Rilevazione di anomalie di rete mediante analisi su serie temporali

Candidato:

Salvatore Costantino

Relatore:

Dott. Luca Deri

Obiettivi

- Realizzazione di un sistema automatico di rilevazione di anomalie di rete su serie temporali
- Implementazione di tecniche per la rilevazione di anomalie, non basate esclusivamente su soglie fisse (difficili da definire a priori da un utente, e non sempre generalizzabili per tutti i tipi di metriche)
- Implementazione di una tecnica di mitigazione del traffico degli host anomali
- Efficienza in spazio ed in tempo, in modo da poter analizzare gli host di un'intera sottorete
- Buona performance in termini di precisione, specificità e sensibilità

Metriche di Rete

- **Metriche a breve termine** (analizzate dal sistema ogni cinque minuti), ovvero coppie di metriche il cui rapporto in situazioni normali si mantiene approssimativamente costante nel tempo e non supera alcuni valori soglia

- **Metriche a medio-lungo termine** (analizzate dal sistema ogni ora), ovvero singole metriche aventi solitamente trend e stagionalità

Metriche a Breve Termine

Data la coppia di contatori (x, y) relativi alle metriche $(M_{x,}\ M_y)$ da analizzare, consideriamo

$$\frac{\Delta x}{\Delta y} = \frac{x_{t_f} - x_{t_i}}{y_{t_f} - y_{t_i}} = r_{t_i}, \quad \Delta r = r_{t_f} - r_{t_i}$$

- In totale sono state analizzate 14 metriche a breve termine tra cui:
  - $\frac{risposte\ DNS\ ricevute}{richieste\ DNS\ inviate + risposte\ DNS\ ricevute}$
  - $\frac{bytes\ protocollo\ DNS\ ricevuti}{pacchetti\ DNS\ ricevuti}$
  - $\frac{flussi\ sospetti\ come\ client}{flussi\ totali\ come\ client}$

Metriche a Medio-lungo Termine

- Bytes inviati, Bytes ricevuti, flussi come client, flussi come server

- Ulteriore controllo sulle seguenti categorie di protocolli/eventi:
  - Protocolli di accesso remoto
  - Protocolli sconosciuti
  - Malware
  - Mining

Tecniche di Rilevazioni di Anomalie

In questo lavoro di Tesi sono state considerate le seguenti tecniche/algoritmi:

- **Threshold**, per la rilevazione di valori anomali delle metriche a breve termine
- **RSI**, per la rilevazione di variazioni anomale dei valori delle metriche a breve termine
- **Prophet**, per la rilevazioni di comportamenti delle metriche a lungo termine non in linea con la loro storia passata (in termini di trend e stagionalità)

Threshold

Soglie utilizzate:

- Soglia fissata a **576** per le metriche relative alla dimensione media dei pacchetti DNS inviati e ricevuti
  - Data exfiltration/infiltration
- Soglia fissata a **0.50** per le altre metriche a breve termine
  - Il valore del rapporto risulta anomalo

RSI

- Indicatore statistico utilizzato per effettuare analisi su mercati finanziari, in grado di rilevare la velocità del movimento dei prezzi
- **Idea chiave**: utilizziamolo per misurare la velocità con cui variano i valori legati alle metriche di rete
- Periodo lungo **50**: numero di punti della metrica che contribuiscono al calcolo dell'RSI

