ftag=6F3AE5EFC05F2D19D5C37FFE2BBC1800>.

Via: SIP/2.0/UDP 198.33.1.12:5070;branch=z9hG4bK253e.ec4a3cc6.0.

Via: SIP/2.0/UDP 198.33.1.198:50060;branch=z9hG4bK253e.dd224b.0.

To: <sip:123456789@oigaa.com:5070>.

From: "Jes Rez"<sip:1234567890@oigaa.com>;
   tag=6F3AE5EFC05F2D19D5C37FFE2BBC1800.
```

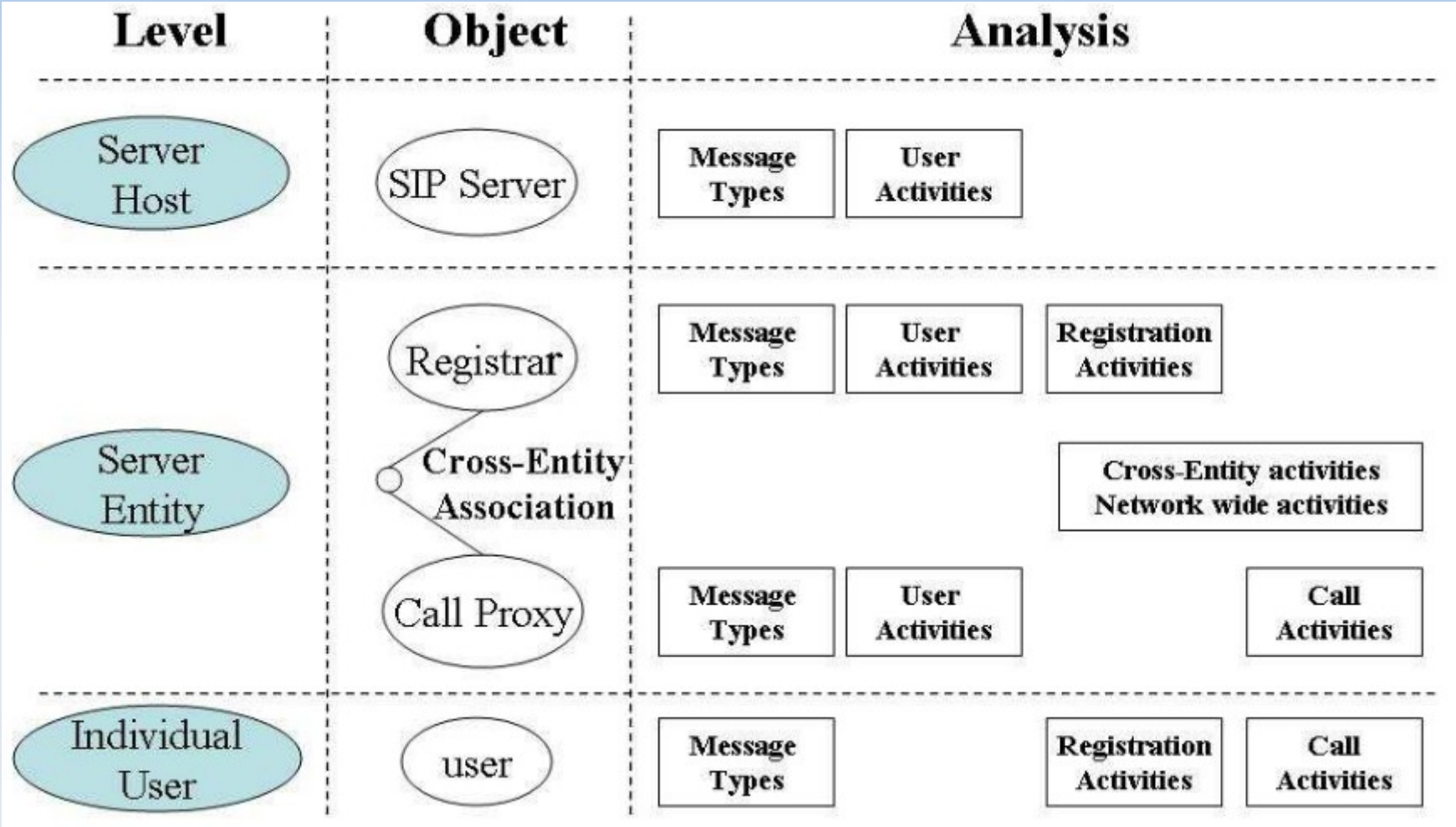
# Identificación de Servidores

**Algorithm 1** Baseline Algorithm for SIP Call Proxy Discovery

1: Parameters: message set $M$, threshold $\alpha$;
2: Initialization: $IPSet := \emptyset$; $ProxyIP := \emptyset$;
3: **for each** $m \in M$ **do**
4: **if** $m.\texttt{method} == \texttt{INVITE}$ **then**
5: $x = m.\texttt{sourceIP}$; $y = m.\texttt{destinationIP}$;
6: $from = m.\texttt{FROM}$; $to = m.\texttt{TO}$;
7: **if** $x \notin IPSet$ **then**
8: $x.Out_{\texttt{FROM}} = \{from\}$; $x.Out_{\texttt{TO}} = \{to\}$;
9: $x.In_{\texttt{FROM}} = \emptyset$; $x.In_{\texttt{TO}} = \emptyset$;
10: **else**
11: $x.Out_{\texttt{FROM}} = x.Out_{\texttt{FROM}} \cup \{from\}$;
12: $x.Out_{\texttt{TO}} = x.Out_{\texttt{TO}} \cup \{to\}$;
13: **end if**
14: **if** $[|x.In_{\texttt{FROM}}|, |x.In_{\texttt{TO}}|, |x.Out_{\texttt{FROM}}|, |x.Out_{\texttt{TO}}|] > [\alpha, \alpha, \alpha, \alpha]$ **then**
15: $ProxyIP = ProxyIP \cup \{x\}$
16: **end if**
17: **if** $y \notin IPSet$ **then**
18: $y.In_{\texttt{FROM}} = \{from\}$; $x.In_{\texttt{TO}} = \{to\}$;
19: $y.Out_{\texttt{FROM}} = \emptyset$; $y.Out_{\texttt{TO}} = \emptyset$;
20: **else**
21: $y.In_{\texttt{FROM}} = y.Out_{\texttt{FROM}} \cup \{from\}$;
22: $y.In_{\texttt{TO}} = y.In_{\texttt{TO}} \cup \{to\}$;
23: **end if**
24: **if** $[|y.In_{\texttt{FROM}}|, |y.In_{\texttt{TO}}|, |y.Out_{\texttt{FROM}}|, |y.Out_{\texttt{TO}}|] > [\alpha, \alpha, \alpha, \alpha]$ **then**
