

# LinuxDay

22 Ottobre 2016



## IDENTIFICAZIONE AUTOMATICA DI MACRO-ATTIVITÀ NEL TRAFFICO DI RETE

*Emanuele Faranda*



# Introduzione: monitoraggio di rete

- Pacchetti di rete
- Flussi di rete



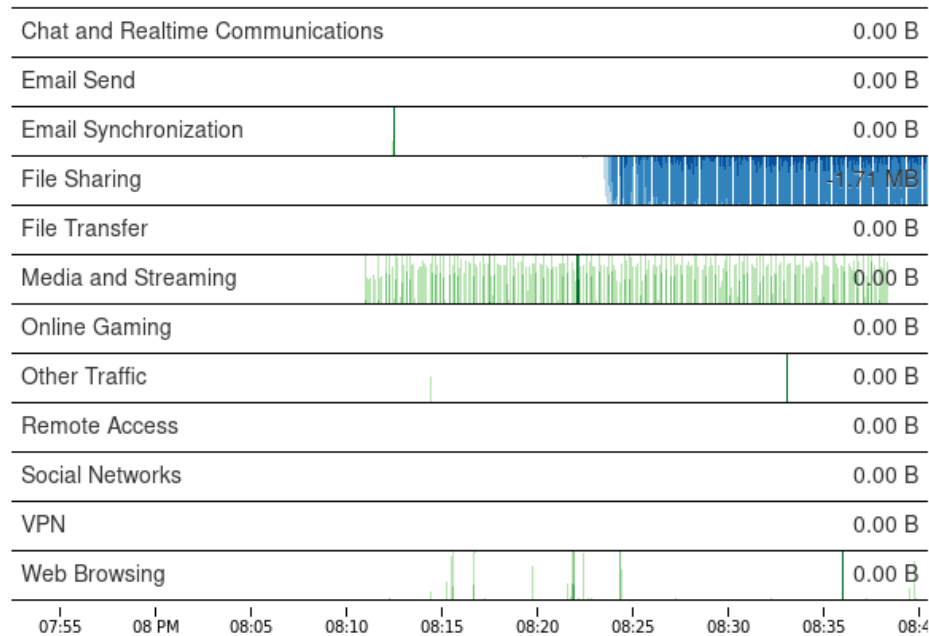
Info	Application	L4 Proto	Client	Server	Duration	Breakdown	Bytes
<a href="#">Info</a>	Unknown	TCP	<a href="#">216.34.181.57:22</a>	<a href="#">192.168.1.92:58356</a>	23 sec	Server	1.12 MB
<a href="#">Info</a>	Unknown	TCP	<a href="#">192.12.193.5:2222</a>	<a href="#">192.168.1.92:61086</a>	23 sec	Client	86.78 KB
<a href="#">Info</a>	SSL	TCP	<a href="#">192.168.1.92:58641</a>	<a href="#">72.233.2.58:443</a>	3 sec	Client	9.79 KB
<a href="#">Info</a>	Unknown	TCP	<a href="#">66.155.11.238:443</a>	<a href="#">192.168.1.92:58607</a>	5 sec	Client	8.83 KB
<a href="#">Info</a>	Google	TCP	<a href="#">192.168.1.92:58638</a>	<a href="#">173.194.35.4:443</a>	1 sec	Client	2.34 KB
<a href="#">Info</a>	Google	TCP	<a href="#">192.168.1.92:58636</a>	<a href="#">173.194.35.4:443</a>	2 sec	Client	2.15 KB
<a href="#">Info</a>	Google	TCP	<a href="#">192.168.1.92:58409</a>	<a href="#">173.194.35.6:443</a>	2 sec	Client	633
<a href="#">Info</a>	Unknown	TCP	<a href="#">2.225.48.185:22515</a>	<a href="#">192.168.1.92:60969</a>	14 sec	Client	612
<a href="#">Info</a>	DropBox	UDP	<a href="#">192.168.1.92:17500</a>	<a href="#">Broadcast:17500</a>	1 sec	Client	516
<a href="#">Info</a>	DropBox	UDP	<a href="#">192.168.1.92:17500</a>	<a href="#">192.168.1.255:17500</a>	1 sec	Client	516

