Intelligent System
Logging and Monitoring
ISyLM Research Lab

On the Path towards Intelligent Logging Practices

Wahab Hamou-Lhadj, PhD

ECE, Concordia University

wahab.hamou-lhadj@concordia.ca

IEA-AIE 2018, Montréal, QC

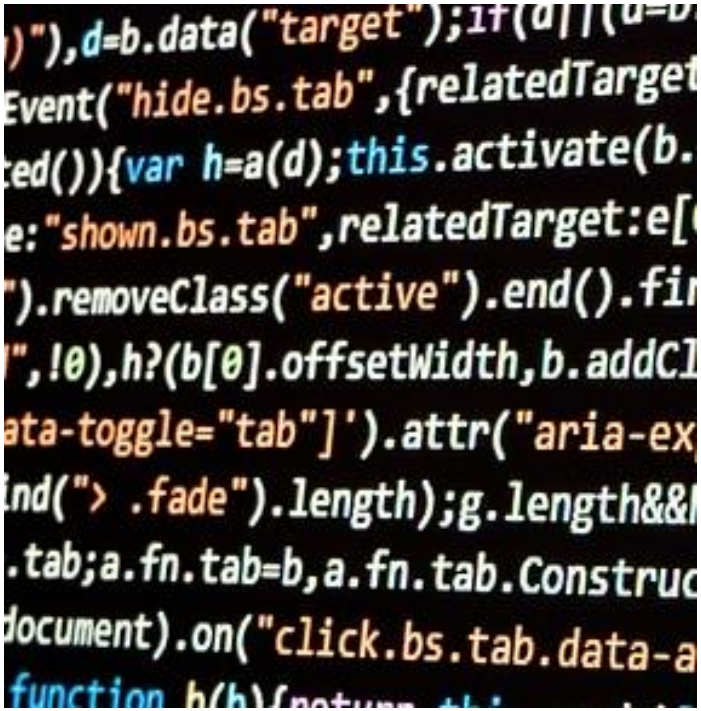
June 28, 2018

Software logging and machine data

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Software Logging

- A practice used by developers to debug and analyze software systems
- Various types of logs: user-defined, execution traces, profiling data, etc.
- Collectively known as **machine data**
- Contrasted with user data, which focus on user behavioral patterns

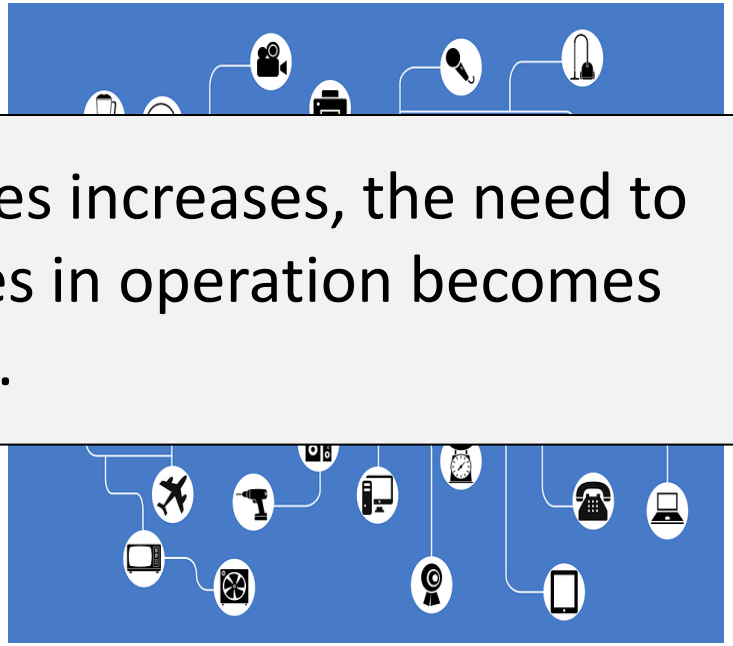


Why is it important? and why now?

- Astronomical increase in software-enabled machines
- Device connectivity and advances in IoT
- Software ecosystem complexity
- Emergence of machine intelligence paradigms

Why is it important? and why now?

- Astronomical increase in software-enabled machines
- Increasing machine complexity
- Emergence of machine intelligence paradigms



As our dependence on machines increases, the need to monitor and observe machines in operation becomes critical.

Why is it important? and why now?

- Astronomical increase in software-enabled machines
- Emergence of machine intelligence paradigms

We must start listening to machines

Listening to machines: The big picture

Industries

- Telecom
- Healthcare
- Energy
- Defense
- Manufacturing
- Finance
- Retail
- Education
- and more

Applications

- Fault Diagnosis
- IT Operations
- Operational Intelligence
- Fraud Detection
- Security Forensics
- Compliance Management

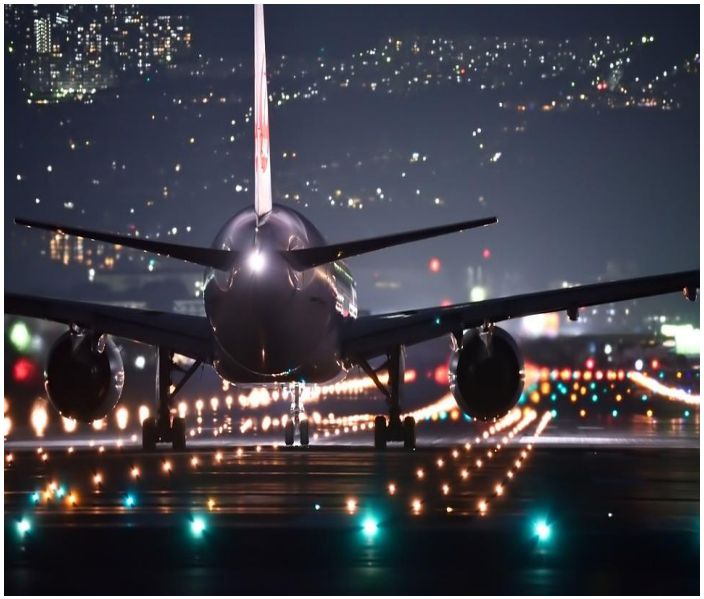
Technology and Processes

- Storage, Processing, Analytics, Visualization, Etc.
- Machine Data
- Logging and Tracing Infrastructure

Examples of industrial projects that use machine data

Using machine data to drive efficiency

- Dubai airport uses machine data to increase airport capacity by 30%
  - without any additional terminal space, infrastructure, or runways
- Machine data sources:
  - Flight schedules,
  - Wi-Fi network data
  - Metal detector data
  - Baggage system
  - Sensor data (bathroom doors, faucets, etc.),
  - Measurement cameras, etc.



