악성코드 그룹 및 변종 관리 시스템 개발†

강홍구*, 지승구*, 정현철*
*한국인터넷진흥원 연구개발팀
e-mail:redball@kisa.or.kr

A Development of Management System of Malware Group and Variant Information

Hong-Koo Kang*, Seung-Goo Ji*, Hyun-Cheol Jeong*
*Team of Security R&D, Korea Internet & Security Agency

요 약

최근 변종 악성코드가 크게 증가하고 하나 이상의 악성코드로 이루어진 그룹 형태의 악성코드들이 빠르게 유포되고 있다. 이러한 그룹 형태의 악성코드와 변종 악성코드에 대한 효과적인 대응을 위해서는 악성코드 그룹 및 변종을 관리하고 안티바이러스 업체와 정보를 공유할 수 있는 시스템이 필요하다. 본 논문에서는 대용량 악성코드 분석 정보로부터 악성코드 그룹 및 변종 정보를 효율적으로 관리하고 공유하는 시스템을 제안하다. 악성코드 그룹 정보는 악성코드 행위를 기반으로 연계된 악성코드 정보들로 생성되고, 악성코드 변종 정보는 CFG 분석을 통한 악성코드간 유사도 정보로 생성된다. 본 논문에서 제안하는 시스템은 악성코드 그룹 및 변종 정보를 쉽게 검색하고 공유할 수 있기 때문에 다양한 악성코드 대응 시스템과 쉽게 연계될 수 있는 장점을 가지고 있다.

1. 서론

악성코드(Malware)란 악성 또는 악용 가능한 소프트웨어의 집합으로서, 바이러스, 웜, 스파이웨어, 악성 애드웨어 등 사용자와 컴퓨터에게 잠재적으로 위험이 되는 모든 소프트웨어를 총칭하는 말이다. 최근 인터넷을 통해 다양한 악성코드가 빠르게 제작, 유포되고 있다[1,2,3].

특히 악성코드가 금전적인 이익을 얻는 수단으로 활용되고 악성코드 변종을 자동으로 제작하는 프로그램까지 등장하면서 변종 악성코드는 하루가 다르게 급속히 증가하고 있다[4,5]. 또한, 악성코드는 하나의 파일로서 모든 악성 행위를 수행하는 경우는 드물고, 둘 이상의 악성코드가 유기적으로 실행되는 그룹 형태의 악성코드가 대부분을 차지하고 있다. 이러한 그룹 형태의 악성코드와 변종 악성코드에 대한 효과적인 대응을 위해서는 악성코드 그룹 및 변종을 관리하고 안티바이러스 업체와 정보를 공유할 수 있는 시스템이 필요하다[2,4].

본 논문에서는 악성코드 분석 정보로부터 악성코드 그룹 및 변종 정보를 효율적으로 관리하고 공유할 수 있는 악성코드 그룹 및 변종 관리 시스템을 제안하다. 본 논문에서는 제안하는 시스템은 악성코드별 행위와 연관되는 악성코드들을 그룹화하여 관리하고 악성코드간의 CFG(Control Flow Chart) 분석을 통한 유사 정도에 따라 변종 관계의 악성코드들을 체계적으로 관리한다.

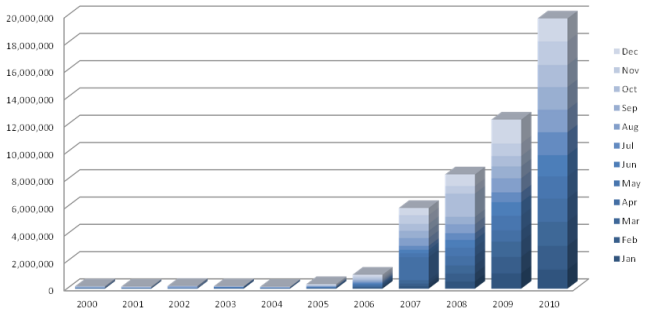
† "본 연구는 한국방송통신전파진흥원의 방송통신기술개발사업의 일환으로 지능형 악성코드 자동 분석 및 경유/유포지 탐지 기술 개발 과제(10914-06001)에 의해 지원되었음"

따라서, 본 시스템을 통해 특정 악성코드에 대한 행위 연관관계가 있는 악성코드 그룹 정보와 그 변종 악성코드들을 신속하게 파악할 수 있기 때문에, 점차 다양해지고 있는 악성코드들에 대한 체계적이고 효과적인 대처가 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대해 기술한다. 그리고 3장과 4장에서 제안하는 악성코드 그룹 및 변종 관리 시스템 설계와 구현을 차례대로 설명한다. 마지막으로 5장에서 결론에 대해 기술한다.

