



Getting back up:

Understanding how enterprise data backups fail

George Amvrosiadis – *University of Toronto*

Medha Bhadkamkar – *Veritas Labs*

# We need to talk about backup

- Backup in the news: rarely a good thing
  - 123-reg erases customer website data. **No backup.** **April**
  - Salesforce loses 4 hours of data. **Backup incomplete.** **May**
- Business surveys: backups fail often
  - 27% have lost data due to backup errors
  - 80% have trouble configuring backup software

**Need a systematic study: why do backups fail?**

# Study goals


- ❑ Understand why backups fail
- ❑ Help resolve backup errors
- ❑ Prevent backup errors

# Collecting the data

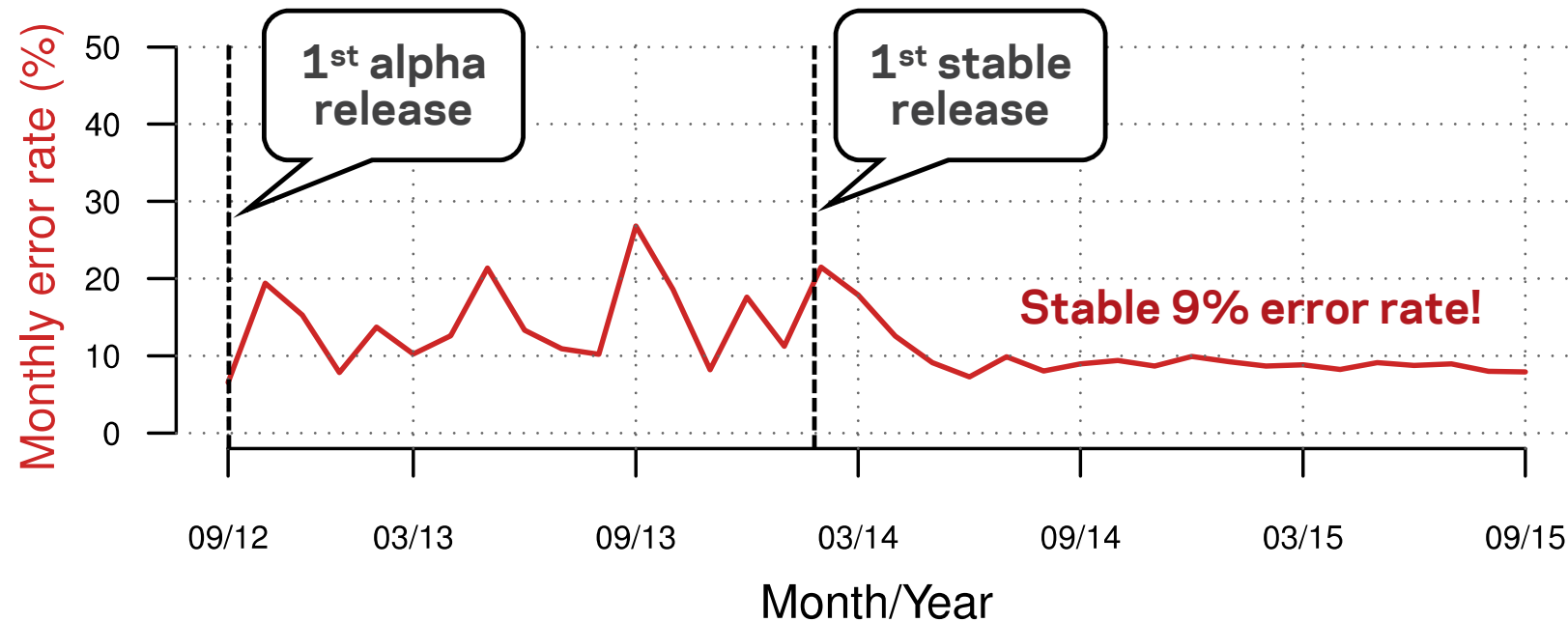
- Telemetry from NetBackup customer installations
  - Weekly runtime and configuration statistics
  - 775M jobs from 20,000 installations in 3 years