Source: https://www.splunk.com/en_us/customers/success-stories/dubai-airports.html

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

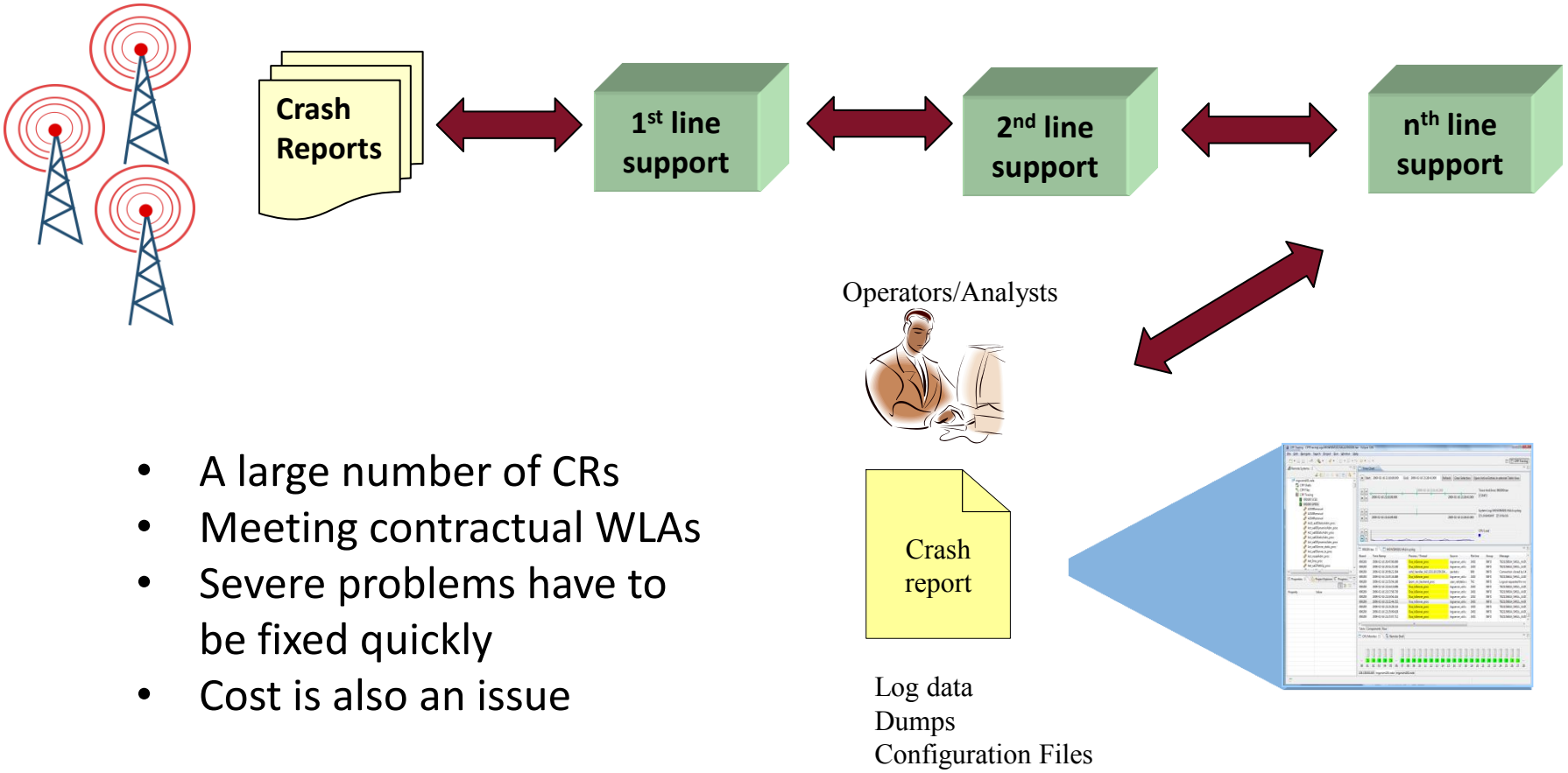
GDPR Compliance and Logs

- GDRP (EU's General Data Protection Regulation) came in effect in May 2018
- Enforces the principle of security and privacy by design
- Impact on PI (personal information) in log files:
  - minimization
  - encryption
  - access controls
  - audit

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Improving IT operations at Ericsson

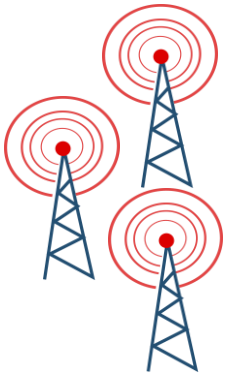
A research project between Concordia and Ericsson



- A large number of CRs
- Meeting contractual WLAs
- Severe problems have to be fixed quickly
- Cost is also an issue

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Improving IT operations at Ericsson



Crash Reports

1st line support

2nd line support

nth line support

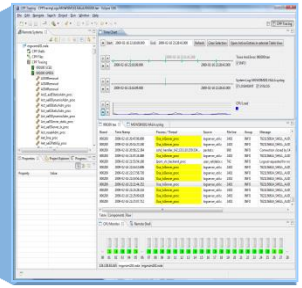
Operators/Analysts

- ML model of CRs
- Domain knowledge

D2K Database

Crash report

Log data
Dumps
Configuration Files



- **ReCRAC: CR Duplicate Detection Tool**
- **TotalADS: Anomaly Detection System**

