Towards Social Botnet Behavior Detecting in the End Host

Yuede Ji, Yukun He, Xinyang Jiang, and Qiang Li
Jilin University

12/17/2014

Outline

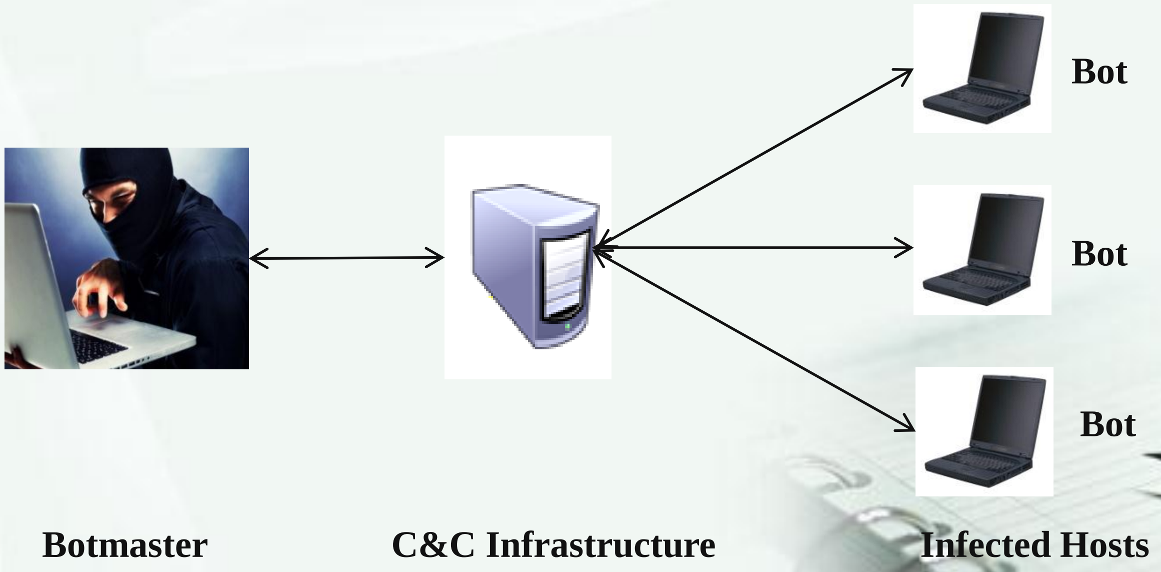
1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

1.1 What is Botnet and Bot?

- A botnet is a network composed by a large scale of infected hosts under the control of a botmaster through Command and Control (C&C) channel.
- Bot is the infected host



1.1 What is Botnet and Bot?

- 3 basic elements:
  - Bot, C&C channel, botmaster
- C&C channel
  - Biggest difference between bot and other malwares
  - Centralized: IRC, HTTP
  - Decentralized: Peer-to-Peer (P2P)
- A major threat to Internet security
  - DDoS, spam, identity theft, phishing

