



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년09월13일
 (11) 등록번호 10-1182793
 (24) 등록일자 2012년09월07일

(51) 국제특허분류(Int. Cl.) (73) 특허권자
G06F 21/20 (2006.01) *H04L 12/22* (2006.01) 고려대학교 산학협력단
 (21) 출원번호 10-2011-0012227
 (22) 출원일자 2011년02월11일 (72) 발명자
 심사청구일자 2011년02월11일 이희조
 (65) 공개번호 10-2012-0092286
 (43) 공개일자 2012년08월21일 이제현
 (56) 선행기술조사문헌
 KR1020100084488 A* 권중훈
 KR1020100098241 A*
 *는 심사관에 의하여 인용된 문헌

(74) 대리인
 특허법인엠에이피에스

전체 청구항 수 : 총 11 항

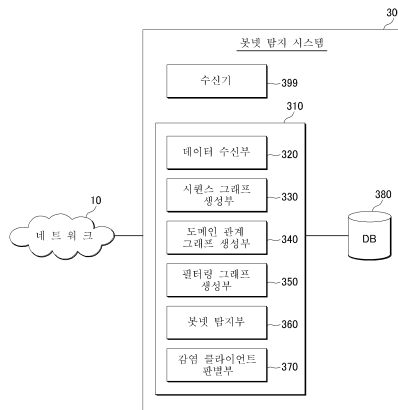
심사관 : 이정은

(54) 발명의 명칭 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법 및 시스템

(57) 요약

봇넷을 탐지하는 방법에 있어서, 도메인 이름 서비스 질의 데이터를 수신하는 단계, 상기 도메인 이름 서비스 질의 데이터를 이용하여 각 클라이언트 단말이 질의한 도메인 이름들의 순차적 관계를 나타내는 클라이언트 단말별 시퀀스 그래프를 생성하는 단계, 다수의 클라이언트 단말들의 시퀀스 그래프를 병합하여 클라이언트 단말들로부터 질의된 도메인 이름들의 전체적인 관계를 나타내는 도메인 관계 그래프를 생성하는 단계, 상기 도메인 관계 그래프로부터 클라이언트 단말들의 정상적인 질의 데이터를 정제한 필터링 그래프를 생성하는 단계 및 상기 필터링 그래프의 연결요소로부터 봇넷을 탐지하는 단계를 포함하는 봇넷 탐지 방법을 제공한다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 WR080951M0211612

부처명 (재)서울특별시시정개발연구원

연구사업명 세계유수연구소유치지원사업

연구과제명 2차년도 [1-B] Blended Services Applications

주관기관 고려대학교 산학협력단

연구기간 2009.12.01 ~ 2010.11.30

특허청구의 범위

청구항 1

봇넷을 탐지하는 방법에 있어서,

- (a) 도메인 이름 서비스 질의 데이터를 수신하는 단계,
 - (b) 상기 도메인 이름 서비스 질의 데이터를 이용하여 각 클라이언트 단말이 질의한 도메인 이름을 노드로, 상기 도메인 이름들의 질의 순서를 간선으로 포함하는, 클라이언트 단말별 시퀀스 그래프를 생성하는 단계,
 - (c) 다수의 클라이언트 단말들의 시퀀스 그래프를 병합하여 클라이언트 단말들로부터 질의된 도메인 이름들의 질의 순서 종속 관계를 나타내는 도메인 관계 그래프를 생성하는 단계,
 - (d) 상기 도메인 관계 그래프로부터 봇에 감염되지 않은 클라이언트 단말로부터의 질의 데이터를 제거하여 산출된 하나 이상의 연결 요소를 포함하는 필터링 그래프를 생성하는 단계 및
 - (e) 상기 필터링 그래프의 연결요소로부터 봇넷을 탐지하는 단계
- 를 포함하되,

상기 연결 요소는 하나 이상의 노드 및 간선을 포함하는 그래프이되, 자신의 외부에 있는 노드와는 연결되어 있지 않은 그래프인 봇넷 탐지 방법.

청구항 2

제 1 항에 있어서,

상기 (a) 단계는,

도메인 이름 서비스 서버로부터 도메인 이름 서비스 질의 데이터를 수신하거나, 네트워크를 통해 전달되는 데이터를 수집하는 수신기를 이용하여 수집된 도메인 이름 서비스 질의 데이터를 수신하는 단계를 포함하는 봇넷 탐지 방법.

청구항 3

제 1 항에 있어서,

상기 시퀀스 그래프는,

도메인 이름을 노드로 표현하고, 노드 1로 표현되는 도메인 이름 1과 노드 2로 표현되는 도메인 이름 2가 연속하여 질의된 경우, 노드 1에서 노드 2 방향으로 연장되는 방향 간선을 갖는 방향 그래프인 봇넷 탐지 방법.

청구항 4

제 1 항에 있어서,

상기 도메인 관계 그래프는,

방향 간선의 내부 값으로 간선에 부착된 도메인들의 연속된 질의 수($|Q_{AB}|$), 간선에 부착된 도메인들로 질의를 한 클라이언트 단말의 수($|C_{AB}|$) 및 상기 간선에 부착된 도메인들의 연속된 질의 수를 간선에 부착된 도메인들로 질의를 한 클라이언트 단말의 수로 나눈 값($|Q_{AB}|/|C_{AB}|$) 중 어느 하나 이상을 포함하는 봇넷 탐지 방법.

청구항 5

제 1 항에 있어서,
 상기 봇넷에 감염된 클라이언트 단말의 정보를 추출하는 단계 및
 상기 감염된 클라이언트 단말로 감염여부를 송신하는 단계를 더 포함하는 봇넷 탐지 방법.

청구항 6

제 1 항에 있어서,
 사용자 단말로부터 상기 사용자 단말의 봇넷 감염여부 확인 요청을 수신하는 단계,
 데이터베이스로부터 상기 사용자 단말의 식별정보와 연관되어 저장된 감염 클라이언트 단말 정보를 검색하는 단계 및
 상기 검색 결과를 이용하여 상기 사용자 단말의 봇넷 감염여부를 상기 사용자 단말로 송신하는 단계를 더 포함하는 봇넷 탐지 방법.

청구항 7

제 1 항에 있어서,
 상기 (e) 단계는,
 상기 필터링 그래프의 연결요소의 간선의 값이 미리 설정된 임계치를 초과하는 경우, 해당 연결요소를 봇넷으로 분류하는 봇넷 탐지 방법.

청구항 8

제 1 항에 있어서,
 상기 (e) 단계는,
 데이터베이스에 저장된 악성도메인 목록을 이용하여, 상기 필터링 그래프의 연결요소에 악성도메인이 포함되었는지 여부를 판별하는 단계 및
 상기 연결요소에 악성도메인이 포함된 경우, 상기 연결요소를 봇넷으로 분류하는 단계를 포함하는 봇넷 탐지 방법.