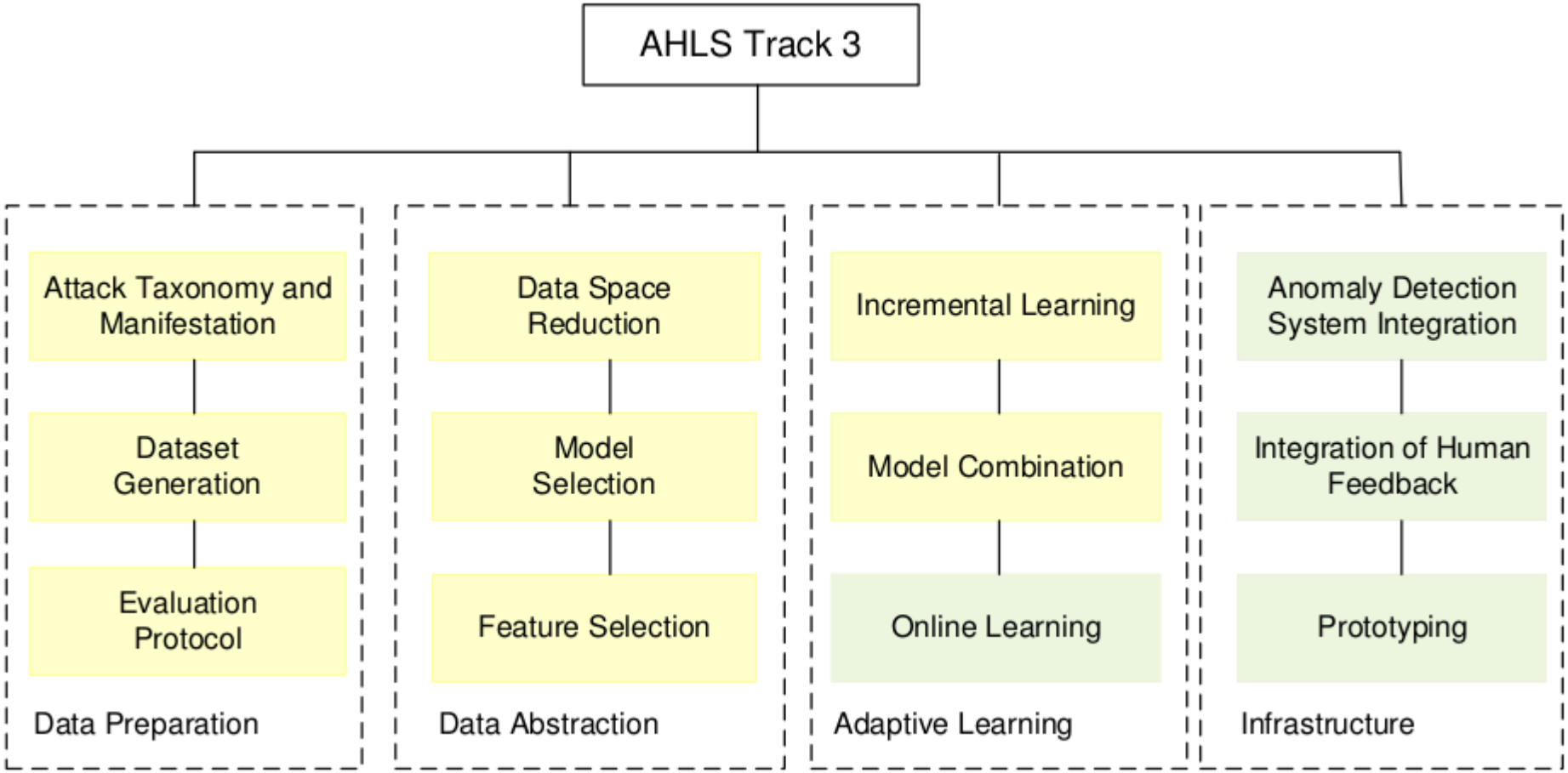
Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Harmonized anomaly detection techniques

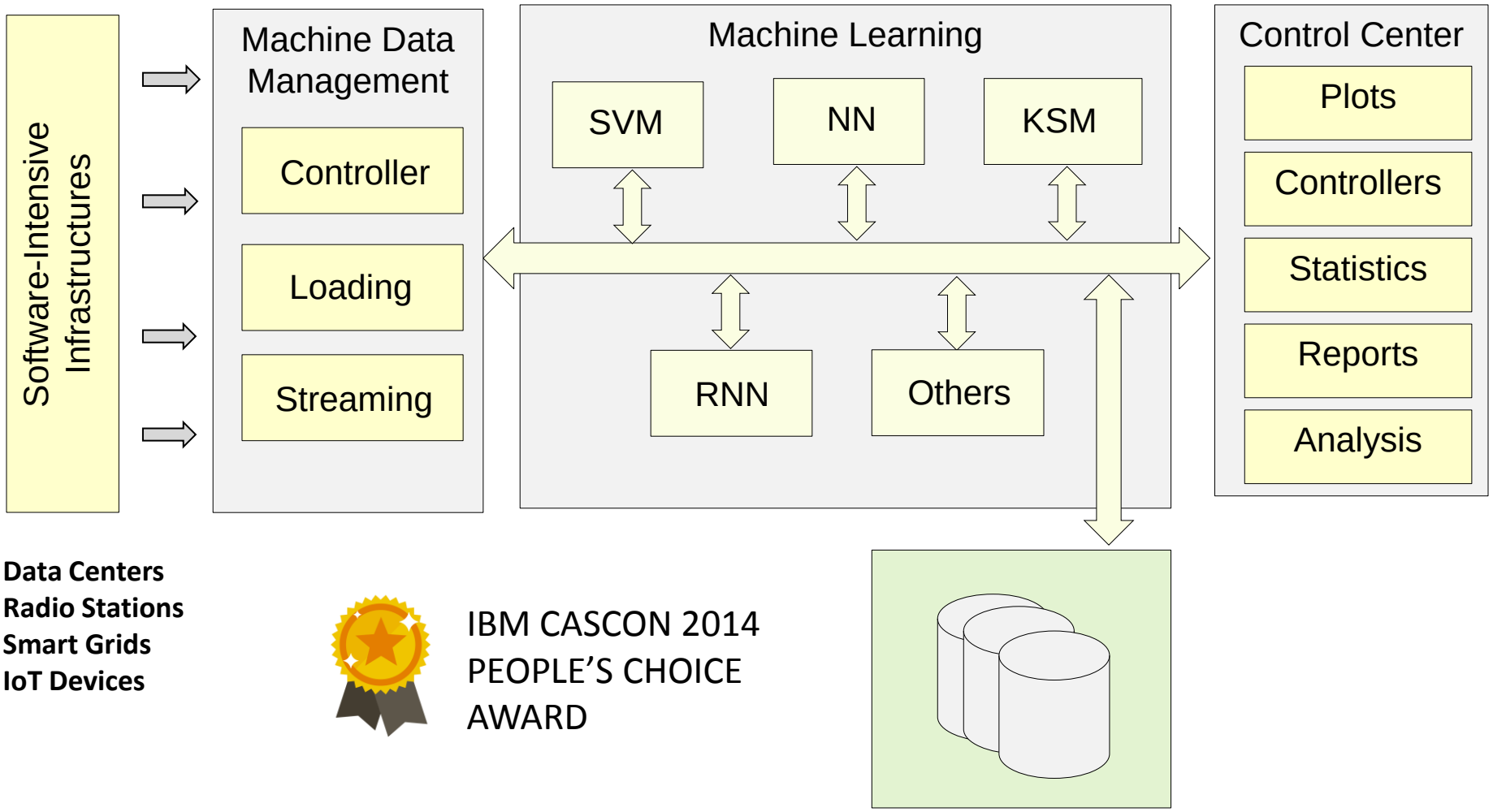
- A research project between Concordia, Defence R&D Canada and Ericsson
- Objectives:
  - Detection of abnormal behavior in computer hosts through the analysis of machine data
  - Combination of multiple machine learning techniques
  - Leverage of data abstraction, model combination, adaptive learning, and online learning
  - Tool development and integration