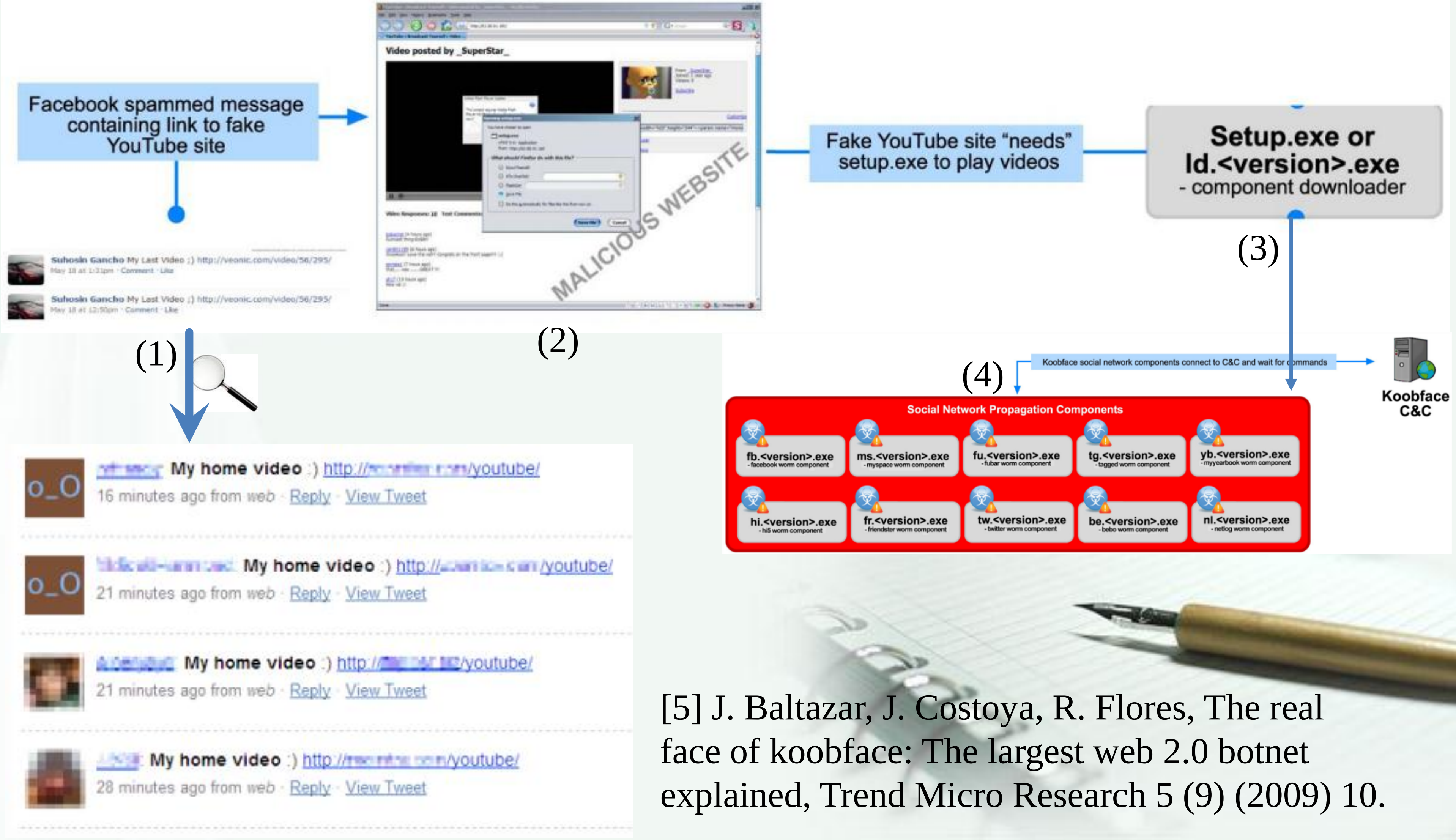
1.2 What is Social Botnet?

- Social botnet utilizes Online Social Network (OSN) as C&C channel.

- Social bot runs on user hosts stealthily, controls user account on OSN site, and communicates with the botmaster.

1.2 What is Social Botnet?

- Example: koobface



Facebook spammed message containing link to fake YouTube site

Suhosin Gancho My Last Video ;) http://veonic.com/video/56/295/
May 18 at 1:31pm · Comment · Like

Suhosin Gancho My Last Video ;) http://veonic.com/video/56/295/
May 18 at 12:50pm · Comment · Like

Fake YouTube site "needs" setup.exe to play videos

Setup.exe or ld.<version>.exe - component downloader

Koobface social network components connect to C&C and wait for commands

Koobface C&C

Social Network Propagation Components

fb.<version>.exe - facebook worm component	ms.<version>.exe - myspace worm component	fu.<version>.exe - fubar worm component	tg.<version>.exe - tagged worm component	yb.<version>.exe - myyearbook worm component
hi.<version>.exe - hi5 worm component	fr.<version>.exe - friendster worm component	tw.<version>.exe - twitter worm component	be.<version>.exe - bebo worm component	nl.<version>.exe - netlog worm component

[redacted] My home video :) http://[redacted]/youtube/
16 minutes ago from web · Reply · View Tweet

[redacted] My home video :) http://[redacted]/youtube/
21 minutes ago from web · Reply · View Tweet

[redacted] My home video :) http://[redacted]/youtube/
21 minutes ago from web · Reply · View Tweet

[redacted] My home video :) http://[redacted]/youtube/
28 minutes ago from web · Reply · View Tweet

[5] J. Baltazar, J. Costoya, R. Flores, The real face of koobface: The largest web 2.0 botnet explained, Trend Micro Research 5 (9) (2009) 10.