25: $ProxyIP = ProxyIP \cup \{y\}$
26: **end if**
27: **end if**
28: **end for**

- M = traza
- m = paquete
- x,y = IPs fuente y destino del paquete IP (layer 3)
- TO, FROM = campos (URI) del mensaje SIP (layer 7)
- Out (mensaje saliente)
- In (mensaje entrante)
- α = valor umbral

# Multilevel Profiling



# CARACTERIZACIÓN

# Caracterización del servidor (host)

- Message Types:
  - # peticiones recibidas y enviadas en un período T (5-15 min.)
  - # respuestas recibidas y enviadas en un período T
  - Se mantienen estadísticas agregadas de ambos tipos y relaciones entre los mismos
- User Activities:
  - # URIs diferentes en campos FROMy TO de las peticiones (Requests)
  - Aggregate User Activity Diversity (UAD)

# User Activity Diversity

$$UAD = \frac{-\sum_{i=1}^{n} p_i \log p_i}{\log m} \in [0,1]$$

- m = # peticiones en el período T
- n = # usuarios distintos
- $p_i = m_i/m$, donde $m_i$ = # peticiones donde aparece el usuario i (campo To/From indistinto)

# Caracterización del Registrar

- Message Types & User Activities: almacenan y procesan datos similares a los anteriores, pero relativos sólo a mensajes REGISTER.
- Registration Activities:
  - Registros correctos y fallidos
  - Δt entre dos peticiones REGISTER de un usuario (tiempo de validez del registro)
  - Δt entre dos peticiones REGISTER consecutivas (tiempo entre llegadas)

# Caracterización del Call Proxy

- # y relaciones entre distintos tipos de peticiones y respuestas (INVITE, CANCEL, BYE, etc.) en el servidor
- Distintas métricas UAD en las que:
  - UAD_llamantes, $m_i$ = # peticiones con el usuario i en el campo From
  - UAD_llamados, $m_i$ = # peticiones con el usuario i en el campo To
  - UAD_pares_llamante-llamado, $m_{ij}$ = # peticiones con el usuario i en el campo From y el j en el To.

