# Rozpoznávanie DDoS útokov pomocou metódy ICA

ŽILINSKÁ UNIVERZITA V ŽILINE
Fakulta riadenia a informatiky

Bc. Patrik Císar (cisar6@stud.uniza.sk)
**Vedúci práce**: doc. Mgr. Juraj Smieško, PhD.

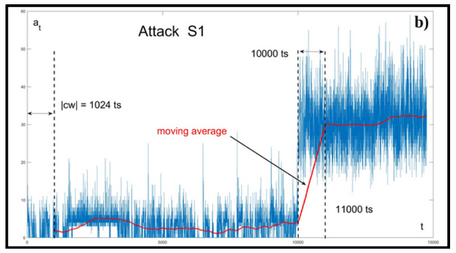
## Cieľ projektu

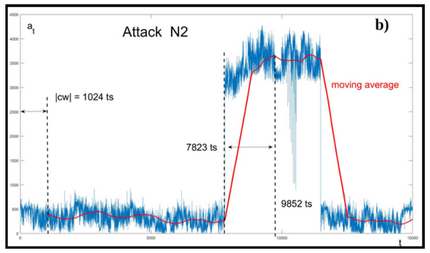
Hlavným zámerom tohto projektu je preskúmať možnosť využitia metódy ICA pri identifikácii DDoS útokov a dosiahnuť to prostredníctvom analýzy a modelovania toku pomocou metódy MMRP.

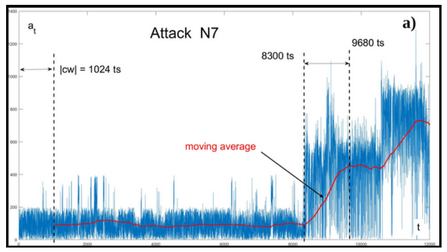
## DDoS útoky

### Distributed Denial of Service

**DDoS** je záplavový tok, ktorý slúži na vyčerpanie zdrojov servera, čo spôsobí odoprenie služby. Na obrázkoch si môžeme všimnúť, že ide o klasické DDoS útoky. Na každom z nich, je skratka "cw". Cw znamená **compute window**, ktorý určuje, po akých veľkých oknách budeme počítať štatistické vlastnosti. "**Moving average**" je priemer prevádzky, v danom časovom okne. Vidíme, že prvá hodnota moving average je práve po skončení prvého prepočítania cw. Následne sa posunie o 1 ts (time slot) a prepočíta sa znovu. Takto prejdeme celú prevádzku. Výsledkom tohto je, že sa nemôžme spoliehať na priemer, pretože nedokáže včas reagovať na útok. V čase útoku sa priemer zvýši len o malú, zanedbateľnú, časť a zvýši sa až príliš neskoro.



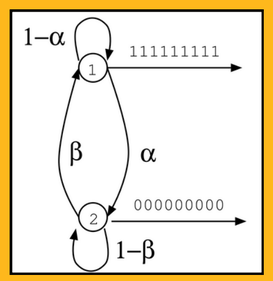




## Modelovanie pomocou MMRP

### (Markov-Modulated Regular Process)

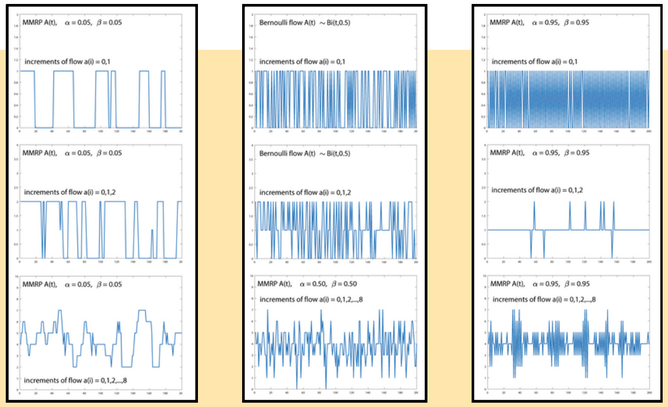
MMRP proces využijeme na **modelovanie** prevádzky. Tento model má dva parametre - **alfa(α)** a **beta(β)**. Tieto pravdepodobnosti určujú, s akou pravdepodobnosťou sa presunieme zo stavu na stav. Napríklad, keď sme v stave 1, tak s pravdepodobnosťou alfa sa presunieme do stavu 2. Keď sme v stave 1, tak generujeme **jednotky**, ak sme v stave 2, tak **nuly**.