1.3 Existing Detection Approaches

- Server-side:
  - mainly use classification methods to identify malicious accounts or messages

- Host-side:
  - mainly monitor the abnormal behaviors on host to determine whether it is infected

1.4 Our Contributions

1. We design a social botnet, named wbbot, based on Sina Weibo.

2. We identify six critical phases based on life cycle, and analyze social bot behaviors based on these phases.

3. We propose a behavior tree-based detection approach, which can get a fairly good detection rate compared with others.

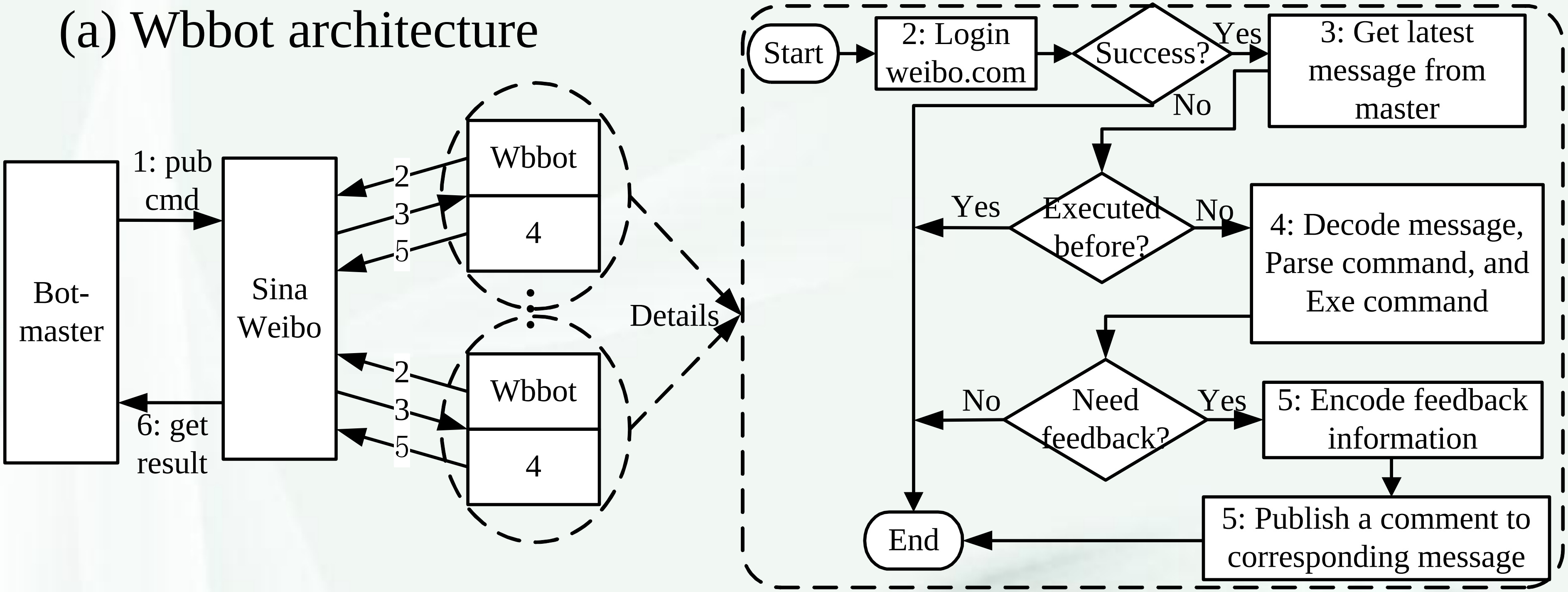
Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