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Research Map



TotalADS: Total Anomaly Detection System Architecture



Data Centers
Radio Stations
Smart Grids
IoT Devices

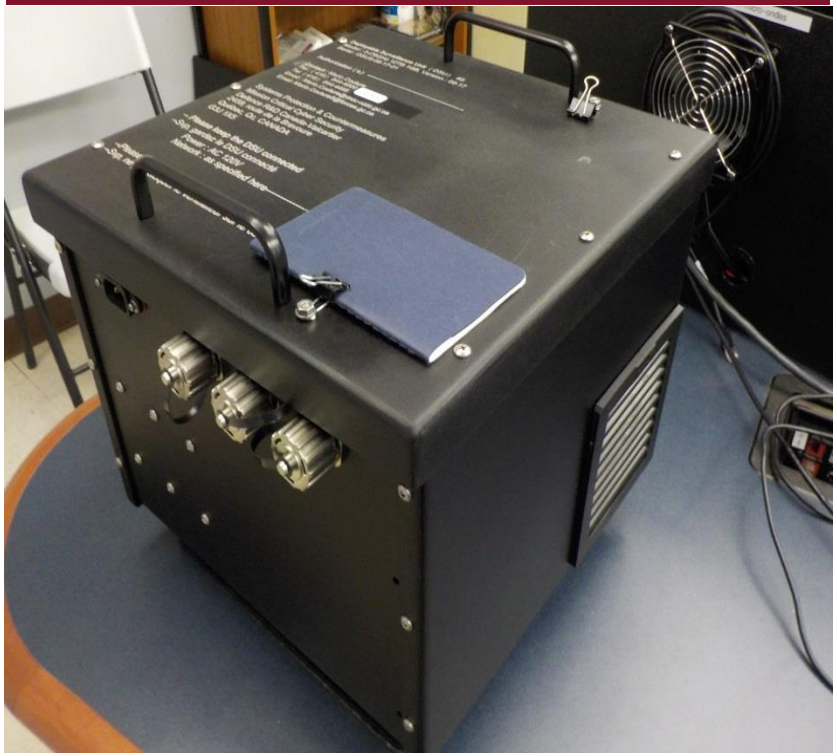
IBM CASCON 2014
PEOPLE'S CHOICE
AWARD

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

TotalADS and Deployable System Units (DSUs)

The 4th DSU prototype (PoC1)

The 5th DSU prototype (PoC2)