# Obiettivo: macro-attività

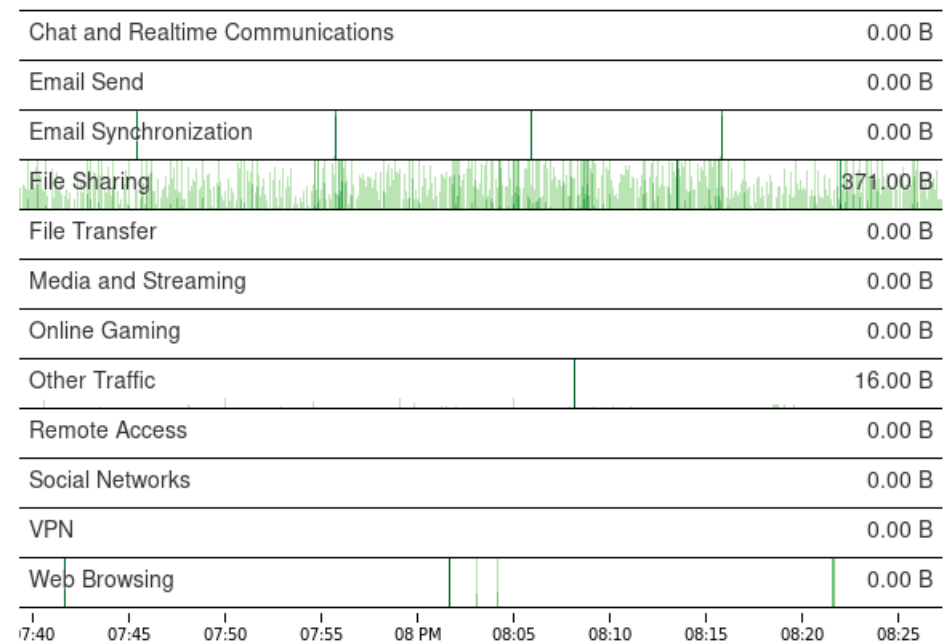
- Sintesi del traffico
- Caratterizzazione del traffico



