SCAN.COVERITY.COM

Open Source Report 2008 and the Architecture Library

July, 2009

David Maxwell
Coverity's Open Source Strategist

For
OSCON 2009

Agenda

- Background on Scan
- Open Source Report Findings from 2008 and 2009
- Rung Promotions
- New Scan Developments
- Architecture Analysis
- Q & A

Background

- Coverity Prevent selected by US Dept. Homeland Security
  - Cyber Security Open Source Hardening Project

- Analyze Open Source codebases with Static Analysis
  - Over 10,000 defects fixed, since project launch, March 2006

- Coverity Prevent is a static code analysis tool that delivers
  - Path Simulation
  - Data Flow Analysis
  - False Path Pruning

By understanding possible code execution paths, defects are identified and eliminated by Open Source developers

Report on Open Source Software 2008

- Original Scan/Coverity research
- Based on the analysis of
  - Over 55 million lines of code
  - From more than 250 open source projects
  - Representing 14,238 individual project analysis runs
  - Totaling nearly 10 billion lines of code analyzed

- All analysis performed with the same tools and configuration
  - Scan Benchmark 2006

Report on Open Source Software 2009

SCAN.COVERITY.COM

- Original Scan/Coverity research
- Based on the analysis of
  - Over **60** million lines of code
  - From more than 250 open source projects
  - Representing **26,181** individual project analysis runs
  - Totaling over **11.5** billion lines of code analyzed

- All analysis performed with the consistent tools and configuration
  - Scan Benchmark 2006, Scan Benchmark 2007

Defect Count or Defect Density

- Defect Counts
  - Absolute number of defects identified in a particular piece of code
    - 314 defects in a particular codebase

- Defect Density will be referred to many times during this session
  - Number of defects per 1,000 lines of code
    - 1.0 = 1 defect in 1,000 lines of code
    - 0.5 = 1 defect in 2,000 lines of code