Job type	Jobs in dataset
Data backup jobs	604.9 Million
Data management jobs	105.8 Million
Data recovery jobs	6.3 Million

# What comes next

-  ☐ Understand why backups fail
- ☐ Help resolve backup errors
- ☐ Prevent backup errors

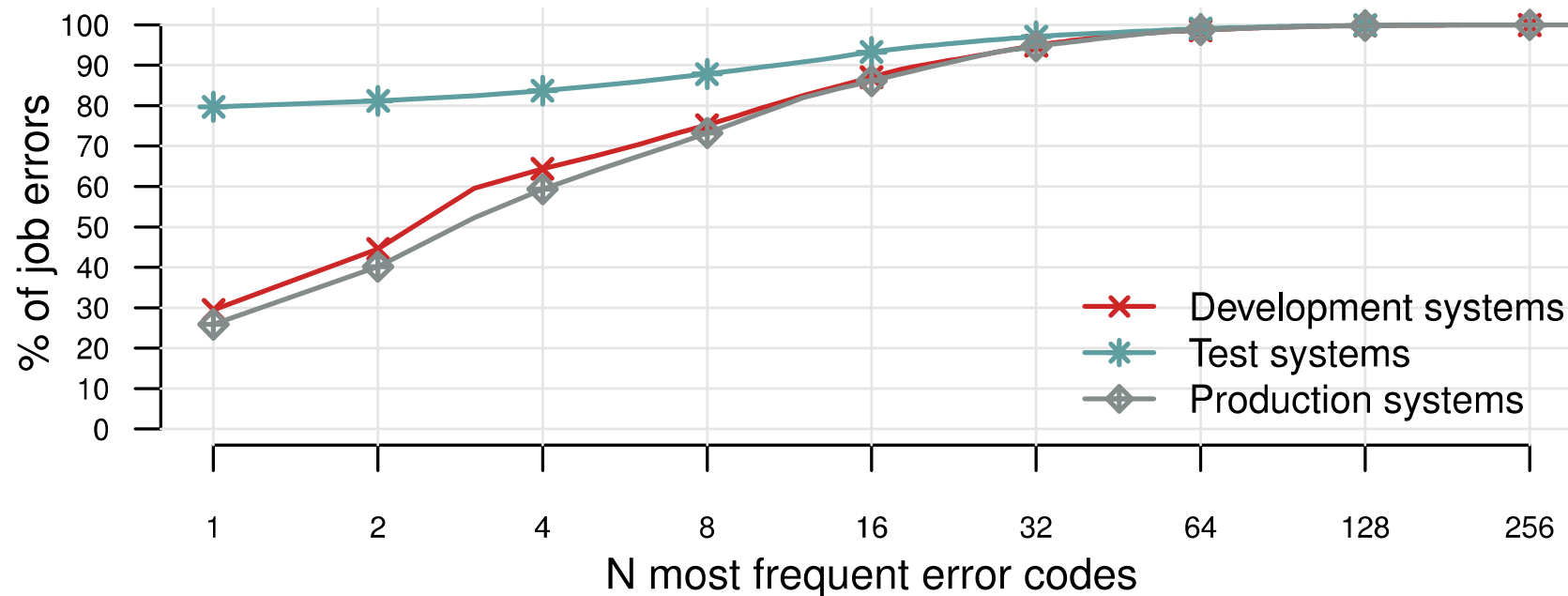
# Jobs fail often



- Not all installations are equal
  - Development systems: feature tests on alpha/beta releases
  - Test systems: configuration testing on stable releases
  - Production systems: long-lived and busy

# Errors are not diverse

- 333 error codes in our dataset (28% of all codes)
  - Testing insufficient: 59 codes only show up in production
- 64% of errors due to same 5 error codes



# Top 5 errors in backup systems

Error description	Jobs affected
Partial backup due to file permissions	25.4%
Invalid filesystem block, or max file size	15.3%
No tapes available in specified robot	11.2%
Device full	7.6%
Backup window too short	4.5%
Total: 64.0%	