☒ User Traffic  
☐ Background Traffic



☐ User Traffic  
☒ Background Traffic



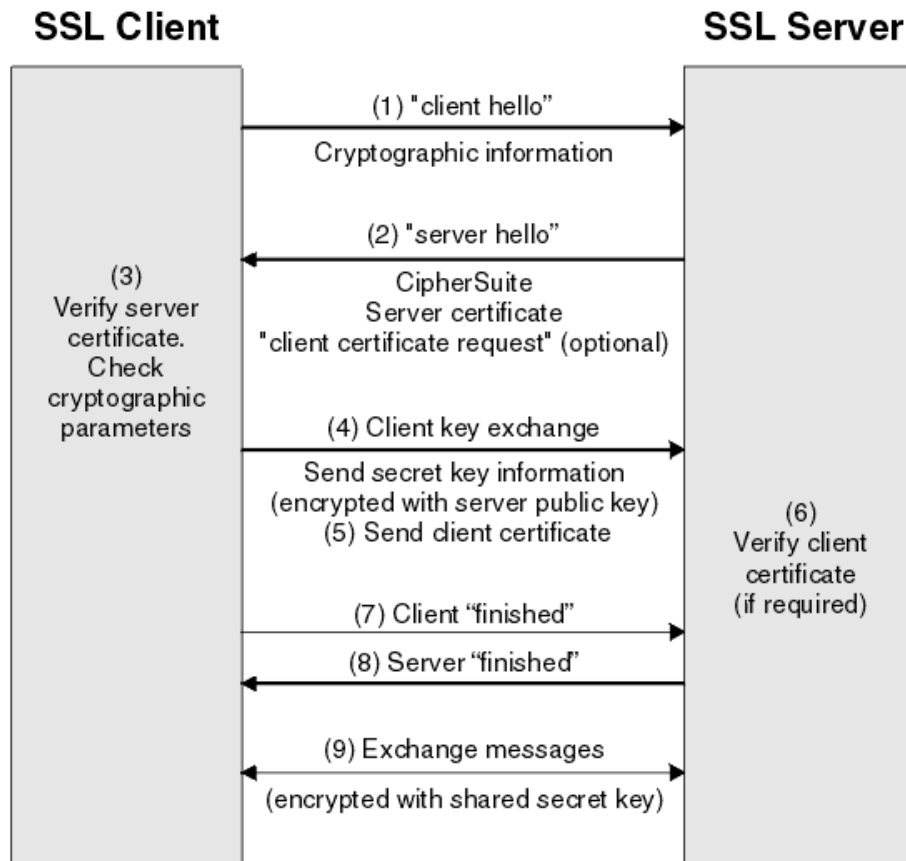
# Problema: cifratura del traffico

- **50% del traffico web è cifrato**
- **Netflix, Youtube, Google Play, Facebook**



- **Difficoltà di caratterizzazione delle attività**
- **Difficoltà di classificazione dei flussi**

# SSL e DPI



## In chiaro

- **Certificati x509**
- **Cypher Suite usati**
- **Porte, indirizzi, SNI**

## Cifrati

- **Payload applicativo**

# Metodologie e strumenti esistenti

- **Non sono conosciuti software opensource per la caratterizzazione del traffico cifrato**
- **In Letteratura non sono conosciute metodologie per la caratterizzazione del traffico cifrato**
- **Le metodologie descritte in Letteratura sono invece orientate all'identificazione dei protocolli nei flussi**

# Idea per caratterizzare il traffico

## **Metriche di rete comuni**

- **Dimensione payload**
- **IAT - Inter packet Arrival Time**
- **Direzione dei pacchetti**

## **Caratteristiche peculiari**

- **Funziona anche su traffico cifrato**
- **Difficilmente camuffabili**

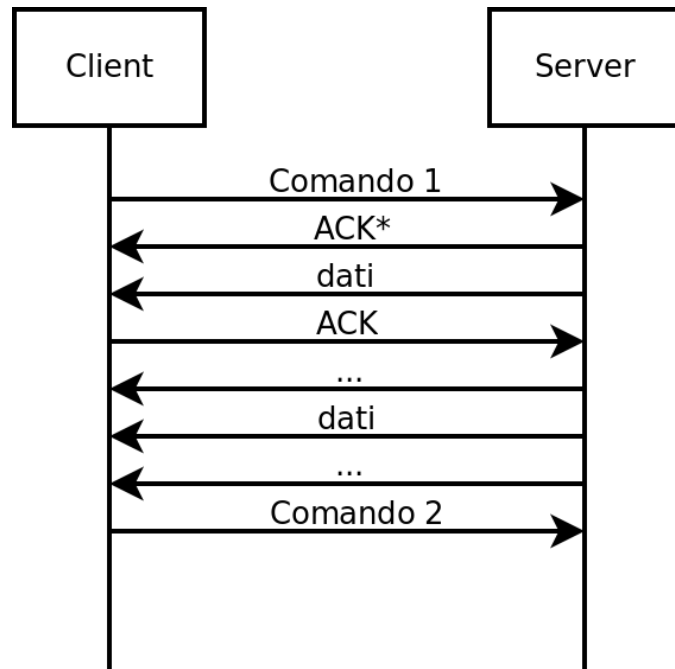
# Proposta: Catalogazione del traffico

Macro-Attività	Traffico attivo	Rumore
Condivisione file	P2P attivo	Collegamento ai peer
Controllo remoto	Invio comandi / ricezione	Keepalive periodico
Giochi online	Gioco	Servizi di supporto
Invio email	Spedizione di email	-
Messaggistica	IM, VoIP, videoconferenza	Traffico di controllo
Multimediale	Audio/video, streaming	Traffico di controllo
Navigazione web	Web browser	Servizi accessori
Sincronizzazione email	Ricezione e sync	Sync periodico
Social Network	Utilizzo portale	Integrazione in siti
Trasferimento file	Scaricamento o sync file	Traffico di controllo
VPN	Traffico offuscato	Keepalive periodico
Altro	Traffico rilevante	Traffico non rilevante