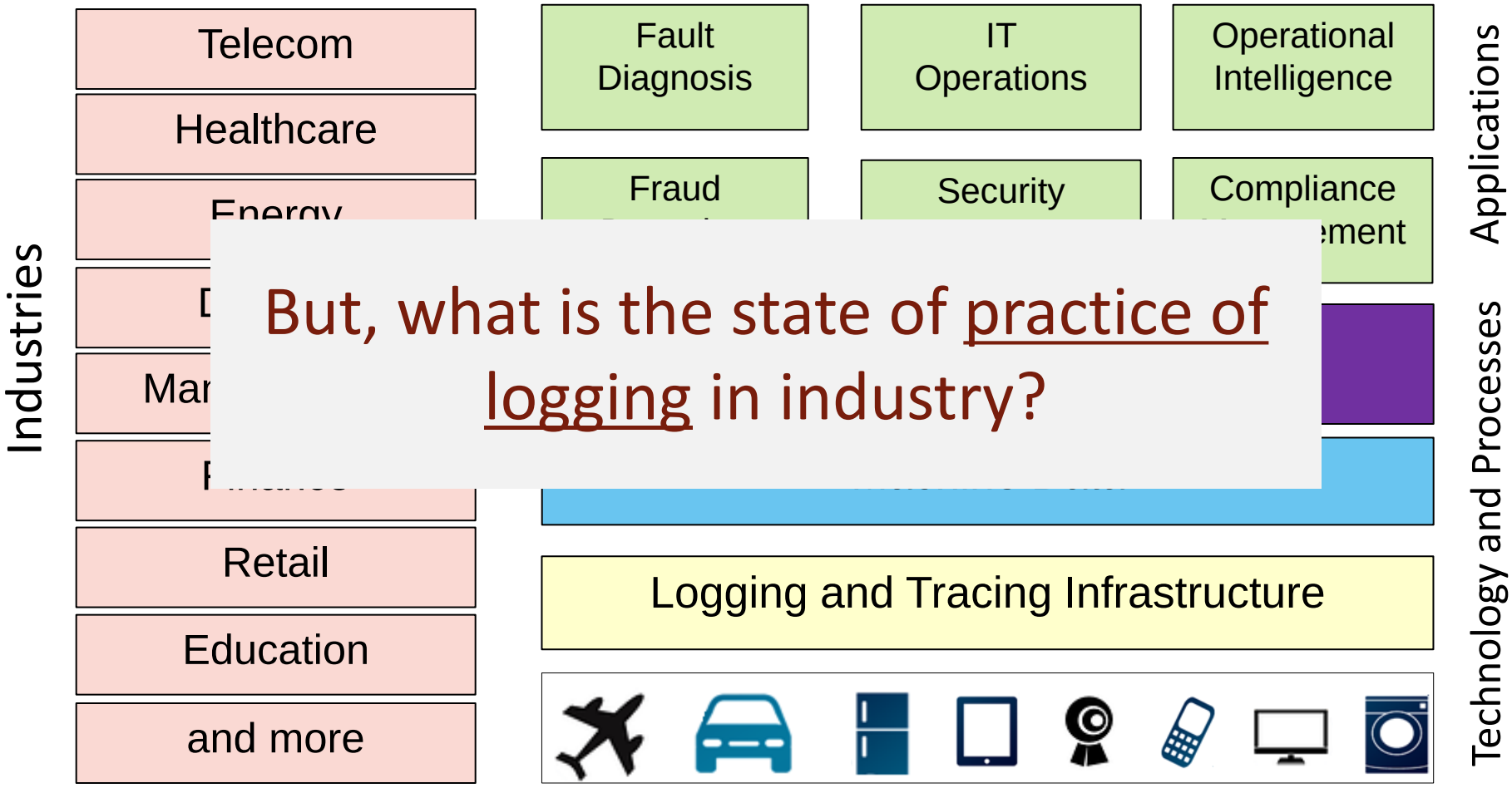


Six Jetway industrial Mini-ITX computers + one manageable GB switch + six 4-TB hard disks
(Intel's Haswell Core i7-4770TE 2.3 GHz processor, 8 GiB DDR3, 6 GB/sec mSATA, dual LAN)
(The whole DSU needs less than 350 watts when used at full capacity)

(Next technology to be considered: the new NVIDIA Jetson TX-2 AI computing board)

Couture, M., Fattahi, J., The use of Deployable Surveillance Units (DSUs) for online cyber surveillance—Proof of concept, Phase 1. Defence Research and Development Canada (DRDC), Report number: DRDC-RDDC-2018-R021, DRDC Valcartier Research Center, April 2018, Unclass.

Listening to machines: The big picture



But, what is the state of practice of logging in industry?

Let's take a look at the published work

Operational-Log Analysis for Big Data Systems

Challenges and Solutions

Andriy Miranskyy, Ryerson University

Abdelwahab Hamou-Lhadj, Concordia University, Montreal

Enzo Cialini, IBM

Alf Larsson, Ericsson

Where Do Developers Log?
An Empirical Study on Logging Practices in Industry

Qiang Fu[1], Jieming Zhu[2], Wenlu Hu[3], Jian-Guang Lou[1], Rui Ding[1], Qingwei Lin[1], Dongmei Zhang[1], Tao Xie[4]

[1]Microsoft Research Asia, Beijing, China {qifu,jlou,juding,qlin,dongmeiz}@microsoft.com

[2]The Chinese University of Hong Kong, HK, China jmzhu@cuhk.edu.hk

[3]Carnegie Mellon University, PA, USA wenlu@cmu.edu

[4]University of Illinois at Urbana-Champaign, IL, USA taoxie@illinois.edu

Characterizing Logging Practices in Open-Source Software

Ding Yuan‡†, Soyeon Park†, and Yuanyuan Zhou†

†University of California, San Diego, ‡University of Illinois at Urbana-Champaign

{diyuan,soyeon,yyzhou}@cs.ucsd.edu

The Game of Twenty Questions: Do You Know Where to Log?

Xu Zhao, University of Toronto

Kirk Rodrigues, University of Toronto

Yu Luo, University of Toronto

Michael Stumm, University of Toronto

Ding Yuan, University of Toronto

Yuanyuan Zhou, University of California San Diego

Characterizing logging practices in Java-based open source software projects – a replication study in Apache Software Foundation

Boyuan Chen[1] · Zhen Ming (Jack) Jiang[1]

Examining the Stability of Logging Statements

Suhas Kabinna[1], Weiyi Shang[2], Cor-Paul Bezemer[1],Ahmed E. Hassan[1]

Software Analysis and Intelligence Lab (SAIL),Queen's University,Kingston, Ontario [1]

Department of Computer Science and Software Engineering Concordia University,Montreal, QC, Canada[2],

Email:{kabinna, bezemer, ahmed}@cs.queensu.ca[1]

shang@encs.concordia.ca[2]

Key findings

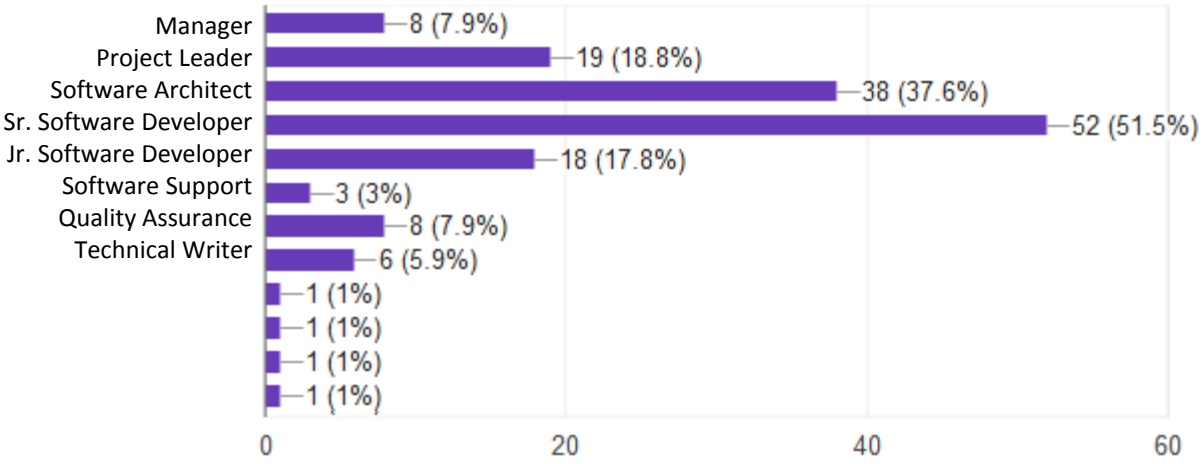
- Every 30 lines of code contains one line of logging code.
- The average change rate of logging code is almost two times compared to the entire code.
- In contrast to its small density, logging code is modified in a significant number of times.
- One third of modifications are after-thoughts.