청구항 9

봇넷을 탐지하는 시스템에 있어서,
 도메인 이름 서비스 질의 데이터를 수신하는 데이터 수신부,
 상기 도메인 이름 서비스 질의 데이터를 이용하여 각 클라이언트 단말이 질의한 도메인 이름을 노드로, 상기 도메인 이름들의 질의 순서를 간선으로 포함하는, 클라이언트 단말별 시퀀스 그래프를 생성하는 시퀀스 그래프 생성부,
 다수의 클라이언트 단말들의 시퀀스 그래프를 병합하여 클라이언트 단말들로부터 질의된 도메인 이름들의 질의 순서 종속 관계를 나타내는 도메인 관계 그래프를 생성하는 도메인 관계 그래프 생성부,
 상기 도메인 관계 그래프로부터 봇에 감염되지 않은 클라이언트 단말로부터의 질의 데이터를 제거하여 산출된 하나 이상의 연결 요소를 포함하는 필터링 그래프를 생성하는 필터링 그래프 생성부 및

상기 필터링 그래프의 연결요소로부터 봇넷을 탐지하는 봇넷 탐지부를 포함하되,

상기 연결 요소는 하나 이상의 노드 및 간선을 포함하는 그래프이되, 자신의 외부에 있는 노드와는 연결되어 있지 않은 그래프인 봇넷 탐지 시스템.

청구항 10

제 9 항에 있어서,

네트워크를 통해 전달되는 데이터를 수집하는 수신기를 더 포함하되,

상기 수신기는 상기 수집된 질의 데이터를 상기 데이터 수신부로 송신하는 봇넷 탐지 시스템.

청구항 11

제 9 항에 있어서,

사용자 단말로부터 상기 사용자 단말의 봇넷 감염여부 확인 요청을 수신하고, 데이터베이스로부터 상기 사용자 단말의 식별정보와 연관되어 저장된 감염된 클라이언트 단말 정보를 검색하며, 상기 검색 결과를 이용하여 상기 사용자 단말의 봇넷 감염여부를 상기 사용자 단말로 송신하는 감염 클라이언트 판별부를 더 포함하는 봇넷 탐지 시스템.

명세서

기술분야

[0001] 본 발명은 봇넷 탐지 방법에 관한 것으로서, 보다 상세하게는 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법 및 그 시스템에 관한 것이다.

배경기술

[0002] 최근, 인터넷 기술의 급속한 발달과 인터넷 사용자의 폭발적인 증가로 인터넷은 정치, 경제, 사회, 문화 전반에 없어서 안될 요소로 자리 잡고 있다. 하지만, 인터넷 기술의 비약적인 발전에서 오는 순기능의 이면에는 인터넷을 악용하는 해킹, 웜바이러스 등 사이버상의 각종 위협 요소 또한 도사리고 있다. 이러한 사이버상의 위협 또는 범죄는 단순한 인터넷 사이트의 마비나 경제적 손실을 넘어 국가안보까지 위협하는 심각한 단계에 이르고 있다.

[0003] 종래의 사이버 공격은 일부 해커들의 호기심 차원에서 이루어져 왔으나, 최근에는 사이트 사용자의 인적 사항 유출, 금융정보 유출, 광고성 스팸메일의 대량 발송, 경쟁사에 대한 DDos(Distributed denial of service) 공격 등 불법 행위를 대행해주고 경제적 이득을 취하려는 목적의 사이버 범죄가 늘고 있다.

[0004] 최근 들어, 봇넷은 사이버 범죄의 가장 위협적인 요소로 주목 받고 있다. 악의를 가진 해커가 불특정 다수의 PC를 감염시켜 자신의 마음대로 조종할 수 있는 봇으로 만들고, 이렇게 감염된 수천, 수만 대의 PC가 네트워크로 연결돼 하나의 봇넷(Botnet)이 된다. 이렇게 형성된 봇넷은 봇에 감염된 PC의 통제권을 가진 봇 마스터(Bot master)에 의해 원격 조종되며, DDOS 공격이나 개인정보 수집, 스팸메일 전송, 피싱과 같은 사이버 범죄를 일으키게 되는 것이다. 최신의 봇은 유저 인터페이스를 통해 쉽게 봇코드를 생성할 수 있고, 제어할 수 있기 때문에 특별한 지식이나 기술이 없는 사람도 쉽게 봇넷을 만들고 이용할 수 있어 특정 해커만의 문제가 아니라는 점에서 더욱 위험하다.

[0005] 이에 봇넷을 탐지하는 다양한 기술이 개발되고, 현재 활용되고 있는 실정이다. 대표적으로, 클라이언트 기반의 봇넷 탐지 기술과 네트워크 기반의 봇넷 탐지 기술이 있다. 클라이언트 기반의 봇넷 탐지 기술은 크게 시그니처 탐지와 이상 행위 탐지 기법으로 나눌 수 있는데, 봇의 감염코드를 이용한 시그니처 탐지 기법의 경우 신종 봇에 대한 탐지가 느리고 소프트웨어적인 패킹기법을 통해 간단하게 회피할 수 있다는 단점이 있다. 또한, 이

상행위 탐지 기술의 경우 시스템 콜의 이상 행위를 이용하여 탐지하는 기술 등이 있으나, 오탐율이 높다는 단점이 있다.

[0006] 네트워크 기반의 봇 탐지에 대한 연구는 네트워크 트래픽을 패턴 분석하여 악성 봇의 감염여부를 판단하는 것으로, 모니터링 하는 네트워크 링크의 트래픽 양이 많으면 데이터 처리가 힘들고, 암호화 통신을 하는 경우 패킷 모니터링이 힘들다는 단점이 있다.

[0007] 이에 급증하는 사이버 범죄를 효율적으로 대처할 수 있는 방안이 시급한 실정이며, 구조적, 기술적으로 간단한 회피 설계만으로 봇넷 탐지 기술을 무력화하지 못하도록 봇의 행동 패턴 등을 이용한 봇넷 탐지 기술을 개발할 필요가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 필요에 의한 것으로, 그래프 구조를 이용하여 봇넷을 효과적으로 탐지할 수 있는 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법 및 시스템을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 봇넷 탐지 방법은 봇넷을 탐지하는 방법에 있어서, 도메인 이름 서비스 질의 데이터를 수신하는 단계, 상기 도메인 이름 서비스 질의 데이터를 이용하여 각 클라이언트 단말이 질의한 도메인 이름들의 순차적 관계를 나타내는 클라이언트 단말별 시퀀스 그래프를 생성하는 단계, 다수의 클라이언트 단말들의 시퀀스 그래프를 병합하여 클라이언트 단말들로부터 질의된 도메인 이름들의 전체적인 관계를 나타내는 도메인 관계 그래프를 생성하는 단계, 상기 도메인 관계 그래프로부터 클라이언트 단말들의 정상적인 질의 데이터를 정제한 필터링 그래프를 생성하는 단계 및 상기 필터링 그래프의 연결요소로부터 봇넷을 탐지하는 단계를 포함한다.