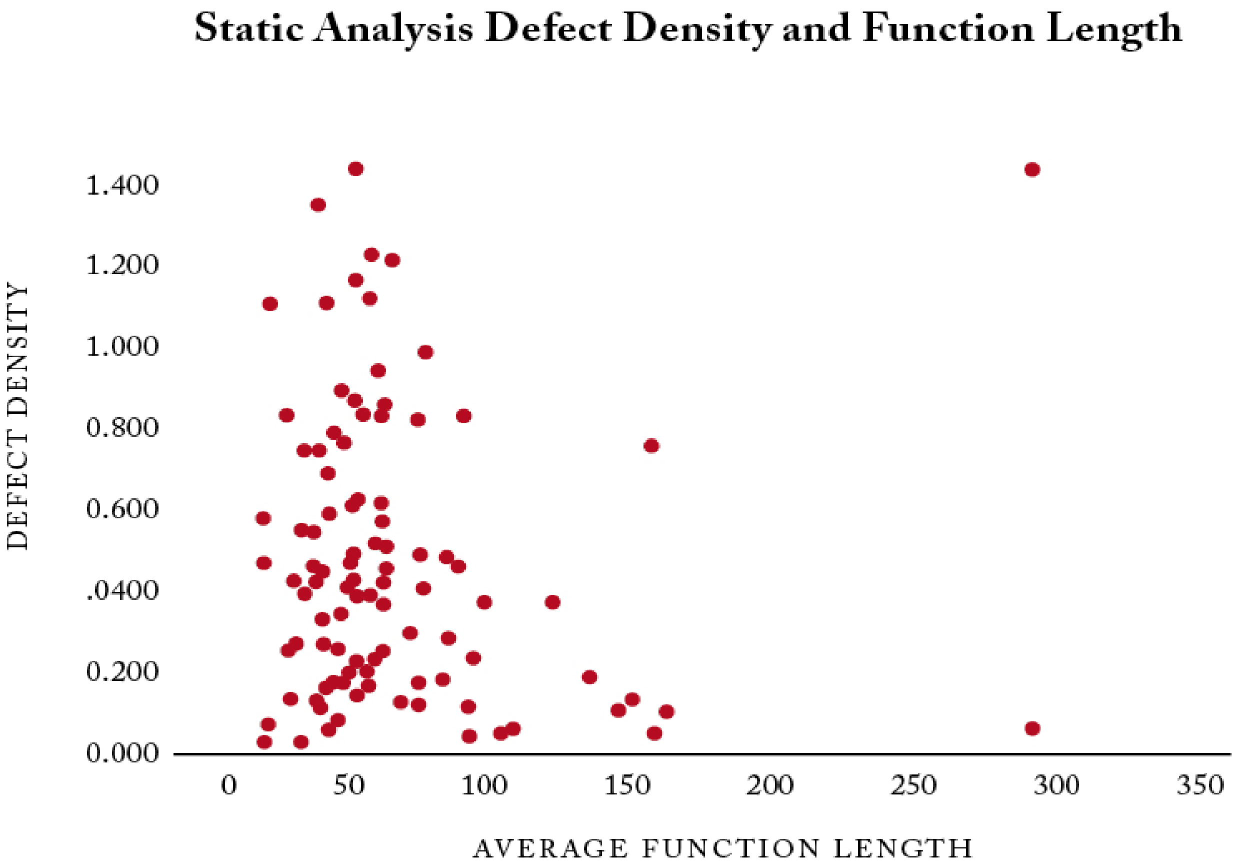
Function Length

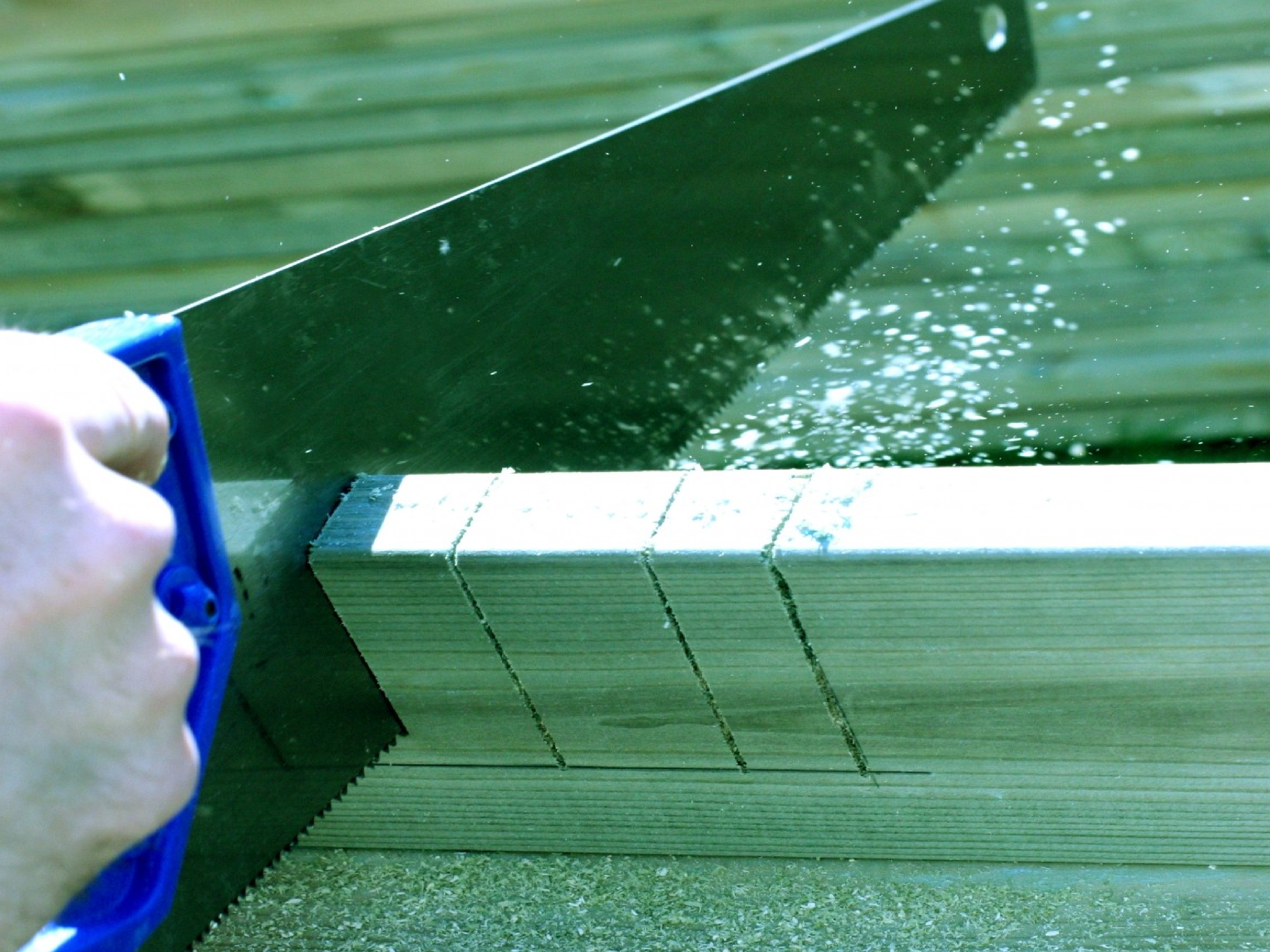
- What makes a function ‘long’?
  - –A single, sequential set of operations
    - •Are those operations common elsewhere in the code?
  - –A large switch statement
    - •Protocol decoding is a common example
  - –A function with many different code paths
    - •Conditional execution – lots of if() statements

