# IOWA STATE UNIVERSITY



**Architecture Recovery** 

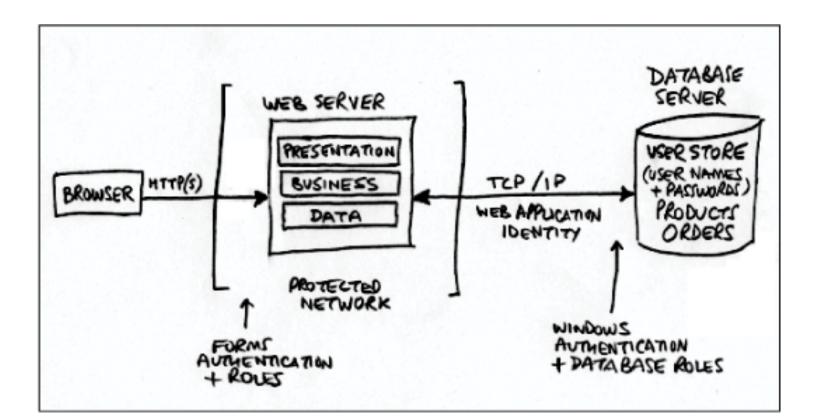
Lotfi ben Othmane Spring 2020

# Architecture Recovery

- What is it?
- Why do we need it?
- How to do it?

### **Prescriptive Architecture**

Prescriptive architecture describes the expected architecture of software (often designed one)



### **Descriptive Architecture**

# Descriptive architecture is the as-implemented architecture of software

Denothmane Update README.n	nd
rasp_pi_tests	Made some fixes that allowed the whole system to work together with a
server_test	Updating project with project files from GitLab
www	Update powerpoint slides
.gitignore	Remove services tests, Add .gitignore, Update website from old repo
Project Plan v1.pdf	Updating project with project files from GitLab
Project Timeline.xlsx	Updating project with project files from GitLab
README.md	Update README.md

# **Descriptive Architecture**

How would you get the descriptive architecture (asimplemented) of a software

- 1. Inspect the code
- 2. See the execution
- 3. Communicate with customers
- 4. Read the documentation
- 5. Other techniques

👮 Ibenothmane Update READM	IE.md
rasp_pi_tests	Made some fixes that allowed the whole system to work together with a
server_test	Updating project with project files from GitLab
www	Update powerpoint slides
Jitignore	Remove services tests, Add .gitignore, Update website from old repo
Project Plan v1.pdf	Updating project with project files from GitLab
Project Timeline.xlsx	Updating project with project files from GitLab
README.md	Update README.md

# Architecture and Software Evolution

 Software are developed Providing secure Authenticating configuration users Preventing in iterations and Handling Protectina parameter sensitive data exceptions manipulation versions Web Application Database Server Server Server Firewall Browser Web Applications Application Database Software architecture describes a given Protecting Authorizing Auditing Encrypting sensitive and logging or hashing users data activity and sensitive iteration/version of transactions data Preventing Validating Authenticating input session and authoring software hijacking and upstream cookie replay identities attacks

### Architecture Gap

- Prescriptive architecture describes the expected architecture of software (often designed one)
- Descriptive architecture is the as-implemented architecture of software
- Descriptive architecture and prescriptive architecture are often different

### Impacts of the Architecture Gap

- Architectural decisions made based on the perspective architecture may be wrong
- Mechanisms for quality attributes could be broken
- The assumption that the software satisfies a set of quality attributes may be wrong

# How to address the gap between prescriptive architecture and descriptive architecture?

# Solutions for the Architecture Gap

Alternatives:

- 1. Update the architecture document and redevelop the architecture diagrams
- 2. Recover the architecture of the software from the code

### **Descriptive Architecture**

- Ground truth the architecture of a software that has been verified as accurate by the architects.
- Architecture recovery is the extraction and analysis of a software architecture

# Architecture Recovery

- Architecture recovery is the extraction and analysis of a software architecture
- Current tools cluster the software code into packages
- Recovery techniques often are based on the call graph of the software

## Architecture Recovery Approaches

- Simple methods Use the folders' structure
- Text-based methods Mine the text code
- Dependency-based methods Use function calls network

#### This is not an exhaustive list

<b>Benothmane</b> Update README.	md
rasp_pi_tests	Made some fixes that allowed the whole system to work together with a
server_test	Updating project with project files from GitLab
i www	Update powerpoint slides
.gitignore	Remove services tests, Add .gitignore, Update website from old repo
Project Plan v1.pdf	Updating project with project files from GitLab
Project Timeline.xlsx	Updating project with project files from GitLab
README.md	Update README.md

# Call Graph

 Call graph represents the calling relationships between nodes. Each node represents a function (or module) and each edge (*f*, *g*) represents calls of function (or module) *f* to function (or module) *g*.