[0010] 또한, 본 발명의 제 2 측면에 따른 봇넷 탐지 방법은, 사용자 단말로부터 상기 사용자 단말의 봇넷 감염여부 확인 요청을 수신하는 단계, 데이터베이스로부터 상기 사용자 단말의 식별정보와 연관되어 저장된 감염 클라이언트 단말 정보를 검색하는 단계, 상기 검색 결과를 이용하여 상기 사용자 단말의 봇넷 감염여부를 상기 사용자 단말로 송신하는 단계를 더 포함한다.

[0011] 또한, 본 발명의 제 3 측면에 따른 봇넷 탐지 시스템은, 봇넷을 탐지하는 시스템에 있어서 도메인 이름 서비스 질의 데이터를 수신하는 데이터 수신부, 상기 도메인 이름 서비스 질의 데이터를 이용하여 각 클라이언트 단말이 질의한 도메인 이름들의 순차적 관계를 나타내는 클라이언트 단말별 시퀀스 그래프를 생성하는 시퀀스 그래프 생성부, 다수의 클라이언트 단말들의 시퀀스 그래프를 병합하여 클라이언트 단말들로부터 질의된 도메인 이름들의 전체적인 관계를 나타내는 도메인 관계 그래프를 생성하는 도메인 관계 그래프 생성부, 상기 도메인 관계 그래프로부터 클라이언트 단말들의 정상적인 질의 데이터를 정제한 필터링 그래프를 생성하는 필터링 그래프 생성부, 상기 필터링 그래프의 연결요소로부터 봇넷을 탐지하는 봇넷 탐지부를 포함한다.

발명의 효과

[0012] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 도메인 이름 서비스 질의 데이터를 그래프 구조를 이용하여 다수의 도메인 그룹으로 분류하고, 분류된 도메인 그룹의 봇넷 여부를 판별할 수 있다.

[0013] 또한, 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 도메인 이름 서비스 질의 데이터를 이용하여 봇넷을 탐지하고, 상기 탐지된 봇넷에 감염된 클라이언트 단말에 봇 감염정보를 송신할 수 있다.

[0014] 또한, 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 사용자 단말의 감염여부에 대한 확인 요청을 수신하고, 상기 수신 요청한 사용자 단말이 감염된 클라이언트인지 여부를 봇넷 탐지 과정에서 데이터베이스에 저장한 감염된 클라이언트 정보를 이용하여 판별하여 상기 사용자 단말로 감염여부를 송신할 수 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법을 설명하기 위한 구성도이다.

도 2는 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 시스템을 도시한 블록도이다.

도 3은 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법을 도시한 순서도이다.

도 4는 본 발명의 일실시예에 따른 클라이언트 단말의 도메인 이름 서비스 질의 데이터를 그래프 구조로 표현한 시퀀스 그래프를 도시한 도면이다.

도 5는 본 발명의 일실시예에 따른 시퀀스 그래프를 생성하는 알고리즘을 도시한 도면이다.

도 6은 본 발명의 일실시예에 따른 시퀀스 그래프로부터 도메인 관계 그래프를 생성하는 과정을 도시한 도면이다.

도 7은 본 발명의 일실시예에 따른 필터링 그래프의 연결요소로부터 봇넷을 탐지한 것을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0016] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 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.
- [0017] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 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.
- [0018] 우선, 봇넷 탐지 방법에 대한 자세한 설명에 앞서 봇넷(Botnet)에 대하여 간략하게 설명을 하기로 한다.
- [0019] 봇넷(Botnet)은 봇들을 조종/통제하는 권한을 가진 봇 마스터(Bot master)에 의해 원격 조종되는 수천에서 수십만 대의 봇(Bot)들의 네트워크(Network)로서, 봇(Bot)은 집단행동 악성코드에 감염된 호스트를 말한다.
- [0020] 봇(Bot)은 집단행동을 하기 위해 목적 서버에 접속하는 과정에서 DNS(Domain Name Service; 도메인 이름 서비스, 이하 "DNS" 라 한다) 질의(query)를 수행한다. 구체적으로 살펴보면, 봇(Bot)은 다운로드한 봇 코드(Code)를 실행하고, 상기 봇 코드 내에 존재하는 C&C(Command and Control; 명령 및 제어 채널 서버, 이하 "C&C" 라 한다) 서버의 도메인 이름을 DNS로 전송하여 C&C 서버의 IP 주소(Internet Protocol Address)를 DNS 서버에 질의 한다. 일부 봇넷은 C&C 서버에 접속하기 위해 IP 주소를 그대로 사용하기도 하지만, 보고된 대부분의 봇넷은 C&C 서버에 접속하기 위해 도메인 이름 또는 도메인 이름과 IP 주소를 모두 사용한다.
- [0021] 봇은 DNS 서버로부터 응답 받은 IP 주소를 이용해 C&C 서버에 접속을 하여 C&C 채널에 합류한다. 봇 마스터는 C&C 서버에 접속하여 C&C 서버를 통해 봇들을 통제하고 봇들에게 악성공격 명령을 전달한다. 상기 악성공격 명령을 받은 봇들은 DDoS, 스팸, 개인정보 유출 등의 악성 봇 행위를 수행한다.
- [0022] 최근의 봇넷은 여러 곳에 분산되어 존재하는 서버에 접속하기 위하여 다중의 도메인 이름(multiple domain name)을 사용하여 봇넷 탐지 시스템(300)의 탐지를 회피하기도 한다. 이는 서버에 접속 실패를 하여도 타 서버를 이용하여 봇 행위를 할 수 있도록 하기 위함이기도 하다.
- [0023] 도 1은 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법을 설명하기 위한 구성도이다.
- [0024] 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법은 도메인 이름 서비스 서버(200), 봇넷 탐지 서버(310), 데이터베이스(380), 수신기(399)에 의해 수행될 수 있다.
- [0025] 클라이언트 단말(100)은 네트워크(10)를 통해 도메인 이름 서비스 서버(200)와 연결되고, 도메인 이름 서비스 서버(200)는 상기 클라이언트 단말(100)의 도메인 이름 서비스 질의에 대한 응답으로 도메인의 IP 주소를 클라이언트 단말(100)에 제공한다.
- [0026] 네트워크(10)는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN) 또는 부가가치 통신망(Value Added Network; VAN) 등과 같은 유선 네트워크나 이동 통신망(mobile radio communication

network) 또는 위성 통신망 등과 같은 모든 종류의 무선 네트워크로 구현될 수 있다.