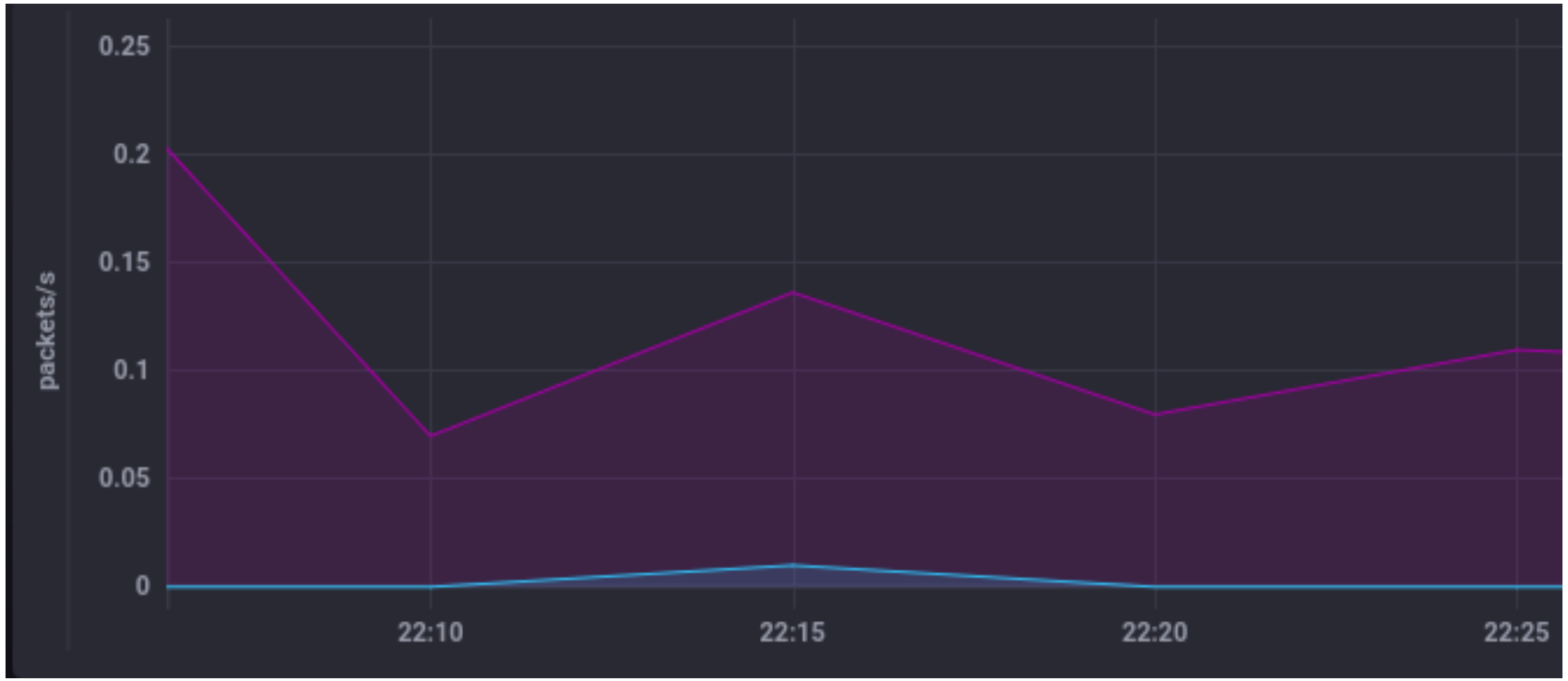
$$\boldsymbol{RSI = 100 * U/(U + D)}$$

  - **U**: media delle differenze positive tra punti consecutivi nel periodo fissato
  - **D**: media delle differenze negative tra punti consecutivi nel periodo fissato

- Oscilla tra due valori: 0 e 100
- Una metrica viene classificata anomala se il relativo valore dell'RSI risulta maggiore di **80**

Soglia RSI: 80 vs 70

- 70 è la soglia superiore utilizzata nell'algoritmo originale, ma essa genera vari falsi positivi
- Falso positivo rilevato di tipo «dns_errors» (valore rapporto circa 0.07)



Prophet

- Modello di **regressione** che assume la seguente forma:
  - $\boldsymbol{y(t) = g(t) * s(t) + h(t) + \varepsilon_t}$ (*modello moltiplicativo*)

- Consente di apprendere il comportamento di una serie temporale, e di predire nuovi punti in base alla storia passata

- Scelta **iper-parametri** (tramite model selection) che determinano la capacità del modello di predire nuovi punti della serie temporale

