(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년03월19일
(11) 등록번호 10-1124615
(24) 등록일자 2012년02월29일

(51) 국제특허분류(Int. Cl.)
H04L 12/26 (2006.01) *H04L 12/22* (2006.01)
(21) 출원번호 10-2008-0006826
(22) 출원일자 2008년01월22일
심사청구일자 2008년01월22일
(65) 공개번호 10-2009-0080841
(43) 공개일자 2009년07월27일
(56) 선행기술조사문헌
논문1:한국정보과학회

(73) 특허권자
고려대학교 산학협력단

(72) 발명자
이희조

최현상

(74) 대리인
유미특허법인

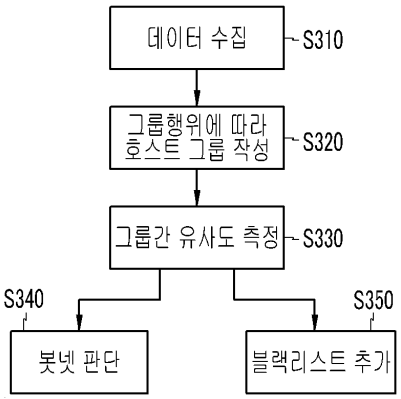
전체 청구항 수 : 총 17 항 심사관 : 김대성

(54) 발명의 명칭 **집단행동 악성코드 검색 방법 및 장치**

(57) 요 약

본 발명은 악성코드에 감염된 호스트들의 그룹행위를 기반으로 집단행동 악성코드를 검색하는 방법에 관한 것으로, 검색 대상 네트워크의 트래픽 데이터를 수집하여 상기 트래픽 데이터를 검토하여 서로 다른 호스트에서 일어나는 공통행위인 그룹행위와 상기 그룹행위의 행위대상을 기준으로 복수의 호스트 그룹을 작성하고, 상기 복수의 호스트 그룹 각각에 속한 호스트가 얼마나 일치하는지를 나타내는 상기 복수의 호스트 그룹간 유사도를 측정하고, 상기 유사도를 이용하여 집단행동 악성코드를 찾는다.

대 표 도 - 도3



특허청구의 범위

청구항 1

검색 대상 네트워크 상에서 송수신되는 트래픽 데이터를 수집하는 단계;

상기 트래픽 데이터를 검토하여, 단위 시간당 서로 다른 호스트들이 소정의 트래픽 데이터에 대하여 동일하게 수행한 행위인 그룹 행위와, 상기 트래픽 데이터를 구성하는 정보 중에서 그룹 행위가 이루어진 정보인 행위대상을 기준으로 복수의 호스트 그룹을 작성하는 단계;

상기 복수의 호스트 그룹 각각에 속한 호스트가 얼마나 일치하는지를 나타내는 상기 복수의 호스트 그룹간 유사도를 측정하는 단계; 및

상기 유사도를 이용하여 집단행동 악성코드를 찾는 단계를 포함하고

상기 그룹 행위는 DNS(domain name server) 룩업과정에서 DNS에 질의를 하는 행위이고,

상기 트래픽 데이터가 DNS 트래픽이며, 상기 행위 대상은 DNS 쿼리의 도메인 네임인,

집단행동 악성코드 검색방법.

청구항 2

검색 대상 네트워크 상에서 송수신되는 트래픽 데이터를 수집하는 단계;

상기 트래픽 데이터를 검토하여, 단위 시간당 서로 다른 호스트들이 소정의 트래픽 데이터에 대하여 동일하게 수행한 행위인 그룹 행위와, 상기 트래픽 데이터를 구성하는 정보 중에서 그룹 행위가 이루어진 정보인 행위대상을 기준으로 복수의 호스트 그룹을 작성하는 단계;

상기 복수의 호스트 그룹 각각에 속한 호스트가 얼마나 일치하는지를 나타내는 상기 복수의 호스트 그룹간 유사도를 측정하는 단계; 및

상기 유사도를 이용하여 집단행동 악성코드를 찾는 단계를 포함하고

상기 그룹 행위는 C&C(Command and Control) 채널에 합류한 호스트들이 자신의 채널 접속 상태를 알리기 위해 IRC(Internet Relay Chat)가 사용하는 핑(Ping)과 퐁(Pong) 메시지를 주기적으로 주고 받는 행위이고,

상기 트래픽 데이터가 와치독 트래픽이며, 상기 행위 대상은 트래픽 데이터의 IP(internet protocol) 주소지인,

집단행동 악성코드 검색방법.

청구항 3

검색 대상 네트워크 상에서 송수신되는 트래픽 데이터를 수집하는 단계;

상기 트래픽 데이터를 검토하여, 단위 시간당 서로 다른 호스트들이 소정의 트래픽 데이터에 대하여 동일하게 수행한 행위인 그룹 행위와, 상기 트래픽 데이터를 구성하는 정보 중에서 그룹 행위가 이루어진 정보인 행위대상을 기준으로 복수의 호스트 그룹을 작성하는 단계;

