Building efficient ML models for ransomware detection in storage systems

Roman Pletka – IBM Research, Zurich

the Future of Memory and Storage

Ransomware cyber security threads

Experienced a cyber attack

75%

experienced an attempted ransomware attack within the last 12 months

Data Loss after attack

85%

not able to fully restore data from backup after an attack

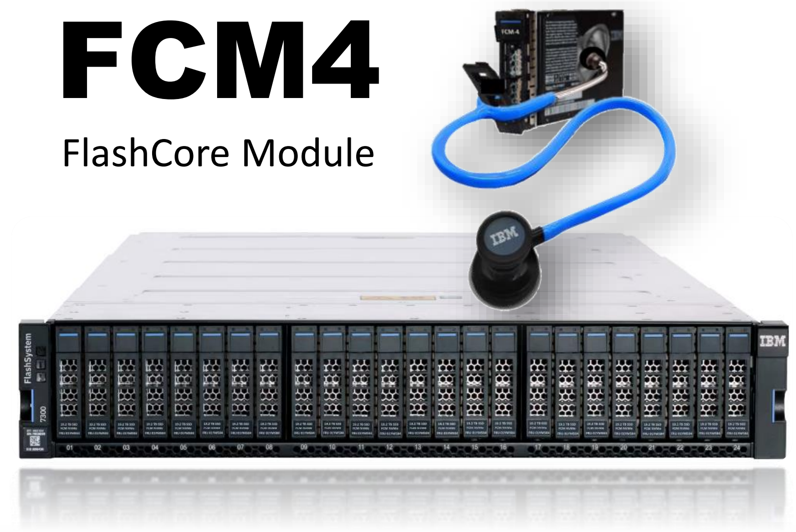
Operational Recovery

89%

More than 1 day to resume normal business operations (MVC)
61% more than 4 days

IBM FlashSystem ransomware threat detection pipeline

1. IBM FlashCore Module collect feature information on IO activity in hardware with no performance impact.
2. IBM Storage Virtualize runs an AI engine on every FlashSystem using ML model trained with real-world ransomware.
3. IBM Storage Insights collects thread information from connected FlashSystem arrays, alerts users, and triggers SIEM/SOAR software to initiate response. Collected statistics are used to improve ML models.



IBM Storage Insights

Block-level ransomware detection in IBM FlashSystem using FCM4

IBM

Current features

- Ransomware detection on 1000 volumes.
- Training with 50+ real ransomware and emulated ransomware strains in 200+ configurations.
- Continuous ML model updates.

Outlook

- Filesystem-aware ML models.
- 32k volumes.
- Volume grouping.
- Multi-variate time series processing.
- ML models for wiperware and exfiltration.

FMS

ML model training challenges

1. Feature Selection

- Features extracted from IO operations
- Summarized in seconds intervals
- Derived features

2. Model Selection

- Decision tree ensembles (XGBoost, Random Forest, …)
- Highly parallelizable (SnapML)

3. Filesystem Type

- EXT4
- XFS
- BTRFS
- NTFS
- VMFS
- ….