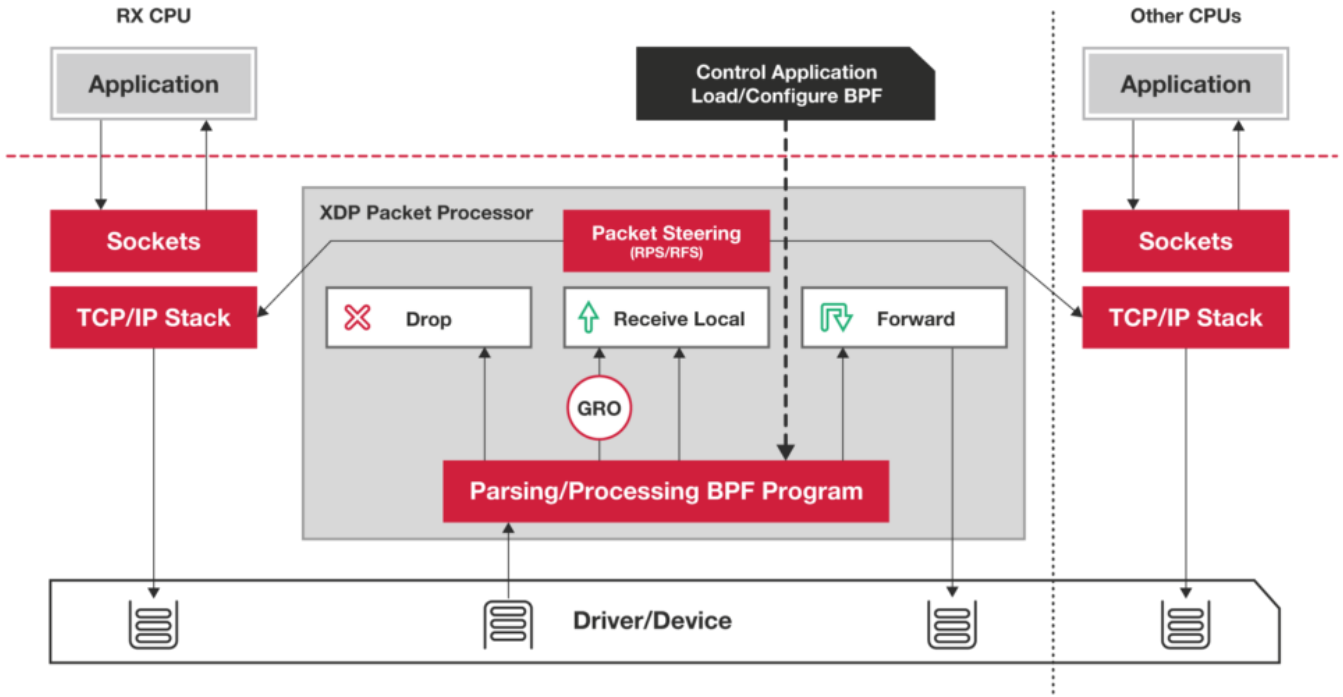
- Tempo impiegato per training e predizione: circa 9 secondi