Key findings (cont.)

- Developers seldom delete or move logging code.
- Developers spend significant efforts on adjusting the verbosity level of log messages.
- Developers are often confused when estimating the cost and benefit of each verbosity level.

What do developers think about this?

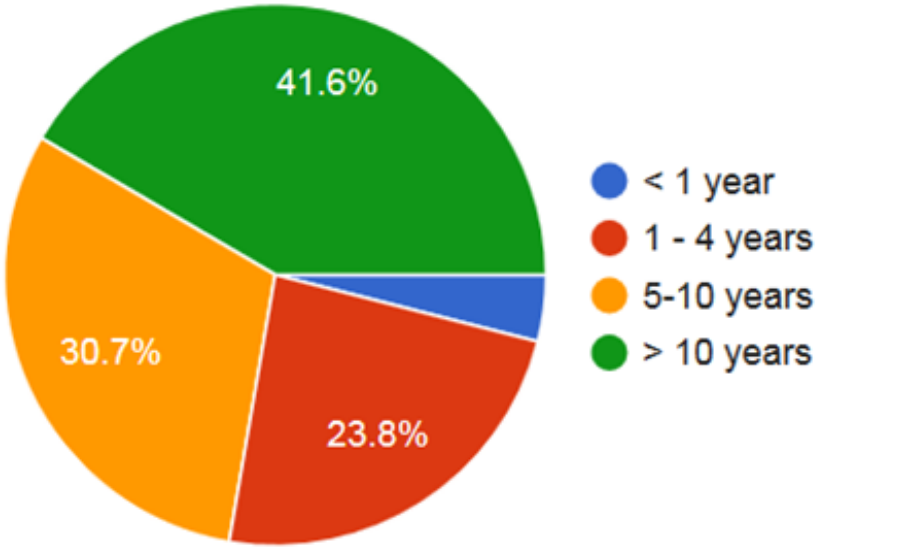
- We conducted an online survey with 102 practitioners from various companies



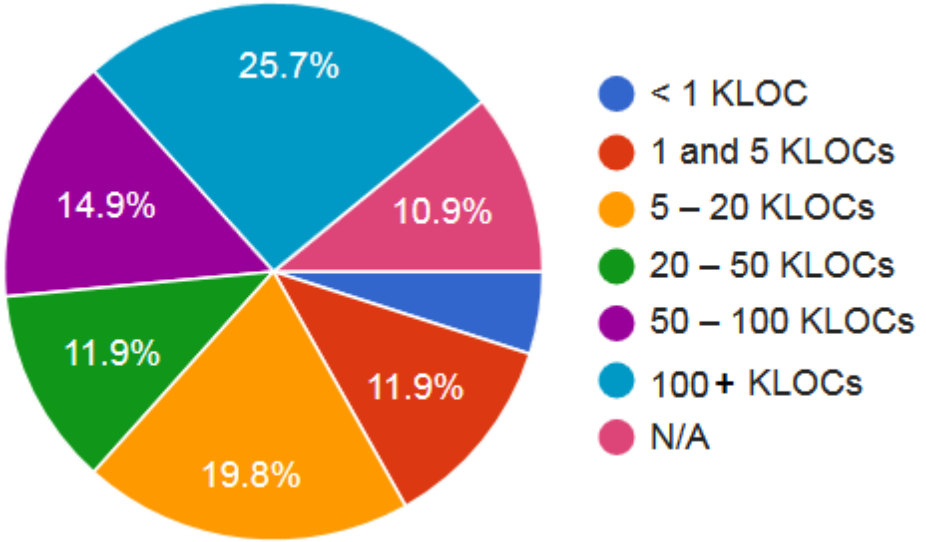
Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Participants' Background

How long have you been working in the software field?

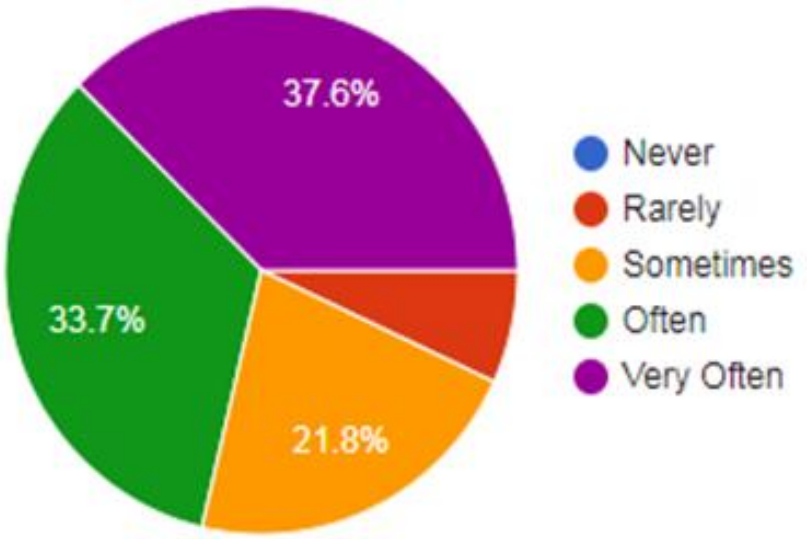


What is the size of your current (or recently completed) project in KLOCs?

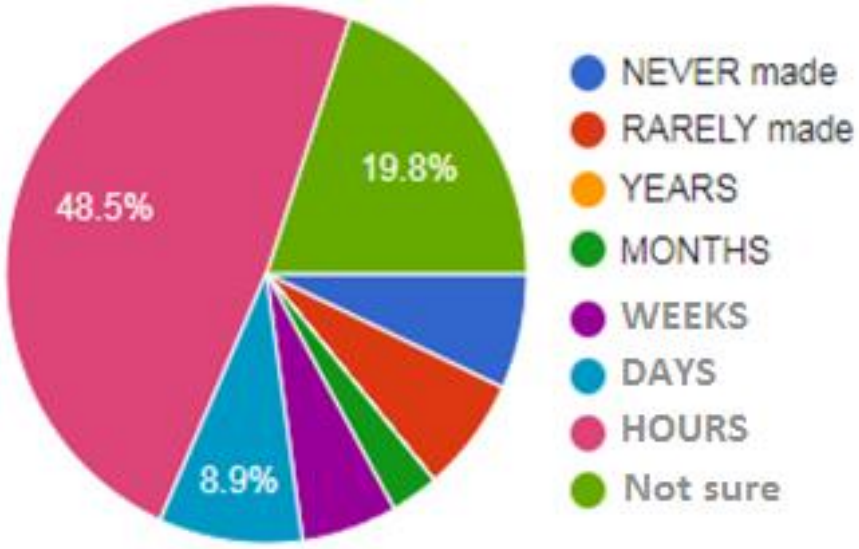


Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

How often do you consult the available logs when working on a software system?