# Caracterización del Call Proxy

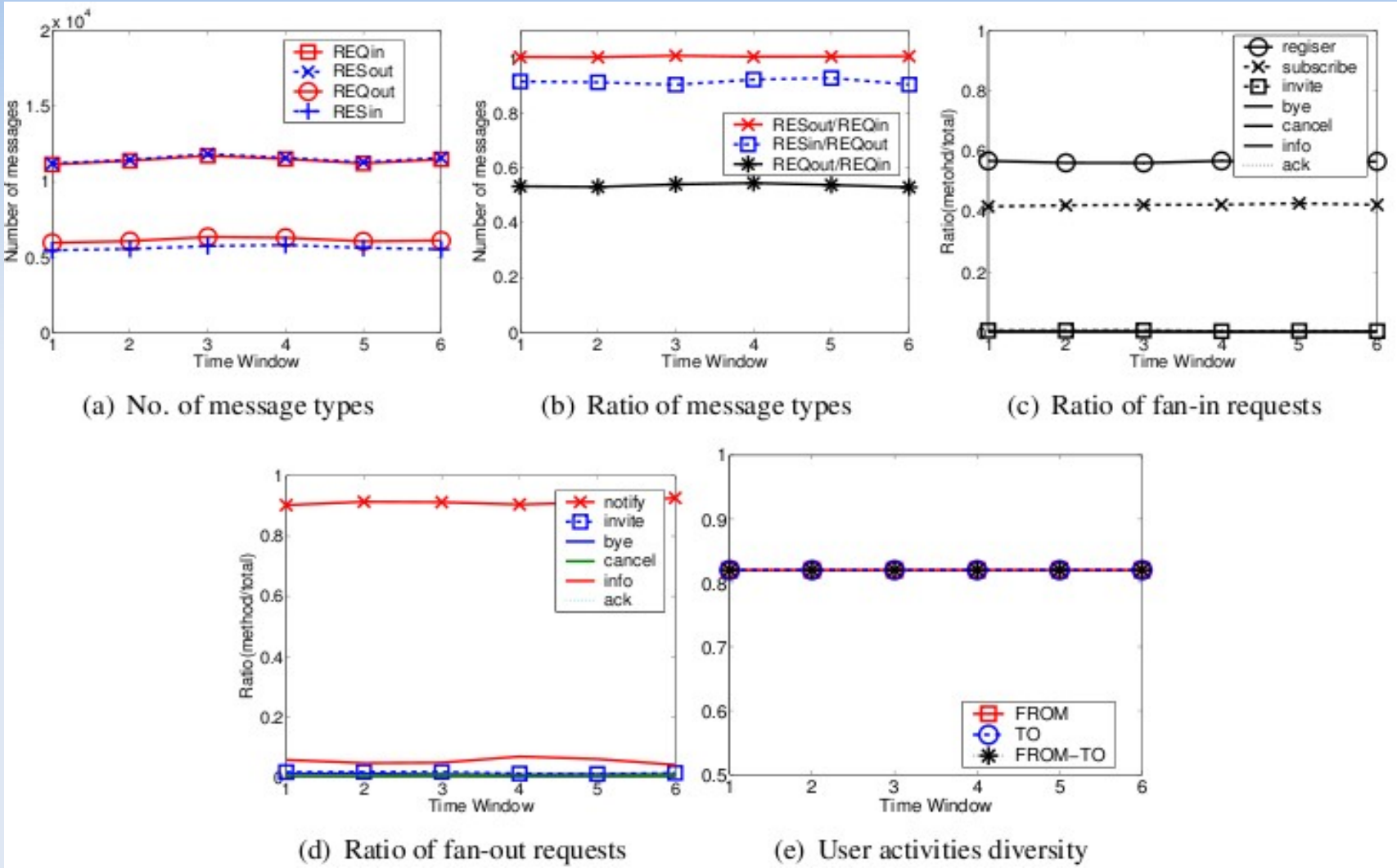
- # llamadas: en curso, finalizadas, canceladas, fallidas, ... e un período T.
- Caracterización (media, varianza, tipo de distribución) de la duración de las llamadas.
- Idem, de los tiempos entre llegadas de las peticiones.

# Caracterización de los usuarios

- # llamadas realizadas/recibidas
- UAD_llamantes, en las llamadas recibidas por un usuario
- UAD_llamados, en las llamadas hechas por un usuario
- Duración de las llamadas
- [...]