2. 관련연구

국내외에서 악성코드 변종이 빠르게 증가하고 동시에 지능적이고 융복합된 악성코드가 등장하면서 안티바이러스 업체에서 악성코드를 개별 분석하는 것은 현실적으로 어렵게 되었다[3,5]. (그림 1)은 2000년부터 2010년까지 신규 악성코드 샘플 수집 통계를 보여준다.



(그림 1) 신규 악성코드 샘플 수집 통계

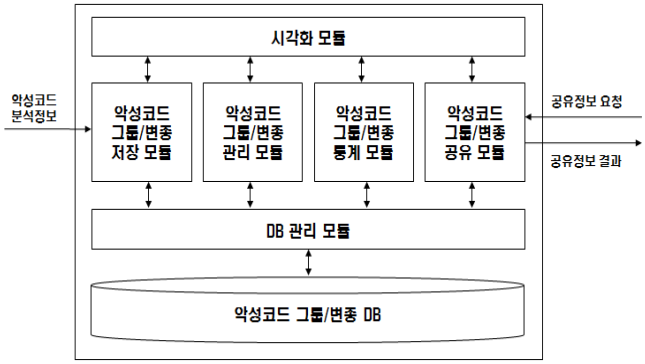
(그림 1)과 같이 신규 악성코드는 최근 몇 년간 빠르게 증가되었으며 2010년에는 매달 150만개 이상의 신규 악성코드가 발생하고 있는 것으로 나타났다[6]. 그리고 최근 발견되는 많은 수의 악성코드는 기존 악성코드의 변종인 경우가 많고 다양한 악성코드들의 집합체 형태를 가지는 악성코드도 많이 증가하고 있는 추세이다. 이러한 상황에서 악성코드 그룹 및 변종을 자동적으로 분석할 수 있는 기술 연구가 활발히 진행되고 있다.

2008년 오스트리아의 Vienna 대학교, 프랑스의 Eurocom 사 등으로 구성된 ISEC lab에서 자동화된 행위 분석 시스템을 개발하여 공개하였으나, Conficker, Storm등과 같은 커널 레벨 실행 은닉 기술을 사용하는 악성코드를 분석하지 못하는 문제점이 있다. 그리고 미국 조지아 공대, 퍼듀 대학교 등에서 자동화된 악성코드 정적 분석 및 행위 분석 기술에 대한 연구를 수행하고 있다. 국내에서는 KISA가 행위 기반 봇넷 자동 분석 기술을 개발하였고 현재 정적/동적 분석이 결합된 악성코드 자동 분석 기술을 개발하고 있다[2,4].

그러나 급증하는 악성코드에 대한 분석 결과에서 악성코드 그룹 및 변종 관리는 미흡한 실정이고 그룹 형태의 악성코드와 변종 악성코드에 대한 효과적인 대응을 위한 공유 연구도 초기 단계에 있다. 따라서 악성코드 그룹과 변종 정보를 효과적으로 관리하고 안티바이러스 업체와 정보를 공유할 수 있는 시스템이 절실히 요구된다.

