Evaluating the Evolution of Small Scale Open Source Software Systems

Chanchal K. Roy and James R. Cordy
School of Computing, Queen's University, Kingston, Ontario, Canada

Abstract

This poster presents an analysis of the evolution behavior of two small size open source software systems and observed that evolution of such systems appears to follow Lehman's laws for software evolution unlike large scale open source software systems.

1 Motivation

Previous work on software evolution has looked primarily at:

- Large real-time software systems [3]
- Large open source software [2], [5] such as the Linux kernel

Surprisingly, little or no work has been concerned with the evolution of smaller size software systems.

Here:

- Examining the evolution behavior of two such small size open source software systems, the *Barcode Library* [1] and *Zlib* [6].
- Put emphasis on the evolution of *Barcode Library*.

2 Target Open Source Software Systems

Two useful small scale open source software systems are used:

- The *Barcode Library*: Mainly a C library for creating bar-code output files.
- The *Zlib*: A compression/decompression tool.

3 Methodology and Metrics

- Collected 9 different releases of the *Barcode Library* and 43 different releases of *Zlib*.
- Calculated the total size, total lines of code (LOC), and number of uncommented non-blank lines of code (UNB-LOC or ULOC for short) for each of the releases of both the software systems using Numlines [4]. For *Barcode Library*, calculated the same for module level too.
- Calculated the number of global functions, variables and macros using Ctags.
- Developed a small java program for drawing the least-squares fit (LSF) of the plotted data.

The following measuring metrics are used:

Size$_i$ = Total size of ith release in bytes

UNB-LOC$_i$ = No. of uncommented non-blank LOC in Release$_i$

Added$_i$ = No. of UNB-LOC of the newly added files in Release$_i$

Changed$_i$ = No. of UNB-LOC of the changed/modified files in Release$_i$

Growing Rate$_i$ =(Added$_i$ X 100) / (UNB-LOC$_{i-1}$)

Changing Rate$_i$ =(Changed$_i$ X 100) / (UNB-LOC$_{i-1}$)

Handled$_i$ = (Changed$_i$ + Added$_i$) - (Changed$_{i-1}$ + Added$_{i-1}$)

UnHandled$_i$ = (UNB-LOC$_i$)- (Handled$_i$)

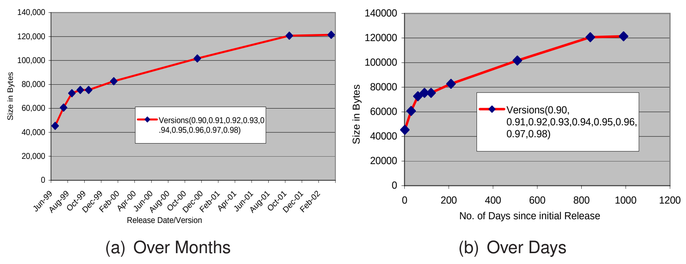
Handling Rate$_i$ =(Handled$_i$ X 100)/ (UNB-LOC$_i$)

4 Observations

During the analysis of the software systems we have made some interesting observations. In the following subsection we concentrate on the global systems and their evolution first. Then we zoom in on the subsystems.

4.1 Overall Growth w.r.t. Size

- Over time the size of the system (Fig. 1(a)) has increased to meet additional user requirements.
- The growth (with some exceptions) follows a monotonic process and therefore we can say the evolution of *Barcode Library* follows Lehman's 6^{th} Law of continuing growth of system size.



(a) Over Months (b) Over Days

Figure 1: Growth of Barcode Over Time (months + days) w.r.t Size of the Released Versions

- This is even clearer in Fig. 1(b). Using LSF we get the linear equation $\mathbf{Y = 65.92X + 63138.28}$.
- This linear growth is in contrast to the super-linear growth of Linux, reported in [2], [5], and can be termed as continuing growth, Lehman's 6^{th} law of evolution.
- Even when plotted against RSN (Fig. 2), system growth seems to be liner in nature and can be fitted to the linear equation $\mathbf{Y = 6568.53X + 175518.76}$.
- Seems the growth of *Zlib* also follows the Lehman's 6^{th} law of continuing growth.

