ΘPAD:

Online Performance Anomaly Detection with Kieker

Tillmann Bielefeld[1]

[1] empuxa GmbH, Kiel

KoSSE-Symposium Application Performance Management (Kieker Days 2012)

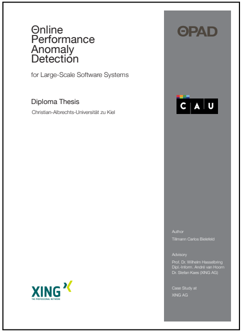
November 29, 2012 @ Wissenschaftszentrum Kiel

Agenda

1. Monitoring at XING
2. OPAD’s Architecture
3. Evaluation
4. Results
5. Conclusion

Thesis Goals

1. Design of online performance anomaly detection concept (ΘPAD)
2. ΘPAD implementation as Kieker plugin
3. ΘPAD integration with case study system
4. Evaluation @ XING



Θnline
Performance
Anomaly
Detection

for Large-Scale Software Systems

Diploma Thesis
Christian-Albrechts-Universität zu Kiel

ΘPAD

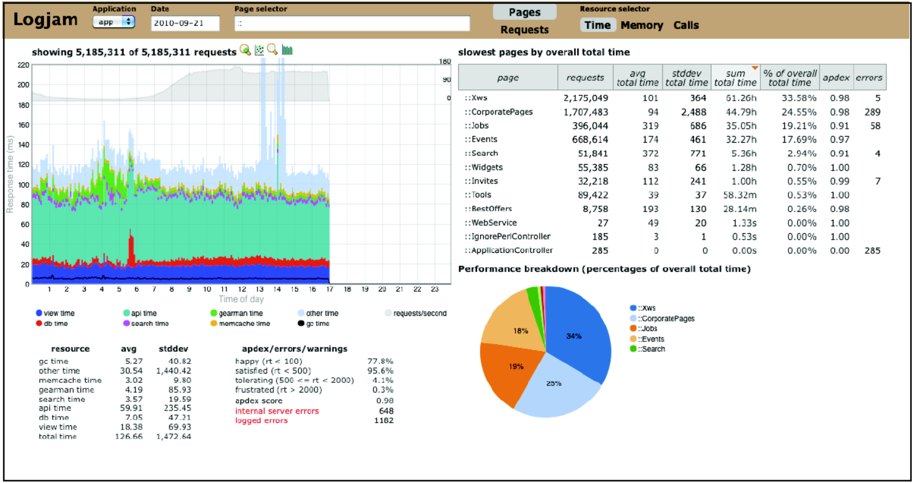
Author
Tillmann Carlos Bielefeld

Advisory
Prof. Dr. Wilhelm Hasselbring
Dipl.-Inform. André van Hoorn
Dr. Stefan Kaes (XING AG)

Case Study at
XING AG

Tillmann C. Bielefeld:
"Online performance anomaly detection for large-scale software systems"
March 2012. Diploma Thesis, Kiel Univ.

Existing Logjam-based Monitoring @ XING



Logjam | Application: app | Date: 2010-09-21 | Page selector: :: | Pages | Requests | Resource selector: Time | Memory | Calls

showing 5,185,311 of 5,185,311 requests

slowest pages by overall total time

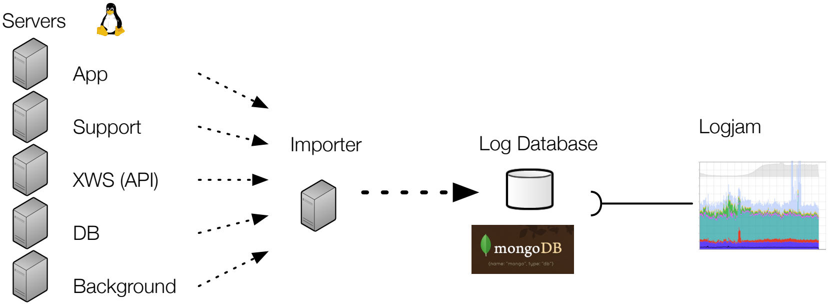
page	requests	avg total time	stddev total time	sum total time	% of overall total time	apdex	errors
::Xws	2,175,049	101	364	61.26h	33.58%	0.98	5
::CorporatePages	1,707,483	94	2,488	44.79h	24.55%	0.98	289
::Jobs	396,044	319	686	35.05h	19.21%	0.91	58
::Events	668,614	174	461	32.27h	17.69%	0.97	
::Search	51,841	372	771	5.36h	2.94%	0.91	4
::Widgets	55,385	83	66	1.28h	0.70%	1.00	
::Invites	32,218	112	241	1.00h	0.55%	0.99	7
::Tools	89,422	39	37	58.32m	0.53%	1.00	
::BestOffers	8,758	193	130	28.14m	0.26%	0.98	
::WebService	27	49	20	1.33s	0.00%	1.00	
::IgnorePerlController	185	3	1	0.53s	0.00%	1.00	
::ApplicationController	285	0	0	0.00s	0.00%	0.00	285