# Caratterizzare la ricezione di email

## Rumore: sincronizzazione automatica delle email

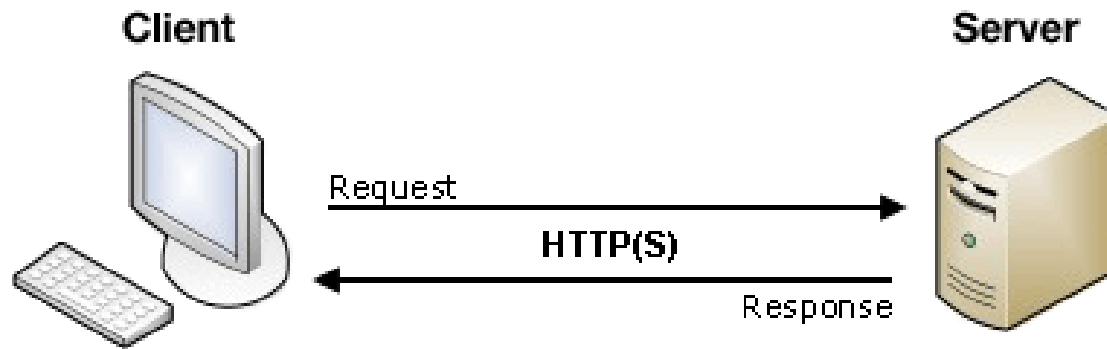


### Caratterizzazione:

- **Invio ACK\* senza dati**
- **Numero pacchetti di risposta**
- **Totale comandi inviati**

# Caratterizzare la navigazione web

**Rumore: servizi che si appoggiano su HTTPS**



## Caratterizzazione:

- Limitare lo IAT massimo
- Numero minimo di pacchetti scambiati
- Direzione prevalente: da client a server

# Caratterizzare Facebook e Twitter

## Rumore: integrazione in siti web

Flusso	Ultimo pacchetto	Totale bytes
[flusso 1]	[timestamp 1]	[bytes flusso 1]
...	...	...
...	...	...

## Caratterizzazione inter-flusso:

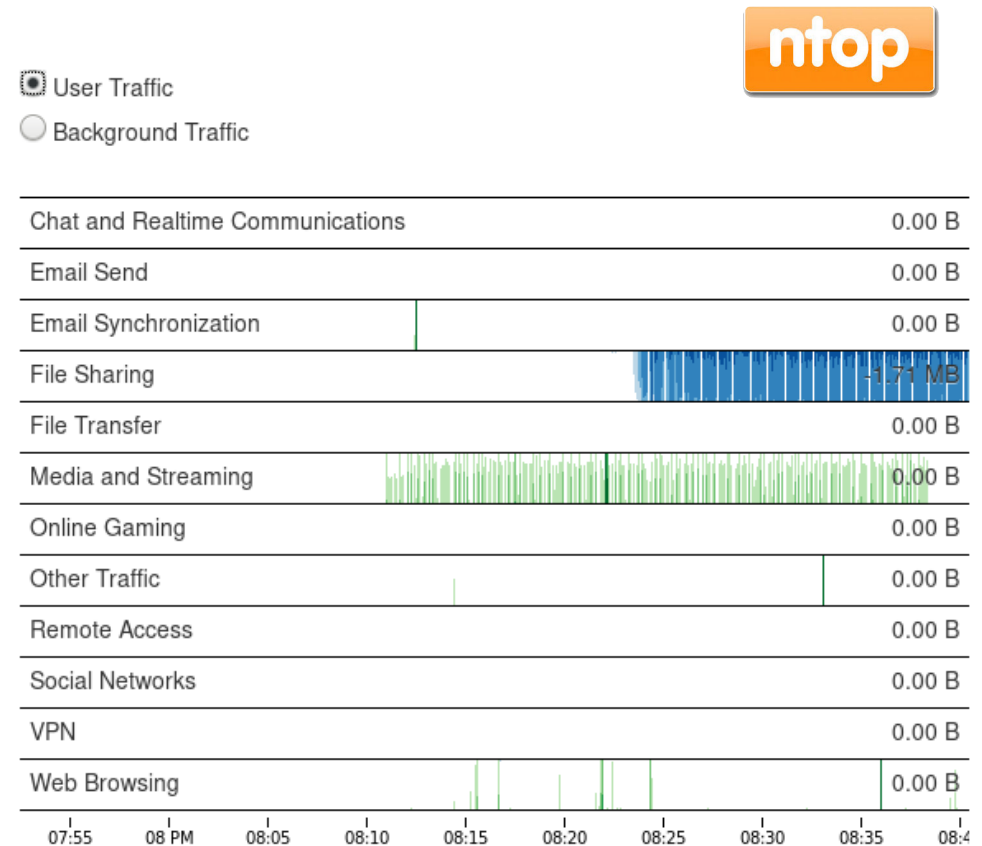
- Flussi attivi contemporaneamente
- Numero flussi attivi
- Numero totale di bytes