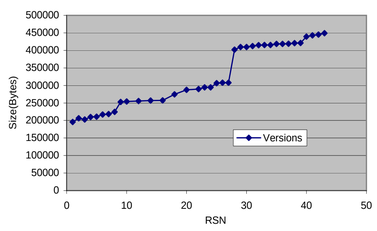
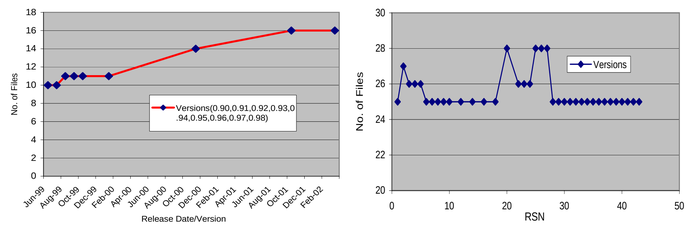


Figure 2: Growth of *Zlib* Over RSN w.r.t Size of the Released Version

4.2 Growth w.r.t. Number of Files

- Measuring growth with w.r.t. the size of released versions might be questionable (cloned or redundant code, comments or blank lines).
- In Fig. 3(a), there has been little significant growth in the number of files (only 6 files added).
- In Fig. 3(b), more interesting results for *Zlib*, the number of files remains relatively constant over all releases (with some exceptions).
- Barcode evolution follows the 6^{th} law but not for Zlib. Contrasts to the literature.

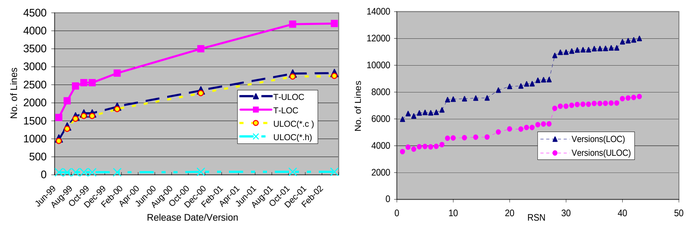


(a) Barcode (b) Zlib

Figure 3: Growth Over Time (Months) w.r.t the Number of Files

4.3 Growth w.r.t. LOC and UNB-LOC

- Linear growth (Fig. 4(a)) over time except in the case of the *.h (header) files
- Main concern is the T-ULOC curve of the releases, fitted to the linear equation $\mathbf{Y = 44.55X + 1115}$ following Lehman's 6^{th} law of evolution. Similarly for Zlib, liner equation $\mathbf{Y = 103.41X + 3356.25}$.



(a) Barcode (b) Zlib

Figure 4: Growth Over Time (months) w.r.t. T-LOC and UNB-LOC

4.4 *Barcode* Growth w.r.t Different Rates

- Very few files are added over time (Fig. 5). Initially the number of changed files is more than in later versions even decreasing later on following some properties of Lehman's 1^{st} and 5^{th} laws.

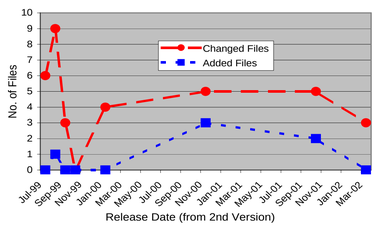
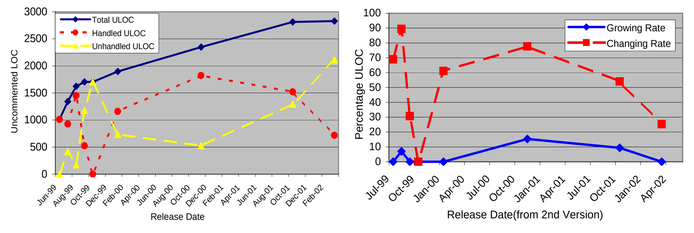


Figure 5: Changed/Added Files Over Time(months) of Barcode

- Except for some early releases(Fig. 6(a)), the *handled ULOC* consistently increased for a time and since then has decreased significantly.
- The *handled ULOC* and *unhandled ULOC* curves can be fitted with the linear equations $\mathbf{Y = 11.58X + 892.20}$ and $\mathbf{Y = 36.48X + 519.19}$ respectively.