- Average function lengths in the Scan database ranged from
  - –Low of 14 lines
  - –High of 345 lines
  - –The longest average is almost 25x the shortest average

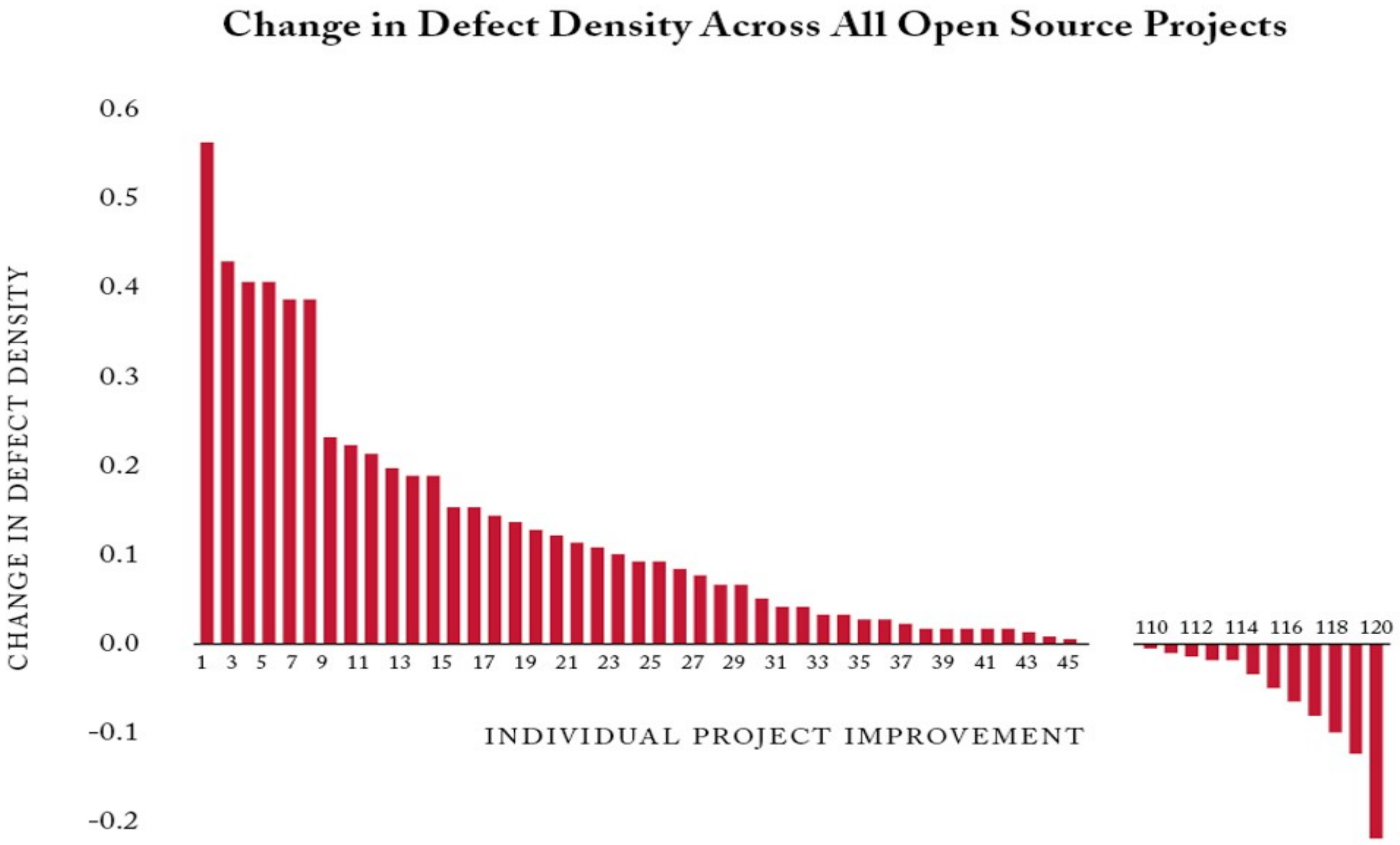
Defect Density/Function Length

SCAN.COVERITY.COM





Overall Project Progress



Software Tools

SCAN.COVERITY.COM

Version Control

Bug Trackers

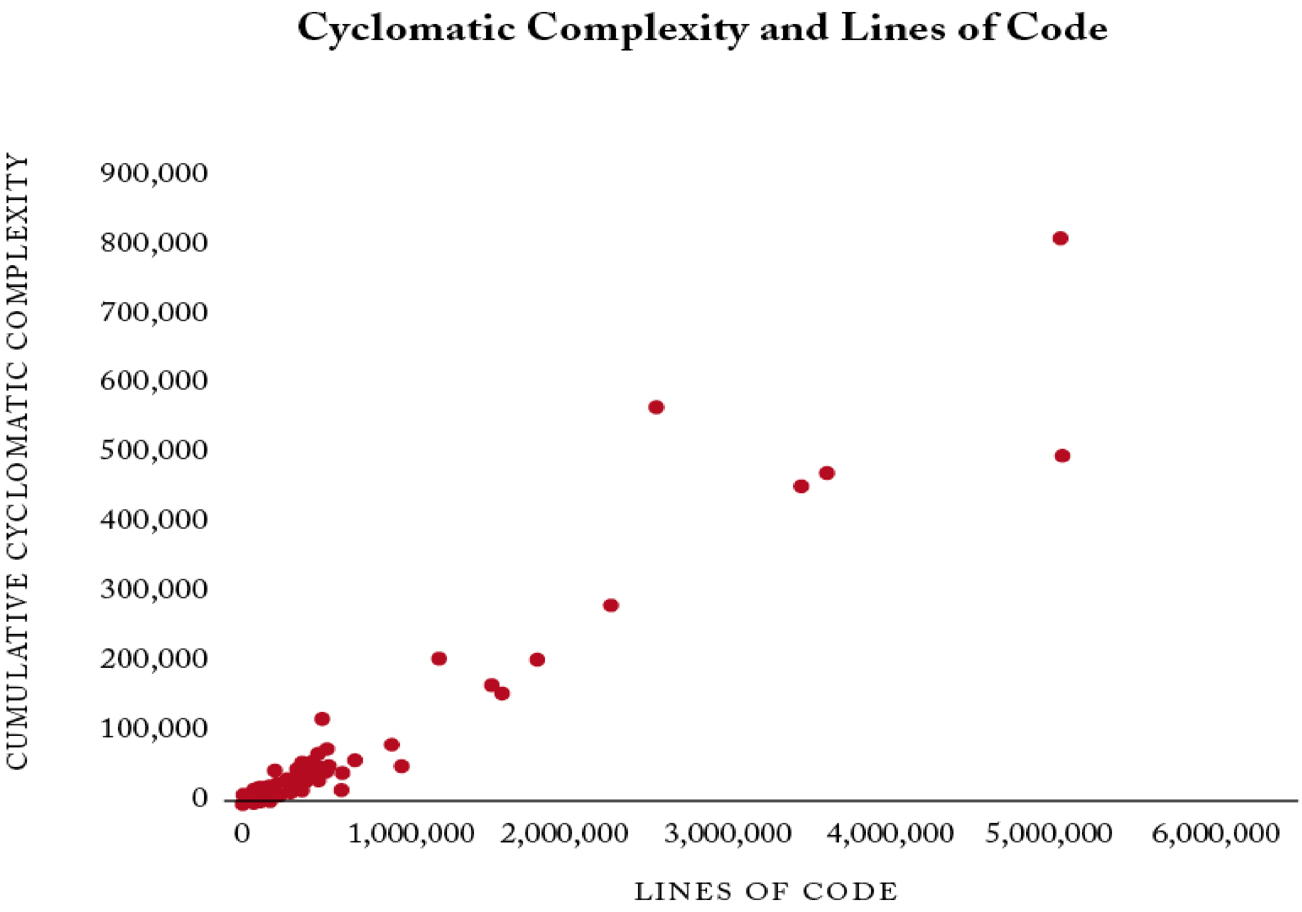
Debuggers

Frequency of Defects

Defect Type	# of Defects	Percentage
NULL Pointer Dereference	6,448	27.95%
Resource Leak	5,852	25.73%
Unintentional Ignored Expressions	2,252	9.76%
Use Before Test (NULL)	1,867	8.09%
Buffer Overrun (statically allocated)	1,417	6.14%
Use After Free	1,491	6.46%
Unsafe use of Returned NULL	1,349	5.85%
Uninitialized Values Read	1,268	5.50%
