Implementation of Netflow based Interactive Connection Traceback System

Jung-Tae Kim/Ik-Kyun Kim
Software Research Division
ETRI
Daejeon, Korea
email: jungtae_kim/ikkim21@etri.re.kr

Koo-Hong Kang
Dept. of Information and Communications
Engineering, Seowon University,
Chengju, South Korea
email: khkang@seowon.ac.kr

***Abstract*— The paper proposes a method and system for finding stepping stones, as well as origins of the advanced cyber attacks based on the interactive connections traceback system. To do so, the traceback system to be installed at the enterprise gateway utilizes distributed netflow collectors that subscribe netflow information from nearby edge routers with command configured to support netflow generation and traceback agents for finding real-time connections from the victim to attacker. By implementing such a system would support the real-time connection traceback of the attack origins for the interactive hacking attacks without any helps of the Internet Service Providers or governmental security organizations.**

Keywords-Netflow; Peer-to-Peer; Connection Traceback; Interactive Hacking; Timing-based Traceback.

I. INTRODUCTION

With developments of Internet technologies and smart devices, information available on the Web and stored on the personal devices are ever valuable than before. Increasing demands for the social network services also triggered usages of internetworked smart devices such as phones, pads and tablets. Consequently, such valuable information resources including personal profiles available on the Social Network Services (SNS) and enterprise resources on the Web need to be protected from the cyber attacks. Although there are many tools and solutions for preventing the cyber attacks based on the static analysis of network behaviors and host processes available, there are still lack of a traceback mechanism for detecting the origins of attacks due to sophisticated hacking techniques as well as the accessibility issues across the closed Internet Service Provider (ISP) networks for network information gatherings.

Since Zhang and Paxson [1] proposed a distinctive method for detecting Stepping Stones based on the packet size and timing of interactive traffics, its theoretical limitation need to be expended to find the origin of hacking connections and to cope with the Network Address Translation (NAT) [2] and IP Spoofing [3][4] issues. In order to overcome the conventional limitations of the timing-based traceback algorithms, we have extended the principle ideas of the Zhang and Paxson to detect interactive stepping stones, such as Internet Relay Chat (IRC), as well as the attack origins. The paper is organized with the related literature reviews in the Section II and introduces details on the proposed Netflow-based Connection Traceback System (NCTS) in the Section III. After describing the implementation details of the proposed system in the Section IV, the paper concludes with requirements and enhancements for the future works.

II. LITERATURE REVIEW

A. Advanced Persistent Threat (APT)

Recent hacking attacks become more and more sophisticated known as Advanced Persistent Threat (APT) [5], which is a set of stealthy and continuous computer hacking processes. The APT attacks usually targets enterprises or national organizations for business or political purposes. The APT processes generally involves a series of hidden attacks for a long period of time. Such targeted attacks use a wide variety of techniques, including drive-by downloads, Structured Query Language (SQL) injection, malware, spyware, phishing, and spam. Nevertheless of such complex hacking techniques used, a hacker need to make a connection to the Command and Control (C&C) servers to take control over the Zombie or Victim PCs.

Therefore, by observing and monitoring the interactive connection processes involved in order to achieve APT attacks, which normally consists of three major processes: advanced, persistent, and threat, we can identify the Stepping Stones used as well as the origin of the attacks. Firstly, the advanced process signifies sophisticated techniques using malware to exploit vulnerabilities in systems. The persistent process suggests that an external command and control is continuously monitoring and extracting data off a specific target. The threat process indicates human involvements in orchestrating the attack. The APT process includes three major phases [6] that occur over a period of months. Beginning with the Phase 1 called Reconnaissance, Launch, and Infect stage, attackers perform reconnaissance, identifies vulnerabilities, launches the attack, and infects target hosts. Then, the final Phase 3 the attacker controls infected hosts, updates code, spreads to other machines, and discovers and collects target data during the Phase 2 called Control, Update, Discover and Persist stage. The final Phase 3, called Extract and Take Action stage, indicates that the attacker extracts data from the target network and takes action to destroy the systems, as well as information disclosures.