상기 복수의 호스트 그룹 각각에 속한 호스트가 얼마나 일치하는지를 나타내는 상기 복수의 호스트 그룹간 유사도를 측정하는 단계; 및

상기 유사도를 이용하여 집단행동 악성코드를 찾는 단계를 포함하고

상기 그룹 행위는 분산 서비스 거부(Distributed denial of service) 공격이고,

상기 트래픽 데이터가 공격 트래픽이며, 상기 행위 대상은 트래픽 데이터의 도착지의 IP(internet protocol) 주소지인,

집단행동 악성코드 검색방법.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서

상기 복수의 호스트 그룹을 작성하는 단계는

단위 시간 동안 검색 대상 네트워크상의 소정 트래픽 데이터에 관련된 정보인 동일한 행위대상에 대하여, 동일한 그룹행위를 한 호스트의 개수를 세는 단계; 및

상기 호스트의 개수를 상기 검색 대상 네트워크에 속한 전체 호스트의 개수로 나눈 값이 임계값 이상이면 상기 동일한 그룹행위를 한 호스트들을 하나의 그룹으로 작성하는 단계를 포함하는 집단행동 악성코드 검색방법.

청구항 5

제4항에 있어서,

상기 임계값은 상기 검색 대상 네트워크에 상기 집단행동 악성코드가 존재할 확률인 집단행동 악성코드 검색방법.

청구항 6

제4항에 있어서,

상기 트래픽 데이터가 DNS(Domain name server) 트래픽이면, 상기 단위 시간은 DNS TTL(time to live)에 따라 결정되는 집단행동 악성코드 검색방법.

청구항 7

제6항에 있어서,

상기 단위 시간은 동적 DNS의 TTL(time to live)의 최소값보다 크고 정상 DNS의 TTL(time to live)의 최대값보다 작은 집단행동 악성코드 검색방법.

청구항 8

제4항에 있어서,

상기 트래픽 데이터가 와치독 트래픽이면, 상기 단위 시간은 핑퐁 주기의 최소값보다 크고 핑퐁 주기의 최대값보다 작은 집단행동 악성코드 검색방법.

청구항 9

제4항에 있어서,

상기 트래픽 데이터가 공격 트래픽이면, 상기 단위 시간은 상기 공격 트래픽이 따르는 분포를 기반으로 결정되는 집단행동 악성코드 검색방법.

청구항 10

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 유사도는 상기 복수의 호스트 그룹 중 제1 그룹과 제2 그룹에 동시에 속한 호스트의 개수를 상기 제1 그룹에 속한 호스트의 개수로 나눈 값과 상기 제1 그룹과 상기 제2 그룹에 동시에 속한 호스트의 개수를 상기 제2 그룹에 속한 호스트의 개수로 나눈 값의 합을 2로 나눈 값이고

상기 제1 그룹과 상기 제2 그룹은 동일한 행위대상에 대한 동일한 그룹행위를 한 그룹인 집단행동 악성코드 검색방법.

청구항 11

제10항에 있어서,

상기 유사도가 미리 결정된 유사도 임계값을 넘으면 상기 제1 그룹 또는 상기 제2 그룹에 속한 호스트들을 집단행동 악성코드에 감염된 호스트들로 판단하는 단계를 더 포함하는 집단행동 악성코드 검색방법.

청구항 12

제11항에 있어서,

상기 유사도가 상기 유사도 임계값과 유사도 오차범위 한계 값의 차보다 크고 상기 유사도 임계값보다 작으면 상기 제1 그룹 또는 상기 제2 그룹에 속한 호스트들을 의심이 되는 그룹으로 판단하는 단계; 및

상기 제1 그룹과 상기 제2 그룹의 동일한 행위대상을 블랙리스트에 추가하는 단계를 더 포함하는 집단행동 악성코드 검색방법.

청구항 13

제12항에 있어서

상기 복수의 호스트 그룹을 작성하는 단계는 상기 블랙리스트를 이용하는 집단행동 악성코드 검색방법.

청구항 14

검색 대상 네트워크 상에서 송수신되는 트래픽 데이터를 수집하는 제1 수단;

상기 제1 수단이 수집한 트래픽 데이터를 검토하여, 단위 시간당 서로 다른 호스트들이 소정의 트래픽 데이터에 대하여 동일하게 수행한 그룹행위 및 상기 트래픽 데이터를 구성하는 정보 중에서 그룹 행위가 이루어진 정보인 행위대상을 기준으로 복수의 호스트 그룹을 작성하는 제2 수단; 및

상기 복수의 호스트 그룹 중에서 제1 그룹과 제2 그룹에 각각 속한 호스트들이 얼마나 일치하는지를 나타내는 유사도를 측정하여 상기 유사도를 이용하여 집단행동 악성코드를 찾는 제3 수단을 포함하고,

상기 제1 그룹과 상기 제2 그룹은 동일한 행위대상에 대하여 동일한 그룹행위를 한 그룹이며,

상기 그룹 행위는 호스트 스캐닝, 포트 스캐닝, 웜 코드 전파, 봇 코드를 다운로드하는 행위, DNS(domain name server) 룩업과정에서 DNS에 질의를 하는 행위, C&C(Command and Control,) 채널에 합류한 호스트들이 자신의 채널 접속 상태를 알리기 위해 IRC((Internet Relay Chat)가 사용하는 핑(Ping)과 퐁(Pong) 메시지를 주기적으로 주고 받는 행위, 분산 서비스 거부(Distributed denial of service) 공격, 스패밍(spamming) 중 하나이고,

상기 트래픽 데이터가 DNS 트래픽인 경우, 상기 행위 대상은 DNS 쿼리의 도메인 네임이며,

상기 트래픽 데이터가 공격 트래픽인 경우, 상기 행위 대상은 트래픽 데이터의 도착지의 IP(internet protocol) 주소지이며,

상기 트래픽 데이터가 공격 트래픽인 경우인 경우, 상기 행위 대상은 트래픽 데이터의 도착지의 IP 주소지이며,

상기 트래픽 데이터가 명령 및 제어 트래픽인 경우, 상기 행위 대상은 트랙픽 데이터의 출발지 IP 주소인,

집단행동 악성코드 검색장치.

