ENSC 835: Communication Networks
Spring 2011
Final Project Presentation

Internet Endpoints Profiling

Seyed Mojtaba Mohammadian Abkenar
http://www.sfu.ca/~smmohamm/current_projects/default.html
mojtaba_abkenar@sfu.ca

Profiling a behavior refers to the act of observing measured data and extracting information which is representative of the behavior or usage patterns. Profiling is useful in developing a model of the behavior and in deriving guidelines of what is normal and abnormal within that context.

Examples of successful uses of profiles include

- profiling of traffic patterns on server links to uncover DoS
- web-server profiling
- power usage profiles for efficient power management
- profiling end-to-end paths to detect performance problems
- profiling of traffic patterns on aggregated gateway and router links to facilitate accurate application classification
- etc

Internet endpoints profiling is study of hosts in Internet based on the following criteria :

- Applications running on the host
- Popularity of the host
- Relationship between the host and the other hosts
- Classifying traffic flows generated by the host

Why it is important?

- Understanding Internet access trends at a global scale
- Understanding shifts in clients' interests for traffic engineering and IT-business arenas.
- Network security, uncover DOS attack, and worm activities, intruder detection.
- Pricing of the Internet

Internet profiling and traffic classifying approaches :

- Port-based approach.
- Payload-based approach.
- Host-behavior-based approach.
- Flow features-based approach**.**
- Search engine-based approach.

BLINC Methodology

BLINC = Blind classification

BLINC is based on observing and identifying patterns of host behavior at the transport layer.

It analyzes these patterns at three levels of increasing detail:

- social level
- functional level
- application level

Classification at the social level

BLINC identifies the social role of each host in two ways.

- It focuses on host’s popularity, namely the number of distinct hosts it communicates with.
- It detects communities of hosts by identifying and grouping hosts that interact with the same set of hosts. A community may signify a set of hosts that participate in a collaborative application, or offer a service to the same set of hosts.

Classification at the functional level

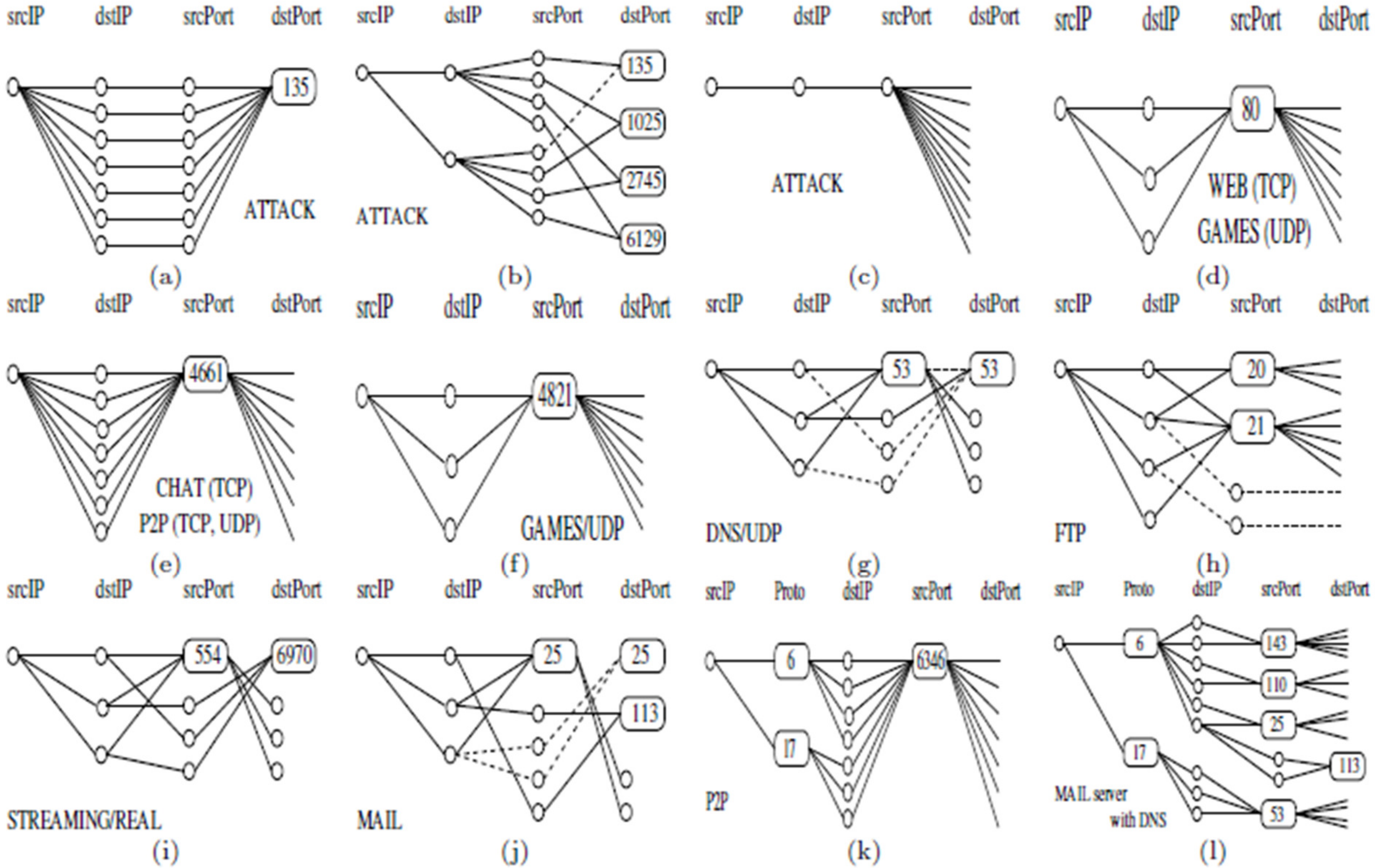
At this level, BLINC identifies the functional role of a host: hosts may primarily offer services, use services, or both.

