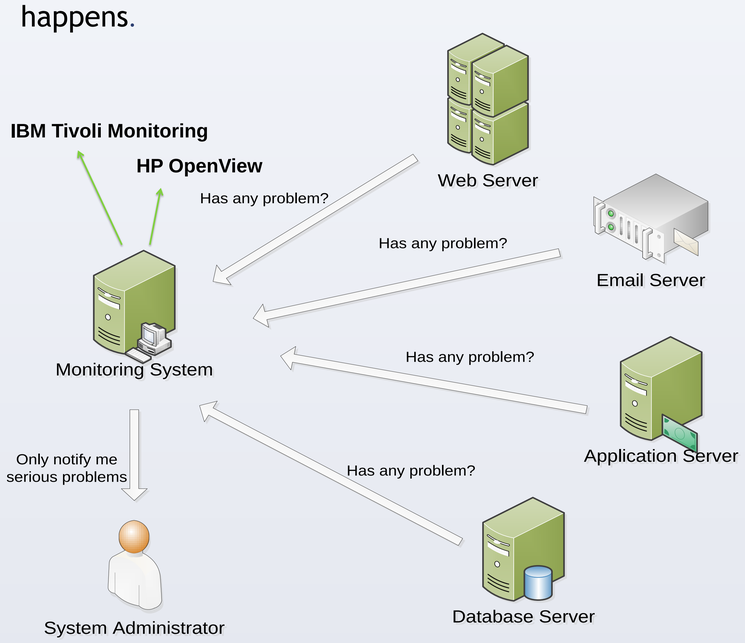
An Integrated Framework for Optimizing Automatic Monitoring Systems in Large IT Infrastructures

Liang Tang, Tao Li, Larisa Shwartz, Florian Pinel, Genady Ya. Grabarnik

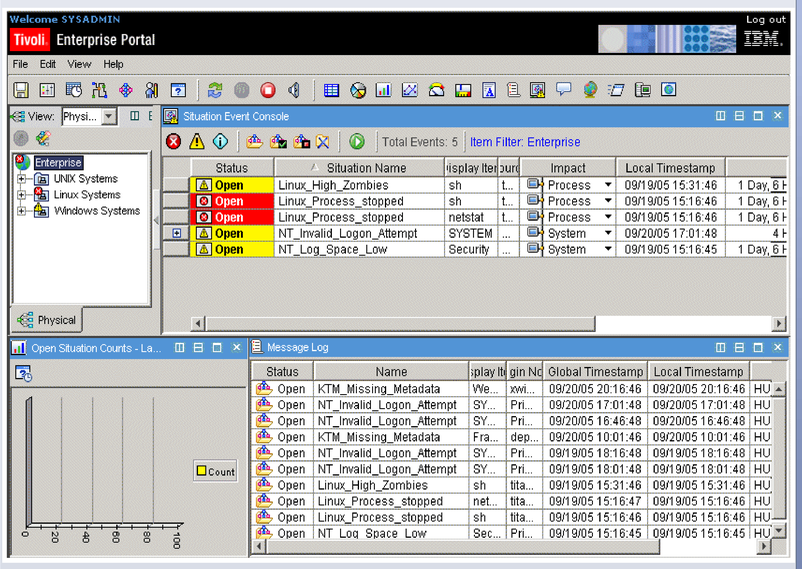
Florida International University, IBM Watson Research Center, St. John's University

Introduction

Automatic Monitoring System: monitor those servers, notify the system administrator only when a problem happens.



Configuration is Complicated



What is Misconfiguration?

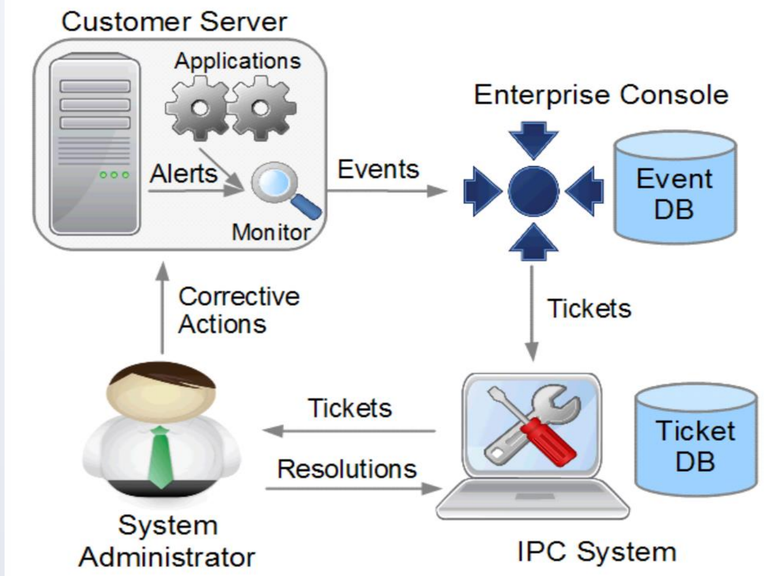
- **False Positive:** A CPU utilization situation with 50% threshold on an intensive DB server (50% is normal for an intensive DB server, it is not a problem).
- **False Negative:** Installed a new web server, but forget to add it into the monitoring situation. If this web server has a problem, it would not be captured by the monitoring system.

