

Getting Back Up: Understanding How Enterprise Data Backups Fail

George Amvrosiadis, *University of Toronto*; Medha Bhadkamkar, *Veritas Labs*

Understanding backups failures

- This is a systematic study of backup errors
 - Weekly reports from NetBackup™ customers
 - 775M jobs from 20,000 installations in 3 years

The bad news

- Backup system jobs fail often
 - 15% of jobs fail in average backup system
- Customer offline testing is insufficient
 - 56 error codes appear only in production

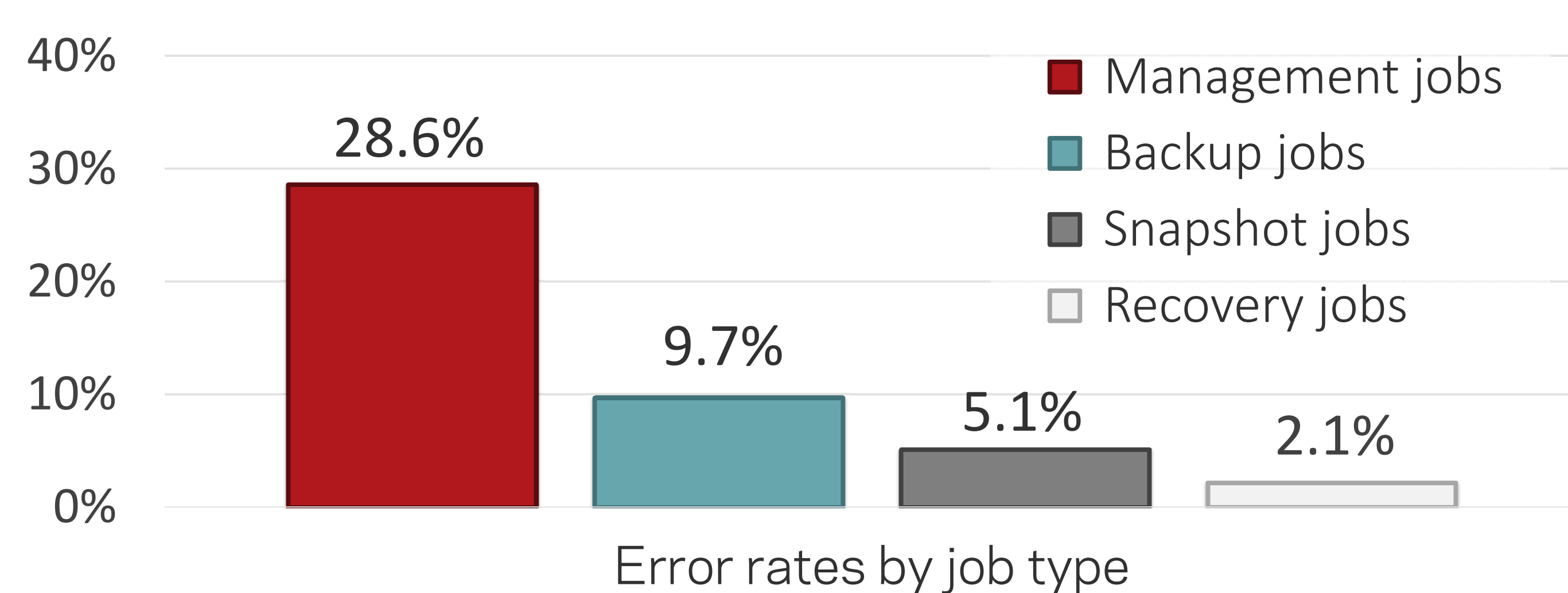
The good news

- Error code distribution is heavy-tailed
 - 64% of errors due to top 5 error codes

Error description	Jobs affected
Insufficient file permissions	25.4%
Invalid file system configuration	15.3%
Tape drive misconfigured	11.2%
Device full	7.6%
Backup window too short	4.5%

Guidelines for resolving backup errors

- Enforce stricter configuration validation
 - 76% of jobs fail due to misconfigurations
- Handle failures according to job type
 - 46% of error codes specific to a single job type



