

UNEVENLY SPACED TIME SERIES FROM NETWORK TRAFFIC

Josef Koumar¹, Tomáš Čejka²

¹Czech Technical University in Prague, Czech Republic ² CESNET, a.l.e.

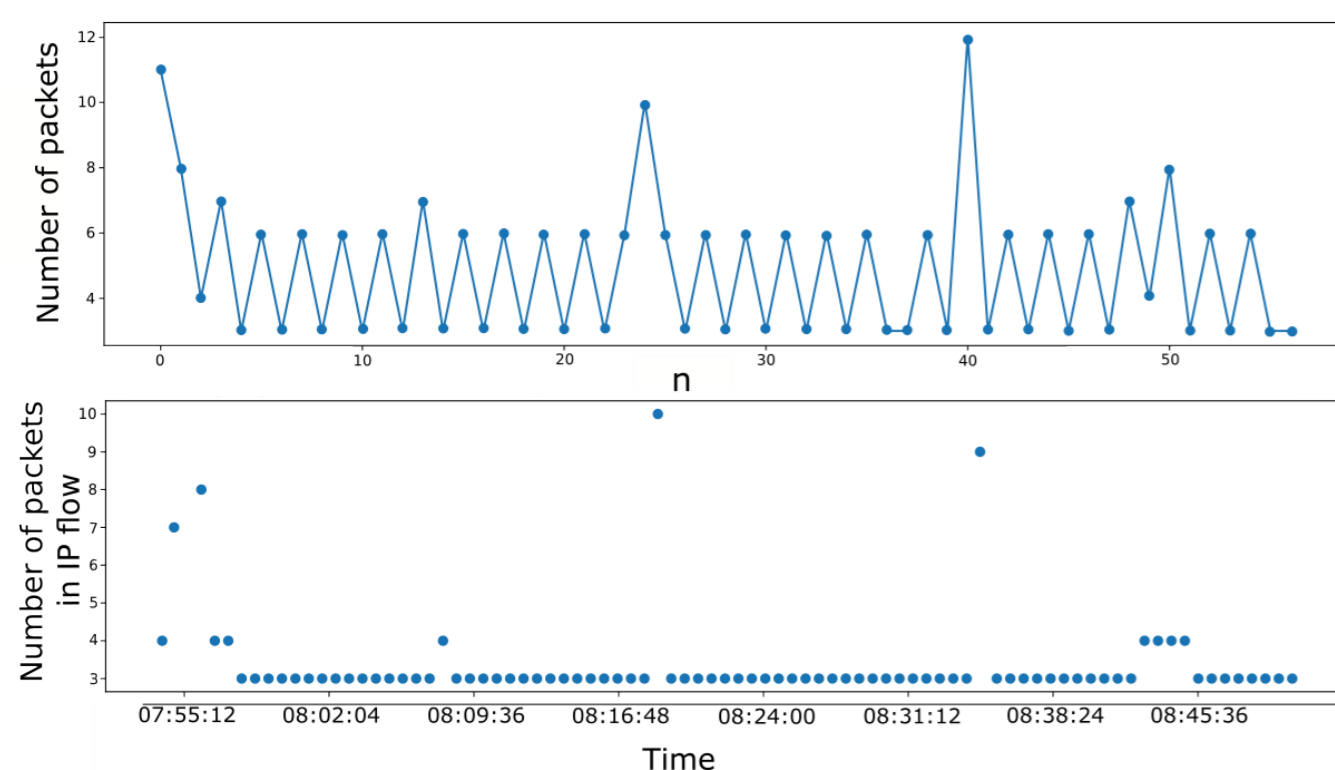


MOTIVATION

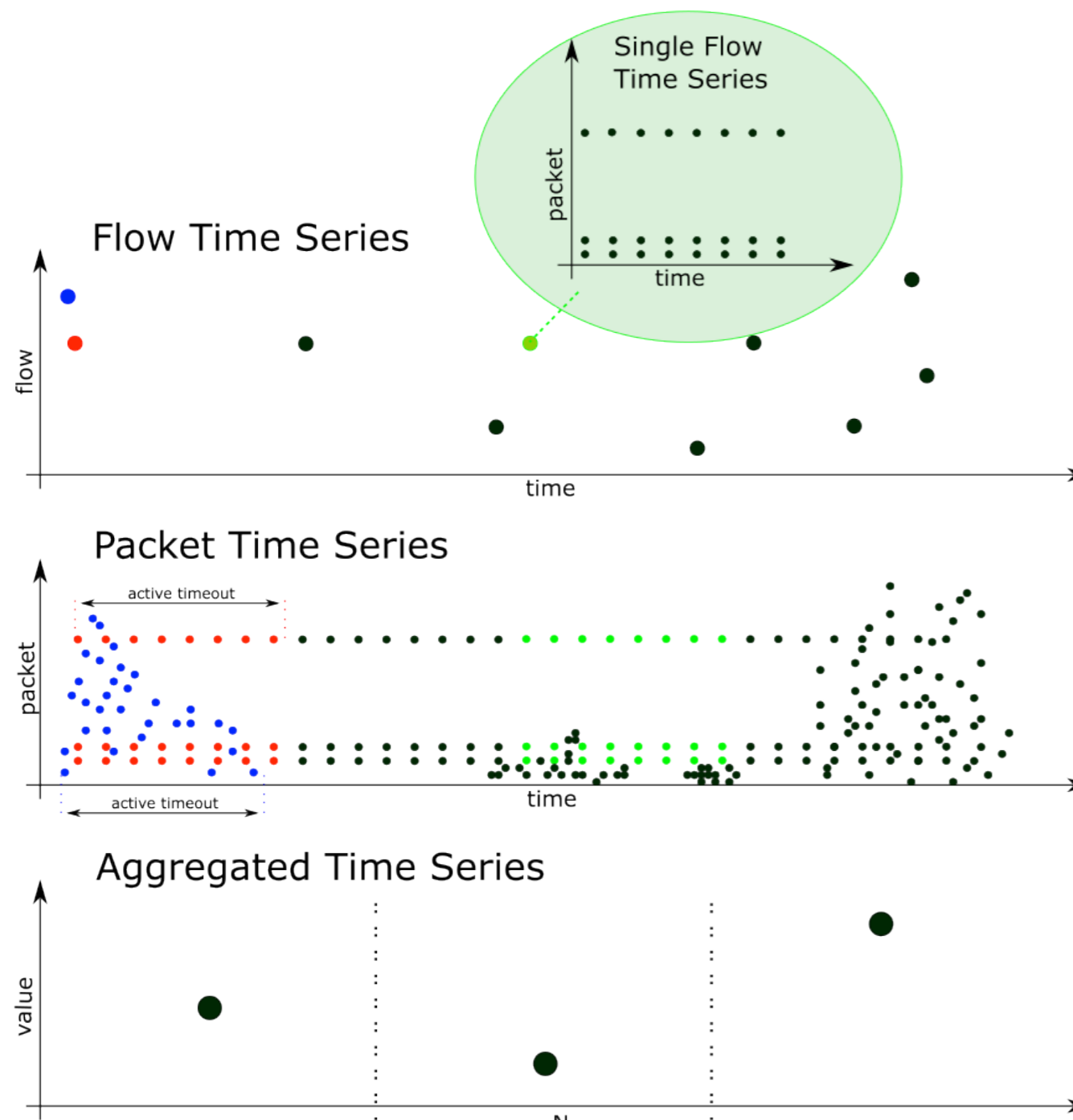
Reliable detection of security events is essential for network security. Therefore, a suitable traffic representation and model are required. The currently used approaches aggregate network traffic into time series by sum all packets/flows values in each interval. However, the size of interval is hard to select universally.