- [0027] 클라이언트 단말(100)은 네트워크(10)를 통해 도메인 이름 서비스 서버(200) 및 봇넷 탐지 시스템(300)에 접속할 수 있는 컴퓨터나 휴대용 단말기로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop) 등을 포함하고, 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, 스마트폰, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet) 단말 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.
- [0028] 도메인 이름 서비스 서버(200)는 도메인 이름 서비스를 제공하는 서버로서, 도메인 이름 서비스(DNS)란 호스트의 도메인 이름을 호스트의 네트워크 주소로 바꾸거나 그 반대의 변환을 수행할 수 있도록 하기 위해 개발되었다. 특정 컴퓨터(또는 네트워크로 연결된 임의의 장치)의 주소를 찾기 위해, 사람이 이해하기 쉬운 도메인 이름을 숫자로 된 식별 번호(IP 주소)로 변환해준다. 본 발명의 일실시예에 따르면, 클라이언트 단말(100)은 서비스를 받고자 하는 목적 서버에 접속하기 위하여, 상기 도메인 이름 서비스 서버(200)로 도메인 이름 서비스 질의를 하고, 상기 도메인 이름 서비스 서버(200)는 이에 대한 응답으로 IP 주소를 반환해준다. 악성 봇에 감염된 클라이언트 단말(100)의 경우, 특정 도메인으로서의 DNS 질의의 빈도수가 잦으며, 동일 봇에 감염된 클라이언트 단말간에는 유사한 질의 패턴을 갖는 등의 특성으로 인해 정상적인 클라이언트 단말과 차이점이 있다.
- [0029] 데이터베이스(380)는 일반적으로 데이터를 송수신하고 저장하기 위한 DBMS(Database Management System)으로 구성될 수 있으며, 상기 봇넷 탐지 서버(310)와 통신 가능하도록 연결되어 있다.
- [0030] 수신기(399)는 네트워크(10)상에서 도메인 이름 서비스 질의 데이터를 수집하는 장치로서, 이하 봇넷 탐지 시스템(300)의 타 구성 요소들과 함께 설명하기로 한다.
- [0031] 도 2는 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 시스템(300)을 도시한 블록도이다.
- [0032] 본 발명의 일실시예에 따른 봇넷 탐지 방법은 봇에 감염된 클라이언트 단말(100)들이 유사한 DNS 질의를 한다는 특징을 이용하여 다중의 C&C(command and control) 서버와 악성도메인을 탐지할 수 있다. 봇넷의 공통된 DNS 질의 행위는 질의를 하는 감염 클라이언트들 사이에서도 나타나지만, 질의를 받는 도메인 간에도 존재한다는 특징을 이용하여 다중의 C&C 서버를 탐지할 수 있는 것이다. 또한 다중의 C&C 서버에 질의를 분산시켜 DNS 유사도 기반 탐지기법(BotSniffer, BotGAD 등)을 회피하는 봇넷의 탐지에도 유용하다.
- [0033] 구체적으로, 봇넷 탐지 시스템(300)은 수신기(399)로부터 도메인 이름 서비스 질의 데이터를 제공 받고, 상기 질의 데이터로부터 클라이언트 단말(100)별 시퀀스 그래프(Sequence Graph)를 생성하고, 다수의 시퀀스 그래프(Sequence Graph)로부터 도메인 관계 그래프(Domain Relation Graph)를 생성한다. 또한 봇넷 탐지 시스템(300)은 상기 도메인 관계 그래프에서 정상적인 질의 데이터 등으로부터 발생하는 노이즈 데이터를 정제하기 위해 상기 도메인 관계 그래프로부터 기 설정된 임계치 이하의 내부 값을 갖는 간선 및 상기 간선에 부속된 노드를 제거한 필터링 그래프(Filtered Graph)를 생성하고, 상기 필터링 그래프로부터 봇넷을 탐지하여, 악성코드에 감염된 개별 클라이언트 단말(100)로 감염 정보를 송신한다.
- [0034] 또한, 상기 악성코드에 감염된 개별 클라이언트 단말(100) 정보는 데이터베이스(380)에 상기 클라이언트 단말(100)의 식별정보와 연관하여 저장하여 관리하고, 추후에 상기 클라이언트 단말(100)의 사용자가 자신의 클라이언트 단말(100)이 악성코드에 감염되었는지 여부를 조회할 때, 상기 데이터베이스(380)에 저장된 감염 클라이언트 단말(100) 정보를 검색하여 활용할 수 있다.
- [0035] 본 발명의 일실시예에 의한 봇넷 탐지 시스템(300)은 봇넷 탐지 서버(310), 데이터베이스(380) 및 수신기(399)를 포함한다. 그리고, 봇넷 탐지 서버(310)는 수신기(399)로부터 도메인 이름 서비스 질의 데이터를 수신하는 데이터 수신부(320), 클라이언트 단말(100)별 시퀀스 그래프를 생성하는 시퀀스 그래프 생성부(330), 상기 시퀀스 그래프를 병합하여 도메인 관계 그래프를 생성하는 도메인 관계 그래프 생성부(340), 상기 도메인 관계 그래프의 노이즈를 제거한 필터링 그래프를 생성하는 필터링 그래프 생성부(350), 상기 필터링 그래프의 각 연결요소들의 봇넷 여부를 탐지하는 봇넷 탐지부(360) 및 봇에 감염된 클라이언트 단말(100) 여부를 판별하는 감염 클라이언트 판별부(370)를 포함한다.
- [0036] 본 발명의 봇넷 탐지 시스템(300)은 그래프 구조(Graph Structure)를 이용하여 봇넷을 탐지한다는 점에서, 기존

의 봇넷 탐지 시스템과의 차이점을 갖는다. 그래프 구조를 이용하여 클라이언트 단말(100)의 도메인 이름 서비스 질의 데이터를 중첩적으로 표현하고, 상기 질의를 받은 도메인 이름을 행위 패턴과 통계 데이터에 따라 다수의 그룹으로 분류한다. 상기 분류된 각각의 그룹은 타 그룹과는 다른 공통적인 특성을 가지며, 상기 공통적인 특성을 갖는 같은 그룹 내에 존재하는 도메인들은 봇넷을 구성할 가능성이 높다는 점을 이용하여 봇넷을 탐지하게 된다. 봇넷 탐지 시스템(300)은 공통적인 특성을 갖는 그룹의 봇넷 여부를 탐지하는 과정에서, 기존에 알려져 있는 악성도메인 정보를 활용할 수 있으며, 이를 블랙리스트(Black List)라고 한다.