청구항 15

제14항에 있어서,

상기 호스트 그룹은

단위 시간 동안 동일한 행위대상에 대해 동일한 그룹행위를 한 임계값 이상의 호스트의 모임인 집단행동 악성코드 검색장치.

청구항 16

제15항에 있어서,

상기 유사도는 상기 제1 그룹과 상기 제2 그룹에 동시에 속한 호스트의 개수를 상기 제1 그룹에 속한 호스트의 개수로 나눈 값과 상기 제1 그룹과 상기 제2 그룹에 동시에 속한 호스트의 개수를 상기 제2 그룹에 속한 호스트의 개수로 나눈 값의 합을 2로 나눈 값인 집단행동 악성코드 검색장치.

청구항 17

제16항에 있어서

상기 제3 수단은 상기 유사도가 미리 정해진 유사도 임계값을 넘으면 상기 제1 그룹 또는 상기 제2 그룹에 속한 호스트들을 집단행동 악성코드에 감염된 호스트로 판단하는 집단행동 악성코드 검색장치.

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명은 집단행동 악성코드 검색 방법에 관한 것으로 특히, 악성코드에 감염된 호스트들의 그룹행위를 기반으로 집단행동 악성코드를 검색하는 방법에 관한 것이다.

배 경 기 술

[0002] 인터넷 기술의 발달과 이용인구의 증가로 인터넷을 구성하는 많은 기반시설들이 급격히 설치되면서 보안적 요소를 치밀하게 준비하지 못한 관계로 인터넷의 많은 취약점이 노출되고, 이런 취약점을 악용하려는 세력들이 등장하고 불특정 다수를 공격하는 사이버 테러가 빈번히 발생한다. 이러한 사이버 공격은 갈수록 지능화 조직화 되고 있으며 경제적 이득을 취득하기 위해 전문적이고 조직적인 네트워크를 형성해 나가고 있다. 이처럼 근래 사이버 공격자들은 불법행위를 대행하여 금전적 이득을 취득하기 위해 자신들의 존재와 위치 추적이 쉽지 않도록 하고, 자신들의 통제 하에 수많은 컴퓨터를 제어하는 기술을 고안하게 되었으며 이러한 목적으로 등장하게 된 것이 봇넷이다.

[0003] 공격자는 수천에서 수만 대의 컴퓨터에 봇을 설치하고 이들을 네트워크를 통해 동시에 제어하여 악성행위를 대행하게 한다. 봇넷의 규모가 커지며 이를 이용하여 공격하는 종류와 방법도 다양하게 나타나고 있으며 그에 따른 피해 또한 점점 심각해지고 있어서 사이버 보안의 최대의 이슈로 대두되고 있다.

[0004] 최근에 행해지는 많은 분산 서비스 거부(Distributed denial of service, 이하 "DDoS"라 함) 공격, 스패밍(spamming), 파밍(Pharming) 등의 다양한 악성공격들은 대부분 봇넷에 의해서 행해지고 있다. 악의를 가진 해커가 PC(personal computer)를 감염시켜 자신이 마음대로 조종할 수 있는 봇으로 만들고 이렇게 감염된 PC들의 수천, 수만대를 네트워크를 통해 일제히 조종하면서 악성행위들을 수행한다. 봇들을 조종, 통제하는 권한을 가진 봇 마스터(Bot master)에 의해 원격 조종되는 수천에서 수십만 대의 봇들이 네트워크로 연결되어 있는 형태를 봇넷이라고 한다.

[0005] 보통 여러 가지 경로를 통해서 PC가 봇에 감염되게 되는데 스팸 메일에 실행코드를 클릭하여 감염되는 경우, 웜에 감염코드를 싣고 취약성을 가진 PC들을 감염시키는 경우, 메신저를 통해 감염되는 경우 등 그 수법이 매우 다양하다. 게다가 루트킷을 이용한 감염을 수행하는 경우가 많아 감염여부를 쉽게 확인하기 힘들다.

[0006] 봇 마스터는 자신과 봇넷 사이의 명령전달 및 제어를 하기 위하여 IRC(Internet Relay Chat) 채널을 주로 이용한다. 1. 감염이 된 PC는 자동으로 2. 최근의 봇 코드를 받아 자신을 업데이트하고 3. 명령/제어 서버로 접속을 한다. 5. 봇넷에서 주로 TCP 6665 ~ 6669 포트를 사용하며, DDoS공격과 피싱, 파밍, 스패밍등 사이버 상의 많은 공격을 수행한다. 현재 대부분의 DDoS공격과 스팸메일은 봇넷을 통해 발생하고 있는 것이 보고 되었고 이를 통해 많은 심각한 피해들이 발생하고 있다.

