中国科学院软件研究所学术年会’2018
暨计算机科学国家重点实验室开放周

学术论文

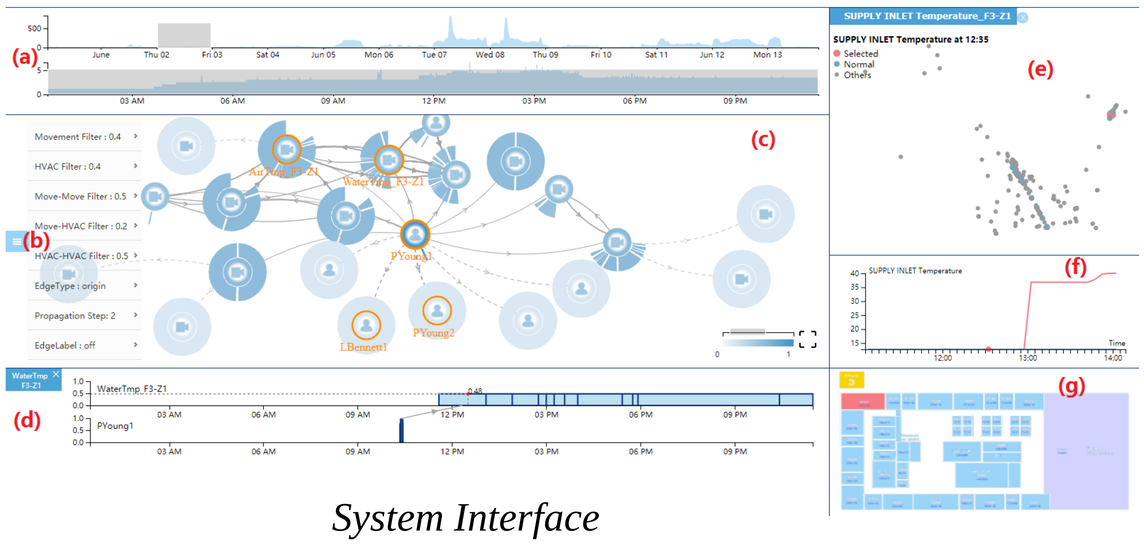
# Visual Analysis of Collective Anomalies Through High-Order Correlation Graph

# 基于高阶关联图的群体异常可视化分析系统

**时磊** 黄聪聪 余如雷

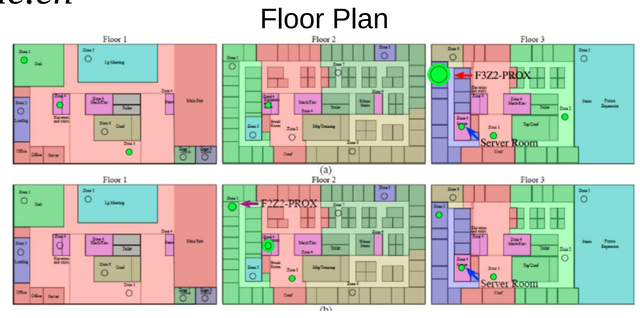
*Accepted by PacificVis(2018): The 11th IEEE Pacific Visualization Symposiu*

*VAST Challenge 2016 Data Set* *联系方式:huangcc@ios.ac.cn*




*System Interface*

Building environment measurement




Employ mobility measurement

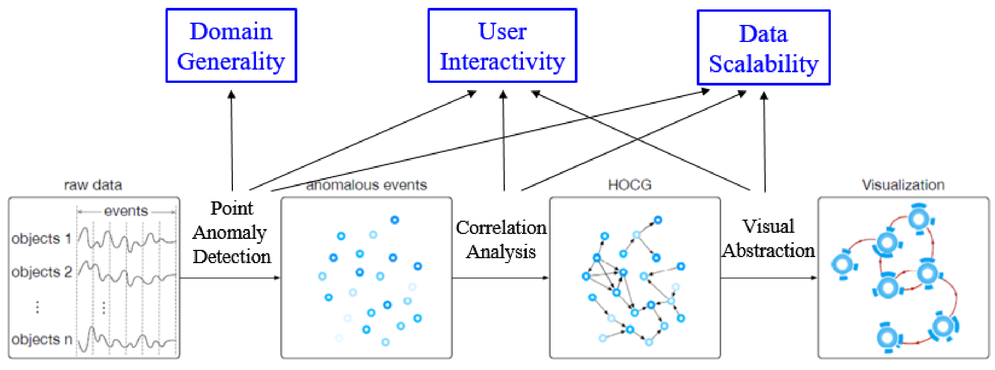
```
"message": {
    "F_1_Z_8A: Thermostat Heating Setpoint": "15.6000",
    "F_1_BATH_EXHAUST:Fan Power": "0.0000",
    "F_1_Z_8B VAV REHEAT Damper Position": "0.0000",
    "F_1_Z_8A REHEAT COIL Power": "0.0000",
    "F_1_Z_2: Thermostat Heating Setpoint": "15.6000",
    "F_1_VAV_SYS SUPPLY FAN:Fan Power": "0.0000",
    "F_1_Z_2 SUPPLY INLET Temperature": "24.4533",
    "F_1_Z_8A RETURN OUTLET CO2 Concentration": "287.5286"
    "F_1_Z_1 SUPPLY INLET Temperature": "24.4533",
    "F_1_Z_8A: Thermostat Temp": "25.0116",
    "F_1_Z_7 RETURN OUTLET CO2 Concentration": "823.2590",
    "F_1_VAV_SYS HEATING COIL Power": "0.0000",
```