Our goal: eliminate false positives and false negatives

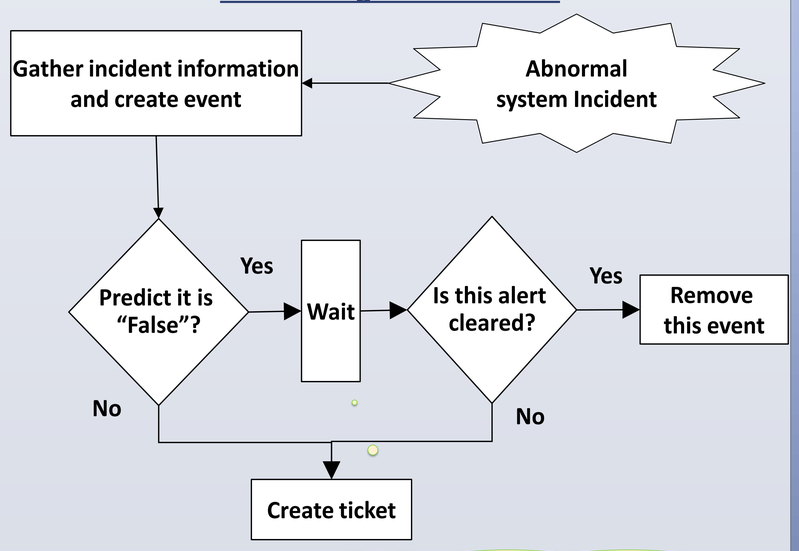
How to Detect False Positives and False Negatives

False positive and false negatives are identified by system administrators and recorded in tickets (ground truth).

Can the monitoring system learn from these tickets?



Eliminating False Positive



Most false positive alerts are **transient alerts (disappear in a short time).**

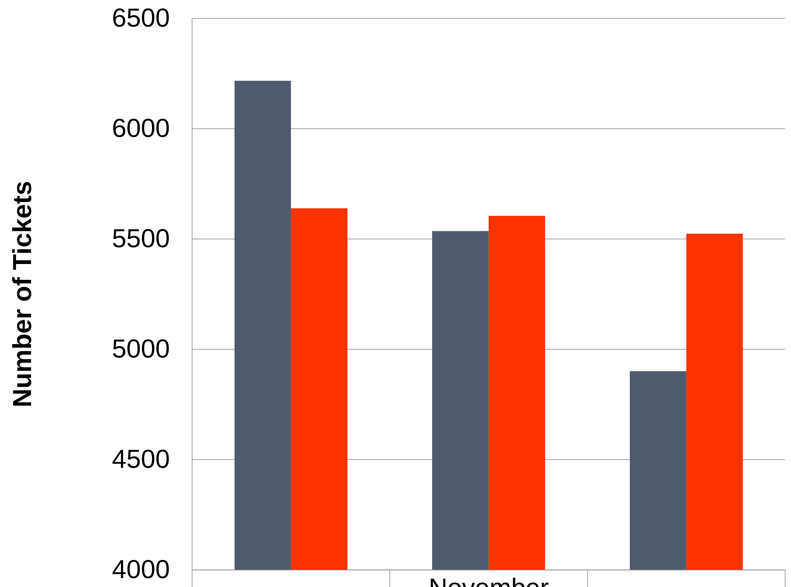
Eliminating False Positive

- Find missed alerts in manual tickets (created by system admins) and report to system admin, then they make corresponding changes on configuration.
- SVM + Selectively Sampling: build binary text classifier

Situation Issue	Words
DB2 tablespace Utilization	DB2, tablespace
File System Space Utilization	space,file
Disk Space Capacity	space,drive
Service Not Available	service,down
Router/Switch Down	router

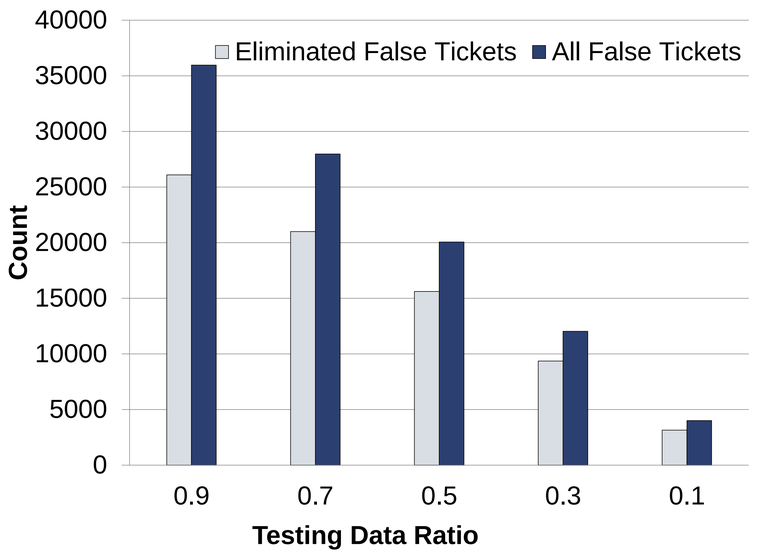
Labeled **Domain** Words (Labeled Features)

Online Evaluation



	October 2011	November 2011	Januray 2012
False Tickets	6217	5536	4901
Real Tickets	5639	5605	5524

Offline Evaluation



Disk Space Issue