• Use code analysis tools to extract the CFG.

# **Graph Clustering**

cluster-4 • Architecture M6 recovery becomes cluster-2 a clustering M4 M8 problem Each method cluster-5 defines its own M2 M5 M7 **M**1 clustering feature M3

### Architecture Recovery Methods

- Algorithm for comprehension-driven clustering cluster code based on dominance
- Weighted combined algorithm cluster based on a distance metric between the clusters
- 3. LIMBO Uses a metric named distributed cluster feature
- 4. Architecture recovery using concerns uses information retrieval and machine learning techniques
- Zone base recovery based on natural language semantics of identifiers found in the code
- 6. Bunch cluster code using modularization metric

https://dl.acm.org/citation.cfm?id=2819022

### **Bunch Metric**

- Assumes: well-designed software systems are organized into cohesive subsystems that are loosely interconnected.
- Interconnectivity dependencies between the modules of two distinct subsystems
- Intra-connectivity dependencies between the modules of the same subsystem
- Modularization Quality trade-off between Interconnectivity and Intra-connectivity

### **Bunch Metric**

• Intra-connectivity - Coefficient of number of edges in the cluster to potential number of NODES in the cluster

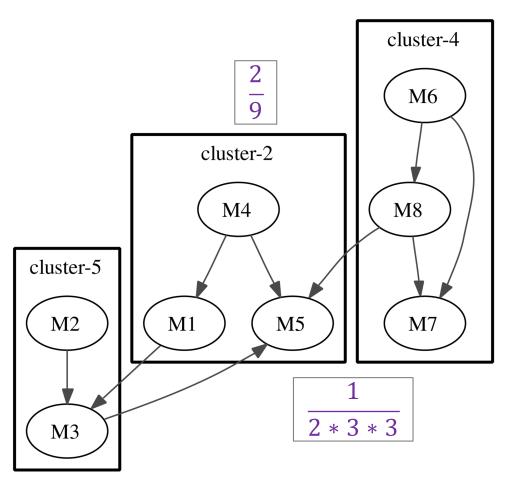
$$A_i = \frac{\mu_i}{N_i^2}$$

 Interconnectivity - Coefficient of number of edges between cluster *i* and cluster *j* to double the number of nodes of <u>cluster *i*</u> multiply by the number of nodes of cluster *j*. (0 if in the same cluster.)

$$E_{i,j} = \frac{\varepsilon_{i,j}}{2 X N_i X N_j}$$
  
Modularization Quality - 
$$\begin{cases} \frac{\sum_{i=1}^k A_i}{k} - \frac{\sum_{i,j=1}^k E_{i,j}}{\frac{k(k-1)}{2}}\\ A_1 & (k=1) \end{cases}$$

### **Bunch Metric**

It uses hill-climbing and genetic algorithms to solve the optimization problem



Compute the:

- 1. Intra-connectivity
- 2. Inter-connectivity

Of the module of the component: Data-Collector

https://github.com/lbenothmane/FleetManagement/t ree/master/rasp\_pi\_tests

### WALA CFG

### T.J. Watson Libraries for Analysis (WALA)

GitHub, Inc. [US]   https://github.com/wala/	WALA		
This repository Search	Pull requests Issues M	Aarketplace Explore	🌲 + - 🕅 -
📮 wala / WALA		● Watch ▼ 21	★ Star 146 <b>%</b> Fork 87
<> Code ① Issues 53 ♀ ♀ Pull r	equests 3 III Projects 0 III Wik	ki 🔟 Insights	
T.J. Watson Libraries for Analysis htt static-analysis java javascript and			
🕞 <b>5,343</b> commits	🔑 3 branches	🛇 11 releases	<b>25</b> contributors
Branch: master 🕶 New pull request		Create new file Upload file	s Find file Clone or download -
juliandolby Merge pull request #265 fro	om liblit/java-8-build-configuration		Latest commit 5ac8bf8 2 days ago
com.ibm.wala-repository	Rename "feature" to "feature" in su	bdirs and features	3 months ago