# RESULTADOS

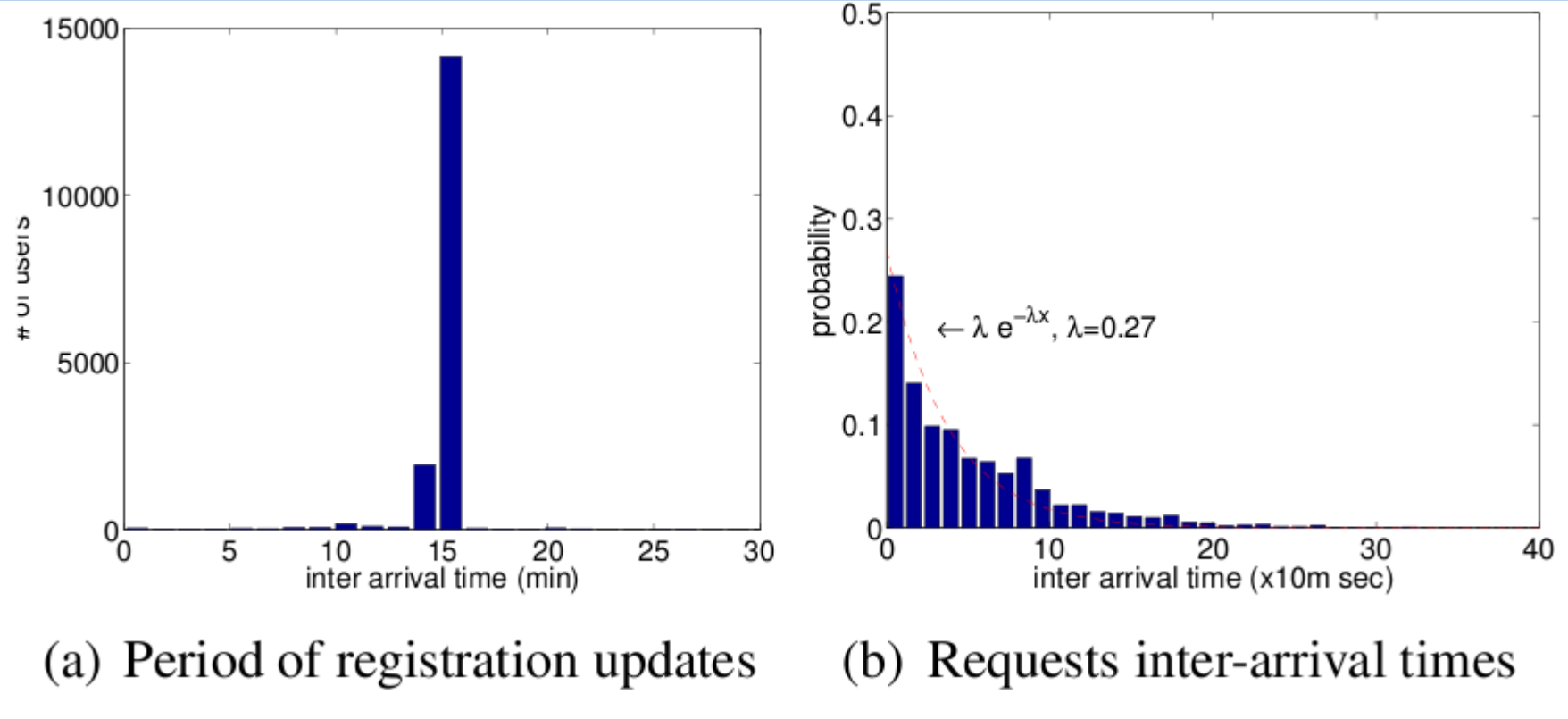
# Características del Servidor




(a) No. of message types (b) Ratio of message types (c) Ratio of fan-in requests

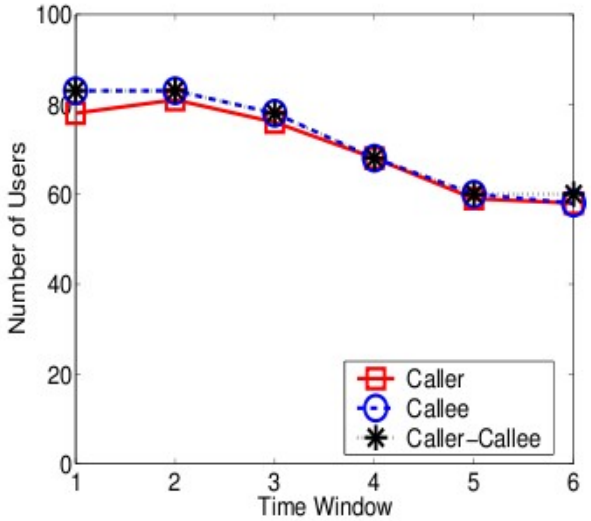
(d) Ratio of fan-out requests (e) User activities diversity

# Comportamiento del *REGISTRAR*

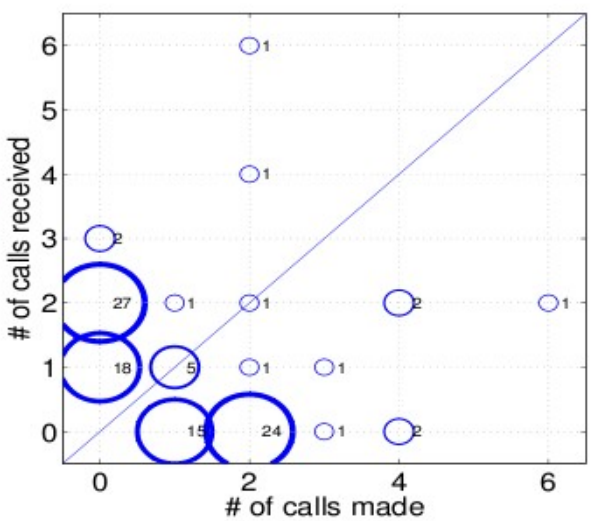



(a) Period of registration updates (b) Requests inter-arrival times