Esempio di allarmi sulle metriche analizzate

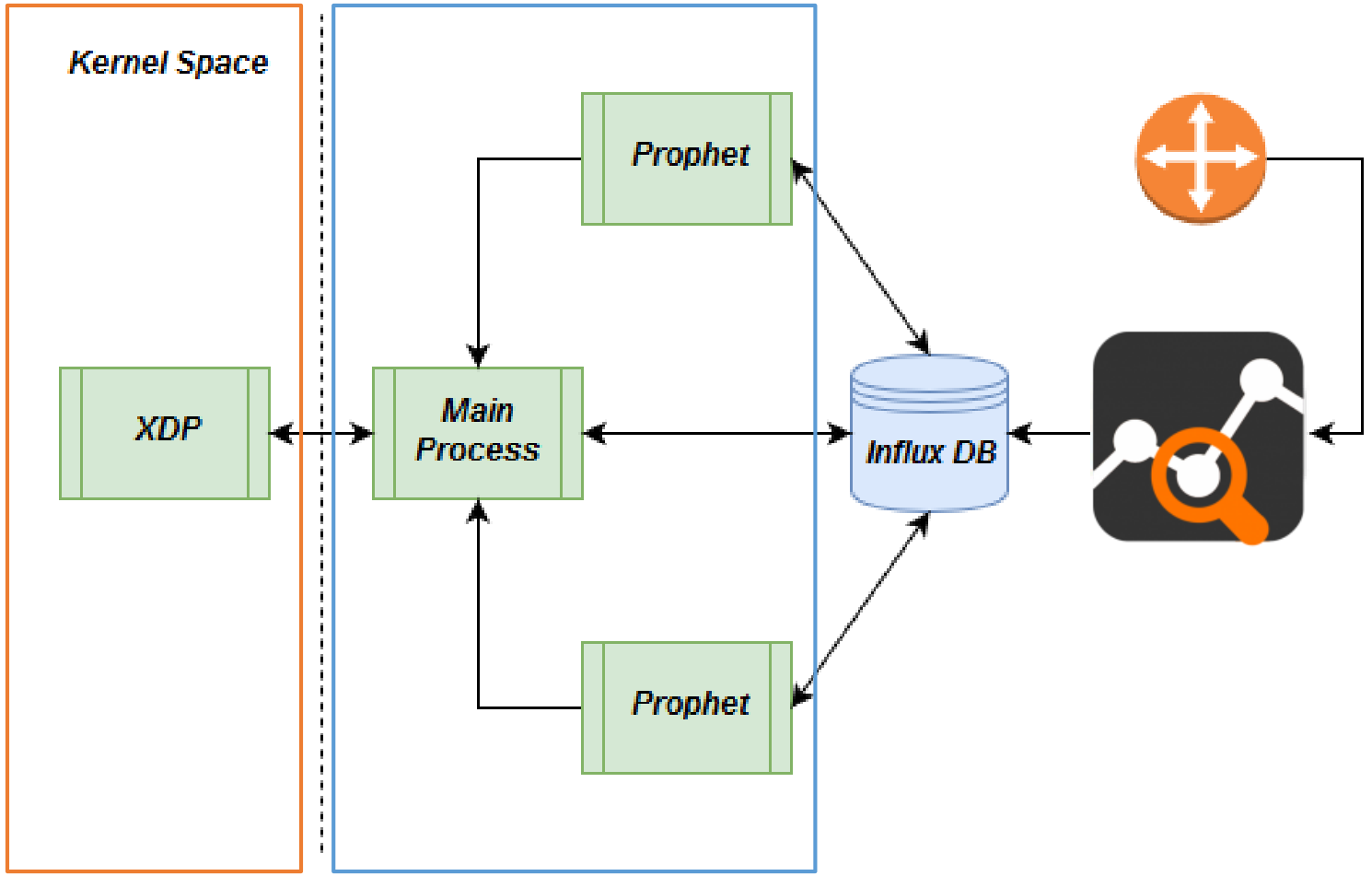
TYPE	ANOMALY	HOST/MAC	IF	DATE	METHOD	VAL
START	dns_errors	[redacted].210@125	0	2019-06-27T21:40:00Z	THRESHOLD	0.6
END	dns_errors	[redacted].210@125	0	2019-06-27T22:05:00Z	THRESHOLD	
START	dns_errors	[redacted].210@125	0	2019-06-27T22:10:00Z	THRESHOLD	0.7
END	dns_errors	[redacted].210@125	0	2019-06-27T22:30:00Z	THRESHOLD	
START	dns_errors	[redacted].210@125	0	2019-06-27T22:35:00Z	THRESHOLD	0.5
END	dns_errors	[redacted].210@125	0	2019-06-27T23:00:00Z	THRESHOLD	
START	dns_errors	[redacted].210@125	0	2019-06-27T23:05:00Z	THRESHOLD	0.6
END	dns_errors	[redacted].210@125	0	2019-06-27T23:45:00Z	THRESHOLD	
START	dns_errors	[redacted].210@125	0	2019-06-27T23:50:00Z	THRESHOLD	0.6
START	dns_errors	[redacted].214@125	0	2019-06-27T16:50:00Z	RSI	85.6
END	dns_errors	[redacted].214@125	0	2019-06-27T16:55:00Z	RSI	
START	dns_errors	[redacted].214@125	0	2019-06-27T22:00:00Z	RSI	92.7
END	dns_errors	[redacted].214@125	0	2019-06-27T22:10:00Z	RSI	
START	dns_errors	[redacted].222@125	0	2019-06-27T12:45:00Z	THRESHOLD	0.6
END	dns_errors	[redacted].222@125	0	2019-06-27T12:50:00Z	THRESHOLD	