2.1 Wbbot Architecture

(b) Wbbot control flow on host

(a) Wbbot architecture


Bot-master
1: pub cmd
6: get result
Sina Weibo
2
3
5
Wbbot
4
Wbbot
4
Details
Start
2: Login weibo.com
Success?
Yes
No
3: Get latest message from master
Executed before?
Yes
No
4: Decode message, Parse command, and Exe command
Need feedback?
No
Yes
5: Encode feedback information
5: Publish a comment to corresponding message
End

(a)

(b)

2.2 Wbbot Behaviors

- Wbbot behaviors can be classified into two categories:
  - host based
  - social network based

	Command	Description
Host	getNetInfo	get host information (MAC, IP, username, etc.)
	getVersion	get the windows system Version
	exeCmd	execute a DOS command
	timeExeCmd	execute a DOS command at a specific time
	visit	force the IE browser to open an URL
	redirect	rebind the domain and IP
Social network	pubWeiboText	order wbbot to publish a message
	postComment	order wbbot to comment a message on a user
	addFollowing	order wbbot to follow an account
	autoAddFollowing	order wbbot to automatically follow others

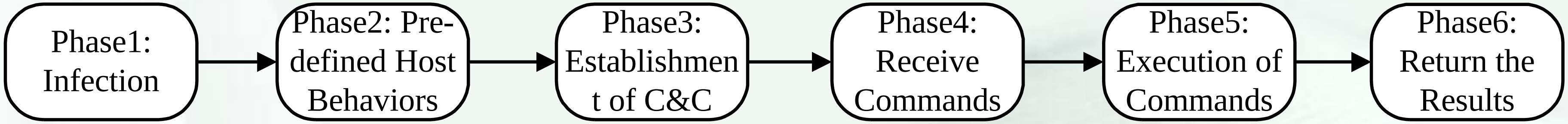
Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

3. Host Behaviors of Social Bots

- Analyze existing social bots:
  - Two samples: koobface, Naz bot;
  - Three laboratory works: stegobot, bot designed by Boshmaf, and facebot.

- Divide their behaviors into six phases:



Phase1: Infection

Phase2: Pre-defined Host Behaviors

Phase3: Establishmen t of C&C

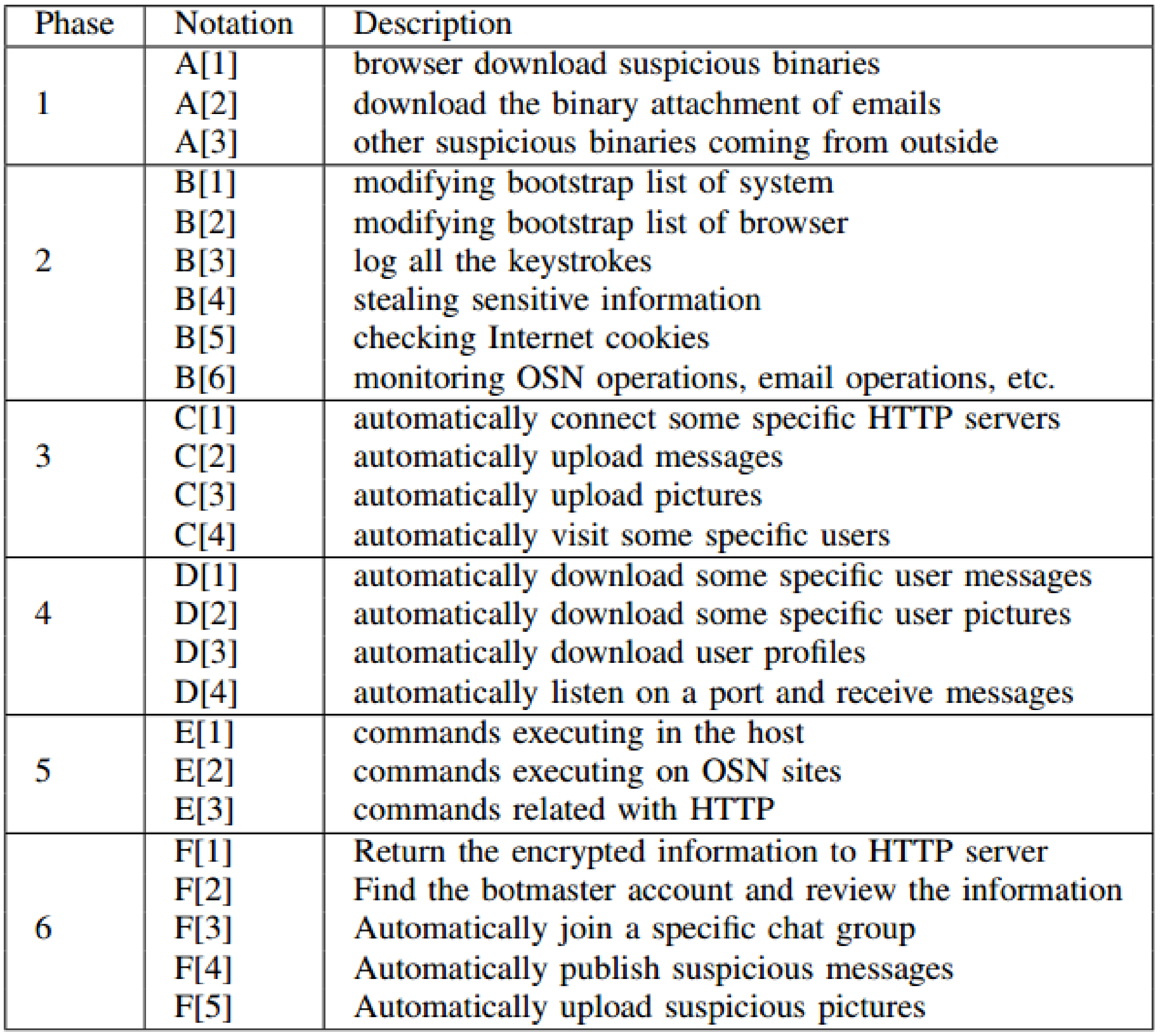
Phase4: Receive Commands

Phase5: Execution of Commands

Phase6: Return the Results

3. Host Behaviors of Social Bots

In each phase, social bots can have several possible behaviors.