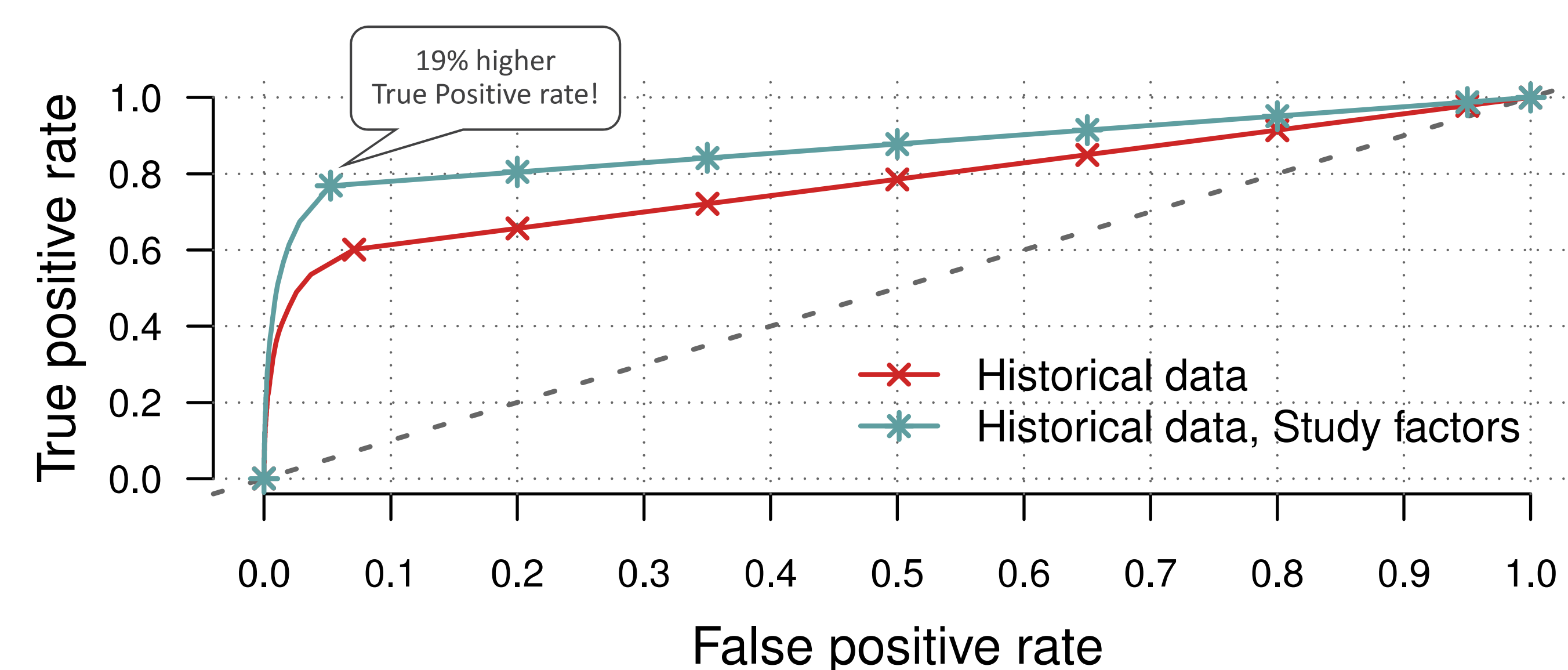
- Design simpler user-facing backup policies
 - More parameters → higher error diversity
- Backup often, and verify large archives
 - Larger jobs fail more frequently



Preventing backup errors proactively

- **Problem:** historical data is insufficient
 - High variability in inter-arrival times of errors
- **Insight:** use context to improve predictions
 - Use **factors** such as: *daily number of jobs, job type, job size, policy complexity*
- **Solution:** use random forest models
 - Generate a model for each error code
 - Classification rules reveal feature importance

- Study factors help predict errors
 - 44% of models rank **study factors** as top feature
 - Most important: *number of jobs, policy complexity*



Contact

George Amvrosiadis

Dept. of Computer Science, University of Toronto

✉ gamvrosi@cs.toronto.edu 🌐 www.cs.toronto.edu/~gamvrosi