Classification at the application level

In this level, BLINC combines knowledge from the two previous levels coupled with transport layer interactions between hosts in order to identify the application of origin. The basic insight exploited by this methodology is that interactions between network hosts display diverse patterns across the various application types.

BLINC provides a classification using only the 4-tuple (IP addresses and ports), and then, this can be refined using further information regarding a specific flow, such as the protocol or the average packet size. BLINC models each application by capturing its interactions through empirically derived signatures using graphlets that reflect the “most common” behavior for a particular application. A sample of application-specific graphlets is presented in the following figure.

Graphlets are small connected non-isomorphic induced subgraphs of a large network.

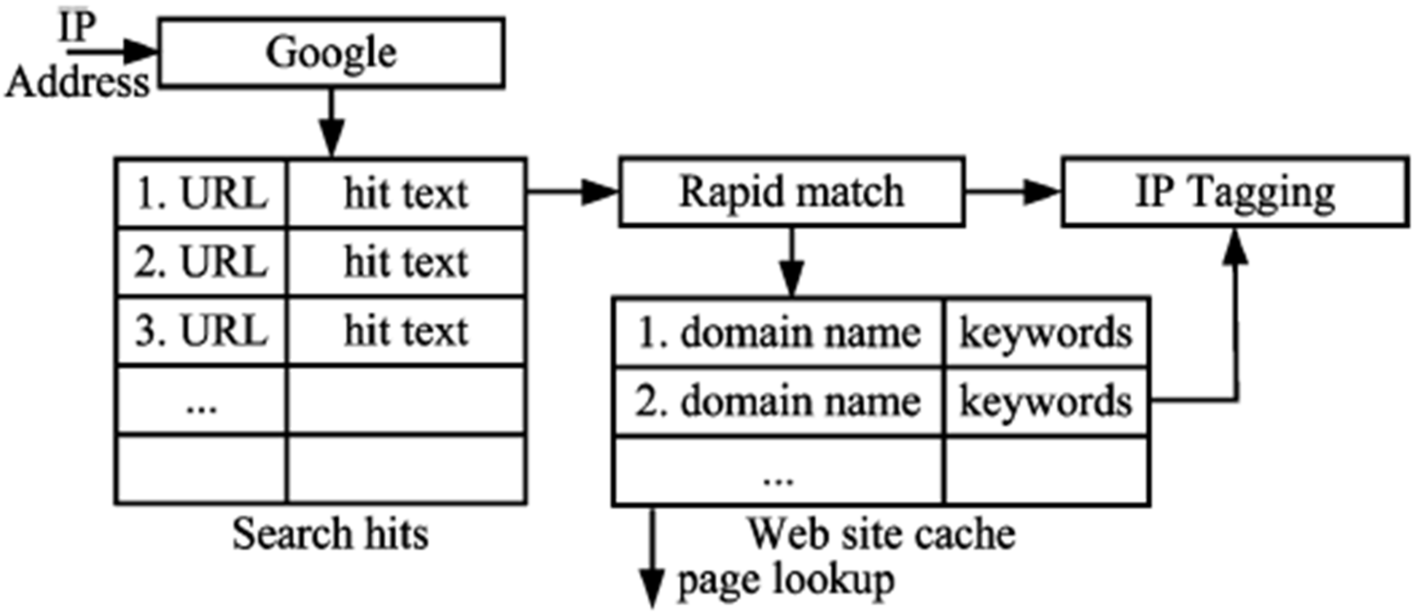

srcIP
dstIP
srcPort
dstPort
135
ATTACK
(a)
srcIP
dstIP
srcPort
dstPort
135
1025
2745
6129
ATTACK
(b)
srcIP
dstIP
srcPort
dstPort
ATTACK
(c)
srcIP
dstIP
srcPort
dstPort
80
WEB (TCP)
GAMES (UDP)
(d)
srcIP
dstIP
srcPort
dstPort
4661
CHAT (TCP)
P2P (TCP, UDP)
(e)
srcIP
dstIP
srcPort
dstPort
4821
GAMES/UDP
(f)
srcIP
dstIP
srcPort
dstPort
53
53
DNS/UDP
(g)
srcIP
dstIP
srcPort
dstPort
20
21
FTP
(h)
srcIP
dstIP
srcPort
dstPort
554
6970
STREAMING/REAL
(i)
srcIP
dstIP
srcPort
dstPort
25
25
113
MAIL
(j)
srcIP
Proto
dstIP
srcPort
dstPort
6
17
6346
P2P
(k)
srcIP
Proto
dstIP
srcPort
dstPort
6
17
143
110
25
113
53
MAIL server
with DNS
(l)

Search engine-based methodology

The key hypothesis of this approach is that most of the information needed to profile the Internet endpoints is already available around us—on the Web.

The goal is to characterize endpoints by strategically combining information available at a number of different sources on the Web. The key is that records about many Internet endpoints’ activities inevitably stay publicly archived. Of course, not all active endpoints appear on the Web, and not all communication leaves a public trace. Still, enormous amounts of information does stay publicly available, and that a ’purified’ version of it could be used in a number of contexts

Unconstrained Endpoint Profiling



At the functional level, the goal is straightforward: we query the search engine by searching on text strings corresponding to IP addresses. We collect search hits returned by search engine, and then extract information about the corresponding endpoint. The output is a set of tags associated with this IP address.