| Person-Id | Prox-Id | Prox-Zone | Second of Timestamp |
|---|---|---|---|
| 85 | pyoung001 | floor 2 prox zone 4 | 6/2/2016 10:20:49 AM |
| 85 | pyoung001 | floor 3 prox zone 4 | 6/2/2016 10:21:31 AM |
| 85 | pyoung001 | floor 3 prox zone 1 | 6/2/2016 10:21:33 AM |
| 85 | pyoung001 | floor 3 prox zone 3 | 6/2/2016 10:21:50 AM |
| 85 | pyoung001 | floor 3 prox Server Room | 6/2/2016 10:22:00 AM |
| 85 | pyoung001 | floor 3 prox zone 3 | 6/2/2016 10:25:01 AM |
| 85 | pyoung001 | floor 3 prox zone 1 | 6/2/2016 10:25:12 AM |
| 85 | pyoung001 | floor 3 prox zone 4 | 6/2/2016 10:25:29 AM |
| 85 | pyoung001 | floor 2 prox zone 4 | 6/2/2016 10:26:11 AM |
| 85 | pyoung001 | floor 2 prox zone 1 | 6/2/2016 10:26:13 AM |
| 85 | pyoung001 | floor 2 prox zone 2 | 6/2/2016 10:26:33 AM |

*Two datasets*

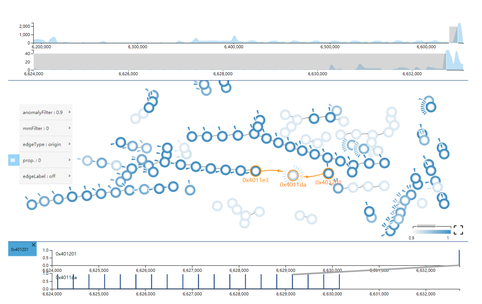
## Background

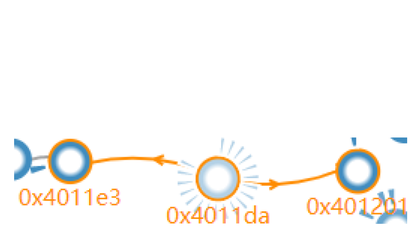
- Anomaly Types
  a) Point anomaly: e.g., an illegal overwrite of memory
  b) Contextual anomaly: e.g., high CPU usage in the resting state
  c) Collective anomaly: e.g., ssh->buffer-overflow-> ftp
- Detection algorithms: collective anomaly is the most challenging
  - –Classification-based
  - –Nearest Neighbor-based
  - –Clustering-based
  - –Statistical model-based
  - –Information-theoretic-based
  - –Spectral-based



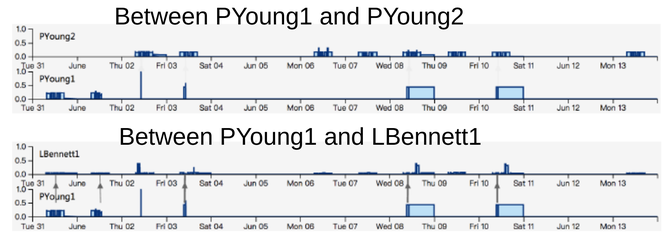

## Research Problem

- Design a visual analytics technique for detecting the collective anomaly on a group of interrelated objects from their observed behaviors

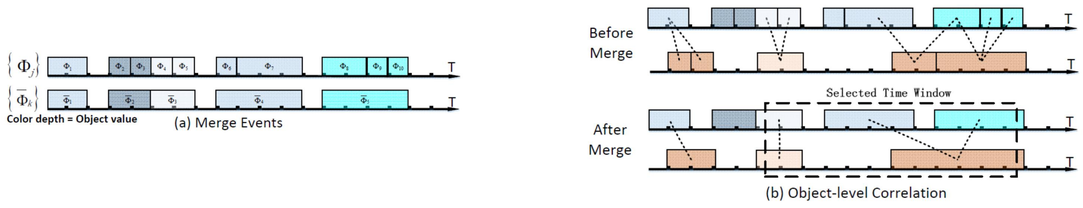





*Case Study: Software Crash Analysis*




*Case Study: Facility Monitoring*




## Visual Abstraction

- Filtering of HOCG
  - –Selection of events based on time and anomaly/correlation threshold
  - –Augment HOCG to include highly related events to the selected ones
- Object-centric abstraction
  - –Aggregate events belonging to the same object into one group node
  - –Aggregate event-level correlations to object-level correlations

## Conclusion

- **We propose the concept of high-order correlation graph**
  - –Generality: detect correlations among objects with heterogeneous data types, incorporate domain knowledge
  - –Scalability: support to scale to huge number of objects, dimensions, and events
  - –Interactivity: customize the HOCG for flexibly analysis anomaly
- **We design a visualization interface to explore HOCG for collective anomaly detection**
  - –Wedge-based visual metaphor
  - –HOCG view, event view, detail view