- [0037] 그래프 생성부(320)는 대규모 네트워크상의 봇넷 탐지를 위하여 그래프 구조를 활용하여 시퀀스 그래프, 도메인 관계 그래프, 필터링 그래프를 생성하는 역할을 한다.
- [0038] 상기 그래프 구조를 이용한 봇넷 탐지 방법은 대형 네트워크에 존재하는 여러 개의 봇넷과 여러 개의 악성도메인을 효율적으로 동시에 탐지할 수 있다는 장점을 갖는다.
- [0039] 도 3은 본 발명의 일실시예에 따른 도메인 이름 서비스 질의 데이터를 이용한 봇넷 탐지 방법을 도시한 순서도이다.
- [0040] 먼저, 봇넷 탐지 시스템(300)은 도메인 이름 서비스 질의 데이터를 수신한다(S300).
- [0041] 상기 질의 데이터의 수신은 DNS 서버로부터 도메인 이름 서비스 질의 데이터를 직접 수신하거나, 별도의 수신기(399)를 통하여 수신할 수 있다. 수신기(399)를 통한 질의 데이터의 수집은 센서를 이용하여 네트워크를 통해 전달되는 데이터를 모으는 것으로서, 상기 센서는 TCP 트래픽 데이터 및 UDP 트래픽 데이터를 모아서 저장하였다가 봇넷 탐지 서버(310)로 전달하거나, 실시간으로 봇넷 탐지 서버(310)로 전달할 수 있다. 상기 센서가 여러 곳에 걸쳐 존재 할수록 봇넷의 탐지 효과를 높일 수 있고, 일반적으로 특정 네트워크의 메인 라우터(router) 또는 보조 DNS(캐쉬서버)의 앞에서 존재할 수 있다. 수집되는 질의 데이터는 수집하여 전달하기에 적합한 데이터 구조를 갖고, UDP 소켓 또는 TCP 소켓으로 전달된다.
- [0042] 상기 수집된 도메인 이름 서비스 질의 데이터는 클라이언트 단말(100)이 도메인 이름 서비스 서버(200)로 도메인 이름 서비스 질의한 데이터를 말하며, 본 발명의 일실시예에 따르면 클라이언트 단말(100)의 IP주소 및 질의된 도메인 이름, 상기 도메인 이름에 대응하는 IP주소, 상기 질의된 시간 정보 등을 포함할 수 있다.
- [0043] 다음으로, 상기 수신된 도메인 이름 서비스 질의 데이터를 이용하여 클라이언트 단말(100)별 시퀀스 그래프를 생성한다(S310).
- [0044] 그래프는 자료 요소에 해당하는 노드(node)와 노드 간의 관계를 표시하는 간선(edge)으로 구성되며, 그래프 $G=(V, E)$ 로 표현한다. 그래프는 자료 요소들의 관계가 비선형 구조로 나타날 때 사용되는 자료 구조이다.
- [0045] 시퀀스 그래프(sequence graph)는 클라이언트 단말(100)로부터 질의된 도메인 이름을 노드(node)로 하고, 연속하여 질의된 도메인 이름의 질의 순서를 방향 간선(edge)으로 하는 방향 그래프이다. 연속하여 질의된 도메인 A와 도메인 B가 있는 경우, 노드 A(도메인 A)와 노드 B(도메인 B)는 노드 A에서 노드 B 방향으로 연장되는 방향 간선에 의해 연결된다.
- [0046] 상기 시퀀스 그래프는 클라이언트 단말(100)별로 생성이 되며, 도 4의 예시에서 클라이언트 단말 1은 도메인 A, 도메인 B, 도메인 C, 도메인 B, 도메인 A, 도메인 C 순으로 도메인 이름 서비스 질의를 하고, 이에 대응하는 시퀀스 그래프를 도 4의 오른쪽 그래프와 같이 표현할 수 있다. 상기 시퀀스 그래프는 클라이언트 단말(100)별 질의 전략을 살펴 볼 수 있는 자료 구조이며, 봇넷은 질의 대상 도메인 집합(Set) 및 질의 전략(query strategy)이 유사하다는 점에서 악성코드에 감염된 클라이언트 단말(100)의 유사성을 판단하는 기준으로 상기 시퀀스 그래프를 활용할 수 있다.
- [0047] 다음으로, 봇넷 탐지 시스템(300)은 다수의 클라이언트 단말(100)들의 시퀀스 그래프를 병합하여 도메인 관계 그래프를 생성한다(S310).
- [0048] 도메인 관계 그래프(Domain Relation Graph)는 도메인 간의 관계를 나타내는 그래프로서, 클라이언트 단말(100)로부터 질의된 도메인 이름을 노드로 하고, 연속하여 질의된 도메인 이름의 질의 순서를 방향 간선으로 한다. 상기 그래프의 간선은 내부 값으로 간선에 부속된 도메인의 연속된 질의 수($|QAB|$), 상기 부속된 도메인으로 질의를 한 클라이언트 단말(100)의 수($|CAB|$) 및 간선에 부속된 도메인들의 연속된 질의 수를 간선에 부속된 도메인들로 질의를 한 클라이언트 단말의 수로 나눈 값 ($|QAB|/|CAB|$)을 가질 수 있다. 이때, AB는 노드 A와 노드 B를 연결하는 간선을 의미한다. 상기 도메인 관계 그래프는 감염된 클라이언트 단말(100)들로부터 질의되는 도

메인의 전체적인 관계를 보여줄 수 있는 자료 구조이다.

