PATTERNIZACE LOGŮ VYSOKÉ ŠKOLY EKONOMICKÉ

Automatizovaná analýza bezpečnostních dat | 48× rychlejší | 2200+ zdrojů | 96% přesnost

Barbora Šandová
Vedoucí: Ing. Karel Maršálek
Konzultant: Ing. Karel Šimeček, Ph.D.
VŠE v Praze, Fakulta informatiky a statistiky, 2025

MOTIVACE A PROBLÉM

2200 zdrojů logů, analyzováno jen 30

98% nezpracovaných dat

12h stahování dat měsíčně

1 pracovník na vše

METODY A ŘEŠENÍ

ENV algoritmus
Frekvenční analýza vzorů

Paralelní zpracování
12 CPU jader současně

Apache Parquet
70-85% redukce velikosti

Streamlit dashboard
Interaktivní vizualizace

VÝSLEDKY EXPERIMENTŮ

48×	96%	15	2200
RYCHLEJŠÍ	PŘESNOST	MINUT	ZDROJŮ

Metrika	Původní	Nové řešení	Zlepšení
Doba zpracování	12 hodin*	**15 minut**	**48× rychleji**
Analyzované zdroje	30	**2 200**	**73× více**
Velikost dat	100%	**25%**	**75% úspora**
F1 skóre detekce	79.1%	**96.0%**	**+16.9%**

** Původních 12 hodin = stahování dat z 30 zdrojů za měsíc*

TECHNOLOGIE

- Python 3.x
- Multiprocessing
- Apache Parquet
- Pandas/NumPy
- SSH/SFTP
- Streamlit
- ENV Algorithm
- LZMA

PRAKTICKÉ DOPADY

- ✓ **Rychlá detekce** útoků
- ✓ **Komplexní** monitoring
- ✓ **Automatizace** procesů
- ✓ **Legislativní** soulad

VYUŽITÍ V PRAXI

- Vysoké školy a univerzity
- Státní instituce
- Střední a velké firmy
- Nemocnice a zdravotnictví
- Finanční instituce
- Kritická infrastruktura

VÝZNAM PRO SPOLEČNOST

Ochrana osobních údajů tisíců studentů a zaměstnanců před kybernetickými útoky. Prevence finančních ztrát a poškození reputace institucí.