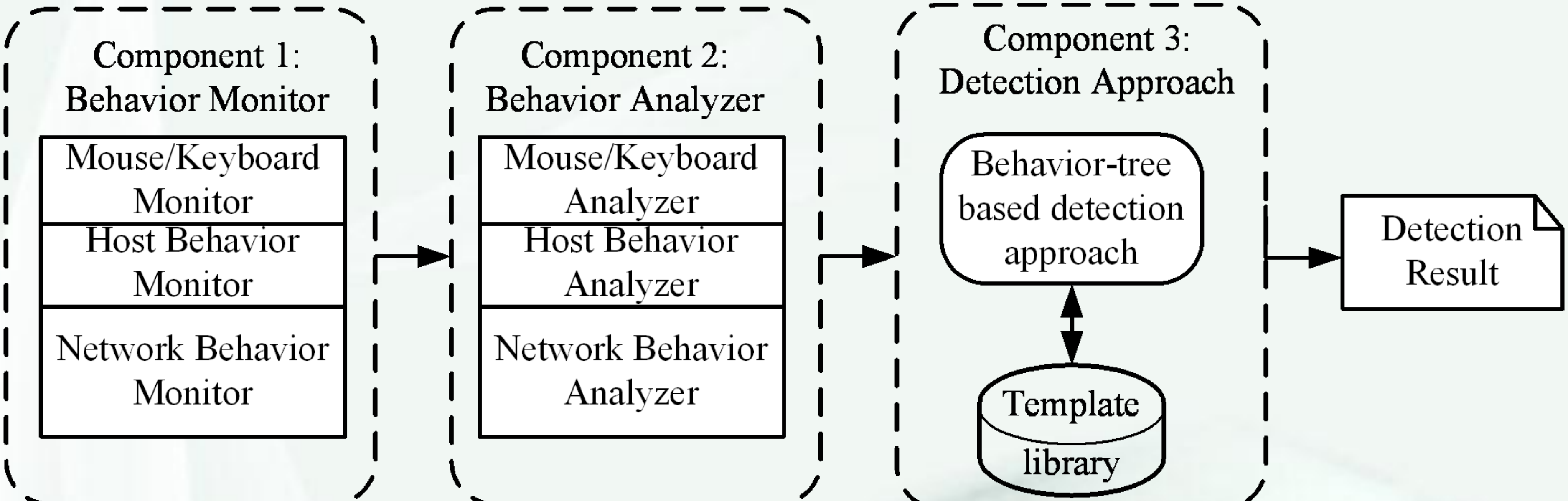


Phase	Notation	Description
1	A[1]	browser download suspicious binaries
	A[2]	download the binary attachment of emails
	A[3]	other suspicious binaries coming from outside
2	B[1]	modifying bootstrap list of system
	B[2]	modifying bootstrap list of browser
	B[3]	log all the keystrokes
	B[4]	stealing sensitive information
	B[5]	checking Internet cookies
	B[6]	monitoring OSN operations, email operations, etc.
3	C[1]	automatically connect some specific HTTP servers
	C[2]	automatically upload messages
	C[3]	automatically upload pictures
	C[4]	automatically visit some specific users
4	D[1]	automatically download some specific user messages
	D[2]	automatically download some specific user pictures
	D[3]	automatically download user profiles
	D[4]	automatically listen on a port and receive messages
5	E[1]	commands executing in the host
	E[2]	commands executing on OSN sites
	E[3]	commands related with HTTP
6	F[1]	Return the encrypted information to HTTP server
	F[2]	Find the botmaster account and review the information
	F[3]	Automatically join a specific chat group
	F[4]	Automatically publish suspicious messages
	F[5]	Automatically upload suspicious pictures

Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

4.1 Detection System Architecture



Component 1: Behavior Monitor

Mouse/Keyboard Monitor

Host Behavior Monitor

Network Behavior Monitor

Component 2: Behavior Analyzer

Mouse/Keyboard Analyzer

Host Behavior Analyzer

Network Behavior Analyzer

Component 3: Detection Approach

Behavior-tree based detection approach

Template library

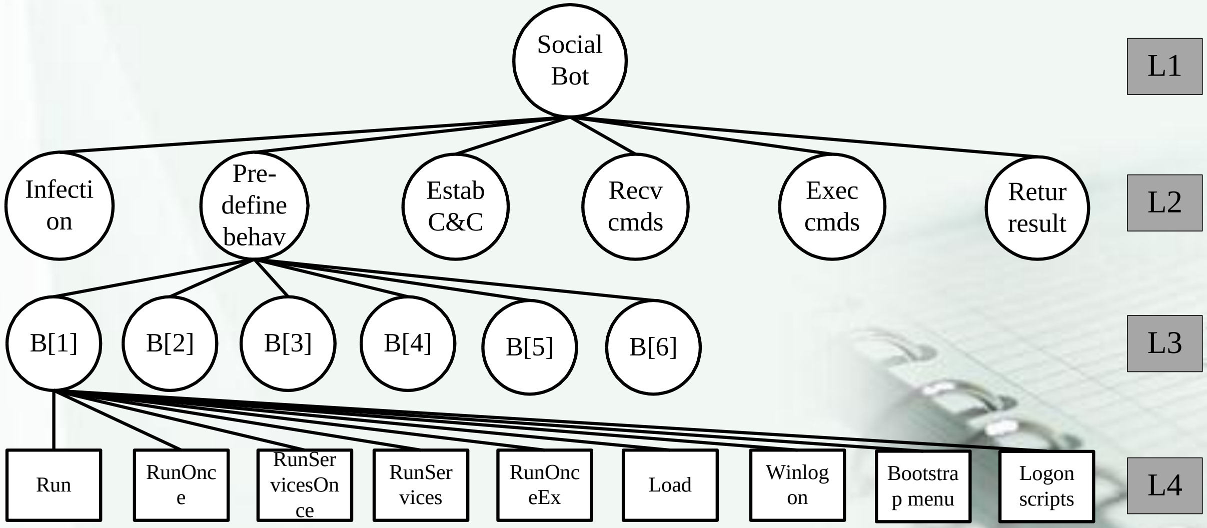
Detection Result

4.2 Behavior Tree-based Approach

(1) Behavior Tree Representation

- T = <V, E>
- L1, root layer, represent the detection result
- L2, six phases based on life cycle
- L3, specific behaviors of L2
- L4, implementations of each behavior