Mitigazione Traffico: XDP

- Analisi dei pacchetti direttamente all'interno del Kernel Linux, grazie alla tecnologia **eBPF** (Extended Berkeley Packet Filter)
  - Blocco totale del traffico degli host anomali, per sospetta attività d'attacco



Architettura software



Validazione

- Validazione effettuata sugli host di un ISP locale avente le seguenti caratteristiche:
  - 256k host totali
  - 16k host attivi
  - Velocità media di rete pari a 600 Mbit/s
- Metriche a breve termine analizzate per circa due giorni
- Metriche a medio-lungo termine analizzate per circa tre settimane

Anomalie Rilevate

Metriche a breve termine

Metriche a medio –lungo termine

TYPE	TOTAL_CHECK	ANOMALIES	METHOD
ping_packets	2602	1684	THRESHOLD
dns_packets	35406	33420	THRESHOLD
dns_errors	35299	4586	THRESHOLD
port_unreach_srv	4237	33	THRESHOLD
port_unreach_clt	9355	8	THRESHOLD
host_unreach_clt	9054	480	THRESHOLD
host_unreach_srv	1891	0	THRESHOLD
TCP_client_iss	1133	377	THRESHOLD
TCP_server_iss	1119	19	THRESHOLD
dns_size_srv	474	0	THRESHOLD
dns_size_clt	444	0	THRESHOLD
anmls_flows_srv	29644	949	THRESHOLD
anmls_flows_clt	33714	2040	THRESHOLD
dns_errors	14666	15	RSI
port_unreach_srv	2555	9	RSI
port_unreach_clt	5285	13	RSI
host_unreach_clt	4273	12	RSI
host_unreach_srv	1191	4	RSI
TCP_client_iss	186	0	RSI
TCP_server_iss	573	0	RSI
dns_size_srv	295	0	RSI
dns_size_clt	265	0	RSI
anmls_flows_srv	13461	0	RSI
anmls_flows_clt	14844	0	RSI
flows_as_client	28	2	PROPHET
flows_as_server	199	0	PROPHET
bytes_sent	28	0	PROPHET
bytes_rcvd	199	1	PROPHET

Valutazione Threshold

- Sono stati etichettati manualmente 25 host: 15 host risultati anomali e 10 host risultati non anomali

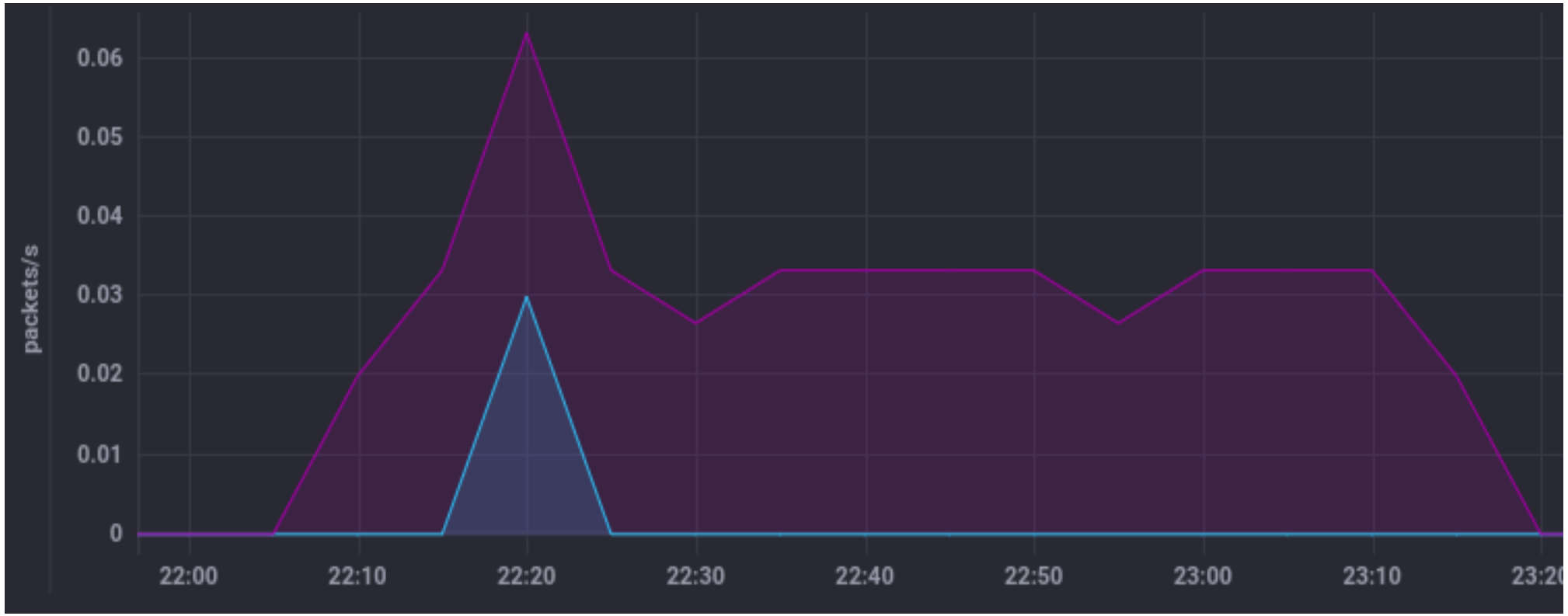
$$PRECISIONE = \frac{TP}{TP + FP} = \frac{13}{13 + 2} = 87\%$$