Performance breakdown (percentages of overall total time)

resource	avg	stddev
gc time	5.27	40.82
other time	30.54	1,440.42
memcache time	3.02	9.80
gearman time	4.19	85.93
search time	3.57	19.59
api time	59.91	235.45
db time	7.05	47.21
view time	18.38	69.93
total time	126.66	1,472.64

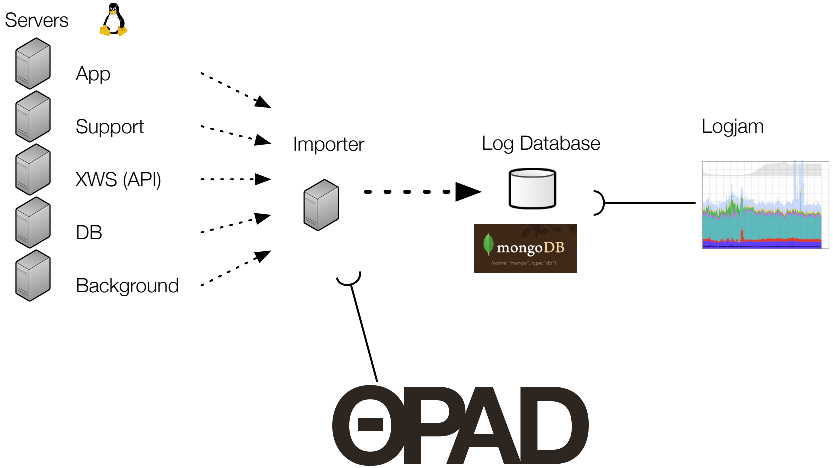
apdex/errors/warnings	
happy (rt < 100)	77.8%
satisfied (rt < 500)	95.6%
tolerating (500 <= rt < 2000)	4.1%
frustrated (rt > 2000)	0.3%
apdex score	0.98
internal server errors	648
logged errors	1182

Logjam-based monitoring already in place @ XING



XING's logging/monitoring architecture

Integration of ΘPAD in XING's Architecture

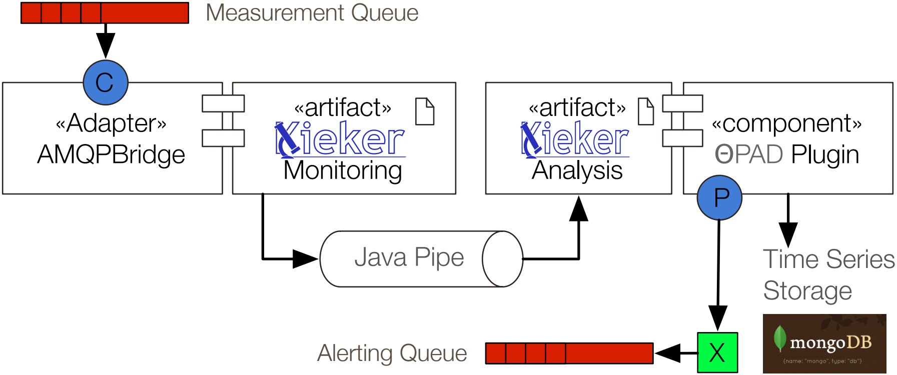


Example JSON Logging Message

```
{
    "count": 5204.903527993169,
    "memcache_time": 6505.196318140181,
    "api_time": 2207.0271495891297,
    "db_time": 5004.8727338680155,
...
    "view_time": 3936.1623304929153,
    "total_time": 1586.8188192888886,
    "api_calls": 5546.250545491678
}
```

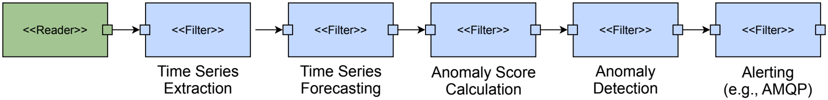
Input data received via AMQP and processed by ΘPAD