Unsafe use of Returned Negative	859	3.72%
Type and Allocation Size Mismatch	144	0.62%
Buffer Overrun (dynamically allocated)	72	0.31%
Use Before Test (negative)	49	0.21%

Cyclomatic Complexity/Lines of Code

SCAN.COVERITY.COM



Summary Findings

- •Open source benefits from static analysis
  - –Overall defect density dropped 16% over the past two years
- •Prevalence of individual defect types
  - –Defect frequency may directly relate to frequency of types of operations
- •False positives identified to date are a reasonably small percentage of results
  - –Currently below 14%

Continuing Improvement in 2009

- A Number of Projects have been promoted to higher Rungs
  - They have resolved all defects identified on their current Rung

Continuing Improvement in 2009

- A Number of Projects have been promoted to higher Rungs
  - They have resolved all defects identified on their current Rung

- Promoted to Rung 2
  - claws-mail, clusterit, Courier-authlib, Courier-maildir, curl, dialog, freeradius, gphoto2, iksemel, libexif, libsndfile, libvorbis, libwpd, mksh, ntp, ruby, parrot, squidGuard, speex, tcl, tor, vim

Continuing Improvement in 2009

SCAN.COVERITY.COM

- A Number of Projects have been promoted to higher Rungs
  - They have resolved all defects identified on their current Rung

- Promoted to Rung 2
  - claws-mail, clusterit, Courier-authlib, Courier-maildir, curl, dialog, freeradius, gphoto2, iksemel, libexif, libsndfile, libvorbis, libwpd, mksh, ntp, ruby, parrot, squidGuard, speex, tcl, tor, vim

