페이지 입상도 기반의 MIPS 펌웨어 베이스 주소 자동추출 기법

문석주[O], 장대희[*]
[O]영남대학교 공업기술연구원,
[*]성신여자대학교 융합보안공학과
e-mail: hooohly@naver.com[O], djang@sungshin.ac.kr[*]

Automated extraction of MIPS firmware image base using page-granularity

Seok-Joo Mun[O], Daehee Jang[*]
[O]Institute of Industrial Technology, Yeungnam University,
[*]Department of CSE, Sungshin W. University

● 요 약 ●

본 논문에서는 MIPS 아키텍쳐 기반 펌웨어에 대한 페이지 단위의 이미지 베이스 주소 탐색 방안을 제안한다. 이 방법은 MIPS 기반 임베디드 기기의 펌웨어를 대상으로, 대상 내의 분석 대상의 이미지 베이스 주소 계산 알고리즘을 효율적으로 개선하여 이미지 베이스 주소탐색 시간을 최소화하는 것을 목표로 한다. 이 방법은 펌웨어 내 문자열의 주소를 기준으로 세그먼트 시작 주소를 유추, 페이지 단위인 4KB 단위로의 이미지 베이스 주소 후보군을 계산하여 이미지 베이스 주소 후보군을 선별하는 것을 그 원리로 한다. 본 논문에 적용된 방법은 기존의 경험적 방법을 통한 펌웨어 베이스 탐색 방안에 비해 정확도면에서 우수함을 보인다.

키워드: MIPS, 베이스주소(base address), 펌웨어(Firmware), 펌웨어 분석(Firmware Analysis)

I. Introduction

최근 IoT (Internet of Things) 시장의 발달로 인해 다양한 종류의 임베디드 시스템에 대한 개발 또한 급증하고 있다. IoT 기기의 활용도가 높아지면서 이러한 기기들의 시스템 보안 분석의 필요성도 증대되었고, 이러한 보안적인 위협을 분석하기 위해서는 가장 선행적으로 임베디드 기기 내에 적용된 펌웨어 바이너리를 분석해야한다. 펌웨어는 임베디드 기기에서 사용되는 모든 바이너리 파일들을 담고 있는 파일로써, 실행코드 및 데이터 등 임베디드 기기를 구동시키기 위한 핵심 요소들을 포함하고 있다. 따라서 임베디드 기기에 대한 위협요소의 분석을 위해서는 펌웨어에 대한 역공학 (Reverse Engineering) 과정이 기본적으로 요구된다.

펌웨어는 일반적으로 여러 개의 이미지 파일들로 구성된 패킹(Packing)된 형태를 지닌다. 소스코드가 없이 역공학으로 펌웨어를 분석하는 경우 이렇게 패킹된 바이너리에서 분석대상 이미지를 강제로 추출하게 되는데, 그 과정에서 정적분석에 필요한 이미지 베이스 주소 (Image Base Address) 가 손실되어 문제가 된다. 일반적인 경우 이러한 베이스 주소는 역공학 분석가의 다양한 경험에 의한 추론 및 무차별적 대입시도 등을 통해 복원된다. 이러한 어려움에 대해서 문자열 매칭 알고리즘을 기반으로 자동화된 복원을 시도하는 연구가 최근에 발표되었는데[1], 본 논문에서는 기존의 연구에서 고려되지 않은 페이지 입상도의 효과를 추가적으로 고려하여 탐색의 정확도 및 속도를 향상하는 것을 목표로 한다.

II. Preliminaries

1. Backgrounds

1.1 MIPS 아키텍쳐

MIPS 아키텍처는 임베디드 장비 시장에서 가장 널리 사용되는 RISC 아키텍쳐 중의 하나로, 주로 소형 임베디드 장비에 사용된다[2]. 해당 아키텍쳐에 사용되는 명령어 세트는 아키텍쳐가 처리할 수 있는 비트 수와 동일한 고정된 명령어 길이를 가지며, 데이터의 처리방법에 따라 R-Type, J-Type, I-Type의 세 종류로 구성된다.

1.2 이미지 베이스 주소

이미지 베이스 주소란 펌웨어가 로드되는 메모리의 첫 주소로, .text, .data 등의 세그먼트들이 메모리에 매핑되기 시작하는 기준주소를 의미한다. 이 주소를 기준으로 오프셋을 가산하여 메모리 상 절대주

소를 산출하기 위해 사용된다. 따라서 정적 분석 수행 시, 이미지 베이스 주소를 정확하게 설정하지 않는다면 이미지 베이스 주소를 기반으로 참조관계가 구성된 함수, 문자열 등의 상호참조가 정상적으로 구성되지 않으며, 반대로 이미지 베이스 주소를 찾는다면 정상적인 상호참조 구성을 통해 펌웨어 에뮬레이션에 활용할 수 있다.