B. Netflow

Information sources for the security analysis on the APT attack processes are based on the netflow information as the attack connections pass through various internetworking devices including routers and switches [7].

 ISBN: 978-1-61208-439-8

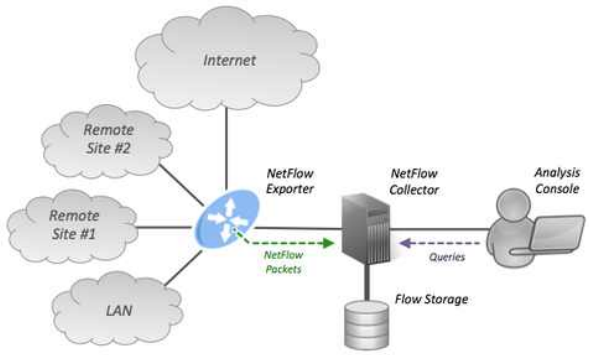

Figure 1. Example of a NetFlow Architecture [9].

Generally routers support and generate a flow information for each unidirectional layer 4 (transport layer) connections that are routed and maintained in its cache [8]. As shown in the Figure 1, the netflow helps to analyze the IP network traffic information as it enters or exits an (ingress or egress) switch interfaces. By analyzing the data that is provided by netflow, a network administrator can monitor the source and destination, class of service, and the cause of traffic congestions. The Cisco standard netflow version 5 defines a flow as a unidirectional sequence of packets that all share the following seven values including the following information; Ingress interface (Simple Network Management Protocol ifIndex), Source & Destination IP address, IP protocol, Source port for (User Datagram Protocol) UDP or (Transmission Control Protocol) TCP, Destination port, type and code, and IP Type of Service (ToS). The netflow enabled routers or switches will output a flow record when it determines that the flow is finished.

TABLE I. NETFLOW HEADER AND RECORD INFORMATION

Netflow Information	
Components	***Details***
Headers	. Version number (v5, v8, v9, v10) . Sequence number to detect loss and duplication . Timestamps at the moment of export, as system uptime or absolute time. . Number of records (v5 or v8) or list of templates and records (v9)
Records	. Input interface index used by SNMP . Output interface index or zero if the packet is dropped. . Timestamps for the flow start and finish time, in milliseconds since the last boot. . Number of bytes and packets observed in the flow . Layer 3 headers: - Source & destination IP addresses - Source and destination port numbers - ICMP Type and Code - IP protocol & Type of Service (ToS) value - IP address of the immediate next-hop - Source & destination IP masks

Routers can also be configured to output a flow record at a fixed interval even if the flow is still ongoing. Netflow records are traditionally exported using UDP and collected using a netflow collector. The IP address of the netflow collector and the destination UDP port must be configured on the sending router. All netflow packets begin with version-dependent header that contains at least four fields as shown in the Table I. A netflow record can also contain a wide variety of information about the traffic in a given flow such as a netflow version 5, which is one of the most commonly used versions, followed by version 9, contains information described in the Table I.

C. P2P(Peer-to-peer) system

In order to collect the netflow information from network, any existing netflow exporters (router configuration in Appendix) required to be configured in a way to export netflow information to the netflow collector as shown in the Figure 1. As the available numbers of exporters increase, there should a solution to manage distributed collectors in a systematic manners. For this purposes, we propose a peer-to-peer (P2P) network which is a type of decentralized and distributed network architecture.

As shown in the Figure 2, it generally consists of individual nodes in the network called "peers" that act as both suppliers and consumers of resources, in contrast to centralized client–server model where client nodes request access to resources provided by central servers. In other words, networks in which all computers have equal status are called peer-to-peer or P2P networks.

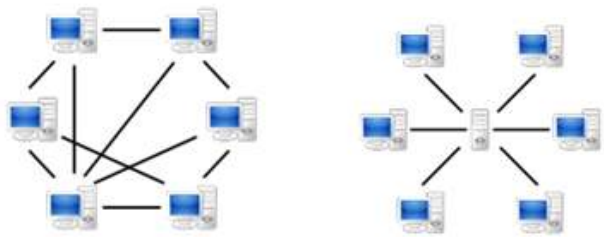

Figure 2. P2P Network vs Client-Server Model [10].