High-Level ΘPAD Architecture



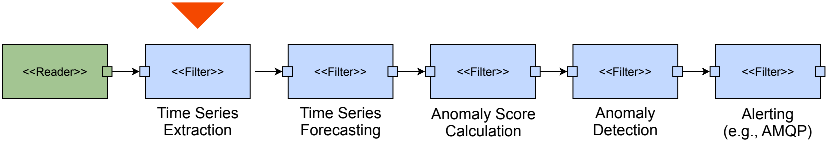
1. AMQP messages transformed into Kieker monitoring records
2. ΘPAD: pipes-and-filters processing of records
3. ΘPAD results passed to alerting queue and time-series storage

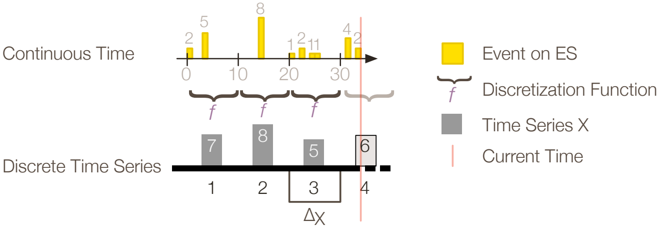
ΘPAD Processing Steps



Step 1: Time Series Extraction

ΘPAD Processing Steps (cont'd)



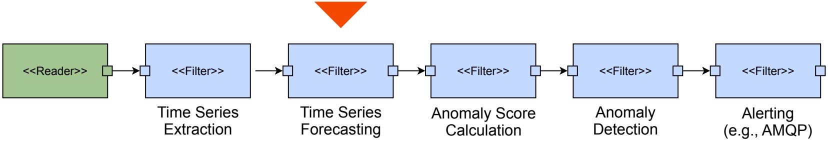


```
select sum(value) as aggregation
from MeasureEvent.win:time_batch( 1000 msec )
```

Step 2: Time Series Forecasting

ΘPAD Processing Steps (cont'd)

OPAD's Architecture



<<Reader>>

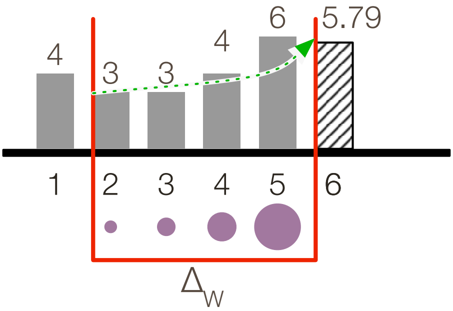
<<Filter>>
Time Series Extraction

<<Filter>>
Time Series Forecasting

<<Filter>>
Anomaly Score Calculation

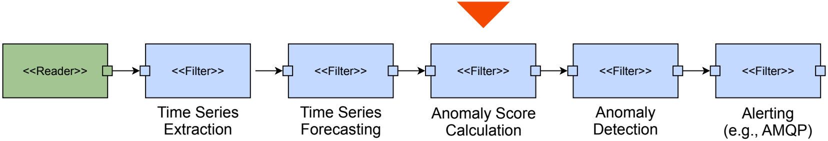
<<Filter>>
Anomaly Detection

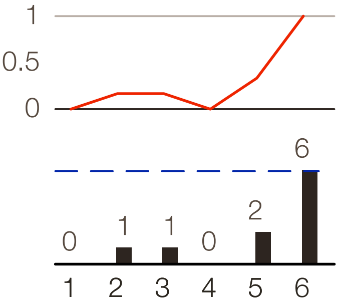
<<Filter>>
Alerting (e.g., AMQP)



Step 3: Anomaly Score Calculation

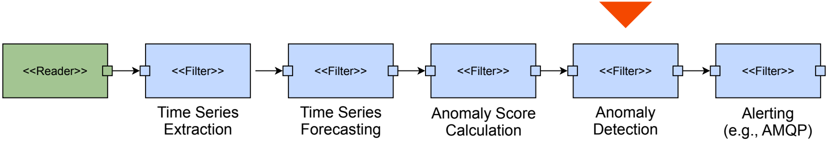
ΘPAD Processing Steps (cont'd)