Steps

1) **Rule Generation**
2) **Web Classifier**
3) **IP Tagging**

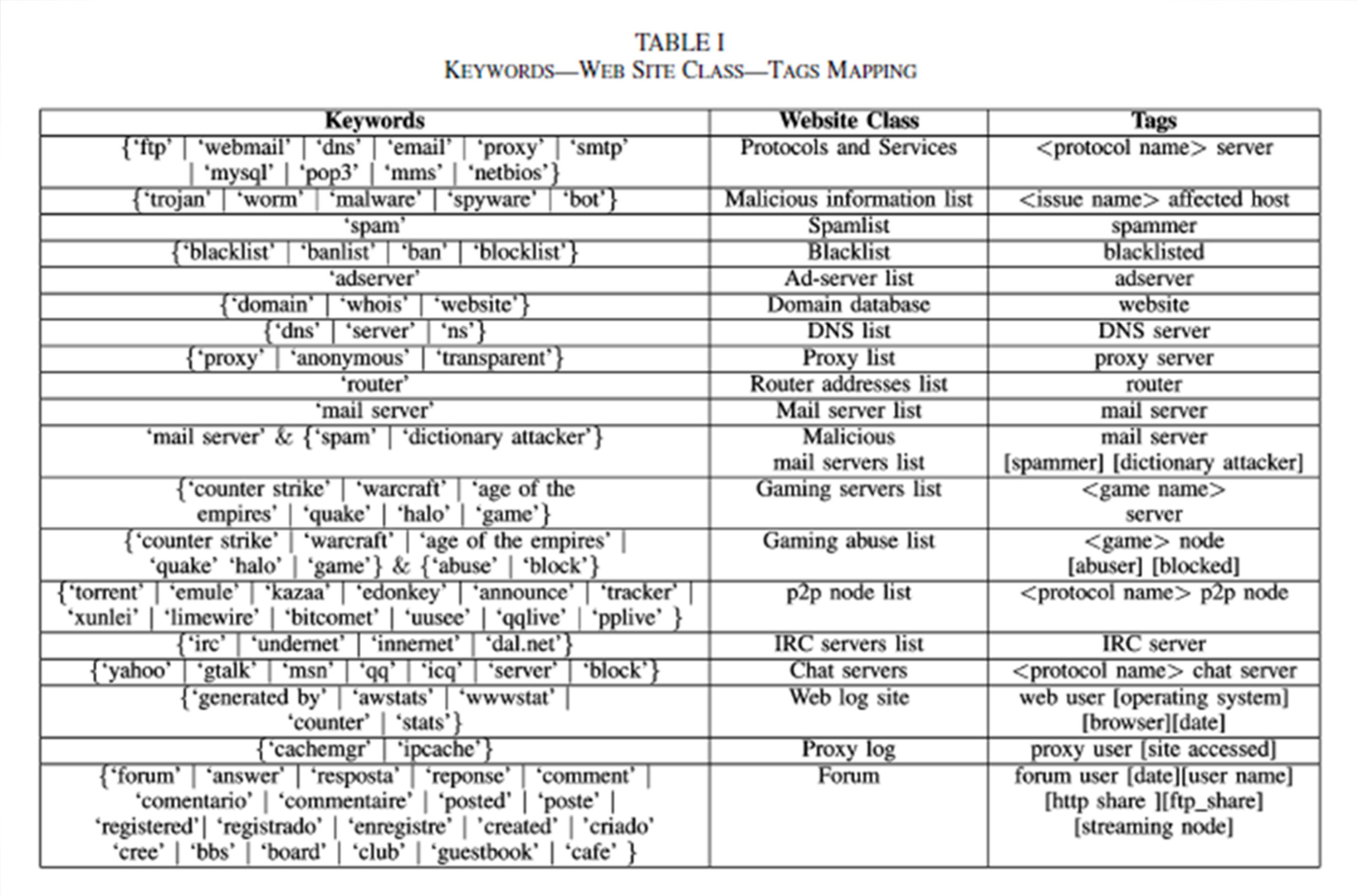
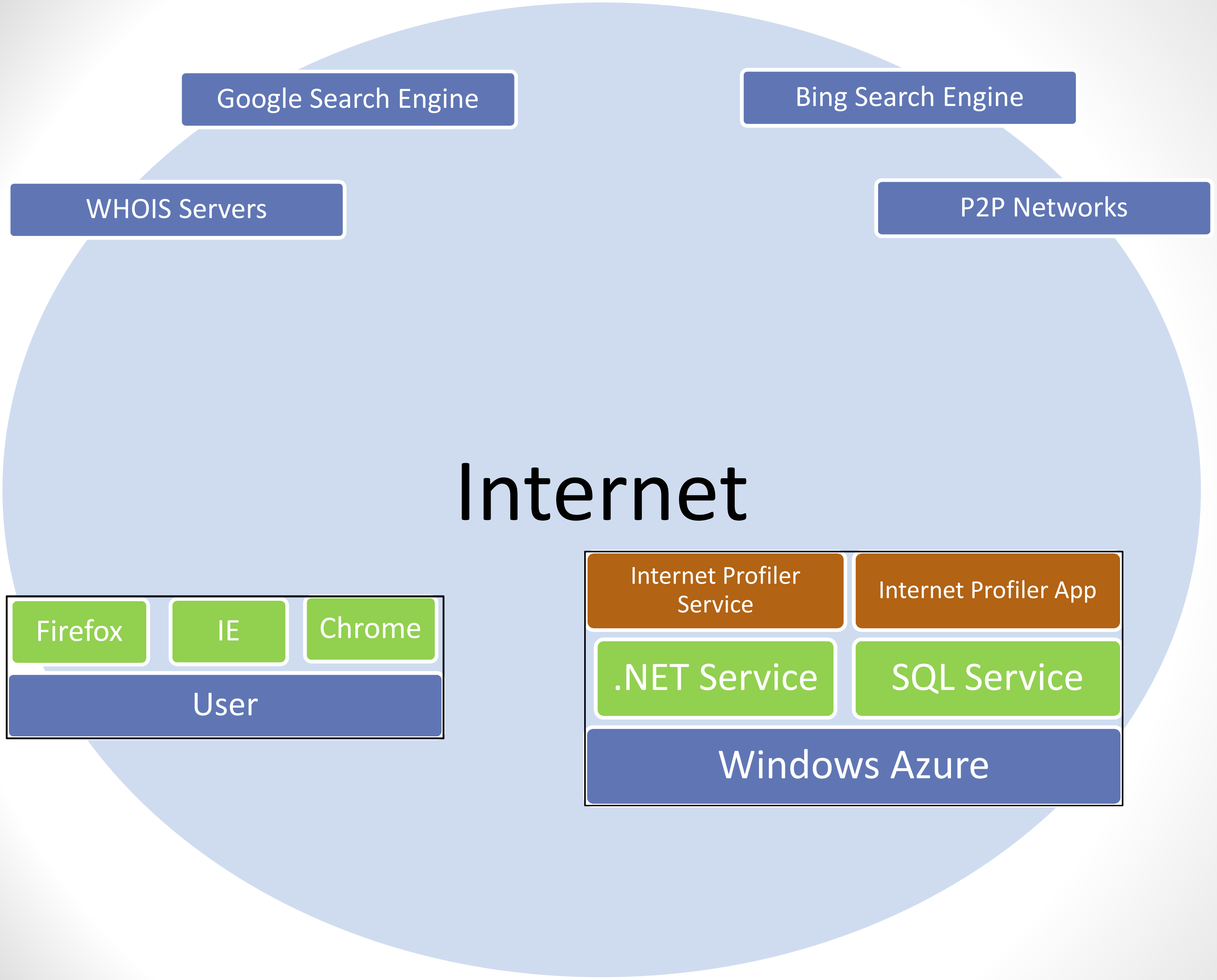