In a peer-to-peer network, tasks such as searching for files or streaming audio/video are shared amongst multiple interconnected peers who each make a portion of their resources including processing power, disk storage or network bandwidth directly available to other network participants, without the need for a centralized coordination by servers. A peer-to-peer network is designed around the notion of equal peer nodes simultaneously functioning as both clients and servers to the other nodes on the network.

By applying such concept into the netflow collectors and traceback manager, the proposed system significantly increases performances for querying and exchanging of a netflow information to search a target connection information among distributed collectors.

 ISBN: 978-1-61208-439-8

III. NETFLOW BASED CONNECTION TRACEBACK SYSTEM

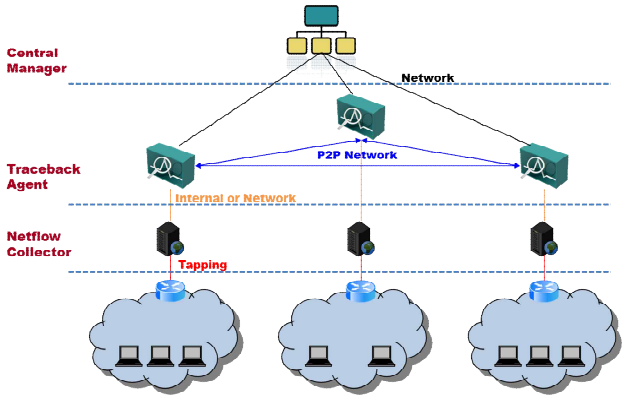

Figure 3. Configuration of the Netflow-based Connection Traceback System (NCTS).

The proposed Netflow-based Connection Traceback System (NCTS) has three major components; Central P2P Manager, Traceback Agent and Netflow Collector as shown in the Figure 3. The bottom layer with the Netflow Collectors (NC) are tapped to the existing network infrastructures, such as routers and switches in order to collect no-sampled netflow information. Especially the NCs collect the netflow v5 information from nearby routers and manage its headers and flow records separately in the database.

The Traceback Agents (TA) requests the netflow data stored in the distributed NCs and obtains traceback results by matching the ON/OFF patterns of a session. It connects NCs to calculate and retrieves a particular session time and other related information including src & dest_ip, src & dest_port, and protocol information for identifying correlations among sessions available within a given time. The TAs also provide a web-based GUI for obtaining victim related information including IP address, port number and connection time as well as displaying connection traceback results for users. Finally, the Central P2P Manager acts as a TA connection server which manages the distributed TAs in peer-to-peer (P2P) manner.

TABLE II. COMPONENTS OF THE NCTS SYSTEM

NCTS Components	
Components	***Description***
User	Input Victim IP Address, Port Number and Attack Time based on TA' Web UI.
Traceback Agent	Manage connections among distributed NCs and generate the Fingerprint information from the Target Connection.
Netflow Collector	Collect netflow from the edge routers and Search Flow Information
Edge router	NetFlow v5 Information Generation on the Ingress Ports
Central P2P Manager	Manage Network Connections among Distributed TAs in P2P Manner

The above Table II summarizes the physical components of the NCTS with a brief description. Also the detailed NCTS software block design is shown in the below Figure 4 with interfaces in Table III.

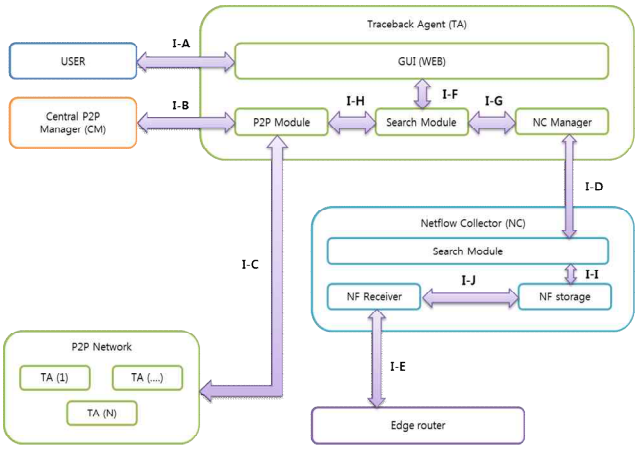

Figure 4. Interfaces and SW Block of the NCTS System.