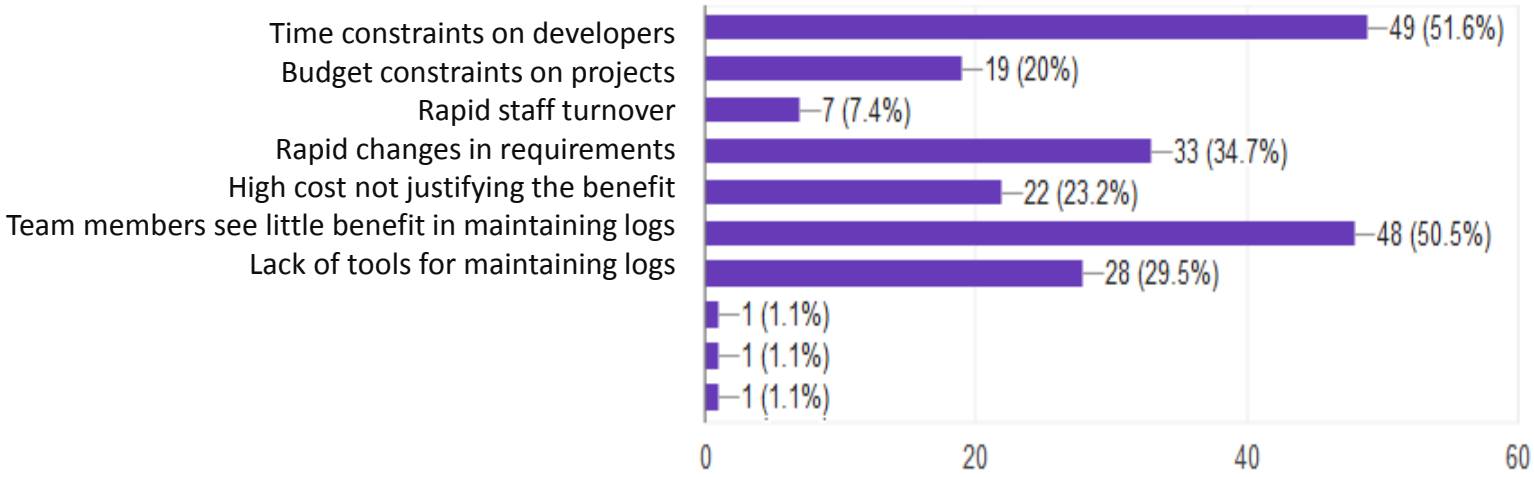


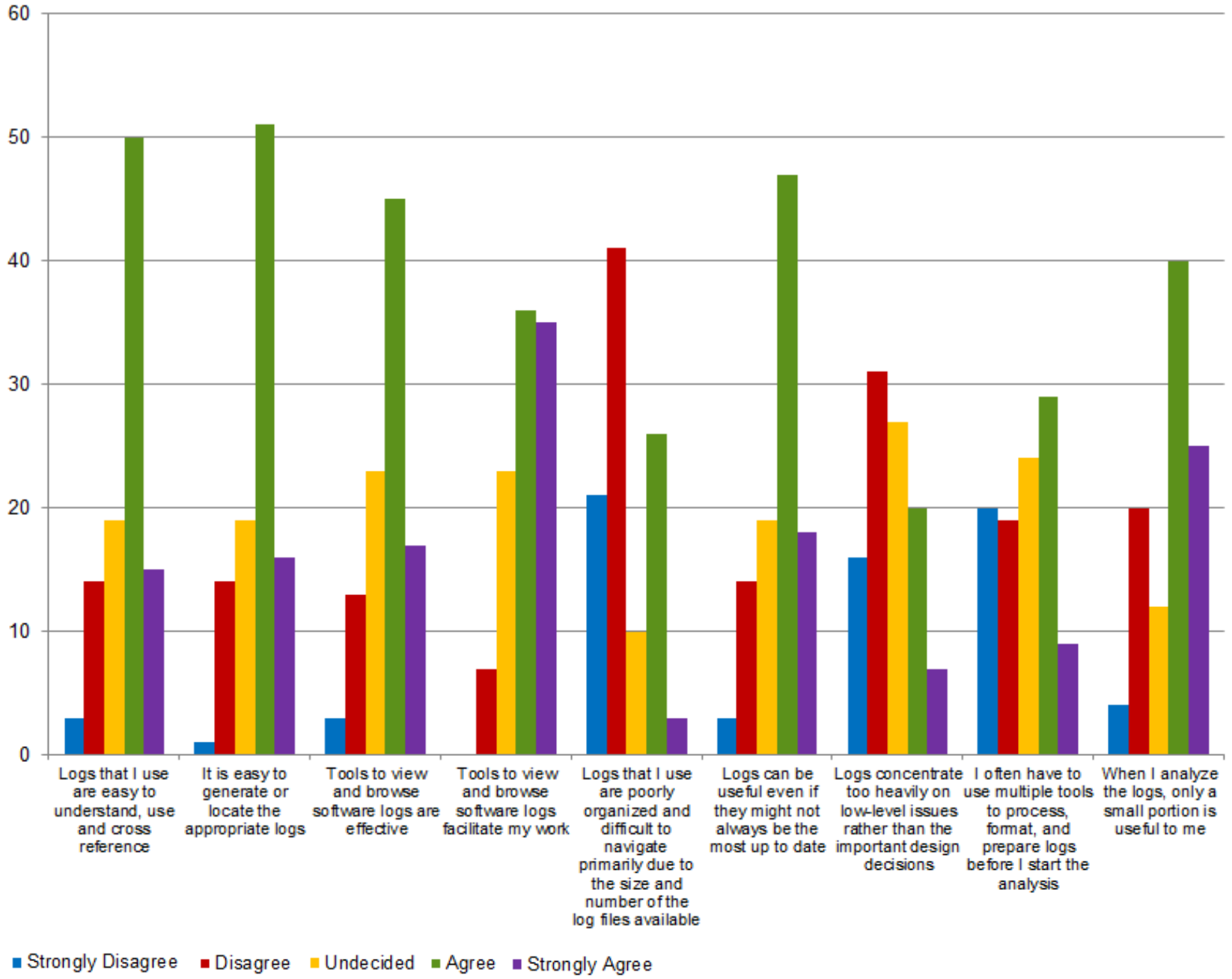
When changes are made to a system, how long does it take for the existing logs to be updated?