### WALA CFG

digraph "DirectedGraph" { graph [concentrate = true];center=true;fontsize=6;node [ color=blue.shape="box"fontsize=6.fontcolor=black.fontname=Arial];edge [ color=black,fontsize=6,fontcolor=black,fontname=Arial]; "Node: synthetic < Primordial, Lcom/ibm/wala/FakeRootClass, fakeRootMethod()V > Context: Everywhere" [ label="Node: synthetic < Primordial, Lcom/ibm/wala/FakeRootClass, fakeRootMethod()V > Context: Everywhere"] "Node: synthetic < Primordial, Lcom/ibm/wala/FakeRootClass, fakeWorldClinit()V > Context: Everywhere" [ label="Node: synthetic < Primordial, Lcom/ibm/wala/FakeRootClass, fakeWorldClinit()V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/SecureEngine, main([Ljava/lang/String;)V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/util/SecureEngine, main([Ljava/lang/String;</pre> )V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/SecureEngine, <clinit>()V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/util/SecureEngine, <clinit>()V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/Util, main([Ljava/lang/String;)V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/util/Util, main([Ljava/lang/String;)V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/Util, <clinit>()V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/util/Util, <clinit>()V > Context: Everywhere"] "Node: < Application, Lorg/compiere/plaf/CompierePLAF, main([Ljava/lang/String;)V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/plaf/CompierePLAF, main([Liava/lang/String; )V > Context: Everywhere"] "Node: < Application, Lorg/compiere/plaf/CompierePLAF, <clinit>()V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/plaf/CompierePLAF, <clinit>()V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/Language, main([Ljava/lang/String;)V > Context: Everywhere" [ label="Node: < Application, Lorg/compiere/util/Language, main([Ljava/lang/String;)V > Context: Everywhere"] "Node: < Application, Lorg/compiere/util/Language, <clinit>()V > Context: Everywhere" [ label="Node: < Application lorg/compiere/util/language <clinits()V > Context: Everywhere"

# Case Study - Compiere

### Overview

Compiere ERP+CRM is the leading open source ERP solution for...

Read More

Compiere ERP + CRM Business Solution Web Site »

iona Fielb				and the second succession of the		ew Result						
	VX AB Q 2 T D	Cherl Cardenting		Organization (say		Inalise ated Receipts	Organization.	Product	Busidess Partner A	ecount Date   P	eded. Account	
Tent Control of the local designation of the l	The second in the second is the	Decommend the appeal		Order Pathonnes Broke Mathematics 1204		Intervalie - Trade	HO		Joe Block	02/04/2005 F	6-04	
and the second se	100 (Del	Description Contar Searchpton from Business P	when Darksburk Johnson			leceivable - Trade	HQ.		Joe Block	01/01/2005.Ju		
Client Qardeninund	Organization *	Target Document Taxe   PDB Drate	101			leceivable - Trade	(HQ		Joe Block	01/05/2005.14		
Name (GardersWorld Admin		Target Decaheret Tare (PCR Dealer				teceivable - Trade	HO		Joe Block	01/26/2005.Ja		
Description or ardenmond Admin Decription		Business Parket (care Construction	42690300	mine Patiet Care Care Conductor		lecelvable - Trade	HQ		Joe Block	02/04/2005 F a		
		Patter London (Bended	1.81	Invoice Catalon   Larv Care Conduction	-	tecelvable - Trade	HQ		Joe Block Joe Block	01/25/2005.Ja		
P Active						tecetvable - Trade tecetvable - Trade	H9		Joe Block	01/25/2004.Ju 02/04/2005.Fe	5-04	
User Level Ctent-Organization	C Manual	UnderControl ( circle in the second		motors Contact [-] or Pret		leceivable - Trade	00		Joe Block	11/01/2003 N		
	La Martine	Wardhans had wardened				Ancohrable . Trade					•	
Currency USO *	Approval Amount	Banking .				Accession of the second	HO		Joe Block	11/01/2003 No	w-03	
P Approve own Docum Record Indo 1/3 - Bale		Price Lind (Purchase)		Carmers 1.60								1 0
		Eater Representative during Administrative		M Discout Points		set	00	Azalea Bush	Joe Block	12/20/2003 D4	0-03	
	40 4, 2005 10 15 32 PM EET	Bin Con.				iset						
rence Level Chant		Particular Statement				enue .		Azales Bush		11/01/2003 /v		
(AD_Rote - AD_Rote_D+1	10	Print				inve	940		Joe Block Joe Block	01/01/2005.Ja		
Column Column	enve Value Ond Value Ubr	Carlorage				Incia Incia	HQ H0		Joe Block	01/26/2005.34		
Description Oarder/Von	# Admin Decription VoardenWorld Admin (Supe	Baba				1000	HO		Joe Block	01/25/2004.Ju		
😥 Can Report		Totar Union	6.00	Orana Tatal	4.04	inue	HQ		Joe Block	01/05/2005.Ja		
Personal Lock		Document Status Courtee		Converse free (** says **		TOTAL OF	1					- 0
C. C	and the second se	Convictoria		Al Company								1,999
						Later in provide the later						
									and the second sec		splay criteria an	