[0007] 그 동안 봇넷은 기업에 큰 피해를 입혀 왔다. 악성 봇이 감염시킬 대상을 찾기 위해 네트워크를 스캔하면서 트래픽을 크게 증가시키기 때문에 네트워크가 매우 느려지거나 장애를 일으켜서 기업에서는 거의 업무를 할 수 없게 된다. 게다가 요즘에는 봇넷이 다른 보안 공격에 악용되면서 그 위험과 피해는 더욱 크게 증가하는 실정이다.

[0008] 봇넷을 통해 악성코드나 스파이웨어가 배포되고, 스팸 메일이 많이 뿌려진다. 특히 봇넷을 통해 이뤄지는 DDoS 공격의 피해는 매우 크다.

[0009] 2007년 2월에 발생한, 전세계 13개 루트 도메인 네임 서버(domain name server, 이하 "DNS"라 함) 서버에 대한 DDoS 공격이나, 2006년 하반기부터 늘어나기 시작한 국내 화상 채팅 서비스 등 성인용 서비스 업체들에 대한 공격, 2007년 10월에 있었던 게임 아이템 중개 사이트에 대한 공격에 이르기까지 봇넷을 통한 공격은 막강한 위력을 발휘하였다.

[0010] 국내, 해외 할 것 없이 봇넷은 매우 광범위하게 퍼져 있다. 한국정보보호진흥원(KISA)에서는, 어떤 PC가 봇 제

어 서버에 연결하려고 할 때 DNS 서버에서 실제 IP가 아닌 특정 IP를 돌려줌으로써 봇 제어 서버와의 연결을 차단하는 DNS 싱크홀 방식으로 국내 봇에 감염된 PC의 통계를 내고 있는데, 2007년 10월 현재 국내 악성 봇 감염율은 11.7%에 이른다. (한국정보보호진흥원, 인터넷침해사고 동향 및 분석월보, 2007.10. 여기에서는 국내 악성 봇 감염율을 전세계 악성 봇 감염 추정 PC 중 국내 봇 감염 PC가 차지하는 비율로 정의한다.)

[0011] 인터넷의 아버지라고 불리우는 구글의 빈트 서브 부사장은 2007년 1월 스위스 다보스에서 열린 세계경제포럼에서 봇넷의 확산을 전세계적인 유행병이라고 비유하며 인터넷에 접속하는 6억 대의 PC 중 1억~1억5000만대가 봇넷으로 이용되고 있고, 이 컴퓨터의 이용자들은 대개 봇넷에 참여하기를 원하지 않는 피해자들이라고 밝혔다.

[0012] 통계나 추정에 따라 편차가 있긴 하지만 전세계적으로 봇에 감염되어 봇넷의 좀비 PC로 활동하는 PC의 수가 매우 많다는 것은 분명해 보인다.

[0013] 최근 봇넷에 대응하기 위해 봇을 탐지하는 여러 기술들이 제안이 되고 있으나 아직까지 제안된 기술들이 원론적 기술 제안 수준의 한계점을 나타내고 있다.

발명의 내용

해결 하고자하는 과제

[0014] 본 발명이 이루고자 하는 기술적 과제는 효율적인 집단행동 악성코드 검색 방법을 제공하는 것이다.

과제 해결수단

[0015] 상기 과제를 달성하기 위한 본 발명의 하나의 특징에 따른 집단행동 악성코드 검색 방법은 검색 대상 네트워크의 트래픽 데이터를 수집하여 상기 트래픽 데이터를 검토하여 서로 다른 호스트에서 일어나는 공통행위인 그룹행위와 상기 그룹행위의 행위대상을 기준으로 복수의 호스트 그룹을 작성하고, 상기 복수의 호스트 그룹 각각에 속한 호스트가 얼마나 일치하는지를 나타내는 상기 복수의 호스트 그룹간 유사도를 측정하고, 상기 유사도를 이용하여 집단행동 악성코드를 찾는다.

[0016] 상기 과제를 달성하기 위한 본 발명의 다른 특징에 따른 집단행동 악성코드 검색장치는 검색 대상 네트워크의 트래픽 데이터를 수집하는 제1 수단, 상기 제1 수단이 수집한 트래픽 데이터를 검토하여 그룹행위와 상기 그룹행위의 행위대상을 기준으로 복수의 호스트 그룹을 작성하는 제2 수단 및 상기 복수의 호스트 그룹 중에서 제1 그룹과 제2 그룹에 각각 속한 호스트들이 얼마나 일치하는지를 나타내는 유사도를 측정하여 상기 유사도를 이용하여 집단행동 악성코드를 찾는 제3 수단을 포함하고, 상기 제1 그룹과 상기 제2 그룹은 동일한 행위대상에 대한 동일한 그룹행위를 한 인접한 단위 시간의 그룹이다.

효 과

[0017] 이상과 같이 본 발명에 의하면, 그룹행위를 기반으로 집단행동 악성코드를 검색함으로써 검색의 정확성을 높일 수 있고 검색율을 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

[0018] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0019] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "…부", "…기", 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0020] 먼저, 봇넷의 행동 순서 및 봇넷의 특성에 대해 도 1을 참조하여 설명한다. 도 1은 봇넷의 라이프 사이클을 나타낸 도면이다.