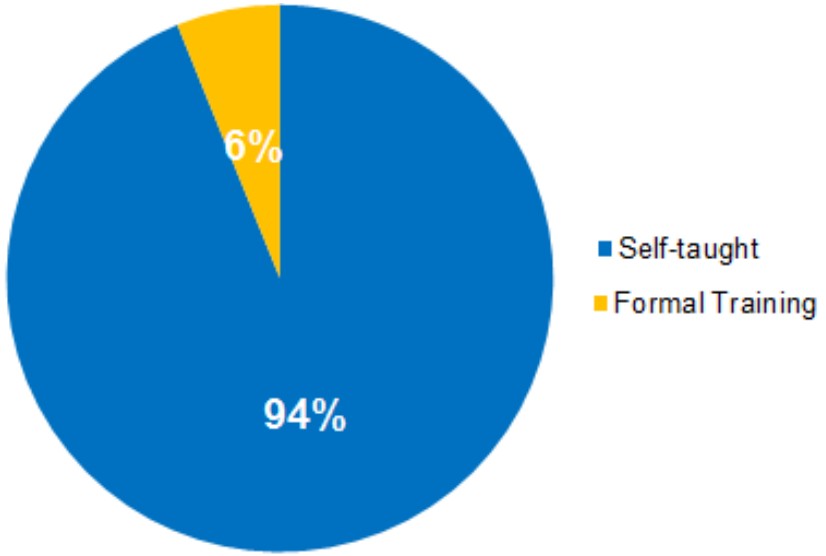
Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)

Which factors cause software logs to be out of sync with the system?

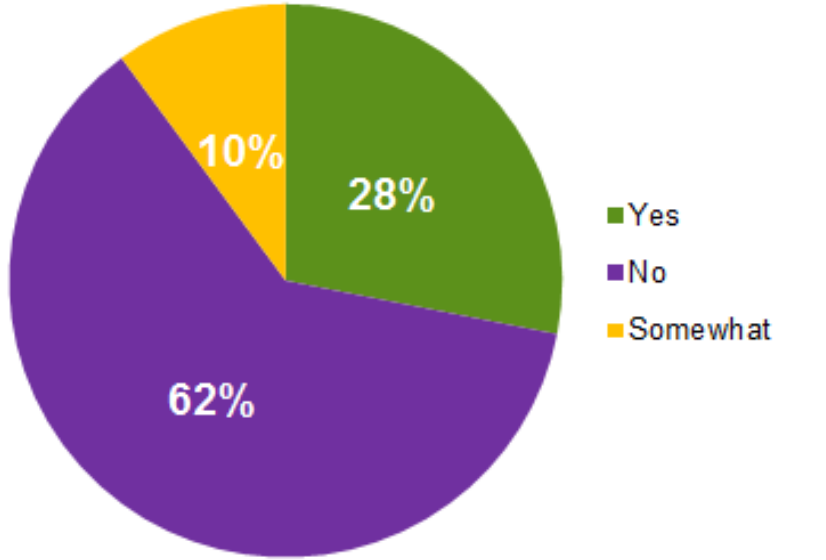



60
50
40
30
20
10
0
Logs that I use are easy to understand, use and cross reference
It is easy to generate or locate the appropriate logs
Tools to view and browse software logs are effective
Tools to view and browse software logs facilitate my work
Logs that I use are poorly organized and difficult to navigate primarily due to the size and number of the log files available
Logs can be useful even if they might not always be the most up to date
Logs concentrate too heavily on low-level issues rather than the important design decisions
I often have to use multiple tools to process, format, and prepare logs before I start the analysis
When I analyze the logs, only a small portion is useful to me
Strongly Disagree
Disagree
Undecided
Agree
Strongly Agree

Where have you learned the practice of logging?



Does your company have guidelines on how to write logs? To what extent they are followed?



Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs

Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs
- Logging statements are not always updated,
  - raising questions as to the validity of analysis

Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs
- Logging statements are not always updated,
  - raising questions as to the validity of analysis
- Logging is thought of during implementation
  - design and architectural decisions are not taken into account.

Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs
- Logging statements are not always updated,
  - raising questions as to the validity of analysis
- Logging is thought of during implementation
  - design and architectural decisions are not taken into account.
- Logging is pervasive in software development
  - adding to the cost of software projects

Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs
- Logging statements are not always updated,
  - raising questions as to the validity of analysis
- Logging is thought of during implementation
  - design and architectural decisions are not taken into account.
- Logging is pervasive in software development
  - adding to the cost of software projects
- Logging takes a significant part of software evolution
  - despite its relatively small presence

Implications

- The current logging practice is ad hoc
  - impacting log quality, hindering the benefits of logging, and increasing log analysis costs
- Logging statements are not always updated,
  - raising questions as to the validity of analysis
- Logging is thought of during implementation
  - design and architectural decisions are not taken into account.
- Logging is pervasive in software development
  - adding to the cost of software projects
- Logging takes a significant part of software evolution
  - despite its relatively small presence
- The current verbosity levels of logging tools are confusing
  - leading to inconsistencies

Where should we go from here?

Awareness

Awareness

Sense of Urgency

Awareness

Sense of Urgency

Proven Practices and Standards

Awareness

Sense of Urgency

Proven Practices and Standards

Design for Observability

Awareness

Sense of Urgency

Proven Practices and Standards

Design for Observability

Intelligent Logging

Awareness

Sense of Urgency

Proven Practices and Standards

Design for Observability

Intelligent Logging

Education

THANK YOU!

CONCORDIA.CA