4. Volume state

- Current space utilization of volume
- Data fragmentation over time

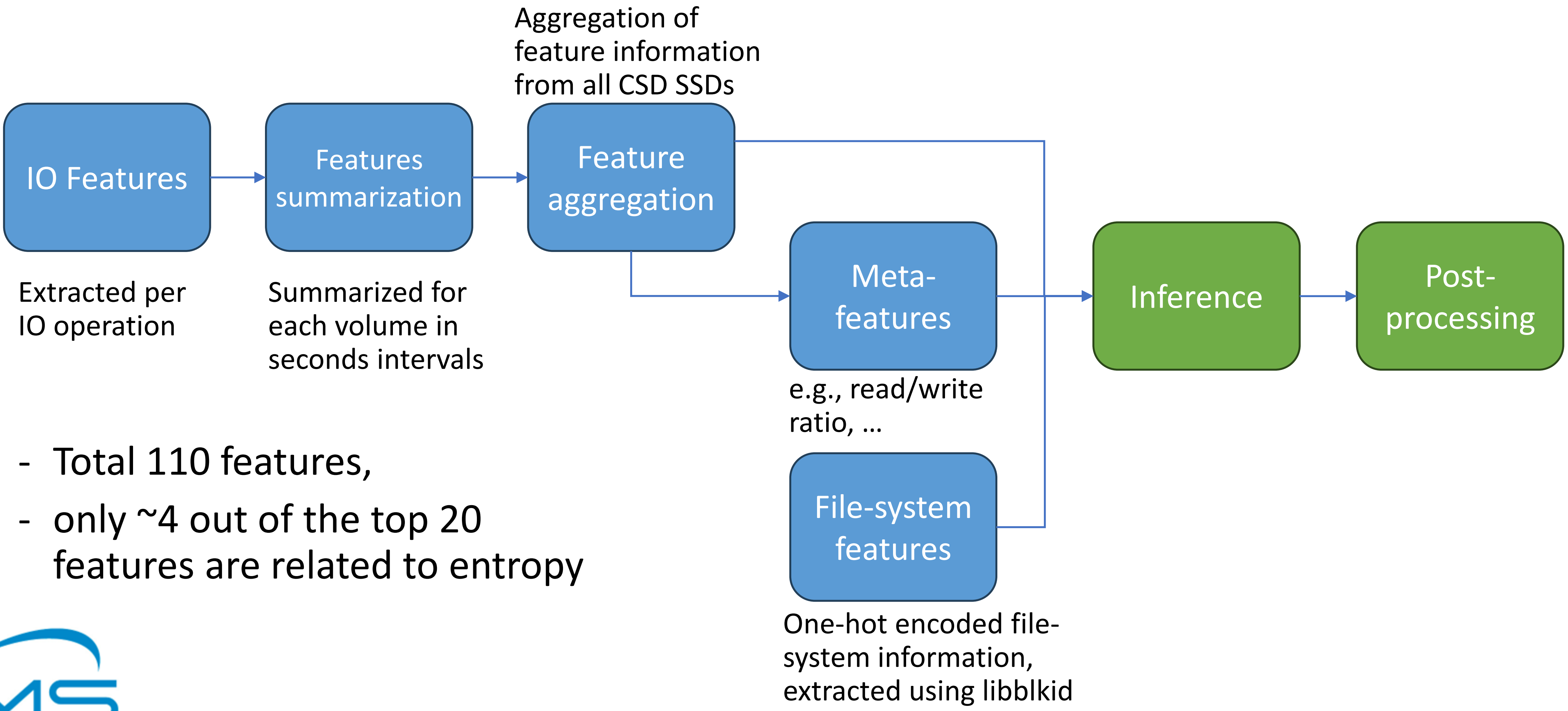
5. Datasets and Workloads

- Databases, Mail server, Filebench, Virtualized environments
- Large-scale datasets
- Real and emulated ransomware

6. Post-processing

- Voting window (sub-minute scale)
- Time-series analysis
- Autoencoders

Feature extraction and processing



- Total 110 features,
- only ~4 out of the top 20 features are related to entropy

Filesystem type and volume state analysis

Random Forest models

- Model 1: using 12 aggregated features
  - Entropy (mean, MAD, slope, Kurtosis, rewrite)
  - LBA (MAD, Kurtosis for reads + writes)
  - Transfers size (reads + writes)
  - Rewrite rate
- Model 2:
  - Adding file system information as one-hot encoded feature
  - Replace computationally expensive features (slope and Kurtosis) with histograms

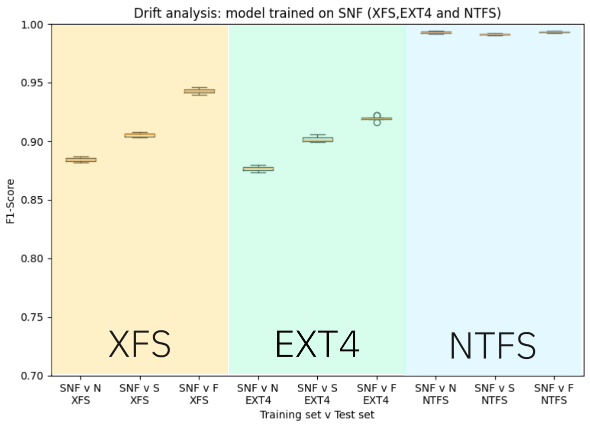
Training setups

- 3 File system types (XFS/EXT4/NTFS)
- Various ransomware and benign workloads
- Volume states (1TB)
  - Normal (N):
    Overall volume utilization 52%
  - Fill (F):
    Overall volume utilization 77%
  - Shuffle (S):
    Same as N, but 10% of the files are copied within the test directories and old data is deleted before using volume to collect traces

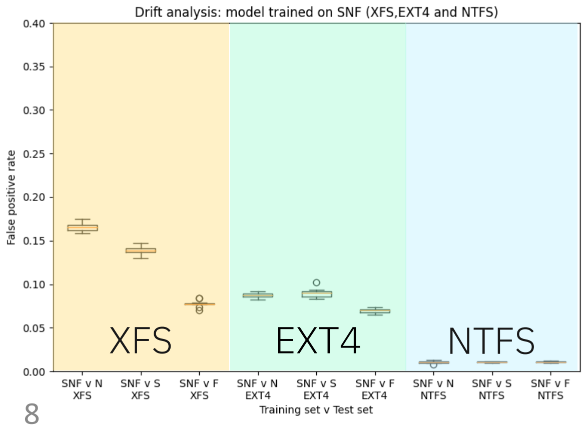
Model evaluation

- For the 3 different volume states, the F1 score as well as the false positive rate varies significantly in EXT4 and XFS.
- Using file system information and histograms in the model improves accuracy (3-8%) and reduces the false positive rate (40-47%).
- Computationally expensive features can be efficiently replaced with histogram.