TABLE III. NCTS SW BLOCK INTERFACES

NCTS Components	
Interfaces	***Description***
I-A	USER and GUI
I-B	CM and TA P2P module
I-C	P2P Network and TA P2P module
I-D	NC Manager and NC Search module
I-E	Edge router and NF Receiver
I-F	GUI and TA search module
I-G	TA search module and NC manager
I-H	TA search module and P2P module
I-I	NC search module and NF storage
I-J	NF storage and NF receiver

Basically, the system has three sub-system blocks of Central P2P Manager (CM), Traceback Agent (TA) and Netflow Collector (NC).

Firstly, the Netflow Collectors (NC) has a netflow receiver which collects no-sampled netflow information from nearby edge router and store them in the database called the netflow storage. The netflow storage can be either commercial databases or file systems depending on the total volumes of collected netflow which varies according to the total bandwidth available as well as the number of flows per second [11]. Generally, Cisco systems defined the amount of netflow export data being about 1.5% of the switched traffic in the router [8]. Currently, the NC collects netflows from the edge router with a default active and inactive timer for 30 and 1800 seconds respectively via a UDP communication. The NC also provides search functions to identify target connection flows from the netflow storage. Those distributed NCs are managed by the NC manager in Traceback Agents (TA) which requests netflow data stored in the distributed NCs and obtains traceback results on the web based GUI as shown in the Figure 5.

 ISBN: 978-1-61208-439-8

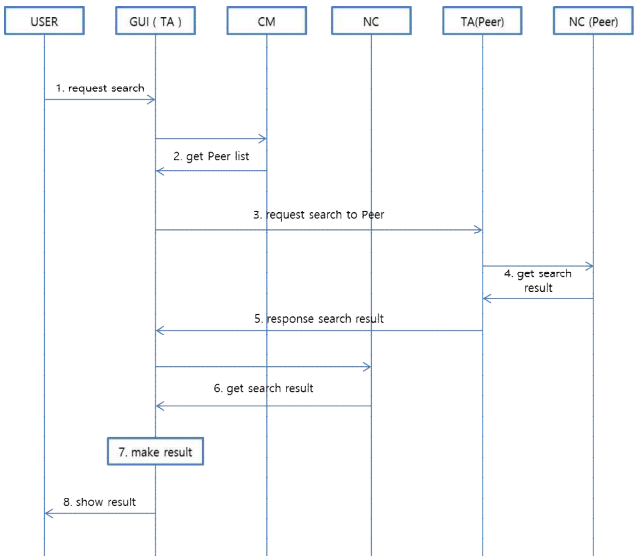

Figure 5. Message Sequence Charts for the NCTS System.

Finally, the Central P2P Manager (CM) acts as a TA connection server which manages status and connections of the distributed TAs in peer-to-peer (P2P) manner.

IV. IMPLEMENTATION

To extend the Zhang and Paxson's works [1], which was to detect the Stepping Stones based on the packet size and timing of interactive traffics, the proposed testbed was designed to overcome the theoretical limitations to find the origin of hacking connections regardless of the NAT and IP Spoofing. Consequently, a real-time evaluation of an interactive connection traceback were setup up according to the below Figure 6.

The Attacker (HA) attacks a Victim (HV) via connection made through a Stepping Stone (HS) with a telnet or SSH sessions. In addition, each of the edge nodes (attacker, stepping stone and victim) were under the Internet line sharer with NAT with unknown private IP addresses.