$$SPECIFICITA' = \frac{TN}{TN + FP} = \frac{10}{10 + 2} = 83\%$$

$$SENSIBILITA' = \frac{TP}{TP + FN} = \frac{13}{13 + 0} = 100\%$$

Vero Positivo Threshold

- «ping_packets»: presenti risposte ICMP echo (curva viola), senza richieste ICMP echo (curva blu)



Valutazione RSI

- Sono stati etichettati manualmente 20 host: 10 host rilevati anomali, 10 host non rilevati anomali

	Host anomali	***Host non anomali***
Rilevati	**7**	**3**
Non rilevati	**2**	**8**

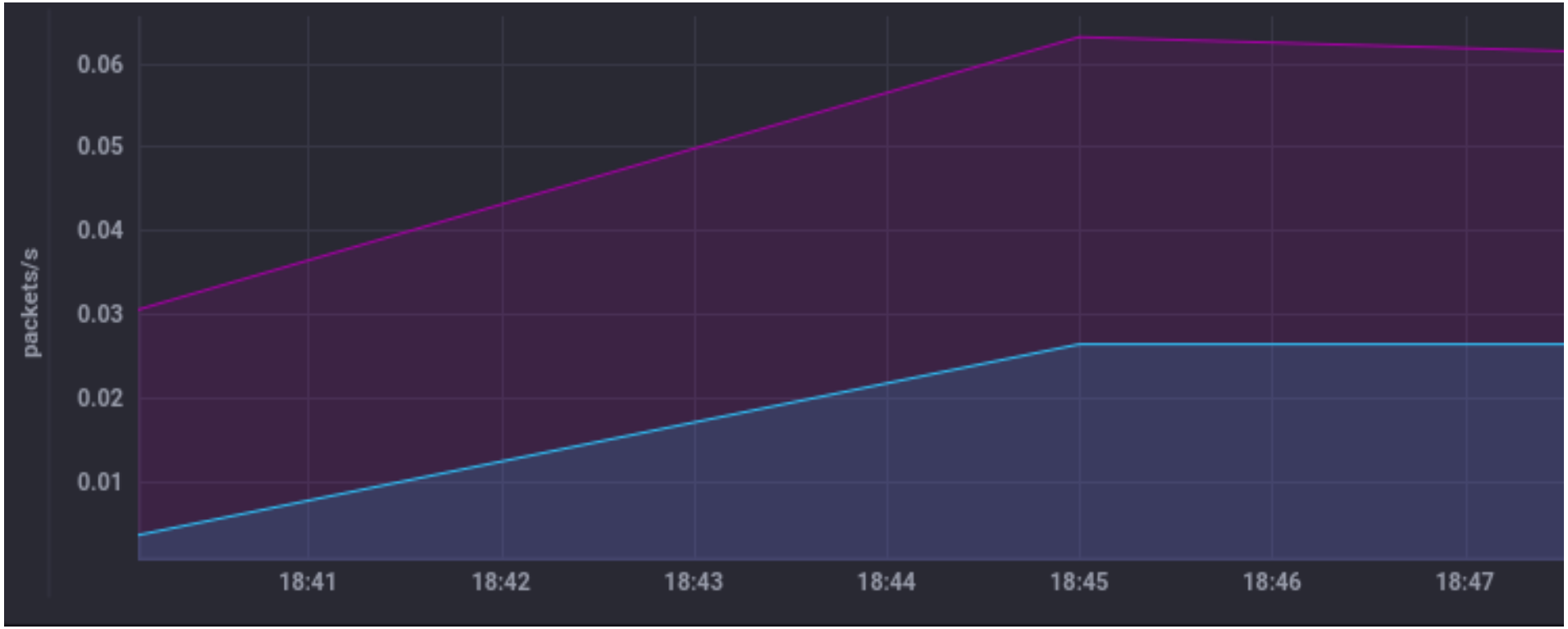
$$PRECISIONE = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} = \frac{\boldsymbol{7}}{\boldsymbol{7} + \boldsymbol{3}} = 70\%$$