https://sourceforge.net/projects/compiere/

# Case Study - Compiere

Suppose that you are asked to develop a new feature for Compiere, how would you proceed?

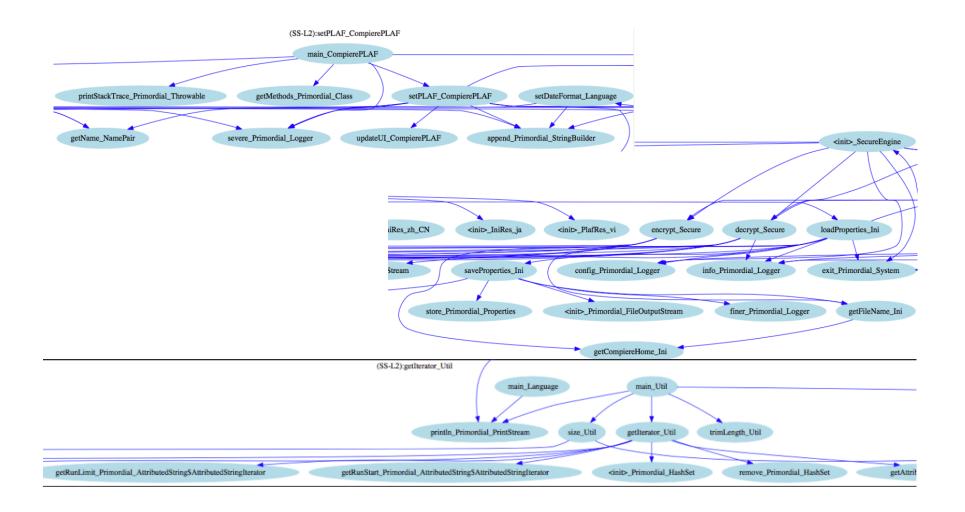
- 1. Read the source code
- 2. Give it to an expert with the software
- 3. Read the documentation
- 4. Say you cannot do it
- 5. Recover the architecture of the software

### Case Study - Compiere

How would you recover the architecture of Compiere?

- 1. Use the folders structure
- 2. Use a cluster-based approach
- 3. Use a text-based approach

### Case Study – Compiere - Clusters



# Example of Challenges

- 1. Code analysis tools add fake nodes.
- Code includes dependency e.g., Java-based code includes JRE java methods.
- 3. Methods may have similar names but in different modules may confuse code analysis.
- 4. The number of clusters is huge, e.g., 2000

### Case Study – Compiere - Clusters

- 1. Util main class for utilities
- 2. New Instance when user logs in, creates new instance with respective look & feel and language
- 3. Logger generation and storage of logs
- 4. Secure Engine responsible for implementing security policies within the application and initializing security
- 5. NameValuePair stores/retrieves/modifies user related data
- 6. List Resources lists user resources on login
- 7. Main\_CompierePLAF provides look & feel
- 8. GetLanguage retrieves the language for user
- 9. Encrypt\_SecureEngine provides data encryption capabilities
- 10. Hashing stores hashmap of user data

# Summary

- Prescriptive architecture describes the expected architecture of software (often designed one)
- Descriptive architecture is the as-implemented architecture of software
- Descriptive architecture and prescriptive architecture are often different
- Architecture recovery is the extraction and analysis of a software architecture

Thank you

**Questions?**