3. 악성코드 그룹 및 변종 관리 시스템 설계

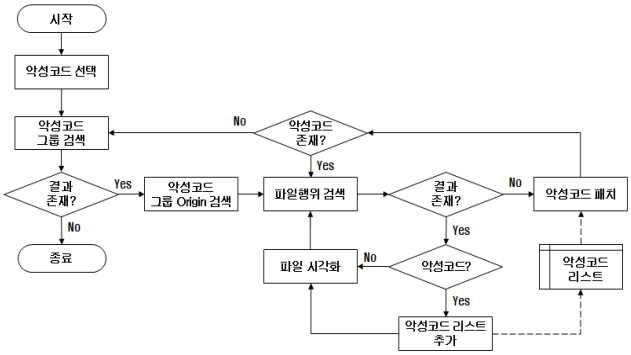
본 논문에서 제안하는 악성코드 그룹 및 변종 관리 시스템의 주요 기능은 다음과 같다. 먼저 악성코드 분석 정보로부터 악성코드 그룹 및 변종 정보를 생성하고 생성된 결과를 DB에 저장한다. 그리고 저장된 악성코드 그룹 및 변종 정보로부터 관리자가 쉽게 악성코드 그룹 및 변종을 검색하고 결과를 파악할 수 있도록 편리한 GUI를 제공한다. 또한 주기적으로 악성코드 그룹 및 변종에 대한 통계를 생성하고 안티바이러스 관련 외부 업체로부터 악성코드 그룹 및 변종 정보를 요청받고 해당 정보를 제공한다. 이를 위해 본 논문에서 제안하는 악성코드 그룹 및 변종 관리 시스템 구성도는 (그림 2)와 같다.



(그림 2) 악성코드 그룹 및 변종 관리 시스템 구성도

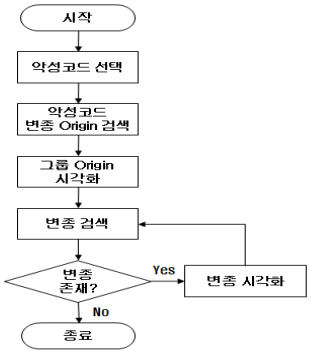
(그림 2)와 같이 악성코드 그룹 및 변종 관리 시스템은 악성코드 그룹/변종 저장 모듈, 악성코드 그룹/변종 관리 모듈, 악성코드 그룹/변종 통계 모듈, 악성코드 그룹/변종 공유 모듈, 시각화 모듈, DB 관리 모듈과 악성코드 그룹/변종 DB로 구성된다. 먼저 각 모듈의 주요 기능은 다음과 같다.

악성코드 그룹/변종 저장 모듈은 악성코드 분석 정보로부터 악성코드 그룹 및 변종 정보를 추출하고 추출된 악성코드 그룹 및 변종 정보를 DB 관리 모듈을 통해 악성코드 그룹/변종 DB에 저장한다. 악성코드 그룹/변종 관리 모듈은 악성코드 그룹/변종 DB에서 관리자가 요청하는 악성코드 그룹 및 변종 정보를 검색하고 검색 결과를 시각화 모듈을 통해 화면에 보여준다. (그림 3)과 (그림 4)는 악성코드 그룹과 변종을 시각화하기 위한 처리 흐름도를 순차적으로 보여준다.



(그림 3) 악성코드 그룹 시각화 처리 흐름도

(그림 3)과 같이 먼저 관리자가 선택한 악성코드를 포함하는 악성코드 그룹들을 검색한다. 만일 악성코드 그룹이 존재하면 해당 악성코드 그룹의 최초 악성코드인 기원(Origin) 악성코드를 찾고 해당 기원 악성코드 행위와 관계된(예를 들어, 파일 다운로드 또는 생성 등) 악성코드를 악성코드 리스트에 추가하고 시각화한다. 만약 기원 악성코드 행위와 관계된 파일이 악성코드가 아니면 단순히 시각화만 수행한다. 악성코드 그룹 시각화 과정은 악성코드 리스트에 존재하는 모든 악성코드를 대상으로 수행된다. (그림 4)는 악성코드 변종을 시각화하기 위한 처리 흐름도를 보여준다.