# What comes next

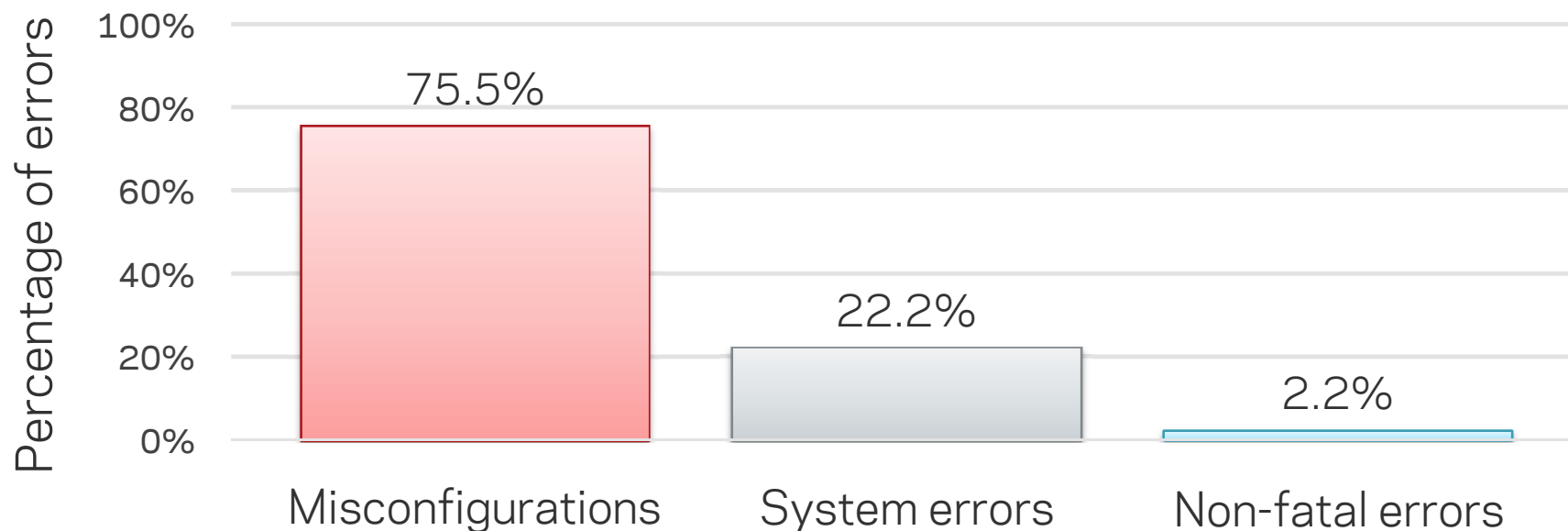
 Understand why backups fail

  Help resolve backup errors

 Prevent backup errors

# Configuration errors prevail

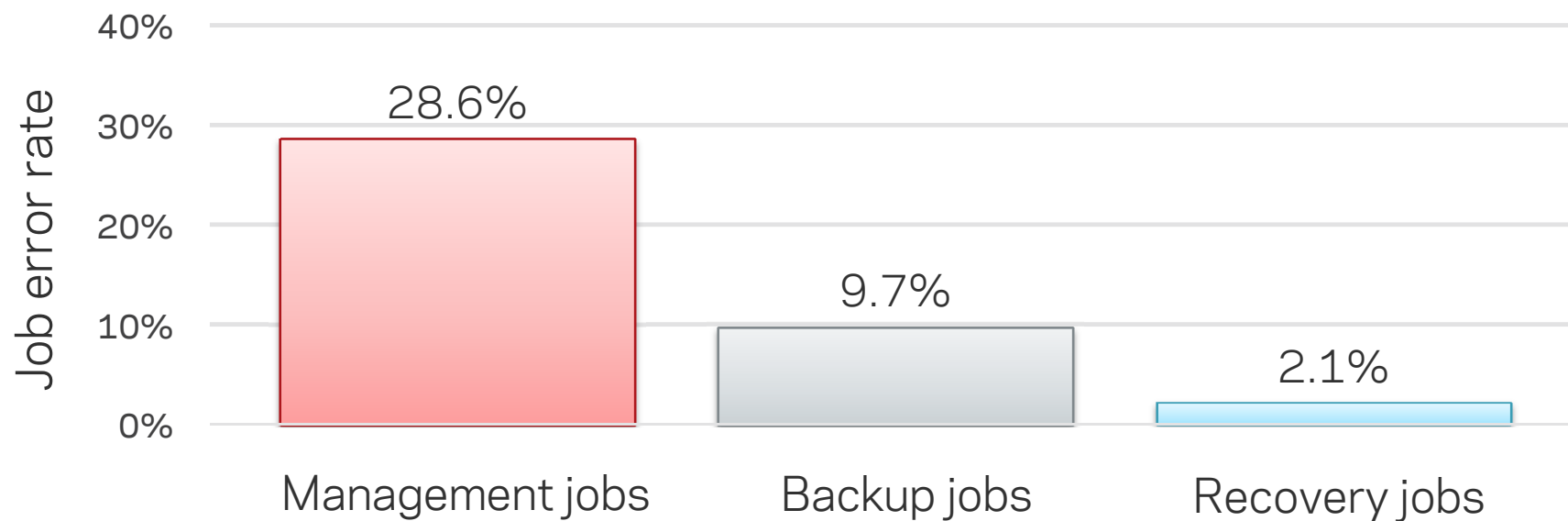
- We manually categorized error codes by cause