TIME SERIES ANALYSIS

Most often, TS are considered with evenly spaced time between observations. This type of TS is called *Evenly spaced time series (ESTS)*, also called regularly sampled or uniformly sampled. They are defined as the sequence of observation $\{X_n\} = \{x_1, \dots, x_n\}$ taken in times $\{T_n\} = \{t_1, \dots, t_n\}$ which satisfy the equation $t_{j+1} - t_j = t_j - t_{j-1}, \forall j \in 2, \dots, n-1$. There are also TS which do not have the times which satisfy $t_{j+1} - t_j = t_j - t_{j-1}$. That means the times are, in general, not regularly spaced, that means, $\delta_j = t_{j+1} - t_j, \forall j \in \{1, \dots, n-1\}$, is not constant. They are called *Unevenly spaced time series (USTS)*, also called unequally spaced, or irregularly sampled.

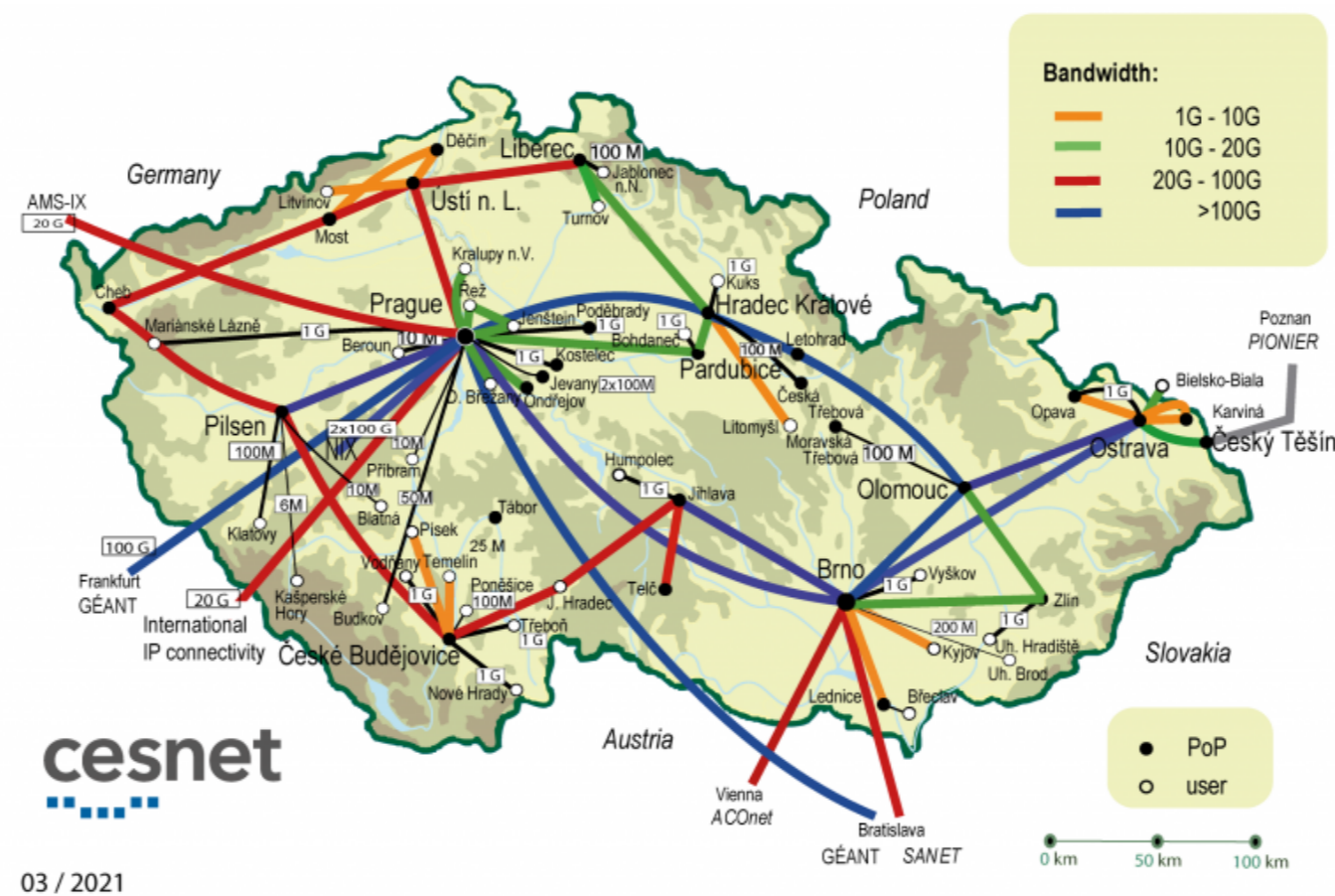


TYPES OF USTS



DATASET CREATION

We have created three datasets for experiments with the USTS from network traffic. The first dataset contains 2,6 million FTS created from 259 million flows, 19 million PTS created from 110 million packets, the second dataset contains, and the third dataset contains 15 million SFTS created from 160 million packets. These datasets were created by traffic capture on the ISP infrastructure of the CESNET2.