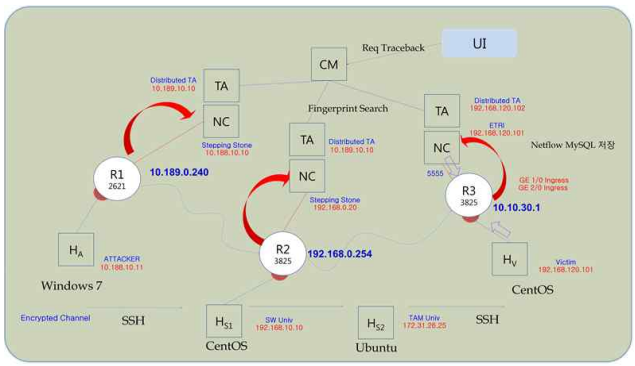

Figure 6. Testbed configuration of the NCTS System.

Also especially, one of the Stepping Stone (HS2) was configured without NC and TA. Then, each connections from Attack-Stepping Stone and Stepping Stone-Victim were lasted until the edge routers (R1~3) exports corresponding netflow records to the distributed NCs. Upon identification of an attack connection (target connection) with victim IP, port and time, the web-UI provides target lists. It also shows the fingerprint information (a set of vector values that representing On & OFF time of flows for a target session) of the target connection [Figure 7-b] by calculating On and Off time values with comparing the time intervals between the netflow records. Consequently, each fingerprint information contains a time series of ON and OFF values for a target connection. Therefore by matching the fingerprint information of a target connection with others connections helps to identify the related connections that are maintained and shown a similar time intervals with the target attack connections. To do so, a time series analysis method called the Correlation Point Function (CPF) were introduced to measure a ratio between the summation of the minimum fingerprint elements and the summation of the maximum fingerprint elements. By matching candidate connections in (Correlation Value) CV rank orders [Figure 7-c] that is collected from the distributed NCs helps to verifies that the CPF values over 0.8 shown a clear distinction of the attack connections among many others connections that exist within a connection time zone. Also by sorting the connection information based on the CV ranks and connection time order, the traceback results [Figure 7-a] are shown from source to destination IP and port numbers of the Internet line sharer as well as the unknown private IP addresses of the Attacker (H_A), Stepping Stone (H_{S1}) and Victim (H_V). The system also founds a connection information of the Stepping Stone (H_{S2}) which was configured without the NC and TA.

Consequently, we have obtained 8 connection traceback results rather than 4 because of the nodes were configured with an unknown private IPS with NAT. Furthermore, the traceback results [Figure 7-a] were sorted according to the time order from Attacker to Victims due to the nature of an interactive communication sessions that start from origin to destination and terminate in an exact reverse order.

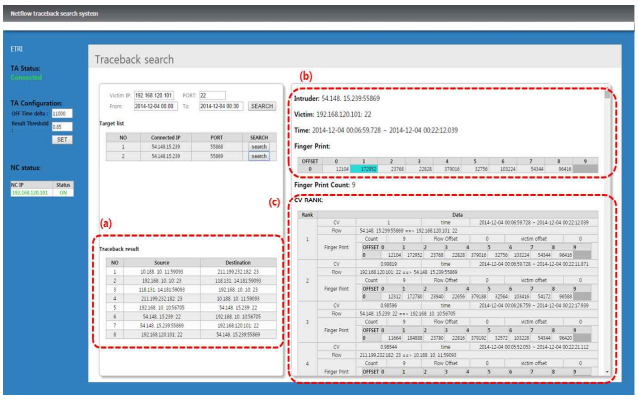

Figure 7. Web based Traceback Search UI.

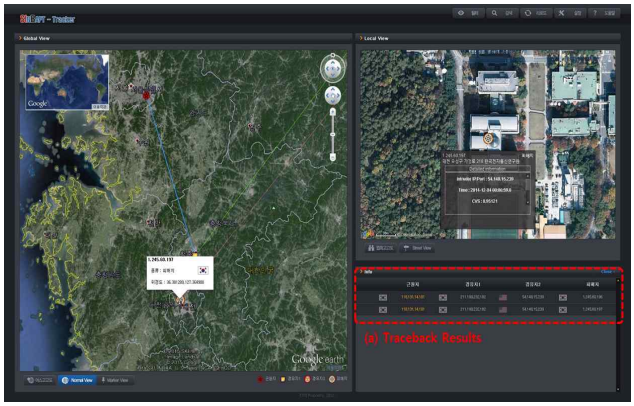

Figure 8. Tracker UI of the Netflow-based Connection Traceback System.