$$SPECIFICITA' = \frac{\boldsymbol{TN}}{\boldsymbol{TN} + \boldsymbol{FP}} = \frac{\boldsymbol{8}}{\boldsymbol{8} + \boldsymbol{3}} = 73\%$$

$$SENSIBILITA' = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} = \frac{\boldsymbol{7}}{\boldsymbol{7} + \boldsymbol{2}} = 78\%$$

Vero Positivo RSI

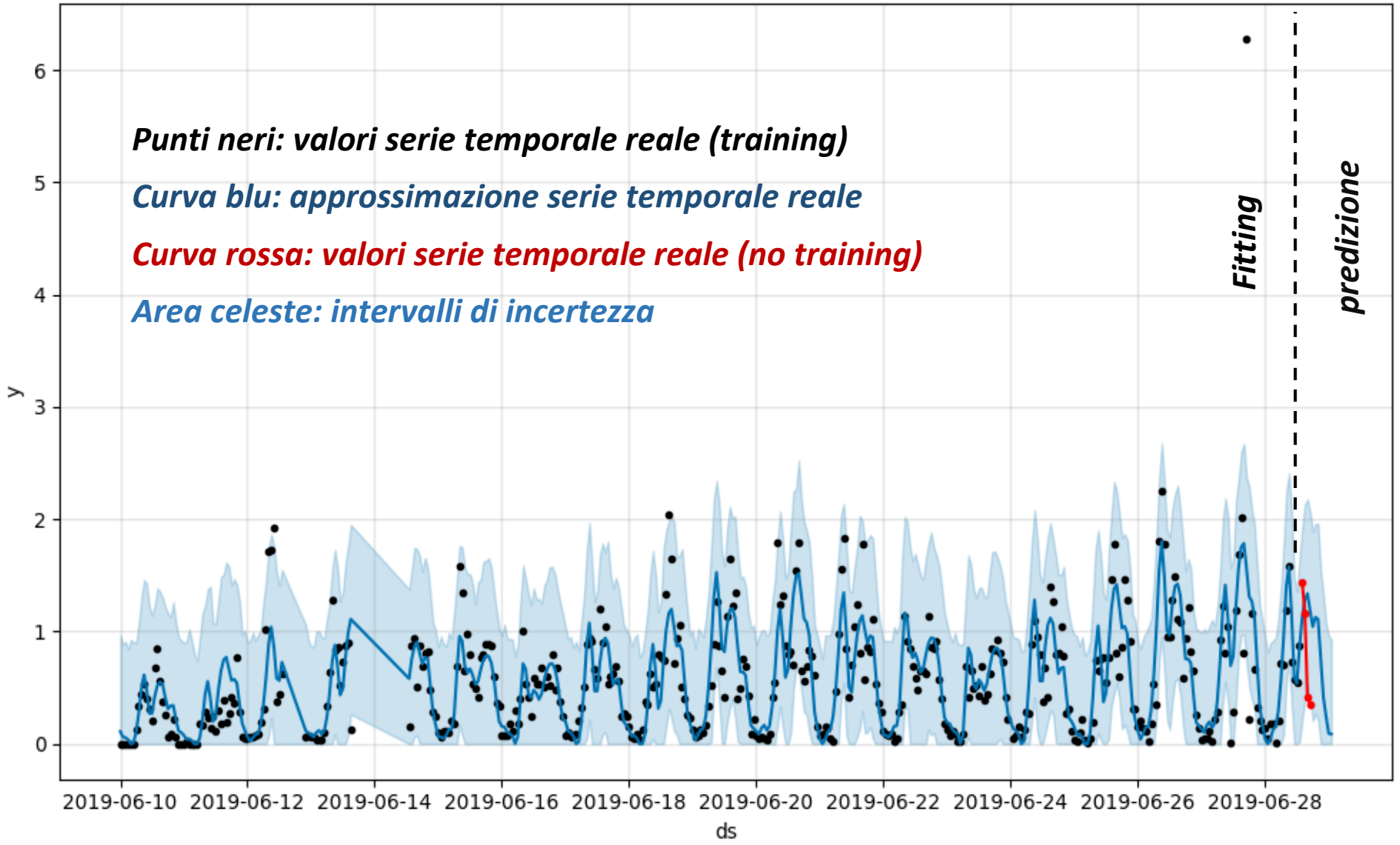
- «dns_errors»: il valore del rapporto risulta anomalo, e registra un incremento di circa 0.50 rispetto ai 5 minuti precedenti