- [0049] 도 5는 본 발명의 일실시예에 따른 시퀀스 그래프를 생성하는 알고리즘을 도시한 도면이다.
- [0050] 상기 알고리즘은 수집된 도메인 이름 서비스 질의 데이터를 이용하여 시퀀스 그래프를 만드는 알고리즘으로, 질의 세트(Query Set)에 포함된 모든 질의 데이터를 처리(LOOP 이용)하며, 시퀀스 그래프에 특정 클라이언트 단말(100)로부터 질의된 질의(query)가 포함되어 있는지를 검토하여, 없을 경우 노드와 간선을 생성하여 추가하고, 있는 경우 간선의 값(weight)을 증가시켜주는 구조이다.
- [0051] 다음으로, 봇넷 탐지 시스템(300)은 도메인 관계 그래프로부터 필터링 그래프(Filtered Graph)를 생성한다(S330).
- [0052] 상기 필터링 그래프는 도메인 관계 그래프에서 정상적인 도메인 이름 서비스 질의 데이터를 제거하여 정제된 그래프를 얻기 위해 생성하는 그래프이다. 상기 필터링 과정을 통해, 도메인 관계 그래프는 여러 개의 연결요소로 구분될 수 있다. 상기 연결요소에 포함된 도메인들은 봇넷 또는 강력히 연관된 봇넷들의 연합으로 분류된다. 그래프의 연결요소(connected component)란 적어도 한 개 이상의 간선으로 연결된 노드들로 구성된 부분 그래프(subgraph)를 뜻하는 용어로, 여기에서는 도메인 관계 그래프 중 악성 봇에 감염되지 않은 단말로부터의 질의 부분을 제거하여 산출한 하나 이상의 부분 그래프를 뜻한다. 따라서 각 연결요소는 하나 이상의 노드 및 간선을 포함하는 그래프로, 자신의 외부에 있는 노드와는 연결되어 있지 않은 그래프이다.
- [0053] 도 6은 본 발명의 일실시예에 따른 시퀀스 그래프로부터 도메인 관계 그래프 및 필터링 그래프를 생성하는 과정을 도시한 도면이다.
- [0054] 클라이언트 단말 1, 클라이언트 단말 2 및 클라이언트 단말 3의 시퀀스 그래프를 병합함으로써 도메인 관계 그래프를 생성하게 된다. 봇넷 탐지 시스템(300)은 상기 도메인 관계 그래프로부터 정상적인 질의 데이터를 제거하는 정제 작업을 거쳐 필터링 그래프를 생성하며, 상기 필터링 그래프는 간선의 값(weight)이 임계치(도 6의 예시에서는 2) 이하인 간선 및 이에 부속되는 노드를 제거하여 생성하며, 블랙리스트에 등록된 악성도메인은 검은색 노드로 표현하여 나타낼 수 있다. 필터링 그래프의 각 연결요소에 포함된 도메인 이름은 통계적으로 유사한 연결을 가지며, 그룹별로 분류되고, 동일한 그룹에 속한 도메인 이름들은 동일하거나 유사한 속성을 갖는다는 점에서 동일한 연결요소 내에 악성도메인이 존재할 경우, 상기 연결요소는 봇넷 또는 강력히 연관된 봇넷들의 연합으로 추정할 수 있다.
- [0055] 다음으로, 봇넷 탐지 시스템(300)은 필터링 그래프를 이용하여 봇넷을 탐지 한다(S340).
- [0056] 먼저, 필터링 그래프의 연결요소의 간선의 값이 미리 설정한 임계값을 초과하는지 여부를 조사하고, 임계값을 초과하는 경우에는 해당 연결요소를 봇넷으로 분류하는 방법을 사용할 수 있다.
- [0057] 또한, 봇넷 탐지 과정은 그래프 필터링을 통해 분류된 도메인그룹이 유사한 속성을 갖는다는 점과 기 존재하는 악성도메인 정보를 이용하여 수행될 수 있다.
- [0058] 도 7은 본 발명의 일실시예에 따른 필터링 그래프의 연결요소로부터 봇넷을 탐지한 것을 도시한 도면이다.
- [0059] 본 발명의 일실시예에 따른 봇넷 탐지 시스템(300)의 성능을 측정하기 위해, 실제 네트워크 시스템의 DNS 서버 전단(front of DNS server)에서 약 2주간에 걸쳐 DNS 질의 데이터를 수집하였다. 실험을 위해 약 2주 동안 데이터를 수집하였으며, 낮 시간을 기준으로 시간당 약 1,000만개의 질의 데이터가 수집되었다. 상기 질의 데이터에는 시간당 약 100만개의 도메인 이름과 약 20만개의 IP 주소가 포함되었다.
- [0060] 본 실험을 통해 블랙리스트에 속하는 악성도메인이 포함된 필터링 그래프의 구성 요소들은 다중 C&C 서버 또는 악성 도메인 그룹을 구성한다는 사실을 확인할 수 있었다.
- [0061] 다음으로, 봇넷 탐지 시스템(300)은 탐지된 봇넷으로부터 감염된 클라이언트 단말(100) 정보를 추출하고(S350), 상기 추출된 클라이언트 단말(100)로 감염여부를 송신한다(S360).
- [0062] 상기 감염된 클라이언트 단말(100)의 정보는 악성도메인을 추출하고, 추출된 악성도메인들로 DNS 질의를 하는 클라이언트 단말(100)을 역으로 추출하는 과정을 통해 이루어진다. 상기 클라이언트 단말(100)은 봇넷 탐지 시스템(300)으로부터 봇 감염여부에 대한 정보를 제공 받고, 부가적인 분석 작업을 통하여 감염된 악성코드의 종류 등을 제공 받을 수도 있다.
- [0063] 다시 도 1과 도 2를 참고하여 사용자가 자신의 사용자 단말이 봇넷에 감염되었는지 여부를 봇넷 탐지 시스템

(300)을 통하여 확인하는 방법에 대해 살펴보기로 한다.

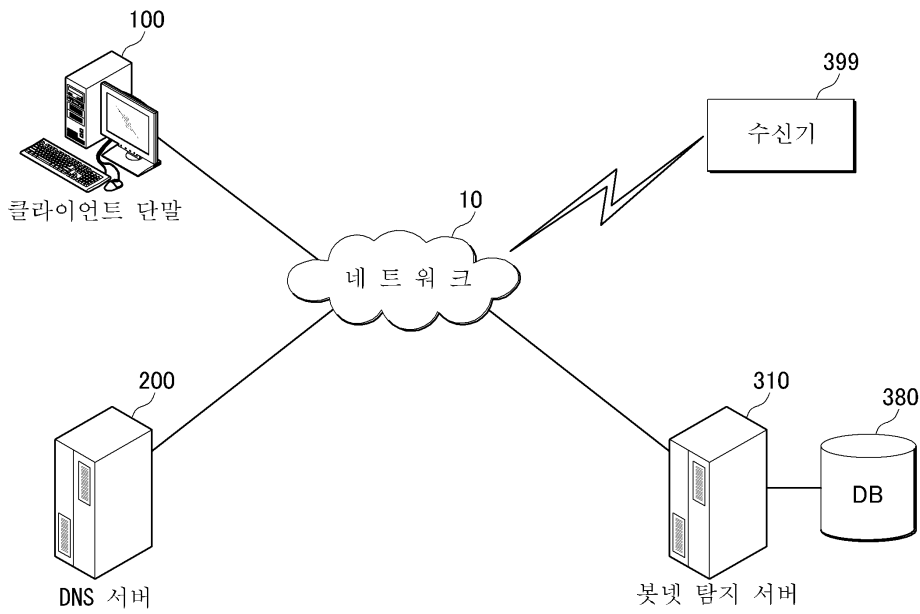
- [0064] 최근 들어, 악성 봇에 감염된 클라이언트 단말(100)로 인한 피해가 속출하고 있으며, 이에 따라 일반 사용자의 사용자 단말이 감염된 클라이언트인지 여부에 대한 확인 작업의 필요성이 대두되고 있다.
- [0065] 감염 클라이언트 판별부(370)는 사용자 단말로부터 봇넷 감염여부에 대한 확인 요청을 수신하고, 데이터베이스(380)에서 상기 사용자 단말의 식별정보와 연관되어 저장된 감염 클라이언트 단말(100) 정보를 검색한다. 상기 데이터베이스(380)는 봇넷 탐지 과정을 통해 탐지된 봇넷의 정보를 저장하며, 이와 관련된 악성도메인 정보를 저장한다. 또한, 상기 봇넷에 감염된 클라이언트 단말(100)의 정보를 저장하여, 사용자 단말의 감염여부에 대한 확인 작업을 수행하게 된다. 사용자 단말의 식별정보로는 IP 주소 등이 사용될 수 있으며, 유동 IP 주소를 사용하는 사용자 단말의 경우, IP 주소가 변경될 가능성이 있으므로 사용자 단말을 나타낼 수 있는 별도의 정보를 부가하여 식별정보로 활용할 수 있을 것이다.
- [0066] 본 발명은 감염된 클라이언트 단말에 의한 사이버 범죄, 불필요한 자원의 낭비 등으로 인한 사회 경제적인 문제를 해결하는데 일조할 수 있는 발명으로서, 봇넷들이 다중 도메인 이름을 활용하여 종래의 탐지 시스템을 회피하여 지속적으로 사이버 범죄 행위에 동원되고 있는 실정에서 봇넷을 탐지하고 감염된 클라이언트 단말을 판별할 수 있는 새로운 대안으로 활용될 수 있다. 또한, 그래프 구조를 이용하여 대규모 네트워크 상에서 다수의 봇넷을 빠르고 효율적으로 탐지할 수 있다는 점에서도 그 효용 가치가 크다고 할 수 있다. 따라서, 본 발명은 DNS 서버의 전단에 위치하여 보안 장치로 활용될 수 있으며, 네트워크 상의 데이터를 효율적으로 수집할 수 있는 수신기(399)와 함께 사용될 경우, 특정 서버(DNS 서버 등)와 독립적으로 활용될 수도 있다.
- [0067] 참고로, 본 발명의 실시예에 따른 도 2에 도시된 구성 요소들은 소프트웨어 또는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)와 같은 하드웨어 구성 요소를 의미하며, 소정의 역할들을 수행한다.
- [0068] 그렇지만 '구성 요소들'은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, 각 구성 요소는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다.
- [0069] 따라서, 일 예로서 구성 요소는 소프트웨어 구성 요소들, 객체지향 소프트웨어 구성 요소들, 클래스 구성 요소들 및 태스크 구성 요소들과 같은 구성 요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다.
- [0070] 구성 요소들과 해당 구성 요소들 안에서 제공되는 기능은 더 작은 수의 구성 요소들로 결합되거나 추가적인 구성 요소들로 더 분리될 수 있다.
- [0071] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0072] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0073] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