Model 1 with 12 aggregated features

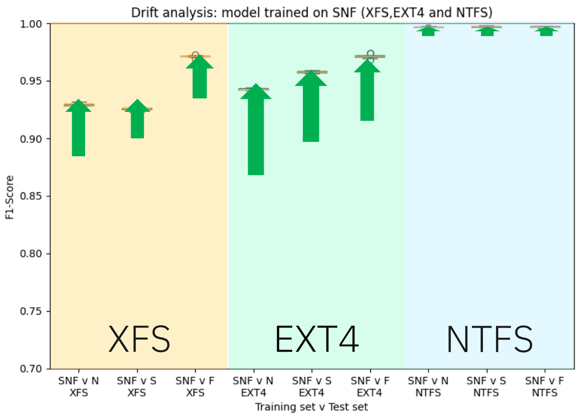


F1 Score

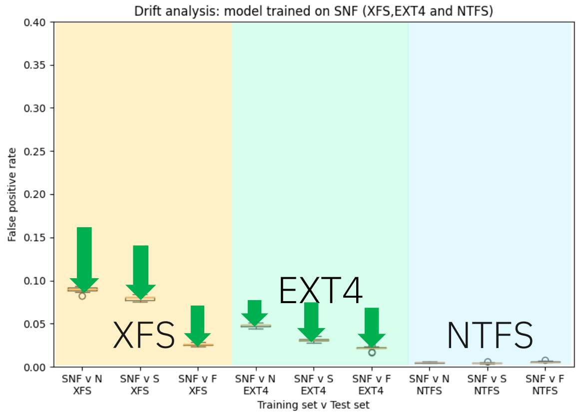


False positive rate

Model 2: 12 aggregated features + file system type



F1 Score



False positive rate

Measured ransomware detection time

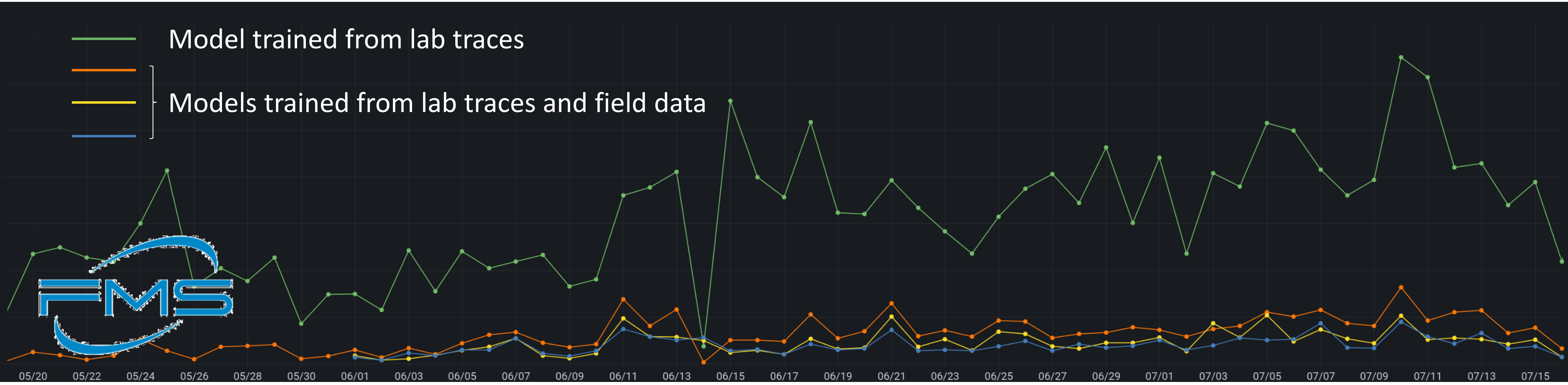
- Results measured while the inference engine is performing the feature vector classification for 1000 volumes in parallel and the evaluation classification results using majority voting.
- Evaluated the ransomware detection time in a KVM setup with a Windows 10 VM where the Conti ransomware was executed.

Detection in less than 1 min

Inference time for 1000 volumes in less than 10 ms

Improving classifier accuracy with field data

- Sets collected from real systems in the field can be used to retrain models. Must ensure correct labeling.
- Here, the FPR of the single-level classifier was reduced by 78.2 – 88.0% with models trained that include field data.



Conclusion

- ML models based on decision-tree ensembles combined with post-processing are well suited for ransomware detection in storage systems. Per-volume inference for thousands of volumes feasible.
  - Large feature set consisting of computationally inexpensive features using more than 100 features.
- Must carefully study the Generalizability of ML models.
  - Inclusion of volume state information, file system type, ransomware strain.
  - Large variety of benign workloads.
- Real world traces from field data help to improve accuracy of ML models.
  - Proper balancing of labeled training set.

Thank you!

Dr. Roman Pletka

Senior Research Scientist
Master Inventor

rap@zurich.ibm.com

Contributors:

Dionysios Diamantopoulos
Nicolas Reategui
Haris Pozidis
Yves Santos
Andrew Walls