TABLE I
KEYWORDS—WEB SITE CLASS—TAGS MAPPING

Keywords	Website Class	Tags
{'ftp' | 'webmail' | 'dns' | 'email' | 'proxy' | 'smtp' | 'mysql' | 'pop3' | 'mms' | 'netbios'}	Protocols and Services	<protocol name> server
{'trojan' | 'worm' | 'malware' | 'spyware' | 'bot'}	Malicious information list	<issue name> affected host
'spam'	Spamlist	spammer
{'blacklist' | 'banlist' | 'ban' | 'blocklist'}	Blacklist	blacklisted
'adserver'	Ad-server list	adserver
{'domain' | 'whois' | 'website'}	Domain database	website
{'dns' | 'server' | 'ns'}	DNS list	DNS server
{'proxy' | 'anonymous' | 'transparent'}	Proxy list	proxy server
'router'	Router addresses list	router
'mail server'	Mail server list	mail server
'mail server' & {'spam' | 'dictionary attacker'}	Malicious mail servers list	mail server [spammer] [dictionary attacker]
{'counter strike' | 'warcraft' | 'age of the empires' | 'quake' | 'halo' | 'game'}	Gaming servers list	<game name> server
{'counter strike' | 'warcraft' | 'age of the empires' | 'quake' 'halo' | 'game'} & {'abuse' | 'block'}	Gaming abuse list	<game> node [abuser] [blocked]
{'torrent' | 'emule' | 'kazaa' | 'edonkey' | 'announce' | 'tracker' | 'xunlei' | 'limewire' | 'bitcomet' | 'uusee' | 'qqlive' | 'pplive' }	p2p node list	<protocol name> p2p node
{'irc' | 'undernet' | 'innernet' | 'dal.net'}	IRC servers list	IRC server
{'yahoo' | 'gtalk' | 'msn' | 'qq' | 'icq' | 'server' | 'block'}	Chat servers	<protocol name> chat server
{'generated by' | 'awstats' | 'wwwstat' | 'counter' | 'stats'}	Web log site	web user [operating system] [browser][date]
{'cachemgr' | 'ipcache'}	Proxy log	proxy user [site accessed]
{'forum' | 'answer' | 'resposta' | 'reponse' | 'comment' | 'comentario' | 'commentaire' | 'posted' | 'poste' | 'registered'| 'registrado' | 'enregistre' | 'created' | 'criado' 'cree' | 'bbs' | 'board' | 'club' | 'guestbook' | 'cafe' }	Forum	forum user [date][user name] [http share][ftp_share] [streaming node]

