Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Q1 Evolution of Linux operating system network

Q2 Guanping Xiao, Zheng Zheng *, Haoqin Wang

School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, PR China

HIGHLIGHTS

- Networks of 62 major releases of Linux operating system are constructed.
- Manifestations of the evolution of network properties are observed.
- The evolution of functionality structures has revealed.
- Seven events are found in the evolution of functional modules.

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form 23 August 2016
Available online xxxx

Keywords:
Complex network
Linux operating system
Network evolution
Evolution events

ABSTRACT

Linux operating system (LOS) is a sophisticated man-made system and one of the most ubiquitous operating systems. However, there is little research on the structure and functionality evolution of LOS from the prospective of networks. In this paper, we investigate the evolution of the LOS network. 62 major releases of LOS ranging from versions 1.0 to 4.1 are modeled as directed networks in which functions are denoted by nodes and function calls are denoted by edges. It is found that the size of the LOS network grows almost linearly, while clustering coefficient monotonically decays. The degree distributions are almost the same: the out-degree follows an exponential distribution while both in-degree and undirected degree follow power-law distributions. We further explore the functionality evolution of the LOS network. It is observed that the evolution of functional modules is shown as a sequence of seven events (changes) succeeding each other, including continuing, growth, contraction, birth, splitting, death and merging events. By means of a statistical analysis of these events in the top 4 largest components (i.e., *arch*, *drivers*, *fs* and *net*), it is shown that continuing, growth and contraction events occupy more than 95% events. Our work exemplifies a better understanding and describing of the dynamics of LOS evolution.

1. Introduction

Many real-world complex systems ranging from nature to human society can be abstracted as networks, where entities are denoted by nodes and interactions between entities are denoted by edges [1–3]. At first, people used regular networks to represent real systems. After that, ER random-graph model was proposed [4], in which the edges are completely randomly connected. However, real-world networks are neither regular lattices nor simple random networks. Since the small-world network model and the scale-free network model were put forward at the end of the last century [5,6], complex networks science is flourishingly developed and greatly prompts the understanding of structures and functions of real-world complex systems, such as protein networks [7–9], Internet [10–12], power grids [13–16], airport networks [17–20] and scientific collaborations [21–23]. Over the past years, researchers primarily focus on several research aspects, involving epidemic

* Corresponding author.
E-mail address: zhengz@buaa.edu.cn (Z. Zheng).

http://dx.doi.org/10.1016/j.physa.2016.09.021
0378-4371/

spreading [24–27], cascading failures [28–30], traffic dynamics [31–33], evolutionary games [34–36], network control [37,38], optimization process [39–42], coupling networks [43,44] etc. One interesting direction is to investigate structures and functions of software system networks [45–47].

Software systems represent one of the most sophisticated man-made systems which can be modeled as networks [48]. Over the past decades, a wealth of researches have been concentrated on software system networks from several aspects. Valverde et al. [49] presented the first evidence for the presence of scale-free and small-world behaviors in software systems from a well-defined local optimization. Cai et al. [50] modeled software dynamic execution processes as an evolving software mirror graph and found that execution processes might illustrate as a small-world network in the topological sense but no longer manifest in the temporal sense. Šubelj et al. [51] investigated the community structure in software networks. In the work, observations on several networks constructed from *Java* and various third party libraries were presented.

Operating system is a special software system which interacts with hardware devices and provides execution environments to the software executing on a computer [52]. Among various operating systems, Linux operating system (LOS) is typical and well deployed in most fields of the society nowadays. Due to its open source characteristic [53], the LOS network draws continuous attentions. In 2008, Zheng et al. [54] developed two new network growth models to better represent Gentoo Linux. Recently, Gao et al. [55] analyzed the core component of LOS as a complex network and showed that the large in-degree nodes providing basic services would do more damage on the whole system at the time of intentional attacks. Wang et al. [56] studied the network of networks in LOS and different manifestations of the coupling correlations among components were observed from the aspects of topologies and functions.

The functionality of LOS is continuously updated and optimized, brings it a strong vitality to adapt changing environments. Since version 1.0 was released in 1994, more than 1300 releases including versions 1.0 to 4.1[1] have been developed, and about every three months, a major version of LOS would be released. Understanding characteristics and processes of LOS evolution are meaningful to the development of operating systems, whereby what metrics could be utilized to evaluate changes of structures and functions of LOS during its evolution and how to reveal underlying principles of the evolution of LOS comprehensively are necessary to be explored. However, the study of the LOS network from the evolution aspect is still rare. To the best of our knowledge, the only work so far is that in 2011, Fortuna et al. [57] investigated packages in the first 10 releases of Debian GNU/Linux operating system and found that the modularity of the network of dependencies increased over time. In this paper, we systematically study the network evolution of 62 major releases of LOS ranging from version 1.0 released in 1994 to version 4.1 released in 2015, from the aspects of topological properties and functionality structures.

The paper is organized as follows. The following section describes the research data and the network modeling of LOS. Afterwards, results of the evolution of the LOS network are presented and discussed. Finally, the conclusion is given.

2. Methods

2.1. Research data

LOS is originally developed by Linus Torvalds in 1991. Over the past 20 years, more than 1300 releases ranging from versions 1.0 to 4.1 were put out, and the development of LOS has gone through three stages which can be classified by development methods [58]. The first stage contains versions 1.0 to 2.5, the second stage includes versions 2.6 series and the third stage contains version 3.0 and its subsequent releases. In the first stage, developers utilized 3 digits which labeled as "a.b.c" to represent the releases: the first digit "a" represents the LOS version, the second digit "b" denotes the major revision of LOS, and the third digit "c" indicates the minor revision of LOS. In this stage, a distinction was made between even and odd second major revision digits: odd digits were used to represent development releases, whereas even digits correspond to stable releases [59]. Differently, in the second stage, releases were not developed by dividing into development or stable versions, and 4 digits were used to denote the releases as a result of the change of development methods. The major revision of LOS was denoted by the third digit while the minor revision of LOS was represented by the forth digit. In the third stage, developers decided to stop the old numbering method which was utilized in versions 2.6, from which the numbering with 3 digits was reused to denote the releases. In this scheme, the development speed was accelerated by mainly separating the releases into mainline versions, stable versions and longterm versions [60].

The research data of LOS are obtained from 62 releases provided by the official website[2] of Linux ranging from versions 1.0 to 4.1, covering the three development stages of LOS, as exhibited in Fig. 1. According to the version numbering method of LOS, all of these releases are major releases.

2.2. Network modeling

LOS is a C language based operating system whose realization mainly depends on the function calls, which can be commonly termed as a call graph [48]. Consequently, a directed network based on the call graph where nodes are functions

[1] The latest release when we started this work.

[2] https://www.kernel.org/.

Please cite this article in press as: G. Xiao, et al., Evolution of Linux operating system network, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.09.021