Z nameranej prevádzky si vieme **odmerať** priemernú intenzitu, maximálnu hodnotu a pravdepodobnosť na maximálnu hodnotu toku. Tieto štatistické vlastnosti nám stačia na to, aby sme si dokázali vygenerovať MMRP tok.

$$\pi_1 = \frac{\lambda_{avg}}{n} = \frac{\beta}{\alpha+\beta} \quad \Rightarrow \quad \beta = \frac{\lambda_{avg}\cdot\alpha}{n-\lambda_{avg}}$$

$$p_{peak} = \pi_1(1-\alpha)^{n-1} \quad \Rightarrow \quad p_{peak}\cdot\frac{n}{\lambda_{avg}} = (1-\alpha)^{n-1} \quad \Rightarrow \quad \alpha = 1-\left(p_{peak}\cdot\frac{n}{\lambda_{avg}}\right)^{\frac{1}{n-1}}$$




Nastavovaním a hýbaním parametrami alfa a beta dokážeme **simulovať** rôzne toky. V riadkoch sa postupne zvyšuje veľkosť vzorky a v stĺpcoch sa menia parametre.

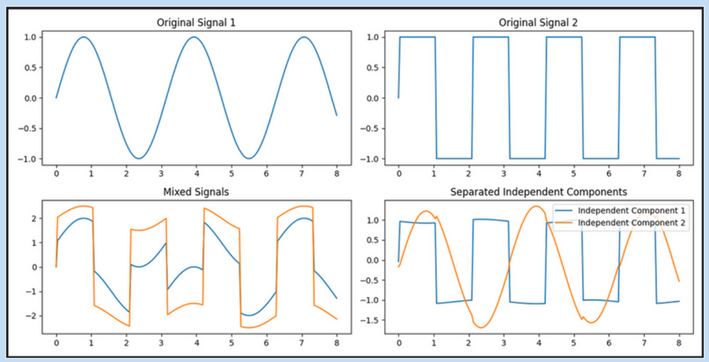
- V prvom stĺpci vidíme, že keď nastavíme oba parametre na nízke hodnoty, tak si môžeme všimnúť že prechod do stavov je **málo frekvenčný**.
- V druhom stĺpci sa simuluje **Bernoulli**-ho tok. Bernoulli tok má len jeden parameter, čo znamená, že je podobný hodu mincou.
- V poslednom stĺpci sú nastavené parametre na vysoké hodnoty, čo spôsobuje **neustále prepínanie** zo stavu na stav.

## ICA

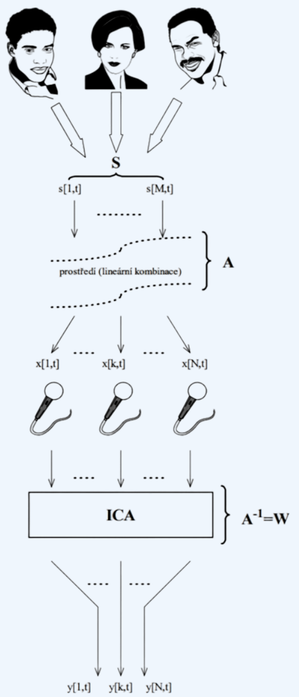
### (Independent Component Analysis)

ICA je štatistická metóda používaná v oblasti signálového spracovania a analýzy dát. Jej cieľom je **rozložiť** komplexný signál na niekoľko **nezávislých** zložiek, nazývaných nezávislé komponenty. Princíp ICA spočíva v tom, že sa snaží nájsť transformačné matice, ktoré zobrazia pôvodný signál na nový priestor, kde sú komponenty čo najviac nezávislé. Týmto spôsobom je možné oddeliť zmiešané signály na ich základné zložky. Medzi najpopulárnejšie algoritmy patria: **FastICA**, Infomax, JADE,...

Na obrázku vidíme v prvom riadku **vstupné signály**. Prvý je sínusový, druhý obdĺžnikový. Následne tieto dva signály **zmiešame** tak, aby nebolo z nich vedieť, čo reálne predtým predstavovali (graf vľavo dole). Ako posledné sme použili **FastICA** algoritmus, ktorý tieto signály **oddelil**. Môžeme si všimnúť, že to nie je úplne ideálne, ale veľmi sa podobajú na pôvodné signály.




Ďalším typickým príkladom pre algoritmus ICA je takzvaná "**Cocktail Party**". Predstavte si, že na párty sa zúčastňuje viacero ľudí a každý hovorí s niekým iným. Zvukové signály z týchto rôznych rozhovorov sa miešajú a vytvárajú komplexný signál. Cieľom ICA je **oddeliť** tieto zvukové zložky tak, aby sme mohli počuť jednotlivé rozhovory nezávisle od seba.




Zoznam literatúry je k dispozícií u autora