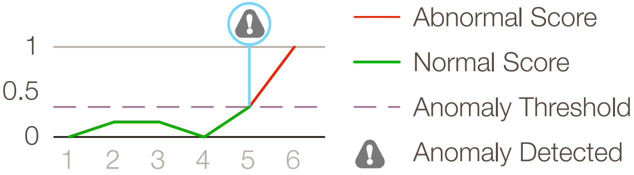




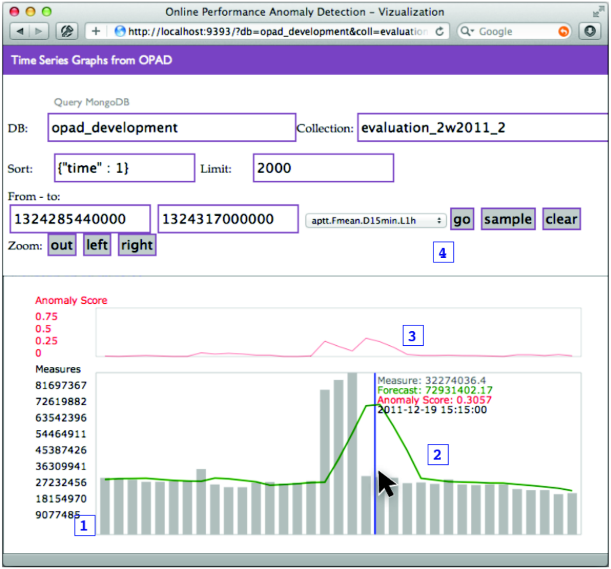
Step 4: Anomaly Detection

ΘPAD Processing Steps (cont'd)





ΘPAD Web Interface



Online Performance Anomaly Detection – Vizualization

http://localhost:9393/?db=opad_development&coll=evaluation

Time Series Graphs from OPAD

Query MongoDB

DB: opad_development

Collection: evaluation_2w2011_2

Sort: {"time" : 1}

Limit: 2000

From - to: 1324285440000 1324317000000

aptt.Fmean.D15min.L1h

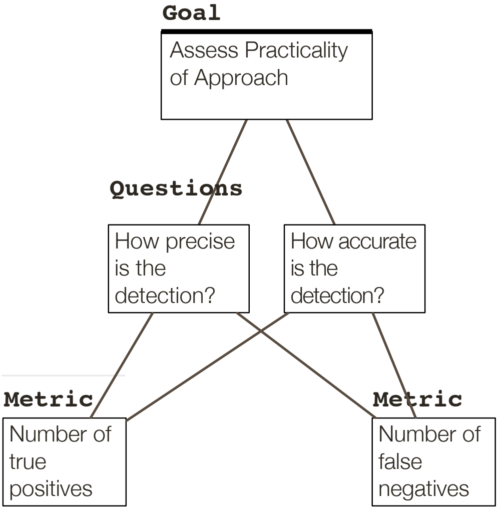
Measure: 32274036.4

Forecast: 72931402.17

Anomaly Score: 0.3057

2011-12-19 15:15:00

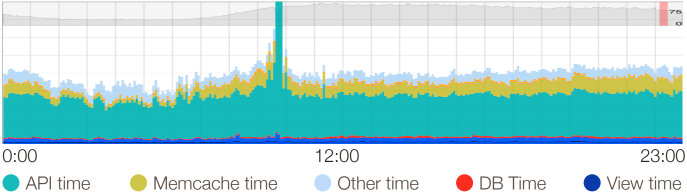
Evaluation Methodology: GQM


Goal
Assess Practicality of Approach
Questions
How precise is the detection?
How accurate is the detection?
Metric
Number of true positives
Metric
Number of false negatives

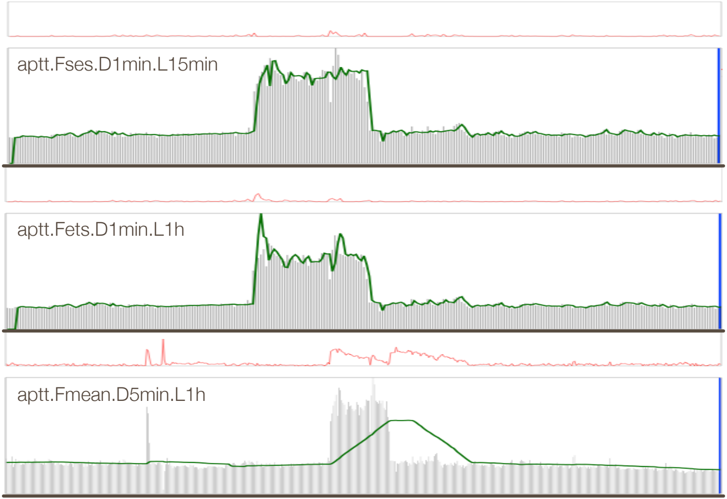
Goal/Question/Metric (GQM) plan (excerpt)