- Ready for Rung 2
  - libpcap, nmap, OpenLDAP, theora, Transmission

Continuing Improvement in 2009

- A Number of Projects have been promoted to higher Rungs
  - They have resolved all defects identified on their current Rung

- Promoted to Rung 2
  - claws-mail, clusterit, Courier-authlib, Courier-maildir, curl, dialog, freeradius, gphoto2, iksemel, libexif, libsndfile, libvorbis, libwpd, mksh, ntp, ruby, parrot, squidGuard, speex, tcl, tor, vim

- Ready for Rung 2
  - libpcap, nmap, OpenLDAP, theora, Transmission

- Ready for Rung 3 (forthcoming)
  - Samba, tor, OpenPAM, ruby

DHS Contract Expiry

- Three year contract is over

Scan and DHS Contract

- Three year contract is over

- Coverity is committed to improving open source security
- Scan will continue, and continue to grow...

- Starting with the following 2 initiatives:

Scan Ladder - Rung 3

SCAN.COVERITY.COM

- Scan Ladder denotes success of defect elimination
  - Open Source projects progress as they resolve defects

- Supports common analysis configuration on a given Rung
  - Allows comparison and statistical analysis

- Rung 2 announced January 2008

- Rung 3 to be announced Fall 2009
  - New Prevent version
  - More Security checkers
  - Concurrency checkers
  - Boolean Satisfiability for False Path Pruning

The Architecture Library

- A public database of implementation diagrams, for over 2,500 open source projects

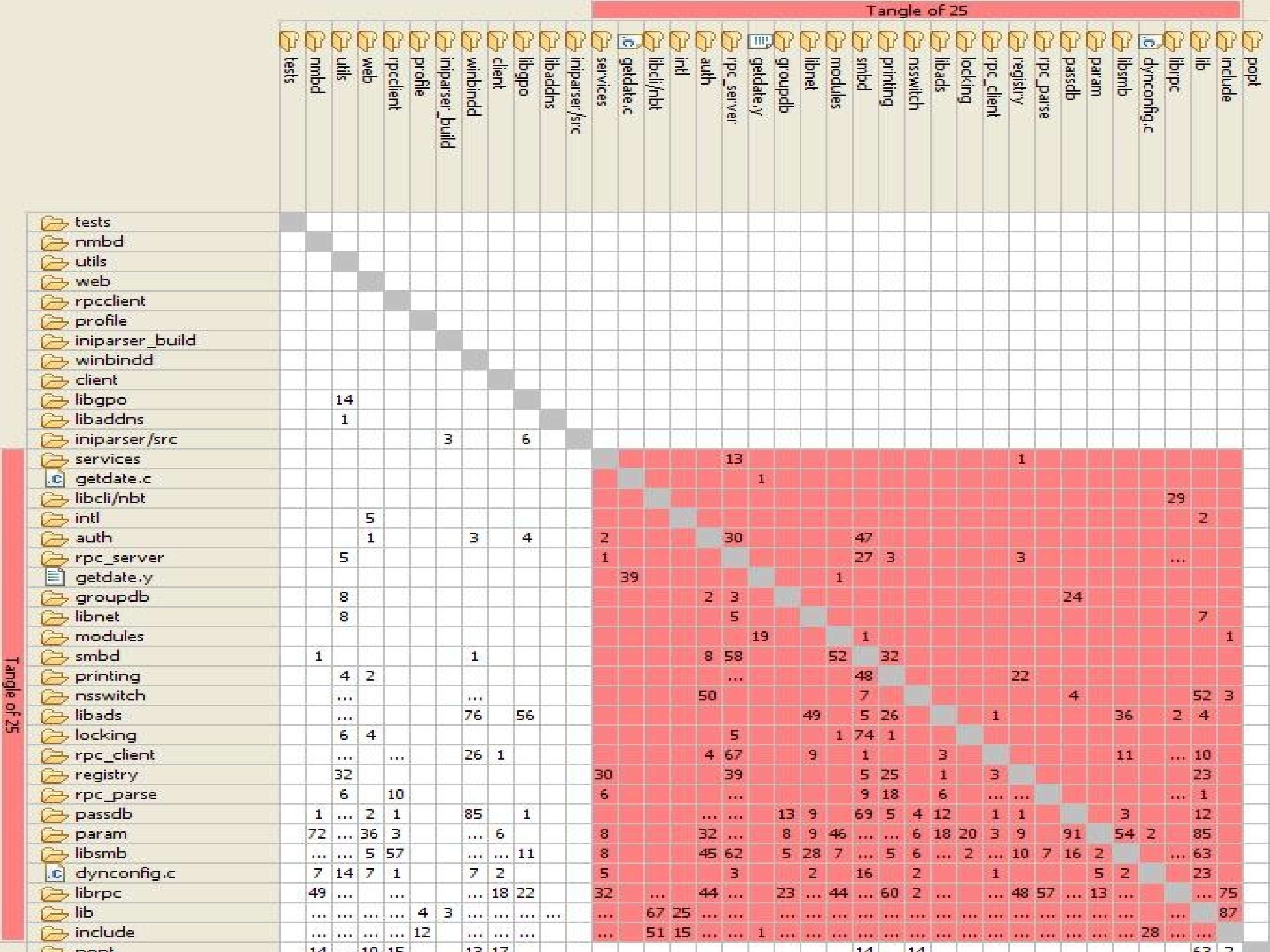
- Published under a Creative Commons license
  - Reusable by anyone