- Configuration fixes can resolve top 5 errors



**We need better configuration validation, self-healing mechanisms**

# Job type is indicative of error rate

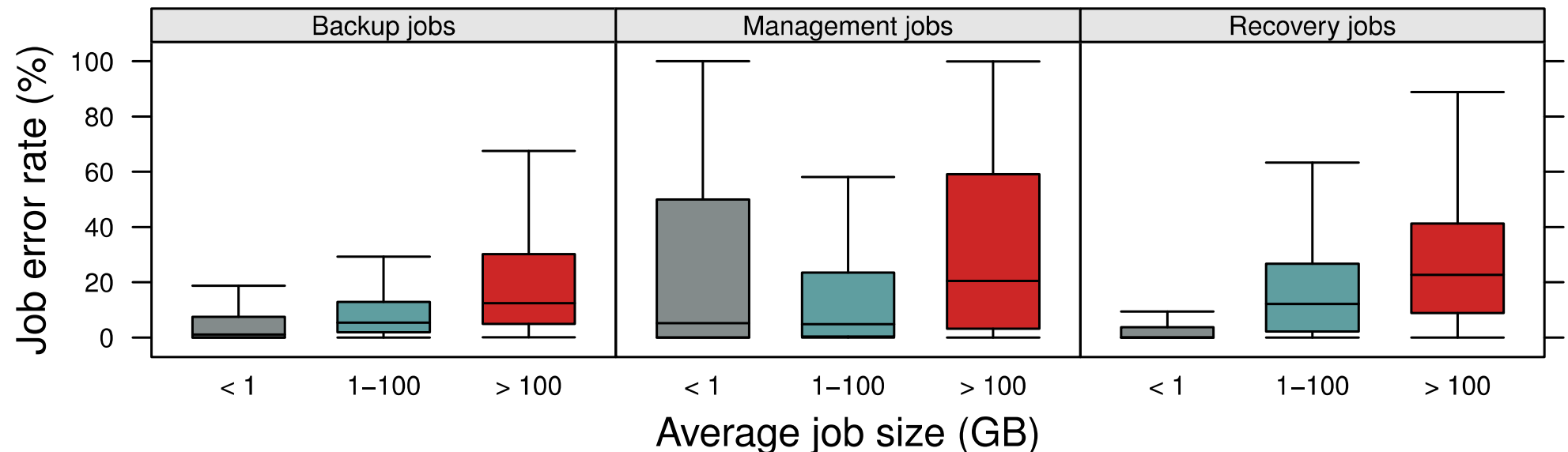
- 46% of error codes specific to job type
  - Type-specific errors usually refer to misconfigurations