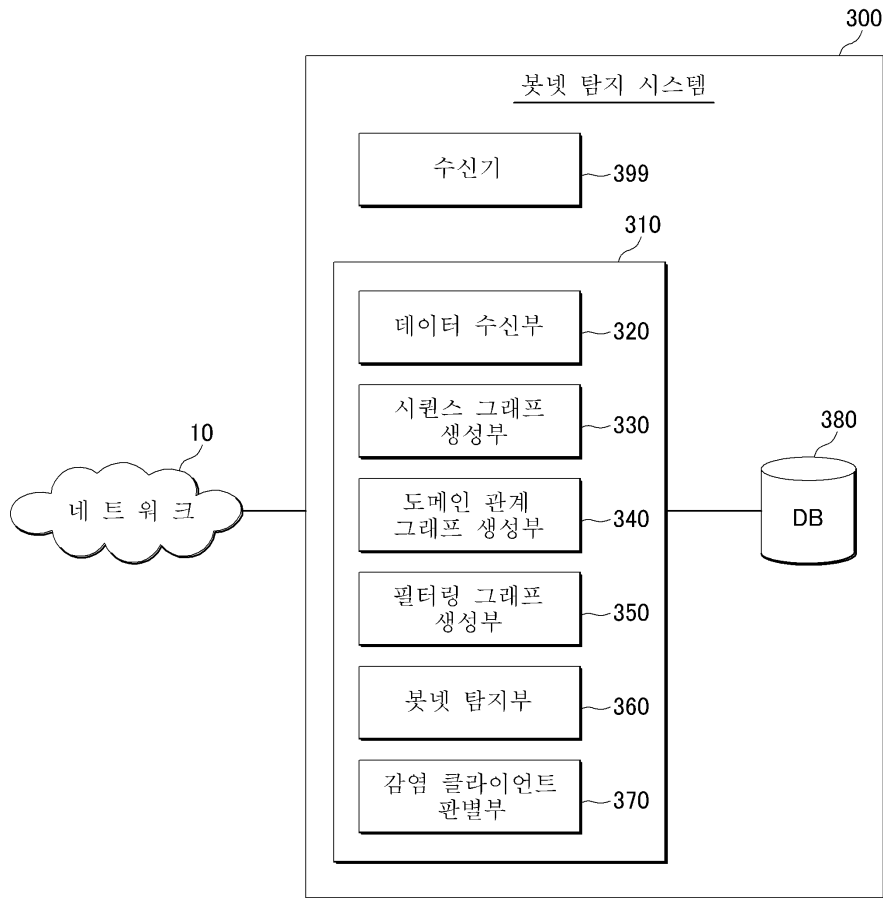
- [0074]
- | | |
|--------------------|---------------------|
| 10: 네트워크 | 100: 클라이언트 단말 |
| 200: 도메인 이름 서비스 서버 | 300: 봇넷 탐지 시스템 |
| 310: 봇넷 탐지 서버 | 320: 데이터 수신부 |
| 330: 시퀀스 그래프 생성부 | 340: 도메인 관계 그래프 생성부 |
| 350: 필터링 그래프 생성부 | 360: 봇넷 탐지부 |
| 370: 감염 클라이언트 판별부 | 380: 데이터베이스 |
| 399: 수신기 | |