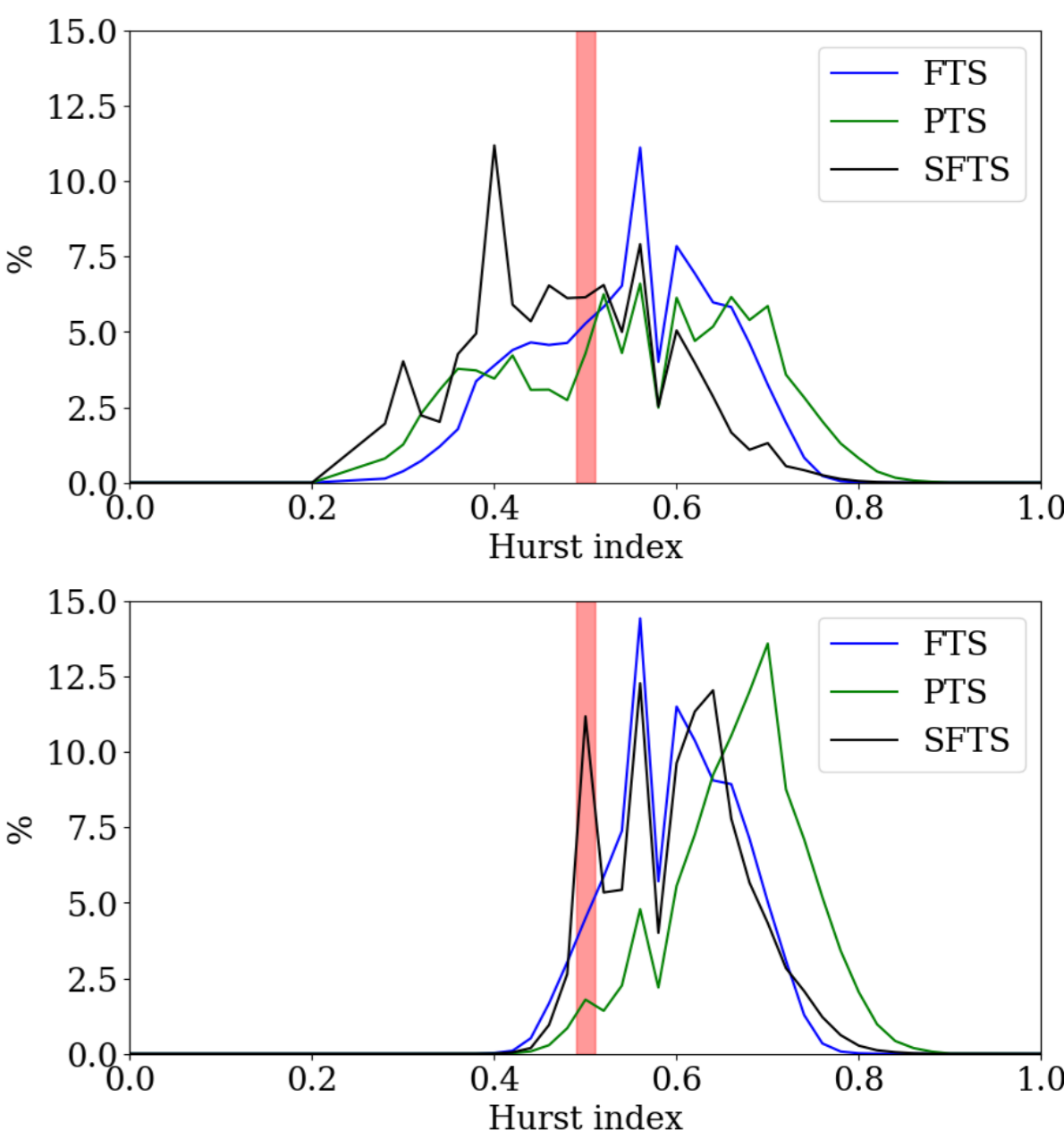
STATIONARITY

The properties of a stationary TS do not depend on the time of observation. The TS with a trend or with seasonality is not stationary, but the TS with periodic (or cyclic) behavior can be stationary. Results of our experiments are below:

	Time [min]	Number of data points					
		all	<25	25-100	100-500	500-1000	≥ 1000
FTS	all	85.8	72.2	91.3	97.3	99.6	99.6
	< 1	67.0	66.3	86.6	85.7	100	NaN
	1-10	56.7	55.6	76.8	93.9	91.6	88.8
	10-60	65.6	62.2	83.2	91.6	97.1	100
PTS	all	83.8	82.5	85.3	90.7	95.3	95.3
	< 1	76.4	74.9	83.5	89.6	92.6	89.0
	1-10	85.6	86.2	83.3	87.8	93.1	94.9
	10-60	95.8	98.2	91.4	93.1	97.7	98.5
SFTS	all	54.9	50.3	83.9	89.0	94.5	97.2
	< 1	45.2	41.0	81.7	87.1	94.7	96.9
	1-2	72.4	71.1	70.5	81.1	92.5	97.5
	2-4	69.1	66.9	70.7	83.7	90.5	97.7
	≥ 4	82.0	78.2	91.4	94.0	95.6	97.5

HURST EXPONENT

We performed tests using the Hurst exponent. If the Hurst exponent $H \in \langle 0; 0.5 \rangle$, then it indicates a long-term switching between high and low values in adjacent pairs and the TS is anti-persistent. If $H \sim 0.5$, then this indicates a random (uncorrelated) TS. Furthermore, if $H \in \langle 0.5; 1 \rangle$, then it indicates a long-term positive autocorrelation in the TS and the TS is persistent.



Application of USTS on Real Traffic

1. Network Traffic Classification based on Periodic Behavior Detection (published, CNSM 2022)
2. Network Traffic Classification based on Single Flow Time Series Analysis (submitted, CNSM 2023)
3. Enhancing DeCrypto: Finding Cryptocurrency Miners based on Periodic Behavior (submitted, CNSM 2023)
4. NetTiSA: Extended IP Flow with Time-series Features for Universal Bandwidth-constrained High-speed Network Traffic Classification (submitted, COMNET)

Conclusion

The results of our experiments show that USTS are feasible for network traffic analysis and exhibits significant advantages over ESTS, which are as follows:

1. TS distribution are not affected by aggregation interval,
2. we know what data points and their values represent,
3. it is not necessary to set aggregation time interval, which is hard to select,
4. there are no zero values (times without data points),
5. they contain minimal noise,
6. they are stationary, so there is no need to perform TSD before TSA, which allows automatic procession, and
7. they usually occur with periodic behavior.

This research was funded by the Ministry of Interior of the Czech Republic, grant No. VJ02010024: Flow-Based Encrypted Traffic Analysis and also by the Grant Agency of the CTU in Prague, grant No. SGS23/207/OHK3/3T/18 funded by the MEYS of the Czech Republic.