[0021] 봇은 집단행동 악성코드에 감염된 호스트이고, 봇넷은 봇들을 조정 통제하는권한을 가진 봇 마스터(Bot maste

r)에 의해 원격 조종되는 수천에서 수십만 대의 봇들이 네트워크로 연결되어 있는 형태를 의미한다.

[0022] 도 1에 도시된 바와 같이, 봇넷은 공격자가 만든 개인의 봇 악성코드를 이용해 취약성을 갖는 호스트를 감염시킨다(S110). 감염 시키는 방법은 이메일의 첨부파일을 이용하는 방법, 메신저를 통해 감염시키는 방법, 인터넷 웜을 이용하는 방법 등 다양하다.

[0023] 다음으로, 감염된 호스트는 실제 봇 코드를 다운로드한다(S120). S110 단계에서 직접 봇 코드를 다운로드 하지 않는 이유는 봇 코드 사이즈가 크고, 봇을 전파시키는 공격자가 자신의 봇 코드를 최신 데이터로 자주 업데이트를 하기 때문이다.

[0024] 감염된 호스트는 DNS 룩업(look up)을 수행한다(S130). 감염된 호스트는 다운로드한 봇 코드를 실행하여 봇 코드 내에 존재하는 명령 및 제어(Command and Control, 이하 "C&C"라 함) 채널 서버의 도메인 이름을 DNS로 전송하여 C&C 채널 서버의 IP 주소를 DNS에게 질의 한다. 거의 대부분의 봇은 DNS 룩업 과정을 수행한다. C&C 채널 서버는 봇 마스터가 봇을 통제하기 위한 채널 서버이고, 감염된 호스트의 봇 코드가 C&C 채널 서버의 도메인 이름을 이용하여 자동으로 C&C 채널에 접속한다.

[0025] 감염된 호스트는 DNS로부터 응답으로 받은 IP 주소를 이용해 C&C 채널 서버에 접속을 하여 C&C 채널에 합류한다(S140). 일반적으로 아이알씨(Internet relay chat, 이하 "IRC"라 함)가 C&C 채널서버로 이용된다.

[0026] 봇 마스터는 C&C 채널 서버에 접속하여 C&C 채널 서버를 통해 봇 호스트들을 통제하고 봇 호스트들에게 악성공격 명령을 전달한다(S150, S160). 그러면 악성공격 명령을 받은 봇 호스트들은 DDoS, 스패밍, 개인정보 유출 등의 악성 봇 행위를 수행한다.

[0027] 봇넷은 봇 마스터의 공격 명령을 수행하기 위해 보통 IRC 채널 서버에 상주하여 대기하고 있다. 채널 서버에 접속, 채널 서버에서 대기, 공격의 수행 등 봇넷의 행위는 그룹으로 행위 특성을 나타내는 그룹행위적 특성이 있다는 점에서 정상 호스트들의 행위와 구분된다. 그룹행위란 단위 시간에 서로 다른 호스트에서 일어나는 공통행위를 의미한다. 봇넷의 그룹행위적 특성을 위에서 설명한 봇넷의 행동 순서와 연관 지어 살펴보면 다음과 같다.

[0028] 봇넷은 취약성을 갖는 호스트를 탐지하고, 봇 악성코드를 웜을 통해 전파하는 과정에서 그룹행위 특성을 나타낸다. 즉 봇넷에 속한 호스트들은 호스트 스캐닝, 포트 스캐닝 및 웜 코드 전파의 악성행위를 하는데 이 과정에서 그룹행위적 특성을 확인 할 수 있다.

[0029] 감염된 호스트의 봇 코드 다운로드 과정에서 봇 코드를 다운로드하는 행위도그룹행위적 특성이 있다.

[0030] 봇넷의 구성호스트들이 C&C 채널 서버를 찾기 위한 DNS 룩업과정에서 DNS에 질의를 하는 행위도 그룹 행위적 특성을 갖는다.

[0031] C&C 채널에 합류한 봇넷의 구성 호스트들이 자신의 채널 접속 상태를 알리기 위해 IRC가 사용하는 Ping과 Pong 메시지를 주기적으로 주고 받는 행위도 그룹행위적 특성을 갖는다.

[0032] 봇넷에 속한 호스트들이 악성공격을 수행하는 과정에서 DDoS, 스패밍 등의 악성행위도 그룹행위적 특성을 나타낸다.

[0033] 정상적인 그룹행위와 봇넷의 그룹행위는 차이점을 지니며, DNS 룩업의 경우는 표 1과 같은 차이점이 있다.

표 1

[0034]

	도메인 네임에 접속하는 소스 IP	행위 패턴	DNS 타입
봇넷 DNS 질의 특성	고정된 크기의 그룹(봇넷 멤버들)	그룹행위가 간헐적으로 나타남	주로 DDNS(dynamic DNS)
정상 DNS 질의 특성	정상 유저들	비그룹행위가 무작위적으로 계속 나타남	주로 DNS

[0035] 다음으로, 본 발명의 실시예에 따른 집단행동 악성코드 검색장치에 대해 도 2를 참조하여 설명한다. 도 2는 본 발명의 실시예에 따른 집단행동 악성코드 검색 장치의 구조도이다.