# Contributi Originali

- **Utilizzo delle metriche di rete in un ambito ancora inesplorato**
- **Implementazione delle metodologie proposte in un prodotto opensource**
- **Eliminazione del rumore dal traffico di rete**

# Applicazione e lavoro futuro

- **Comprimere il traffico di rete**
- **Visualizzare in maniera chiara l'attività di un host**
- **Rilevamento anomalie tramite regole specifiche**



# Performance e Test

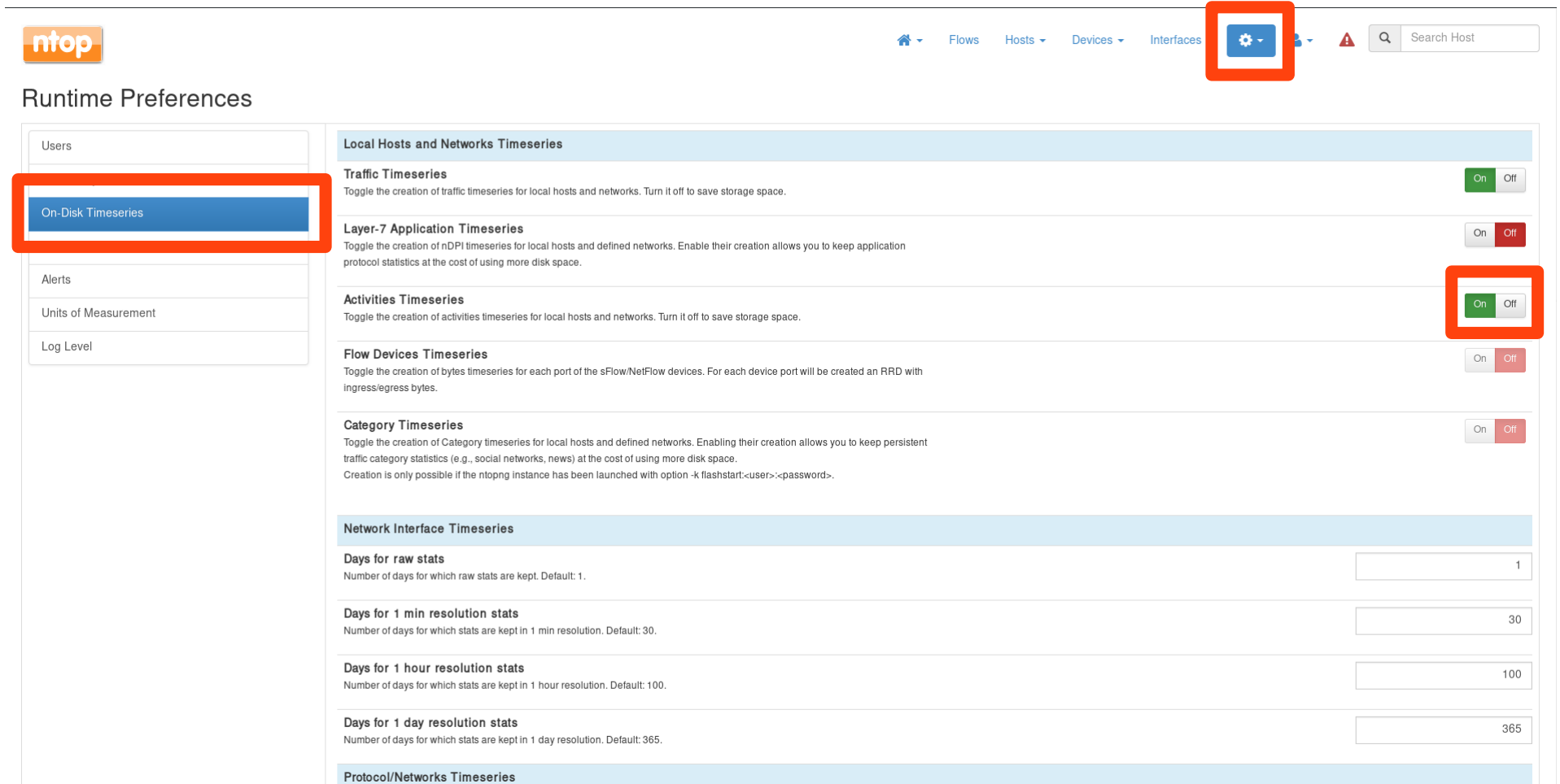
## Test

- **Thunderbird: GMail e Hotmail**
- **Firefox: multimediale e Facebook**
- **Android: IMAPS su HTTPS**

## Performance:

- **rdtsc da 100 a 500 cicli di clock / pacchetto**
- **600 B / host locale, 120 B / flusso**

# Abilitare analisi attività



The screenshot shows the ntopng web interface. In the top right, there is a navigation bar with a home icon, 'Flows', 'Hosts', 'Devices', 'Interfaces', a settings gear icon (highlighted with a red box), a user icon, a warning icon, and a search bar labeled 'Search Host'. On the left, the 'Runtime Preferences' sidebar is visible, with 'On-Disk Timeseries' highlighted (also with a red box). The main content area lists various timeseries settings:

- Local Hosts and Networks Timeseries**
  - Traffic Timeseries**: Toggle set to 'Off'. Description: Toggle the creation of traffic timeseries for local hosts and networks. Turn it off to save storage space.
  - Layer-7 Application Timeseries**: Toggle set to 'Off'. Description: Toggle the creation of nDPI timeseries for local hosts and defined networks. Enable their creation allows you to keep application protocol statistics at the cost of using more disk space.