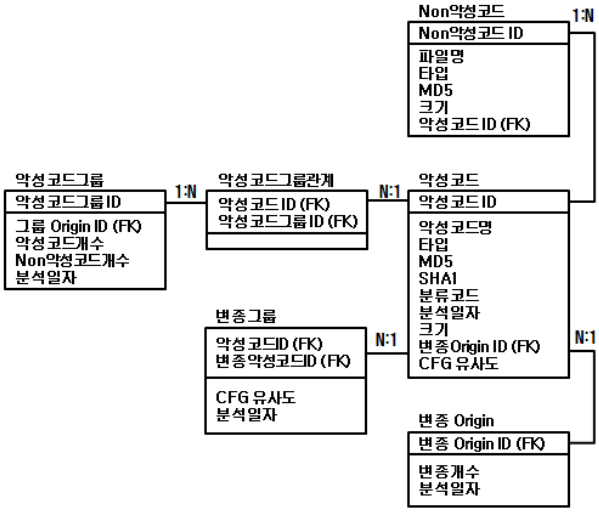


(그림 4) 악성코드 변종 시각화 처리 흐름도

(그림 4)와 같이 관리자가 선택한 악성코드에 대해 최초 변종인 변종 기원을 검색한다. 선택된 악성코드의 변종 기원은 기존에 파악된 변종 기원들과 CFG 분석을 통해 유사도가 가장 높은 악성코드로 선택된다. 변종 기원이 검색되면 이를 시각화 모듈을 통해 화면에 출력한다. 또한, 검색된 변종 기원과 임계치 이상의 유사도를 갖는 악성코드를 변종으로 간주하고 이들 악성코드와 CFG 유사도를 검색하고 결과를 화면에 출력한다. 이때, 변종 악성코드들은 CFG 유사도가 높은 순서대로 출력된다. 변종 악성코드가 더이상 검색되지 않으면 변종 정보 검색을 종료한다.

그리고 악성코드 그룹 및 변종 관리 시스템을 구성하는 모듈로서 악성코드 그룹/변종 통계 모듈은 주기적으로 악성코드 그룹 및 변종 통계를 생성하고 생성된 통계를 저장하는 모듈이고 악성코드 그룹/변종 공유 모듈은 외부 시스템으로부터 요청받은 악성코드 그룹 및 변종 정보를 응답해주는 모듈이다.

마지막으로 악성코드 그룹/변종 DB는 악성코드 그룹 및 변종 정보를 저장하는 저장소로서 주요 테이블은 악성코드 그룹 및 변종 ERD와 악성코드 행위 ERD에 포함된다. (그림 3)은 악성코드 그룹 및 변종 ERD를 보여준다.



(그림 5) 악성코드 그룹 및 변종 ERD

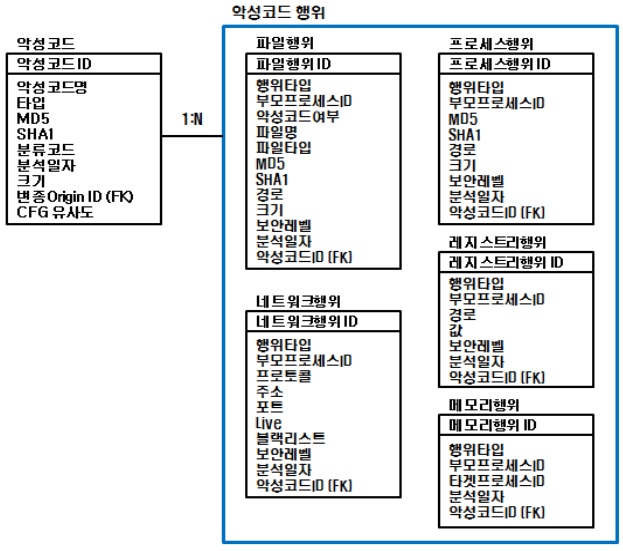
(그림 5)와 같이 악성코드 그룹 및 변종 ERD는 악성코드 테이블, 악성코드 그룹 관계 테이블, 악성코드 그룹 테이블, 악성코드 변종 Origin 테이블, 악성코드 변종 그룹 테이블, Non악성코드 테이블로 구성되어 있다. 악성코드 테이블은 악성코드 ID를 키 값으로 가지며, 악성코드에 대한 정보를 저장하는 테이블이다. 악성코드 그룹 관계 테이블은 악성코드 테이블과 악성코드 그룹 테이블 간의 관계를 설정하는 테이블로서 악성코드 ID와 악성코드 그룹 ID를 키 값으로 갖는다. 하나의 악성코드는 여러 악성코드 그룹에 속할 수 있으므로 악성코드 그룹관계 테이블과 악성코드 테이블은 N:1의 관계를 갖는다.