[0036] 도 2에 도시된 바와 같이, 본 발명의 실시예에 따른 집단행동 악성코드 검색장치는 데이터 수집기(210), 그룹행위 분류기(220), 유사도 분석기(230) 및 봇넷 리포터(240)를 포함한다.

[0037] 데이터 수집기(210)는 센서와 연결되어 있어 센서로부터 네트워크를 통해 전달되는 데이터를 모으는 일을 수행한다. 센서는 TCP 트래픽 데이터 및 UDP 트래픽 데이터를 모아서 저장하였다가 또는 실시간으로 데이터 수집기(210)에게 전달한다. TCP 트래픽 데이터는 IRC 트래픽 데이터(TCP 6667번 포트 등)를 포함하며, UDP 트래픽 데이터는 DNS 트래픽 데이터(UDP 53번 포트 등)를 포함한다. 센서가 여러 곳에 걸쳐 존재 할수록 봇넷의 탐지 효과를 높일 수 있고, 일반적으로 특정 네트워크의 메인라우터 또는 보조 DNS(캐쉬서버)의 앞에서 존재한다. 수집되는 데이터는 데이터 수집기(210)가 전달 받기 적합한 데이터 구조를 갖고, UDP 소켓 또는 TCP 소켓으로 전달된다.

[0038] 그룹행위 분류기(220)는 데이터 수집기(210)로부터 전달받은 데이터를 검토하여 트래픽의 그룹행위에 따라 호스트 그룹을 만들어 자료구조에 저장하고, 그룹에 속한 호스트 리스트를 유사도 분석기(230)로 전달하는 역할을 한다. 즉, 그룹행위 분류기(220)는 행위대상(Object) 및 그룹행위을 기준으로 호스트 그룹을 찾아 그룹에 속한 호스트 리스트를 자료구조에 저장한다. 자료구조로는 연결 리스트 또는 해쉬 테이블(hash table)을 이용할 수 있다. 연결 리스트는 메모리를 적게 사용하는 장점을 지니나 유사도 측정에 시간이 많이 걸리고, 해쉬 테이블은 메모리를 많이 사용하는 단점이 있으나, 유사도 측정에 시간이 적게 걸리는 장점을 갖는다.

[0039] 유사도 분석기(230)는 그룹행위 분류기(220)로부터 전달받은 그룹들 중에서 동일한 행위대상에 대한 동일한 그룹행위를 한 인접한 단위 시간의 두 개의 그룹에 대해서 두 개의 그룹에 속한 호스트들이 얼마나 서로 일치하는지를 계산하여 두 개의 그룹간의 유사도(Similarity)를 측정한다. 측정된 유사도가 유사도 임계값(threshold)을 넘으면 두 개의 그룹 중 적어도 하나의 그룹에 속한 호스트들을 봇넷으로 판단하여 봇넷에 속한 호스트 리스트 및 봇넷의 행위대상을 봇넷 리포터(240)로 보낸다. 그리고, 유사도가 $\lambda_s-\delta<S<\lambda_s$인 경우, 두 개의 그룹 중 적어도 하나에 속한 호스트들을 의심되는 그룹으로 결정하고 의심되는 그룹에 속한 호스트 리스트 및 의심되는 그룹의 행위대상을 봇넷 리포터(240)로 전달한다. 여기서 여기서 λs 는 유사도의 임계값, δ는 임의의 유사도의 오차범위 한계 값이다.

[0040] 봇넷 리포터(240)는 유사도 분석기(230)로부터 전달 받은 봇넷에 속한 호스트 리스트 및 봇넷의 행위대상을 데이터베이스에 저장한다. 탐지 정확성을 높이기 위해 그룹과 행위 대상의 상관관계 분석(Correlation Analysis)를 수행할 수도 있다. 그리고, 봇넷 리포터(240)는 의심되는 그룹으로 판단된 그룹의 행위대상을 그룹행위 분류기(220)의 블랙 리스트로 전달한다.

[0041] 이하, 본 발명의 실시예에 따른 집단행동 악성코드 검색방법에 대해 도 3을 참조하여 설명한다. 도 3은 본 발명의 실시예에 따른 집단행동 악성코드 검색방법의 순서도이다.

[0042] 도 3에 도시된 바와 같이, 데이터 수집기(210)는 센서를 통해 데이터를 수집하여 그룹행위 분류기(220)로 전달한다(S310).

[0043] 그룹행위 분류기(220)는 데이터 수집기(210)로부터 전달받은 데이터를 이용해 호스트 그룹을 작성한다(S320).

[0044] 단위 시간 t 동안 검색 대상 네트워크에서 행위대상 O에 대한 행위 A를 한 호스트가 n_t개 있을 때, 수학식 1을 만족하면 n_t개의 호스트를 그룹 G(O,A)로 작성한다.

수학식 1

[0045] $$\frac{n_t}{N} \geq \lambda$$

[0046] 여기서, N은 검색 대상 네트워크의 전체 호스트의 개수이고, λ는 임의의 임계값이다.

[0047] 따라서, 호스트 그룹을 작성하기 위해서 필요한 요소는 n, λ, t, O이다.

[0048] 탐지하고자 하는 봇넷의 규모 즉, 봇넷에 속한 봇의 개수가 b라면 검색 대상 네트워크에 탐지하고자 하는 봇넷에 속한 호스트가 존재할 확률(P_b)은 b/N이고, λ는 P_b이다.