  - Activities Timeseries**: Toggle set to 'On' (highlighted with a red box). Description: Toggle the creation of activities timeseries for local hosts and networks. Turn it off to save storage space.
  - Flow Devices Timeseries**: Toggle set to 'Off'. Description: Toggle the creation of bytes timeseries for each port of the sFlow/NetFlow devices. For each device port will be created an RRD with ingress/egress bytes.
  - Category Timeseries**: Toggle set to 'Off'. Description: Toggle the creation of Category timeseries for local hosts and defined networks. Enabling their creation allows you to keep persistent traffic category statistics (e.g., social networks, news) at the cost of using more disk space. Creation is only possible if the ntopng instance has been launched with option `-k flashstart-<user>:<password>`.
- Network Interface Timeseries**
  - Days for raw stats**: Input field set to 1. Description: Number of days for which raw stats are kept. Default: 1.
  - Days for 1 min resolution stats**: Input field set to 30. Description: Number of days for which stats are kept in 1 min resolution. Default: 30.
  - Days for 1 hour resolution stats**: Input field set to 100. Description: Number of days for which stats are kept in 1 hour resolution. Default: 100.
  - Days for 1 day resolution stats**: Input field set to 365. Description: Number of days for which stats are kept in 1 day resolution. Default: 365.
- Protocol/Networks Timeseries**

**ntopng --enable-flow-activity**

**Grazie per l'attenzione**