Example: B[1]

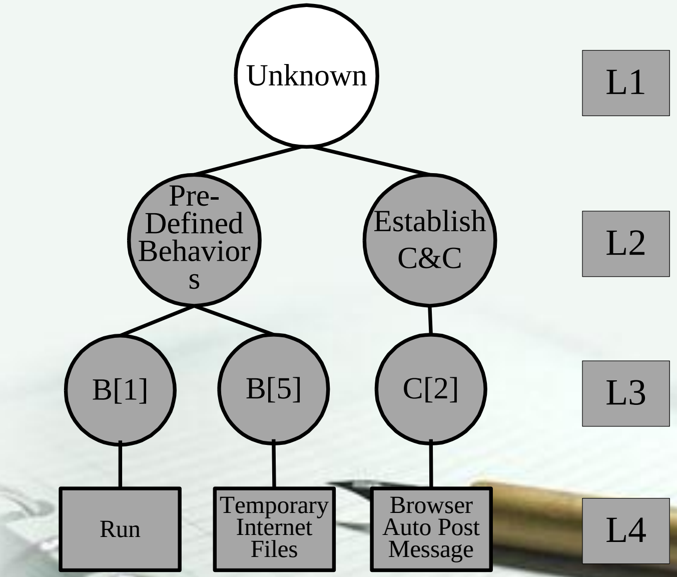


4.2 Behavior Tree-based Approach

(2) Behavior Tree Construction:

- Once the behavior in L4 layer is identified, we will flag the nodes from bottom to top

- Example: the suspicious process has the following behaviors:
  - modify the Registry value of Run,
  - check Internet cookies,
  - automatically upload messages using POST function to OSN sites.



4.2 Behavior Tree-based Approach

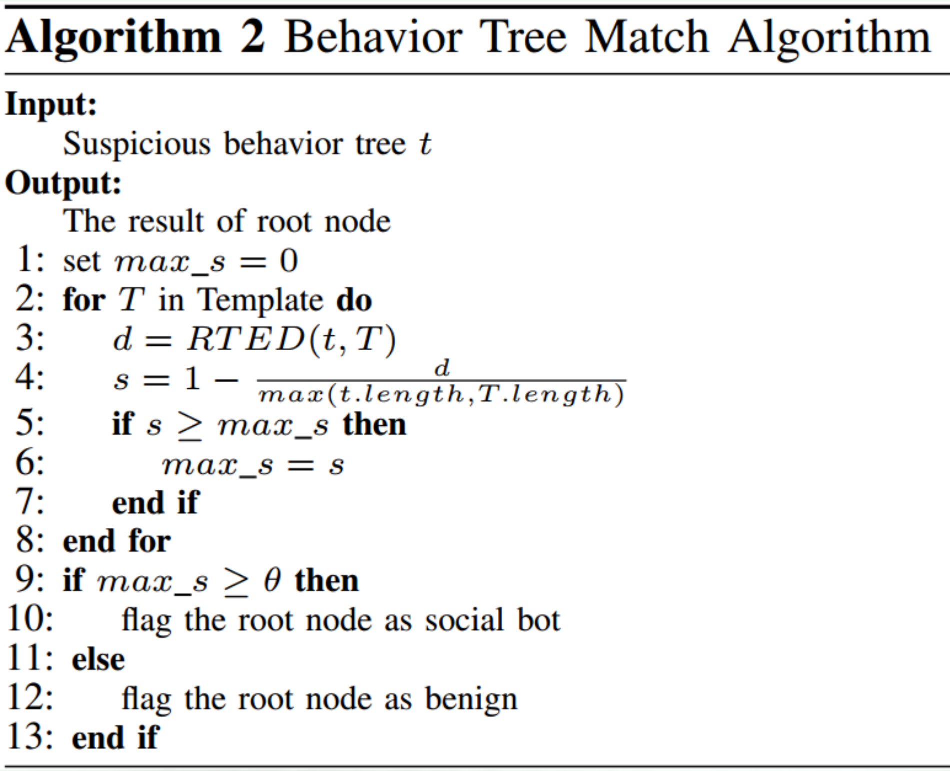
(3) Template Library Construction:

- Off-line process based on three aspects:
  - existing social bot samples,
  - possible social bots of laboratory works,
  - possible implementations from our analysis.

4.2 Behavior Tree-based Approach

(4) Behavior Tree Match :

- Utilize tree edit distance to calculate tree similarity *s*.
- Robust Tree Edit Distance algorithm (RTED)
  - [21] M. Pawlik, N. Augsten, Rted: a robust algorithm for the tree edit distance, Proceedings of the VLDB Endowment 5 (4) (2011) 334–345.
- Calculation of trees similarity
  - $s = 1 - \frac{d}{\max(m,n)}$



Algorithm 2 Behavior Tree Match Algorithm

```
Input:
    Suspicious behavior tree $t$
Output:
    The result of root node
set $max\_s = 0$
for $T$ in Template do
    $d = RTED(t, T)$
    $s = 1 - \frac{d}{max(t.length, T.length)}$
    if $s \geq max\_s$ then
        $max\_s = s$
    end if
end for
if $max\_s \geq \theta$ then
    flag the root node as social bot
else
    flag the root node as benign
end if
```

Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. Conclusion

5.1 Data Collection

Social Bot	Source	Duration	Size
Koobface	Open Malware	24 h	5.32 GB
Twitterbot	The author shared their source code with us	24 h	8.36 GB
TWebot	The author of a social botnet detection approach shared TWebot builders and binaries with us	18 h	2.77 GB
Yazanbot	We reproduced it based on their paper.	24 h	7.36 GB
FixNazbot	We reproduced it based on their paper.	24 h	4.99 GB
Wbbot	We designed.	18 h	11.5 GB
Fbbot	We designed.	5 h	4.65 GB

http://pan.baidu.com/s/1hqvHoSO

5.2 Detection Result

(1) Detection Result

Trace	Avg FP Rate	Avg FN Rate
Koobface	35.9%	31.8%
FixNazbot	34.7%	0%
Yazanbot	35.0%	0%
Twitterbot	15.4%	0%
Fbbot	25.6%	0%
Wbbot	35.3%	0%
TWebot	25.2%	0%
Total	29.6%	4.5%

(2) Result Analysis:

a) FP rate is a little high

- many benign processes perform similar behaviors as social bots
- most social bots mimic user activities or benign application activities

b) Koobface has a high FN rate

- we only have their binaries and cannot configure them

5.2 Detection Result

(1) VirusTotal Detection Result

Trace	Detection Ratio
Koobface	47 / 54
FixNazbot	0 / 54
Yazanbot	1 / 51
Twitterbot	2 / 54
Fbbot	2 / 54
Wbbot	3 / 53
TWebot	2 / 54
Total	15.2%

(2) Result Analysis:

a) Koobface has a high detection ratio

- Koobface has been already in signature database of most antivirus engines

b) Others have a very low detection ratio

c) Compared with them, our detection result is fairly good.

Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. **Discussion**
7. Conclusion

6. Discussion

1. Limitation
   (1) The FP rate of our detection system is a little high.
   (2) The construction of template library is static

2. Future Work
   (1) Try to improve the detection rate
   (2) Try to improve the construction mechanism of template library

Outline

1. Introduction
2. Design and Analysis of Wbbot
3. Host Behaviors of Social Bots
4. Methodology
5. Experiment
6. Discussion
7. **Conclusion**

7. Conclusion

1. Compared with other detection tools, our approach can still get a fairly good result

2. Our research still exists some flaws

3. The topic is interesting and needs a lot of further works…

Thanks for your attention!

Questions?