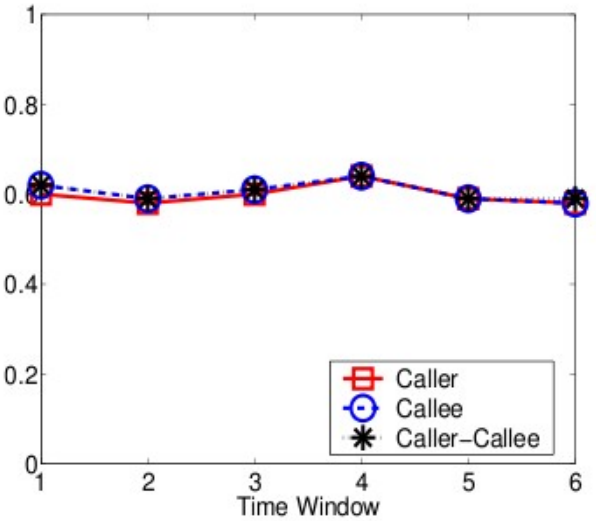
# Comportamiento del *Proxy/User*



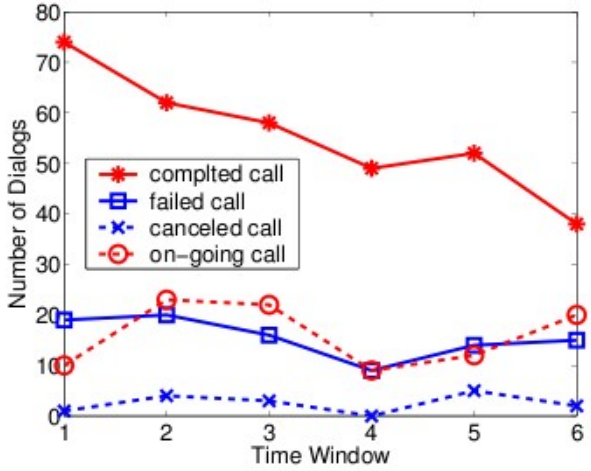

(a) No. of callers and callees



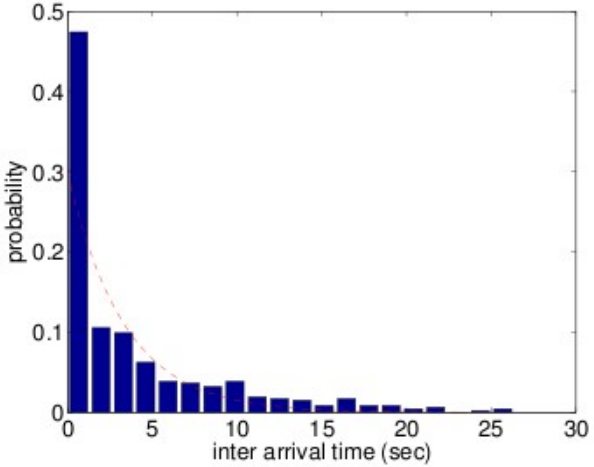

(b) Calls made vs received




(c) User activities diversity



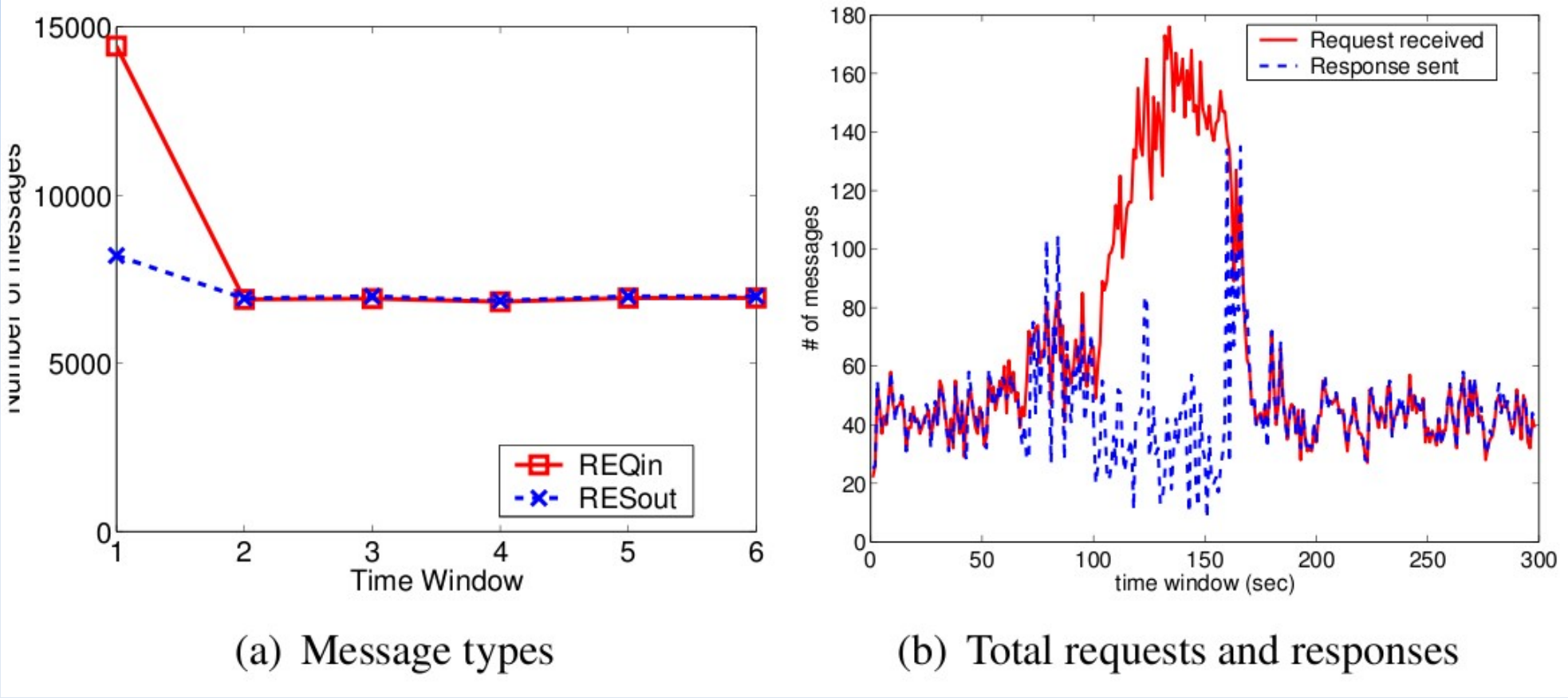

(d) Call types




(e) Call inter-arrival times

# DETECCIÓN ANOMALÍAS

# Traza anómala (*REGISTRAR*)




(a) Message types

(b) Total requests and responses

# [mis] CONCLUSIONES

- ”interesante” aproximación, la caracterización a tres niveles
- Dificultad de la semántica: TCP/IP vs SIP
- Lástima no tener el technical report completo:
  - Extensiones del algoritmo → estudio del efecto del NAT
  - Algoritmo detección de anomalías basado en los resultados de caracterización del tráfico

Ruegos...

Preguntas.....

Sugerencias.....