Manual Identification of Anomalies

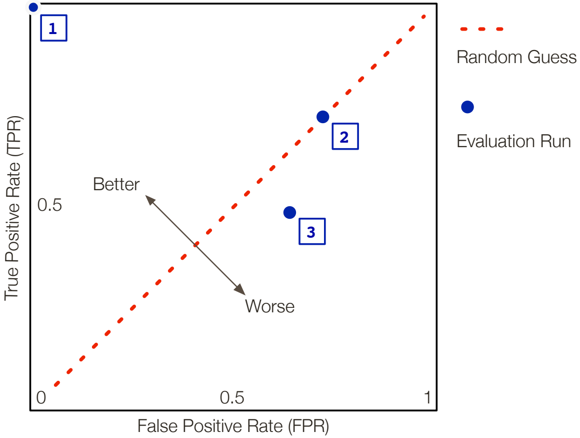
Evaluation Methodology (cont'd)



- Manual detection using the visualization tool
- 8 anomalies were detected


aptt.Fses.D1min.L15min
aptt.Fets.D1min.L1h
aptt.Fmean.D5min.L1h

ROC Curves (Introduction)

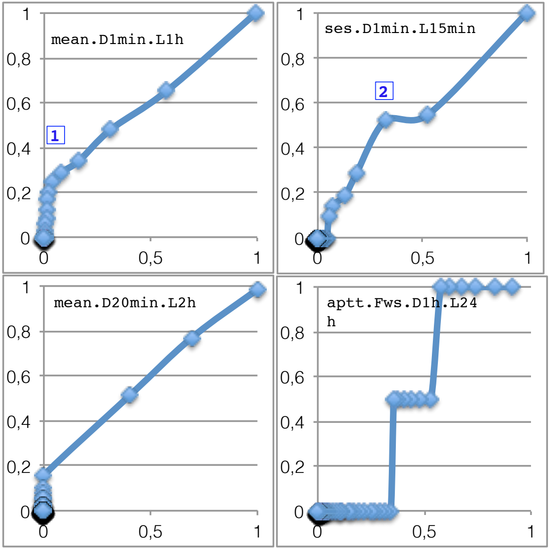


$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = \frac{\mathrm{TP}}{\mathrm{F}} \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} = \frac{\mathrm{FP}}{\mathrm{NF}} \tag{1}$$

ROC Curves (ΘPAD Results)

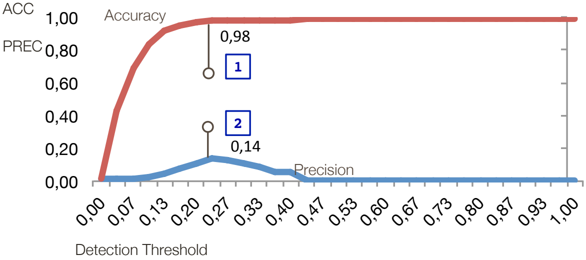
Evaluation (cont'd)

Results



Accuracy and Precision

Evaluation (cont’d)



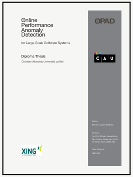
$$\mathrm{PREC} = \frac{\mathrm{TP}}{\mathrm{POS}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad . \tag{2}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \quad . \tag{3}$$

Summary and Outlook

Kieker

http://kieker-monitoring.net



ΘOnline Performance Anomaly Detection

for Large-Scale Software Systems

Diploma Thesis

Christian-Albrechts-Universität zu Kiel

ΘPAD

Author
Tillmann Carlos Bielefeld

Advisory
Prof. Dr. Wilhelm Hasselbring
Dipl.-Inform. André van Hoorn
Dr. Stefan Kaes (XING AG)

Case Study at
XING AG

Tillmann C. Bielefeld: ***"Online performance anomaly detection for large-scale software systems"*** March 2012. Diploma Thesis, Kiel Univ.

Outlook

- ΘPAD to be released as part of Kieker
- Follow-up theses on ΘPAD

Contact Us

- till@empuxa.com

Win a free empuxa Hoodie!
http://freebies.tielefeld.com