Internet Profiler Software

- OS : Windows Azure
- Programming Language : C#
- DBMS : Microsoft SQL Azure
- Used Technologies and APIs : Silverlight, .NET 4, WCF , EF, Prism, LINQ, Parallel LINQ, Google Custom Search API, Bing API, PCAP API
- Client Side Architecture : MVVM design pattern
- Server Side Architecture : SOA design pattern, N-Tire design pattern , DDD pattern , ORM



Google Search Engine

Bing Search Engine

WHOIS Servers

P2P Networks

Internet

Firefox

IE

Chrome

User

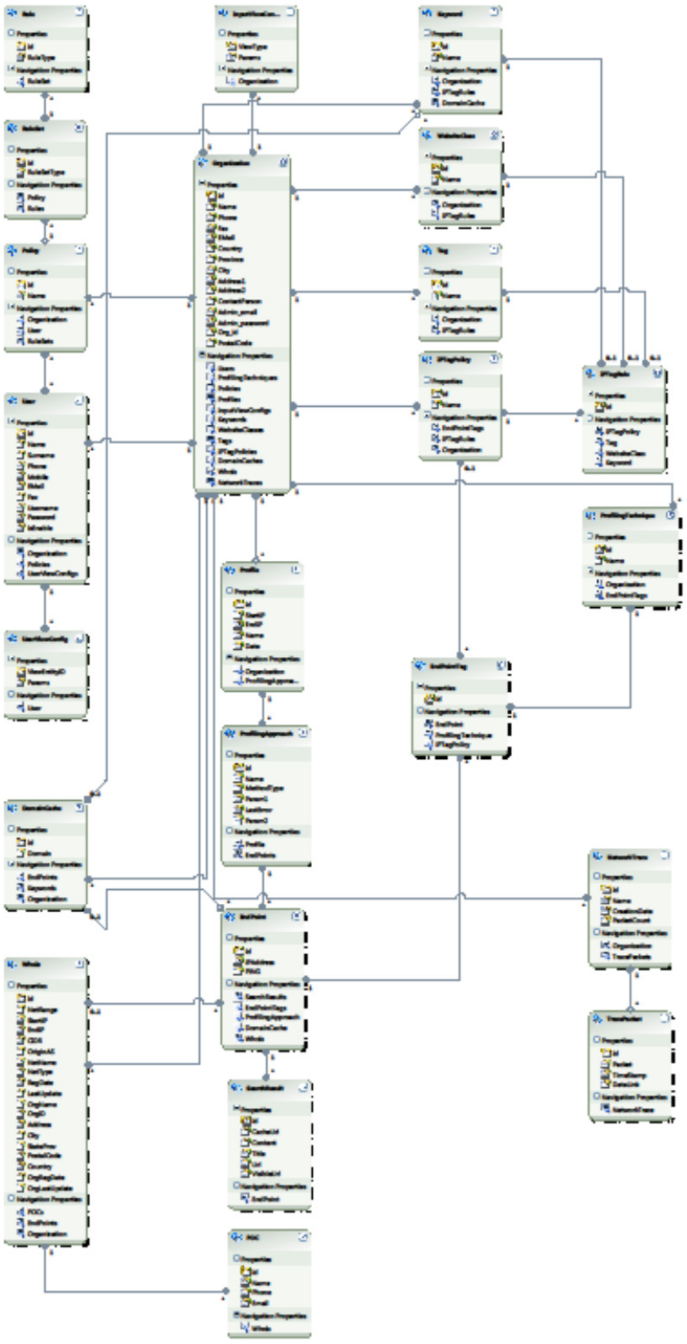
Internet Profiler Service

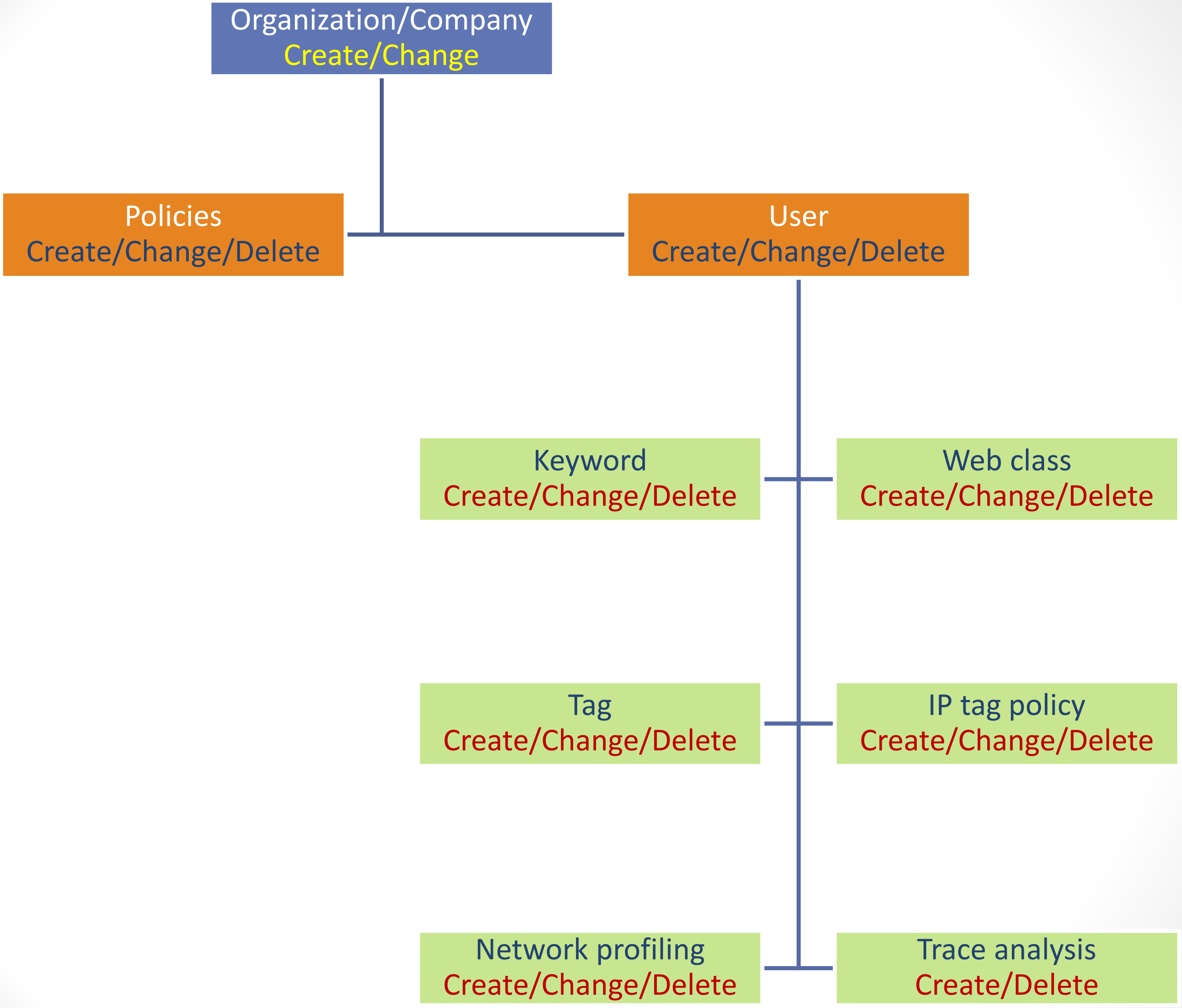
Internet Profiler App

.NET Service

SQL Service

Windows Azure


Rule
Properties
Id
RuleType
Navigation Properties
RuleSet
RuleSet
Properties
Id
RuleSetType
Navigation Properties
Policy
Rules
Policy
Properties
Id
Name
Navigation Properties
Organization
User
RuleSets
User
Properties
Id
Name
Surname
Phone
Mobile
EMail
Fax
Username
Password
IsEnable
Navigation Properties
Organization
Policies
UserViewConfigs
UserViewConfig
Properties
ViewEntityID
Params
Navigation Properties
User
DomainCache
Properties
Id
Domain
Navigation Properties
EndPoints
Keywords
Organization
Whois
Properties
Id
NetRange
StartIP
EndIP
CIDR
OriginAS
NetName
NetType
RegDate
LastUpdate
OrgName
OrgID
Address
City
StateProv
PostalCode
Country
OrgRegDate
OrgLastUpdate
Navigation Properties
POCs
EndPoints
Organization
InputViewCon...
Properties
ViewType
Params
Navigation Properties
Organization
Organization
Properties
Id
Name
Phone
Fax
EMail
Country
Province
City
Address1
Address2
ContactPerson
Admin_email
Admin_password
Org_Id
PostalCode
Navigation Properties
Users
ProfilingTechniques
Policies
Profiles
InputViewConfigs
Keywords
WebsiteClasses
Tags
IPFlagPolicies