**Tune rigor of error prevention mechanisms to job type**

# Larger jobs are more likely to fail

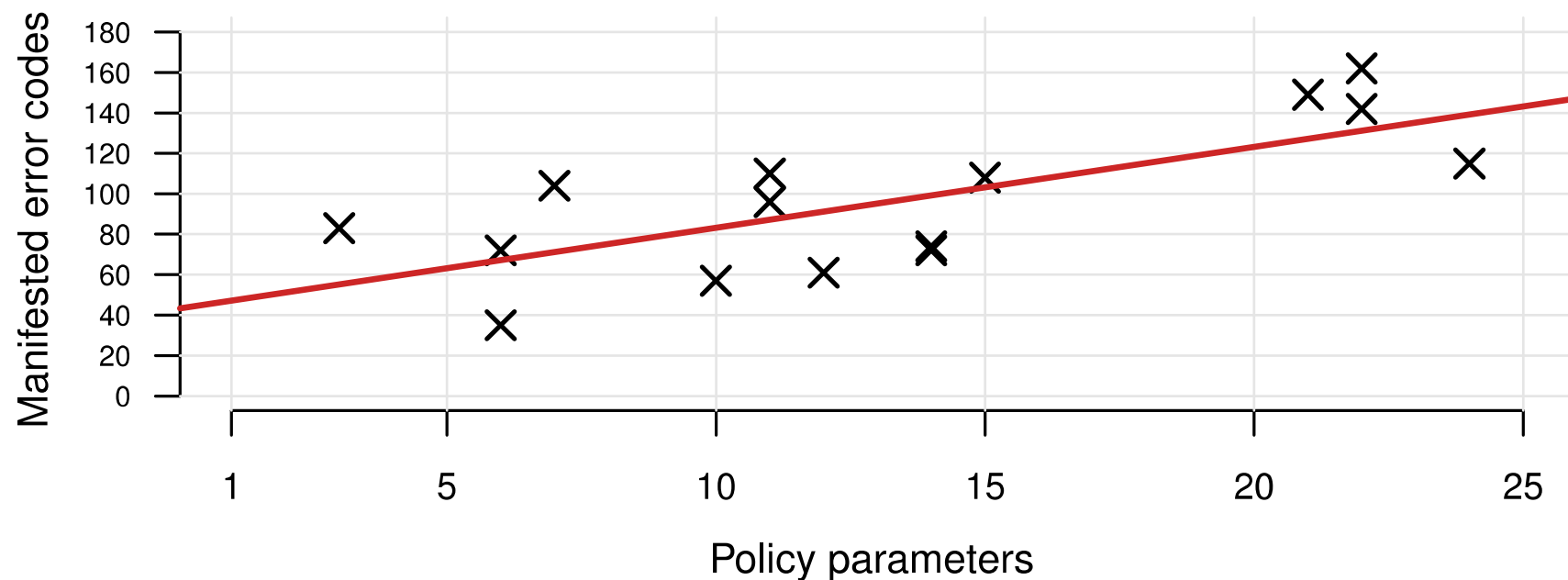
- Systems with larger jobs encounter more errors
- Small management jobs are error-prone
  - E.g. data cleanup, configuration jobs that transfer no data



**Backup often to avoid large jobs, verify large backup images**

# Complexity breeds error diversity

- Backup policies ensure consistent data backups
- Configuration parameters differ by policy
  - Tailored to specific applications, operating environments