III. The Proposed Scheme

본 장에서는 기존 MIPS 기반 펌웨어의 문자열 매칭을 통한 베이스 주소 탐지 알고리즘의 문자열 탐지방법에 페이지 입상도를 적용하여, 베이스 주소의 후보군을 더욱 빠르게 선별하는 방법을 제시하고자 한다.

1.1 페이지 단위 이미지 베이스 주소 검색

본 논문에서 제안하고자하는 이미지 베이스 주소 선별 방법은, 바이너리에서의 시작위치에 가장 가까운 첫 문자열의 주소를 바탕으로 펌웨어 이미지 파일의 전체크기를 제외한 뒤, 4KB 단위로 후보군을 선별하는 것이다.

일반적으로 MIPS 아키텍처의 문자열 주소 지정은 상대주소가 아닌 절대주소를 사용한다. 이 경우 I-type의 명령어, lui-ori, lui-lw, lui-addiu의 세 종류 명령어 쌍을 사용하여 문자열의 주소를 지정한다. 이러한 문자열의 주소는 이미지 베이스 주소를 기반으로 특정 오프셋이 합쳐진 형태이며, I-Type 명령어에서 얻은 주소로부터 “특정 오프셋” 이상의 값을 제외하여 베이스 주소가 포함된 주소 범위를 얻을 수 있다.

Fig. 1. 는 주소를 선별하는 과정을 나타낸 것이다.

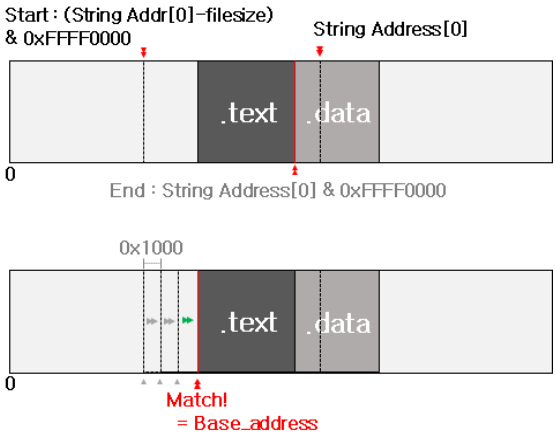

Fig. 1. 4KB 단위의 이미지 베이스 주소 후보군 선별

기존 연구의 방법에서는, 기준 (첫) 문자열의 주소에서 이미지 파일의 크기를 제외한 주소 값부터 마지막 (끝) 문자열의 주소를 수집, 귀납적 추론으로 얻어낸 주소 범위를 1바이트씩 순회하며 문자열의 일치여부를 확인한다. 그러나 이미지 베이스 주소는 페이지 단위인 4KB 단위로 할당되므로 위의 접근에는 비효율성이 존재한다. 본 논문에서는 세그먼트의 시작 주소 시작부터 4KB씩 검색을 수행하고 자 하위 2바이트를 마스킹한다.

따라서 본 논문에서의 방법을 통해서 이미지 베이스 주소의 후보군은 페이지 입상도 단위인 0x1000 단위의 주소만 검증하므로 훨씬 빠른 속도로 베이스 주소에 대한 검증을 수행 할 수 있다.

IV. Conclusions

본 논문에서는 MIPS를 기반으로하는 펌웨어 파일을 기준으로 이미지 베이스 주소를 찾기위한 기존 연구를 살펴보고, 주소를 찾는 과정에 페이지 입상도를 적용하여 탐색시간을 단축시키는 방법에 대해 서술하였다. 페이지 단위 베이스 주소 탐지방법을 사용함으로써 이미지 베이스 주소 후보군을 단시간에 획득할 수 있으며, 이를 통해 정적분석시, 정상적인 상호참조를 구축할 수 있다. 향후 MIPS 기반 펌웨어 내의 레지스터 값 등의 데이터 요소들을 활용하여 이미지 베이스 주소 탐지를 더 정확하고 효과적으로 자동화 할 수 있는 방법에 대해 추가적으로 연구할 계획이다.

REFERENCES

[1] Zhu, Xiaodong, "Determining the base address of MIPS firmware based on absolute address statistics and string reference matching", Computers & Security, Vol. 88, Jan. 2020.

[2] CHEN, Daming D., et al. "Towards automated dynamic analysis for linux-based embedded firmware.", NDSS. p. 1.1-8.1. 2016.

[3] Skochinsky, Igor. "Intro to embedded reverse engineering for PC reversers." REcon conference, Montreal, Canada. 2010.