[0049] 그리고, 그룹행위가 나타나는 각 상황에 따라 다음과 같이 단위 시간 t와 행위 대상 O를 결정할 수 있다.

[0050] DNS 룩업 과정 또는 C&C 채널 이주 과정에서 관찰되는 DNS 트래픽의 경우를 살펴보면, 행위대상 O는 봇 호스트들이 도메인 네임 서버로 질의하는 DNS 쿼리(query)의 도메인 네임(domain name, DN)이고, 단위시간 t는 DNS TTL(time to live) 값을 고려하여 수학식 2과 같이 결정한다.

수학식 2

[0051] Min(TTL of DDNS)< t < Max(TTL of normal DNS)

[0052] C&C 트래픽은 연속적으로 발생하는 와치독 트래픽(Watchdog Traffic)과 일시적으로 발생하는 명령 및 제어 트래픽(Command and Control Traffic)을 포함한다.

[0053] 와치독 트래픽을 살펴보면, 행위대상 O는 TCP 트래픽의 도착지 IP 주소(destination IP address of TCP traffic) 또는 IRC의 핑퐁(Ping/Pong of IRC)이다. TCP 트래픽의 도착지는 IRC 서버이고, IRC의 핑퐁은 보통 90초 주기로 발생한다. 단위 시간 t는 Ping/Pong duration을 고려하여 수학식 3과 같이 결정한다.

수학식 3

[0054] Min(Ping/Pong duration)< t < Max(Ping/Pong duration)

[0055] 일시적으로 발생하는 명령 및 제어 트래픽을 살펴보면, 행위대상 O는 TCP 트래픽의 출발지 IP 주소(source IP address of TCP traffic)이다. TCP 트래픽의 출발지는 IRC 서버이다. 명령 및 제어 트래픽이 어떤 주기로 발생할지 알 수 없으므로 임의의 시간 동안 발생하는 명령 및 제어 패킷의 양을 평가해서 명령 및 제어 트래픽이 어떤 분포를 따르는지 계산하여 이를 기반으로 단위 시간 t를 결정한다.

[0056] 공격 트래픽(Attack traffic)은 DDoS 공격 트래픽과 스패밍 트래픽을 포함한다. 공격 트래픽을 살펴보면, 행위대상 O는 TCP/UDP/ICMP(Internet Control Message Protocol) 트래픽의 도착지 IP 주소(destination IP address of TCP/UDP/ICMP traffic)이다. TCP/UDP/ICMP 트래픽의 도착지는 공격 목표(Victim)이다. 공격 트래픽이 어떤 주기로 발생할지 알 수 없으므로 임의의 시간 동안 발생하는 공격 패킷의 양을 평가해서 공격 트래픽 분포가 어떠한 분포를 따르는지 계산하여 이를 기반으로 단위 시간 t를 결정한다.

[0057] 업데이트 트래픽을 살펴보면, 행위대상 O는 TCP 트래픽의 도착지 IP 주소 또는 도메인 네임이다. TCP 트래픽의 도착지는 업데이트 서버이다. 단위 시간 t는 봇넷의 코드 업데이트 주기를 기반으로 수학식 4와 같이 결정한다.

수학식 4

[0058] Min(Update duration)< t < Max(Update duration)

[0059] 호스트 그룹을 작성할 때, 성능을 높이기 위해서 블랙 리스트 및 화이트 리스트를 이용한다.

[0060] 블랙 리스트는 봇넷으로 탐지되지는 않았으나 의심되는 그룹으로 (Suspicious Group)으로 구분된 그룹의 행위대상이 기록된 리스트로서, 호스트 그룹 작성 시 블랙 리스트의 데이터를 우선적으로 처리한다.

[0061] 화이트 리스트는 널리 알려진 유명한 정상 도메인 및 IP들에 대해서 데이터 처리를 줄이기 위해서 준비하는 행위대상 리스트이다. 화이트 리스트의 행위대상에 대해서는 그룹행위를 구분하는 과정을 수행하지 않는다.

[0062] 유사도 분석기(230)는 작성된 복수의 그룹간 유사도를 측정한다(S330).

[0063] 유사도 분석기(230)는 복수의 그룹들 중에서 동일한 행위대상에 대한 동일한 그룹행위를 한 두 개의 단위 시간의 두 개의 그룹에 속한 호스트들이 얼마나 서로 일치하는지를 계산하여 유사도(Similarity)를 측정한다. 또는 복수의 유사도를 구해 평균값을 이용할 수도 있다.

[0064] 즉, 그룹 $G_1(O,A)$은 단위시간 t_1동안 행위대상 O에 대해 그룹행위 A를 한 그룹이고, 그룹 $G_2(O,A)$은 단위시간 t_2 동안 행위대상 O에 대해 그룹행위 A를 한 그룹이고, 그룹 $G_1(O,A)$과 그룹 $G_2(O,A)$에 동시에 속한 호스트가 얼마나 있는지를 고려하여 유사도를 구한다. 단위시간 t_1과 단위시간 t_2는 서로 인접한 단위시간일 수도 있고, 그렇지 않을 수도 있다.