Finally, the traceback results obtained from the Web-UI can then be interpreted and transferred to the Tracker UI [Figure 8-a], as shown in the Figure 8, which shows a 3D map (Google Earth [12]) interfaces with geographical information. For the domestic location details, including IP address, network and organization name with address and zip code, can be obtained from the WHOIS Open API [13] and the IP2Location™ [14] provides overseas geographical location information based on the IP addresses.

Currently the WHOIS Open API supports as following services;

- APNIC (Asia Pacific Network Information Centre): APNIC Whois Database is an official record that contains information regarding organizations that hold IP address resources and AS numbers in the Asia Pacific.

- ARIN (American Registry for Internet Numbers): ARIN manages the distribution of Ipv4 and Ipv6 address space and Autonomous System Numbers (ASNs), collectively called Internet number resources, for the United States, Canada, and many Caribbean and North Atlantic islands.

- RIPE (Réseaux IP Européens): Regional Internet Registry for Europe, the Middle East and parts of Central Asia which allocates and registers blocks of Internet number resources to Internet service providers (ISPs) and other organizations.

- LACNIN (Latin American and Caribbean Internet Addresses Registry): Assigning and administrating the Internet numbering resources (IPv4, IPv6), Autonomous System Numbers, Reverse Resolution and other resources for the region of Latin America and the Caribbean.

- AFRINIC (African Network Information Center): Regional Internet Registry (RIR) for Africa, responsible for the distribution and management of Internet number resources such as IP addresses and ASN (Autonomous System Numbers) for the African region.

V. CONCLUSION

The proposed Netflow-based Connection Traceback System (NCTS) provides a real-time identifying and tracing of the cyber attack origin by analyzing the fingerprint information of the collected netflow v5 on the interactive communication sessions.

After preconfigured Netflow Collectors (NC) collect the netflow information from the nearby edge routers and store those information, then the distributed Traceback Agents (TA) manage connections among distributed NCs and generate the fingerprint information from the selected target connection at victim. Such distributed TAs are managed in P2P manner by the Central P2P Manager (CM) for establishing connections and sharing netflow information and support to calculate and find a correlations among flows based on the fingerprint information that represents a set of vector values that representing On & OFF time of flows for a target sessions. Finally, the system helps to monitor the traceback results with the inbuilt Web UIs that distributed along with TAs and also provide a 3D User Interfaces for monitoring purposes.

For the future works, the proposed system and traceback need to consider the cases of netflow information subscription losses from routers due to the nature of UDP communication. As the routers also supports a netflow exports via the Stream Control Transmission Protocol (SCTP) [19], the netflow loss and results mis-ordering of the sorting can be solved. Finally, the proposed system needs further works on supporting the various other types and versions of network flow information available including NetFlow v9, sFlow®, CFlow, JFlow as well as supporting tracbacks of the non-interactive connections.

APPENDIX

Configuration Examples for Configuring NetFlow and NetFlow Data Export [20].

Example Configuring Egress NetFlow Accounting configure terminal ! interface ethernet 0/0 ip flow egress
Example Configuring NetFlow Subinterface Support 1. NetFlow Subinterface Support For Ingress (Received) Traffic On a Subinterface

```
configure terminal
!
interface ethernet 0/0.1
 ip flow ingress
!
```

2. NetFlow SubInterface Support For Egress (Transmitted) Traffic On a Subinterface

```
configure terminal
!
interface ethernet 1/0.1
 ip flow egress
!
```

Example Configuring NetFlow Multiple Export Destinations

```
configure terminal
!
ip flow-export destination 10.10.10.10 9991
ip flow-export destination 172.16.10.2 9991
!
```

ACKNOWLEDGMENT

This work was partly supported by the IT R&D program of MSIP/KEIT [No.B0101-15-1293, Cyber targeted attack recognition and trace-back technology based-on long-term historic analysis of multi-source data]

REFERENCES

[1] Y. Zhang and V. Paxson, “Detecting Stepping Stones,” Proc. 9th USENIX Security Symposium, 2000, pp. 4-5.