악성코드 그룹 테이블은 악성코드 그룹 ID를 키 값으로 가지며, 악성코드 행위를 통해 서로 연관된 악성코드들의 집합을 의미한다. 악성코드 그룹 테이블은 악성코드 그룹 관계 테이블과 1:N의 관계를 갖는다. 악성코드 변종 Origin 테이블은 변종 Origin ID를 키 값으로 가지며, 변종 기원과 유사도를 갖는 변종 악성코드에 대한 정보를 저장하는 테이블이다. 하나의 변종 기원과 유사도를 갖는 여러 가지의 악성코드가 존재할 수 있으므로 악성코드 변종 Origin 테이블은 악성코드 테이블과 1:N의 관계를 갖는다.

악성코드 변종 그룹 테이블은 서로 변종간인 악성코드 ID들을 키 값으로 가지며, 서로 변종 관계에 있는 악성코드간 CFG 유사도를 저장하는 테이블이다. 하나의 악성코드는 여러 변종과 CFG 유사도를 측정할 수 있으므로 악성코드 변종그룹 테이블은 악성코드 테이블과 N:1의 관계를 갖는다.

Non악성코드 테이블은 Non악성코드 ID를 키 값으로 가지며 악성코드가 아닌 일반 파일에 대한 정보를 저장하는 테이블이다. 하나의 악성코드는 행위를 통해 여러 일반 파일과 연관될 수 있으므로 Non악성코드 테이블은 악성코드 테이블과 N:1의 관계를 가질 수 있다.

다으므로 (그림 6)은 악성코드 행위 ERD를 보여준다.

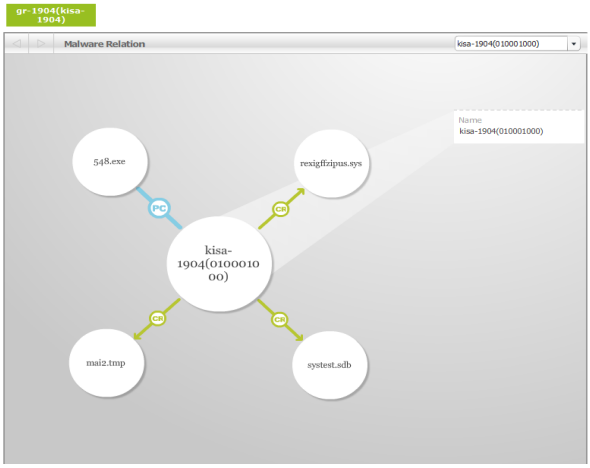


(그림 6) 악성코드 행위 ERD

(그림 6)과 같이 악성코드 행위 ERD는 파일 행위 테이블, 프로세스 행위 테이블, 네트워크 행위 테이블, 레지스트리 행위 테이블, 메모리 행위 테이블로 구성된다. 각각의 테이블은 각 행위에 따른 특성을 저장할 수 있는 다양한 필드를 가지고 있으며, 해당 행위를 수행하는 다른 악성코드 ID 필드도 가질 수 있다. 즉, B라는 악성코드가 C라는 악성코드를 다운로드 하는 행위를 할 경우, B 악성코드의 파일행위와 연관된 다른 악성코드 ID 필드에는 C가 저장된다. 하나의 악성코드는 여러 행위를 수행할 수 있으므로 악성코드 행위 관련 테이블은 악성코드 테이블과 각각 N:1의 관계를 갖는다.

4. 악성코드 그룹 및 변종 관리 시스템 구현

본 논문에서 제안하는 악성코드 그룹 및 변종 관리 시스템은 외부 시스템과 연동이 용이하도록 웹 환경에서 구현되었다. 외부 악성코드 분석 시스템으로부터 전달받는 악성코드 분석 정보와 공유되는 악성코드 그룹 및 변종 정보는 XML 형태로 송수신된다. 그리고 관리자가 쉽게 악성코드 그룹 및 변종을 검색하고 결과를 파악할 수 있도록 시각화 기능을 구현하였다. (그림 7)과 (그림 8)은 악성코드 그룹과 변종을 시각화한 예제를 차례대로 보여준다.