DomainCaches
Whois
NetworkTraces
Profile
Properties
Id
StartIP
EndIP
Name
Date
Navigation Properties
Organization
ProfilingAppro...
ProfilingApproach
Properties
Id
Name
MethodType
Param1
LastError
Param2
Navigation Properties
Profile
EndPoints
EndPoint
Properties
Id
IPAddress
PING
Navigation Properties
SearchResults
EndPointTags
ProfilingApproach
DomainCache
Whois
SearchResult
Properties
Id
CachedUrl
Content
Title
Url
VisibleUrl
Navigation Properties
EndPoint
POC
Properties
Id
Name
Phone
Email
Navigation Properties
Whois
Keyword
Properties
Id
Name
Navigation Properties
Organization
IPFlagRules
DomainCaches
WebsiteClass
Properties
Id
Name
Navigation Properties
Organization
IPFlagRules
Tag
Properties
Id
Name
Navigation Properties
Organization
IPFlagRules
IPFlagPolicy
Properties
Id
Name
Navigation Properties
EndPointTags
IPFlagRules
Organization
IPFlagRule
Properties
Id
Navigation Properties
IPFlagPolicy
Tag
WebsiteClass
Keyword
ProfilingTechnique
Properties
Id
Name
Navigation Properties
Organization
EndPointTags
EndPointTag
Properties
Id
Navigation Properties
EndPoint
ProfilingTechnique
IPFlagPolicy
NetworkTrace
Properties
Id
Name
CreationDate
PacketCount
Navigation Properties
Organization
TracePackets
TracePacket
Properties
Id
Packet
TimeStamp
DataLink
Navigation Properties
NetworkTrace



Organization/Company
Create/Change

Policies
Create/Change/Delete

User
Create/Change/Delete

Keyword
Create/Change/Delete

Web class
Create/Change/Delete

Tag
Create/Change/Delete

IP tag policy
Create/Change/Delete

Network profiling
Create/Change/Delete

Trace analysis
Create/Delete

Internet Profiler Service :

- Query Google search engine
- Query Bing search engine
- Query WHOIS server
- Reverse DNS lookup
- Crawling P2P systems