Valutazione Prophet

- 3 anomalie rilevate, su oltre 400 controlli
- Tutte e 3 sono falsi positivi, riconducibili ad un cambiamento di comportamento fisiologico degli host analizzati
- Specificità prossima al 100%
- I 3 falsi positivi non vengono rilevati, se viene attivato il controllo delle categorie di protocolli

Falso positivo Prophet



Risultati finali

		Host anomali	*Host non anomali*
Threshold	*Rilevati*	13	2
	Non rilevati	0	10
RSI	*Rilevati*	7	3
	Non rilevati	2	8
Prophet + DPI	*Rilevati*	0	0
	Non rilevati	0	454
Totale	*Rilevati*	20	5
	Non rilevati	2	472

$$PRECISIONE = \frac{TP}{TP + FP} = \frac{20}{20 + 5} = 80\%$$

$$SPECIFICITA' = \frac{TN}{TN + FP} = \frac{472}{472 + 5} = 99\%$$

$$SENSIBILITA' = \frac{TP}{TP + FN} = \frac{20}{20 + 2} = 91\%$$

Lavori Futuri

- Correlazione tra serie temporali di host diversi, in modo da generare allarmi più significativi
- Allarmi come input ad un livello di analisi superiore, per esempio un autoencoder
- Miglioramento tecnica di mitigazione, che appare troppo drastica e non in grado di proteggere un eventuale host sotto possibile attacco

Contributo originale

- Uso di algoritmi e tecniche non attualmente presenti nello stato dell'arte relativo alla rilevazione di anomalie di rete:
  - **Indicatore statistico finanziario**, nel breve termine
  - **Modello di predizione** su serie temporale, nel medio-lungo termine

Conclusione

- Il problema della rilevazione di anomalie non ha, ad oggi, una soluzione semplice e universale, e ogni tecnica presenta i suoi punti di forza e debolezza
- In questo lavoro di Tesi vengono analizzati host generici, su cui non è possibile effettuare alcuna assunzione sul tipo di traffico generato (come invece accade in ambiente IoT (Internet of Things)): il problema è risultato particolarmente complesso da affrontare
- Si è realizzato un sistema intelligente, capace di analizzare e mitigare alcune anomalie di rete presenti in un'intera rete, in modo efficiente e con buona precisione