[0065] 또는 그룹 $G_3(O,A)$은 단위시간 t_3동안 행위대상 O에 대해 그룹행위 A를 한 그룹이라고 할 때, 그룹 $G_1(O,A)$와 그룹 $G_2(O,A)$의 유사도를 구하고, 그룹 $G_2(O,A)$와 그룹 $G_3(O,A)$의 유사도를 구하고, 그룹 $G_1(O,A)$와 그룹 $G_3(O,A)$의 유사도를 각각 구하여 평균값을 이용할 수도 있다.

[0066] 그룹 $G_1(O,A)$와 그룹 $G_2(O,A)$의 유사도는 수학식 5를 이용하여 구할 수 있다.

수학식 5

$$\text{Similarity } S = \frac{1}{2}\left(\frac{|G_1 \cap G_2|}{|G_1|} + \frac{|G_1 \cap G_2|}{|G_2|}\right)$$

[0067]

[0068] 여기서, |G|는 그룹 G에 속한 호스트의 개수이고, |G|≠0이다. $G_1 \cap G_2$는 그룹 $G_1(O,A)$와 그룹 $G_2(O,A)$에 동시에 속한 호스트들로 이루어진 그룹을 의미한다.

[0069] 유사도 분석기(230)는 수학식 5를 통해 구한 유사도가 유사도 임계값(λ_s)을 넘으면 그룹 $G_1(O,A)$ 또는 그룹 $G_2(O,A)$에 속한 호스트들을 봇넷으로 판단한다(S340).

[0070] 유사도가 1에 가까울수록 봇넷일 확률이 높고, 0에 가까울수록 정상 그룹일 확률이 높아진다. 일반적으로 봇넷은 정상 그룹은 0~0.3 이내의 유사도 값을 갖는다.

[0071] 유사도 분석기(230)는 수학식 5를 통해 구한 유사도가 $\lambda_s - \delta$보다 크고 λ_s보다 작은 경우, $G_1(O,A)$와 그룹 $G_2(O,A)$에 동시에 속한 호스트들로 이루어진 그룹을 의심되는 그룹으로 결정하고, 의심되는 그룹의 행위대상을 블랙 리스트에 추가한다. 는 임의의 유사도의 오차범위 한계 값이다.

[0072] 본 발명의 실시예는 이상에서 설명한 장치 및/또는 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하기 위한 프로그램, 그 프로그램이 기록된 기록 매체 등을 통해 구현될 수도 있으며, 이러한 구현은 앞서 설명한 실시예의 기재로부터 본 발명이 속하는 기술분야의 전문가라면 쉽게 구현할 수 있는 것이다.

[0073] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면의 간단한 설명

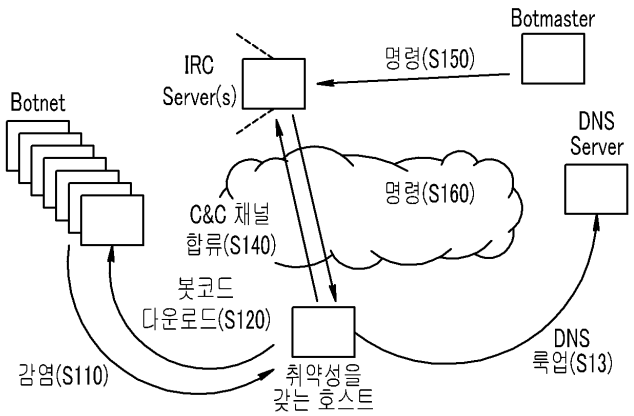
[0074] 도 1은 봇넷의 라이프 사이클을 나타낸 도면이다.

[0075] 도 2는 본 발명의 실시예에 따른 집단행동 악성코드 검색 장치의 구조도이다.

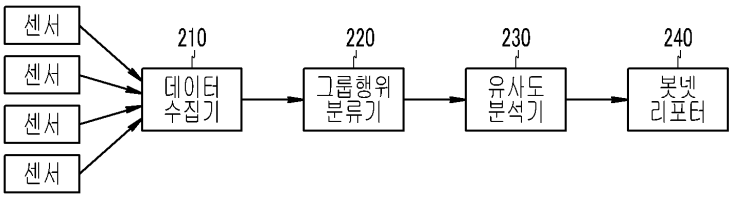
[0076] 도 3은 본 발명의 실시예에 따른 집단행동 악성코드 검색방법의 순서도이다.

도면

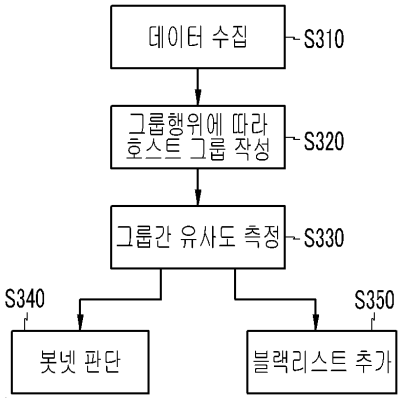
도면1



도면2


센서
센서
센서
센서
210
데이터 수집기
220
그룹행위 분류기
230
유사도 분석기
240
봇넷 리포터

도면3


데이터 수집
S310
그룹행위에 따라 호스트 그룹 작성
S320
그룹간 유사도 측정
S330
S340
봇넷 판단
S350
블랙리스트 추가