(그림 7) 악성코드 그룹 시각화 예제

(그림 7)과 같이 관리자가 악성코드 "kisa-1904"를 선택하면 해당 악성코드가 포함된 악성코드 그룹 "gr-1904"가 검색된다. 악성코드 그룹의 기원 악성코드가 관리자가 선택한 악성코드 "kisa-1904"와 같기 때문에 해당 악성코드의 행위로 생성된 파일 "systest.sdb", "rexigffzipus.sys", "548.exe", "mai2.tmp"가 화면에 시각화된다. 그리고 관리자가 선택된 악성코드 "kisa-1904"와 유사도를 갖는 변종 악성코드 리스트를 보여준다. (그림 8)은 악성코드 변종을 시각화한 예제를 보여준다.

변종 ORIGIN-ID:756(.17)

변종 ORIGIN 정보

변종 ORIGINID	악성코드 ID	CFG 유사도
756	kisa-727	.15
756	kisa-867	.26
756	kisa-928	.3
756	kisa-1014	.5
756	kisa-1025	.5
756	kisa-6	.22
756	kisa-122	.22
756	kisa-168	.16
756	kisa-174	.24
756	kisa-197	.15
756	kisa-270	.23
756	kisa-123	.25
756	kisa-132	.25
756	kisa-133	.25
756	kisa-330	.15
756	kisa-333	.15
756	kisa-341	.3
756	kisa-405	.25
756	kisa-480	.23
756	kisa-310	.25
756	kisa-335	.25
756	kisa-505	.28
756	kisa-578	.23
756	kisa-1904	.17
756	kisa-1278	.5
756	kisa-1449	.5
756	kisa-1455	.16

(그림 8) 악성코드 변종 시각화 예제

(그림 8)과 같이 관리자가 악성코드 "kisa-1904"를 선택하면 해당 악성코드와 가장 유사한 악성코드 기원 "456"과 CFG 유사도 "0.17"를 최상단에 보여준다. 또한, 악성코드 기원 "456"과 CFG 유사도를 갖는 다른 악성코드의 ID와 CFG 유사도 리스트를 보여주기 때문에 변종 정보를 쉽게 파악할 수 있다.

5. 결론

본 논문에서는 악성코드 그룹 및 변종 정보를 관리하고 공유할 수 있는 시스템을 제안하였다. 악성코드 그룹 정보는 악성코드 행위를 기반으로 연계된 악성코드 정보들로 생성되고, 악성코드 변종 정보는 CFG 분석을 통한 악성코드간 유사도 정보로 생성된다. 본 논문에서 제안하는 시스템은 악성코드 그룹 및 변종 정보를 쉽게 검색하고 공유할 수 있기 때문에 다양한 악성코드 대응 시스템과 쉽게 연계될 수 있을 것으로 기대된다.

본 연구에서는 악성코드 행위 중 파일과 프로세스 행위에 국한된 그룹 정보를 관리하고 있다. 따라서 보다 다양한 악성코드 행위를 고려하고 실제 안티바이러스 업계나 기관의 요구사항을 반영한 시스템 보완이 필요하다.

참고문헌

[1] 서희석, 최중섭, 주필환 "윈도우 악성코드 분류 방법론의 설계," 정보보안학회논문지, 2009, pp.83-92.

[2] 강홍구, 오주형, 임채태, 정현철, "PE 기반 악성코드 자동 분석 결과 관리를 위한 DB 설계," 한국정보처리학회 추계학술발표대회 논문집, 2010, pp.1281-1284.

[3] A. Gupta, P. Kuppili, A. Akella, and P. Barford, "An Empirical Study of Malware Evolution," Proc. of the 1st Int. Conf. on Communication Systems and Networks, 2009, pp.1-10.

[4] H.K. Kang, J.H. Oh, C.T. Im, and H.C. Jeong, "Management of the Results of Automated Malware Analysis," 3rd Pacific-Asia Conf. on Web Mining and Web-based Application, 2010, pp.273-276.

[5] M. Apel, C. Bockermann, and M. Meier, "Measuring Similarity of Malware Behavior," The 5th LCN Workshop on Security in Communications Networks, 2009, pp.891-898.

[6] Year-end Malware Stats from AV-Test, http://sunbeltblog.blogspot.com/2011/01/updated-virus-stats-from-av-test.html, 2011.