[2] G. Yao, J. Bi, and A. V. Vasilakos, “Passive IP Traceback: Disclosing the Locations of IP Spoofers From Path Backscatter,” IEEE Transaction on Information Forensics and Security, Volune 10, No. 3, Mar 2015, pp. 476-478.

[3] F. Ali, “IP Spoofing,” The Internet Protocol Journal, Volume 10, No. 4, Dec 2007, pp 2-9.
https://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_10-4/ipj_10-4.pdf

[4] M. Tanase, “IP Spoofing: An Introduction,” Mar 2003.
http://www.symantec.com/connect/articles/ip-spoofing-introduction

[5] Advanced persistent threat, From Online Wikipedia http://en.wikipedia.org/wiki/Advanced_persistent_threat

[6] White Paper “Advanced Persistent Threats and other advanced attacks,” Websense Inc. 2011.
https://www.websense.com/assets/white-papers/whitepaper-websense-advanced-persistent-threats-and-other-advanced-attacks-en.pdf

[7] M. Robertson and B. MacMahon, “Cisco Cyber Threat Defense Solution 1.1 Design and Implementation Guide,” Technical Documents on the Cisco Cyber Threat Defense July 2013.
http://www.cisco.com/c/dam/en/us/td/docs/security/network_security/ctd/ctd1-0/design_guides/ctd_1-1_dig.pdf

[8] NetFlow Services Solutions Guide, from Cisco System. Jul 2001.
http://www.cisco.com/c/en/us/td/docs/ios/solutions_docs/netflow/nfwhite.html

[9] NetFlow, From Wikipedia, the free encyclopedia.
https://en.wikipedia.org/wiki/NetFlow

[10] Peer-to-Peer Network, From Wikipedia, the free encyclopedia.
https://en.wikipedia.org/wiki/Peer-to-peer

[11] Netflow Bandwidth Calculator, From the Plixer Inc.
https://www.plixer.com/Scrutinizer-Netflow-Sflow/netflow-bandwidth-calculator.html

[12] Google 3D Earth Plugin
https://www.google.com/earth/explore/products/plugin.html

[13] Whois Service, KRNIC's Internet Directory
http://whois.kisa.or.kr/eng/
http://wq.apnic.net/apnic-bin/whois.pl
https://www.arin.net/
https://apps.db.ripe.net/search/query.html
http://lacnic.net/cgi-bin/lacnic/whois?lg=EN
http://afrinic.net/

[14] IP2Location, Geolocate IP Address Location
http://www.ip2location.com/

[15] DAUM Map API
http://apis.map.daum.net/
http://www.ip2location.com/

[16] Spring Framework 3.1, Pivotal Software, Inc.
http://projects.spring.io/spring-framework/

[17] MyBatis data mapper framework 3.1, MvnRepository
http://mvnrepository.com/artifact/org.mybatis/mybatis/3.1.1

[18] Java-based document object model (JDOM) 1.1
http://www.jdom.org/

[19] Technical Docement on the NetFlow Reliable Export With SCTP, Cisco Systems, Inc. June 2006.
http://www.cisco.com/c/en/us/td/docs/ios/netflow/configuration/guide/15_1s/nf_15_1s_book/nflow_export_sctp.pdf

[20] Getting Started with Configuring Cisco IOS NetFlow and NetFlow Data Export, Cisco Systems, Inc. 2011.
http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/netflow/configuration/12-4/nf-12-4-book/get-start-cfg-nflow.html#GUID-BBE4C130-DD22-4064-9AE0-EC8D18D5

 ISBN: 978-1-61208-439-8