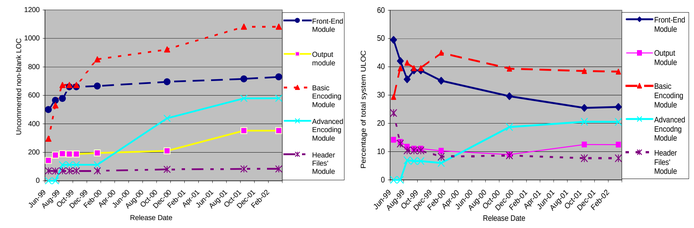
(a) Handled and Unhandled ULOC (b) Growing Rate and Changing Rate

Figure 6: Handled/Unhandled ULOC and Growthing Rate & Changing Rate of Barcode

- The *Barcode Library* (Fig. 6(b)) system is changing continuously over time, with some exceptions in the earlier versions (v0.94 from v0.93) following the Lehman's 1^{st} law.
- Both the *changing rate* and *growing rate* tend to decline, follow the 5^{th} law of evolution, the *Conservation of Familiarity*.

4.5 Module Level Observations of the *Barcode*

- Except for the *Header File module* (Fig. 7(a)), others increasing in the number of ULOC over time, following Lehman's 6^{th} law of continuing growth and signaling the 1^{st} law of continuing change.
- Except Basic Encoding Module, others tend to decline in incremental growth following Lehman's 5^{th} law of evolution on the module level too.



(a) UNB-LOC of the Modules (b) Percentage of ULOC of the Total System

Figure 7: ULOC of the Modules of Barcode

- Except for the *Advanced Encoding Module* (Fig. 7(b)), all other modules either change in parallel or tend to decline in incremental size with respect to the percentage of the total system.
- Except for the *Advanced Encoding Module*, all others are decreasing in their rate with respect to the whole system, demonstrating the predictions of Lehman's 5^{th} law.
- Only 2 bytes of overall size difference (Fig. 8) between versions 0.93 and 0.94 (*code128.c*), same lines modified with complicated code and comments.

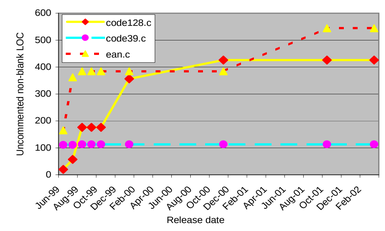
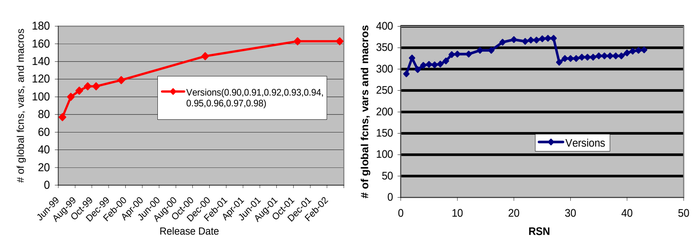


Figure 8: File-Level Growth of the *Basic Encoding Module* Over Time of Barcode

4.6 Observing the Complexity

- The complexity (9(a)) of the *Barcode* system has increased over time following the Lehman's 2^{nd} law of evolution (linear equation $\mathbf{Y = 2.22X + 98.62}$ using LSF).
- More or less same behavior in the case of *Zlib* (9(b)), initially increases and then special work has been done to control the complexity, follows 2^{nd} law of evolution.



(a) Barcode (b) Zlib

Figure 9: Number of Global Functions, Variables and Macros Over Time of Releases

5 Conclusion

- Evolution of small scale open source software systems follows Lehman's laws of software evolution, unlike the reported evolution of larger open source software systems.
- Mainly focused on Lehman's 1^{st}, 2^{nd}, 5^{th} and 6^{th} laws of evolution and found that for the most part the evolution of both the *Barcode Library* and *Zlib* follows these laws with some minor exceptions.

References

[1] The Barcode: http://www.gnu.org/software/barcode/

[2] Godfrey, M. W., Tu, Q.: Evolution in Open Source software: A case study. Proc. of the Intl. Conference on Software Maintenance (2000), pp. 131-142, San Jose, California.

[3] Lehman, M., Ramil, J., Wernick, P., Perry, D.: Metrics and laws of software evolution - the nineties view. Proc. of the Fourth Intl. Software Metrics Symposium(1997), Portland, Oregon.

[4] The Numlines home page. http://www.gammadyne.com, 2005

[5] Robles, G., Amor, J. J., Gonzalez-Barahona, J. M., Herraiz, I.: Evolution and Growth in Large Libre Software Projects. Proc. of the Intl. Workshop on Principles of Software Evolution(2005), pp.165-174, Lisbon, Portugal.

[6] Roelofs, G., Gailly, J., Adler, M.: The Zlib home page. http://www.Zlib.net/