Architectural Analysis

- Separate data, about high level architecture of code, not low level code defects

- Collected by the same analysis mechanisms

- Will be available to Scan projects
  - Starting with Rung 3



Tangle of 25

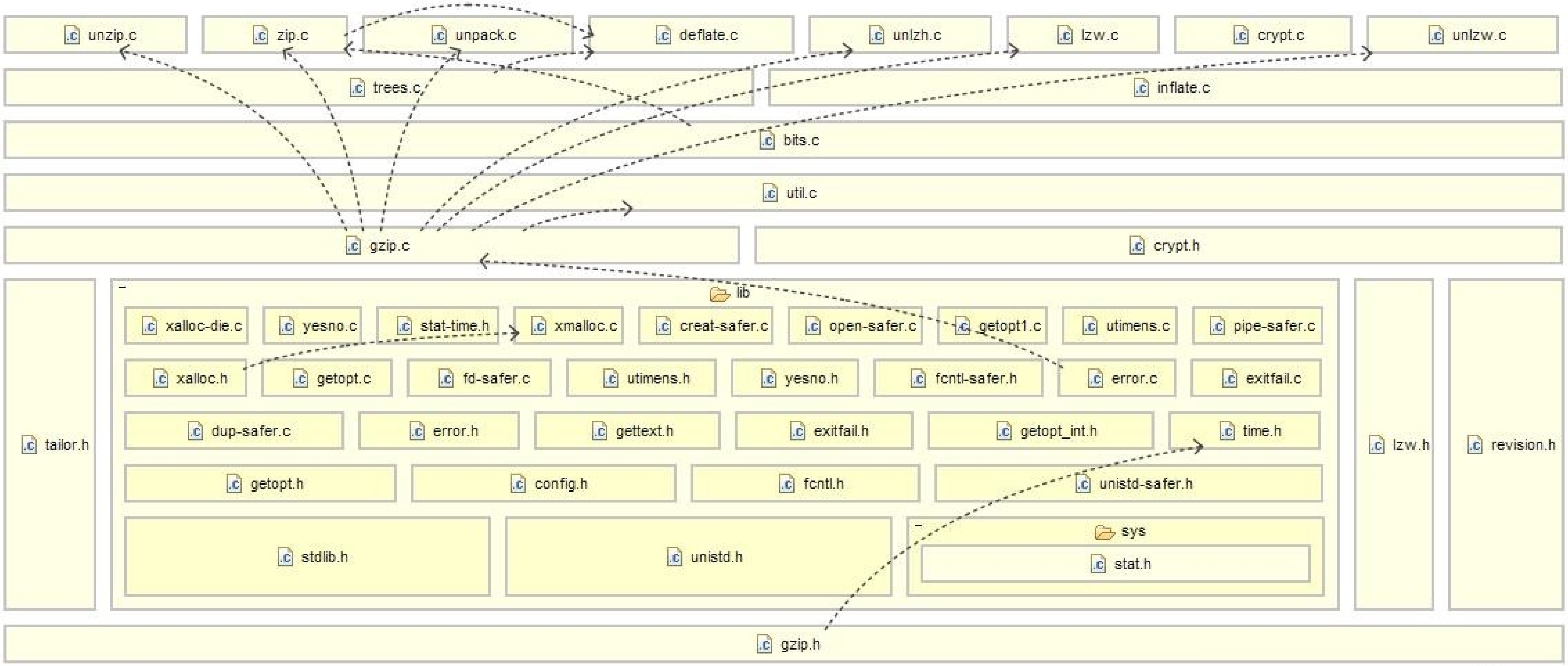
	tests	nmbd	utils	web	rpcclient	profile	iniparser_build	winbindd	client	libgpo	libaddns	iniparser/src	services	getdate.c	libcli/nbt	intl	auth	rpc_server	getdate.y	groupdb	libnet	modules	smbd	printing	nsswitch	libads	locking	rpc_client	registry	rpc_parse	passdb	param	libsmb	dynconfig.c	librpc	lib	include	popt
tests																																						
nmbd																																						
utils																																						
web																																						
rpcclient																																						
profile																																						
iniparser_build																																						
winbindd																																						
client																																						
libgpo			14																																			
libaddns			1																																			
iniparser/src							3			6																												
services																		13											1									
getdate.c																			1																			
libcli/nbt																																			29			
intl				5																																2		
auth				1				3		4			2					30					47															
rpc_server			5										1										27	3					3						...			
getdate.y														39								1																
groupdb			8														2	3													24							
libnet			8															5																		7		
modules																			19				1														1	
smbd		1						1									8	58				52		32														
printing			4	2														...					48						22									
nsswitch			...					...									50						7								4					52	3	
libads			...					76		56											49		5	26				1					36		2	4		
locking			6	4														5				1	74	1														
rpc_client			...		...			26	1								4	67			9		1			3							11		...	10		
registry			32										30					39					5	25		1		3								23		
rpc_parse			6		10								6					...					9	18		6		...	...						...	1		
passdb		1	...	2	1			85		1							...	...		13	9		69	5	4	12		1	1				3			12		
param		72	...	36	3			...	6				8				32	...		8	9	46	...	...	6	18	20	3	9		91		54	2		85		
libsmb		...	...	5	57			...	...	11			8				45	62		5	28	7	...	5	6	...	2	...	10	7	16	2			...	63		
dynconfig.c		7	14	7	1			7	2				5					3			2		16		2			1				5	2			23		
librpc		49	...		...			...	18	22			32		...		44	...		23	...	44	...	60	2	...		...	48	57	...	13	...			...	75	
lib		...	...	...	...	4	3	...	...	...	...		...		67	25	...	...		...	...	...	...	...	...	...	...	...	...	...	...	...	...		...		87	
include		...	...	...	...	12		...	...	...			...		51	15	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	28	...	...		
popt		14	...	10	15			13	17														14		14											63	7	

Tangle of 25

Architectural Analysis

SCAN.COVERITY.COM

- Gzip
  - As implemented, with file granularity



Generated by Coverity Architecture Analyzer coverity™

Q & A

SCAN.COVERITY.COM

David Maxwell - Open Source Strategist

dmaxwell@coverity.com