**Design and prefer simpler backup policies**

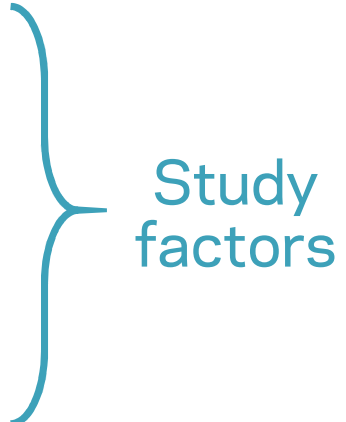
# What comes next

 Understand why backups fail

 Help resolve backup errors

  Prevent backup errors

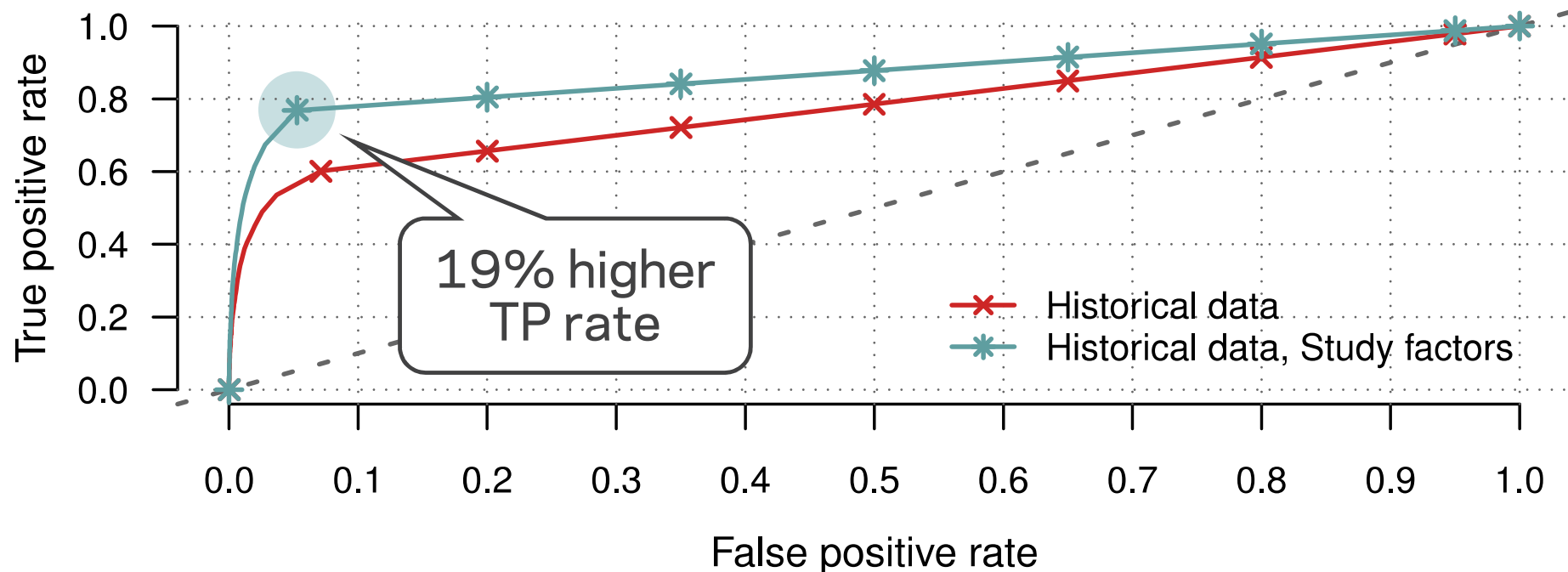
# Towards error prediction

- Historical data insufficient for error prediction
    - High variability in the inter-arrival times of most errors
  - Job type is indicative of error rate
  - Larger jobs are more likely to fail
  - Policy complexity breeds error diversity
- 
- Study factors

**How do we use our study factors to predict errors?**

# A learning approach

- Random forests: decision tree groups
  - Generate a **separate** model for each error code
- 44% of models rank study factors as top feature
  - Most important: number of jobs, policy complexity





# Where do we go from here?

- More targeted **error prediction**
- **Configuration automation** instead of defaults
- Application-specific **configuration validation**
- **Work reduction** to reduce needed downtime



# Thank you!

Copyright © 2016 Veritas Corporation. All rights reserved. Veritas and the Veritas Logo are trademarks or registered trademarks of Veritas Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

This document is provided for informational purposes only and is not intended as advertising. All warranties relating to the information in this document, either express or implied, are disclaimed to the maximum extent allowed by law. The information in this document is subject to change without notice.