도면

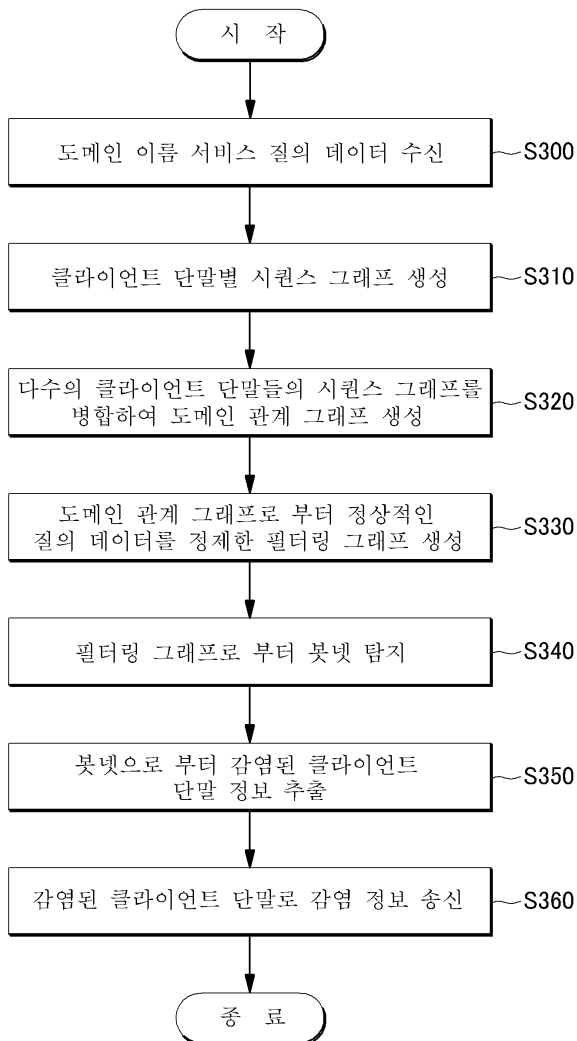
도면1



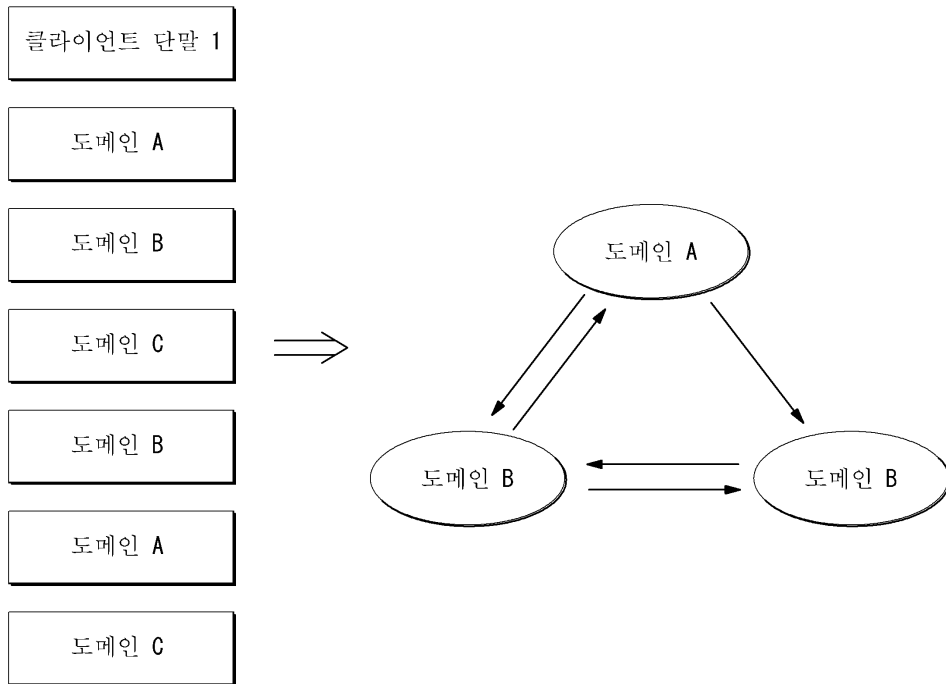
도면2



도면3



도면4



도면5

Algorithm : Query sequence graph generation

Symbols
A: Client Set, D: Domain Set, Q: Query Set
G: Graph, N: Node, E: Edge
 $A = \{a\}$, $D = \{d\}$, $Q = \{q \mid q := (a, d)\}$
 $G = \{N, E\}$, $N = \{n \mid n := d\}$, $E = \{e \mid e := (d_1, d_2, w)\}$

INPUT : Q
OUTPUT : G

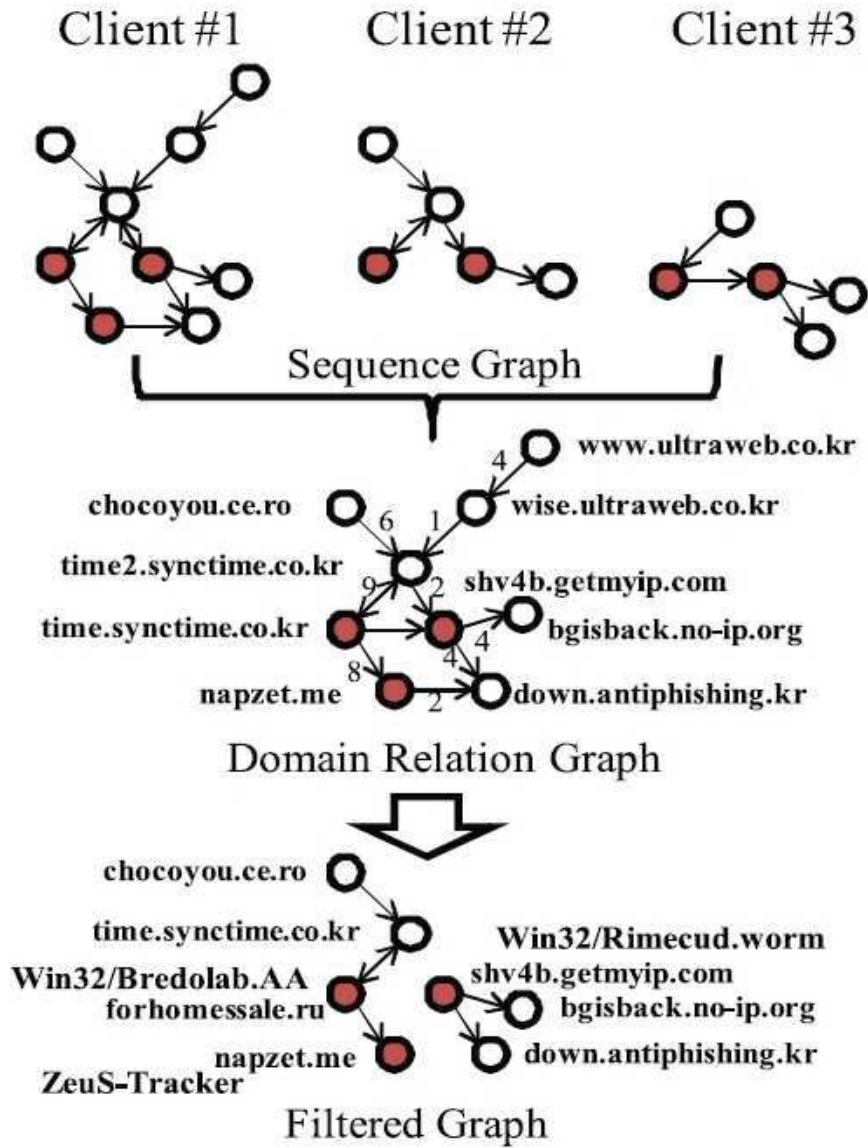
MAP<Key, Value> := < IP, Previous Domain >
N<D, I> := < Domain, Index >
E<N₁, N₂, W > := < Node₁, Node₂, Weight >

LOOP ($q_t, t=0 \sim i, i = |Q|$) {
 IF (MAP.Exist($q_t.a$)) {
 $d_{prev} = \text{MAP.GetValue}(q_t.a)$
 $d_{new} = q_t.d$
 $n_1 = \text{N.GetIndex}(d_{prev})$
 $n_2 = \text{N.Exist}(d_{new})$

 IF (! n_2) { $n_2 = \text{N.Add}(d_{new})$ }
 END IF

 IF ($w = \text{E.Exist}(n_1, n_2)$) {
 $w = w + 1$
 $e = (n_1, n_2, w)$
 E.Modify(e)
 } **ELSE** {
 $w = 1$;
 $e = (n_1, n_2, w)$
 E.Add(e)
 } **END IF**
 } **ELSE** { MAP.Add($q_t.a, q_t.d$) }
 END IF
} **END LOOP**

도면6



도면7

