# Identifying Missed Monitoring Alerts based on Unstructured Incident Tickets


**Liang Tang, Tao Li**
School of Computing and Information Sciences
Florida International University, Miami, FL, USA

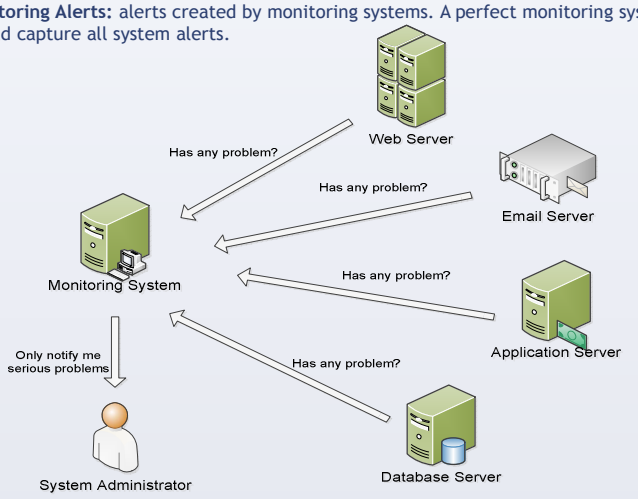
**Larisa Shwartz**
IBM T.J. Watson Research Center
Yorktown, Heights, NY, USA

**Genady Ya. Grabarnik**
Dept. Math & Computer Science
St. John's University, Queens, NY, USA


## Introduction

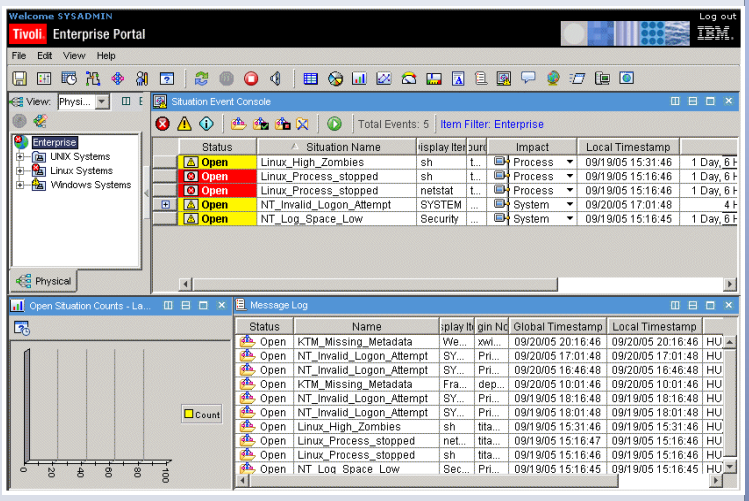
**Monitoring System:** monitor those servers, notify the system administrator only when a problem happens.

**Monitoring Alerts:** alerts created by monitoring systems. A perfect monitoring system should capture all system alerts.




## Why Some Monitoring Alerts are Missed?

**The configuration of Monitoring Systems is complicated**




**Misconfiguration sometimes happens:**

Example:
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.

## What are the Unstructured Incident Tickets?

**Ticket1 (disk space alert):**

*Please clear space from E drive xxxx-fa-ntfwwfdb Please clear space from E drive xxxx-fa-ntfwwfdb.it is having 2 MB free...*

**Ticket2 (service status alert):**

*RFS101681 E2 Frontier all RecAdmin services are down Frontier RecAdmin services are not running on the batch server Kindly logon to the server : xxx.xxx.155.183/xxx ...*

**Ticket3 (database alert):**

*DB2 is not connectable from xxxxx Hi Team@Can you please look into why we are unable to connect to Porfolio XRef DB.Server : xxxx12DB Instance : sec mastId : ipxrbtchWhile...*

## Finding Missed Alerts in Unstructured Incident Tickets

- **Data Source:** Missed monitoring alerts will be captured by system admins in manual tickets, where are textual descriptions.
- **Solution:** finding missed alerts in the tickets. Build a binary text classifier.
- **Challenges:**
  - Highly imbalanced data (very few manual tickets are related system alerts)
  - Not enough labeled tickets (Too many manual tickets for labelling, labeling cost is huge)

## Proposed Method

- **Use domain words to select a subset of tickets for labeling.** Each ticket is scored by the number of contained domain words. High score ticket has a high probability to select. (A randomly selected ticket subset might only contain customer request tickets and it is bad for the classifier training)

| 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 |

- **Use SMOTE to do over-sampling on the false negative tickets.**
- **Apply the SVM algorithm to train the classifier.**

## Evaluation

- **Baseline:** randomly select a subset of tickets for labelling and training.
- **Dataset:** collected from a large customer account in IBM IT service center. This account consists of over 1,000 monitored servers and network devices.