- Web classifier
- IP tagging
- Statistical and analytical analysis
- Trace analysis

  Reading PCAP file

  Packet interpretation : Ethernet , ARP , IPv4 , GRE , ICMP , IGMP , UDP , TCP , HTTP, BGP, other protocols and apps

  Statistical and analytical analysis

Internet Profiler App :

- Creating policies for accessing to forms in application and functionalities of the application.
- Creating organization or company.
- Creating user for the organization or company to login into application.
- Creating IP tagging table.
- Create and run query on search engines.
- View trace packets.
- Statistical and analytical graphs

References

[1] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, “BLINC: Multilevel traffic classification in the dark,” in *Proc. ACM SIGCOMM*, Philadelphia, PA, Aug. 2005, pp. 229–240.

[2] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, “Internet traffic classification demystified: Myths, caveats, and the best practices,” in *Proc. ACM CONEXT*, Madrid, Spain, Dec. 2008, Article no. 11.

[3] T. Karagiannis, K. Papagiannaki, N. Taft, and M. Faloutsos, “Profiling the end host,” in *Proc. PAM*, Louvain-la-neuve, Belgium, Apr. 2007, pp. 186–196.

[4] I. Trestian, S. Ranjan, A. Kuzmanovi, and A. Nucci, “Googling the internet: profiling internet endpoints via the world wide web” in *Proc. ACM SIGCOMM*, New York, NY, USA 2008, pp. 666-679.

[5] Google Custom Search, [Online]. Available: http://code.google.com/apis/customsearch/

[6] Bing API, [Online]. Available: http://www.bing.com/developers/

[7] PCAP.NET, [Online], Available: http://pcapdotnet.codeplex.com/

[8] WinPCAP, [Online], Available: http://www.winpcap.org/

[9] Wikipedia, [Online], Available: http://www.wikipedia.org/