UNIVERSITY OF AMSTERDAM

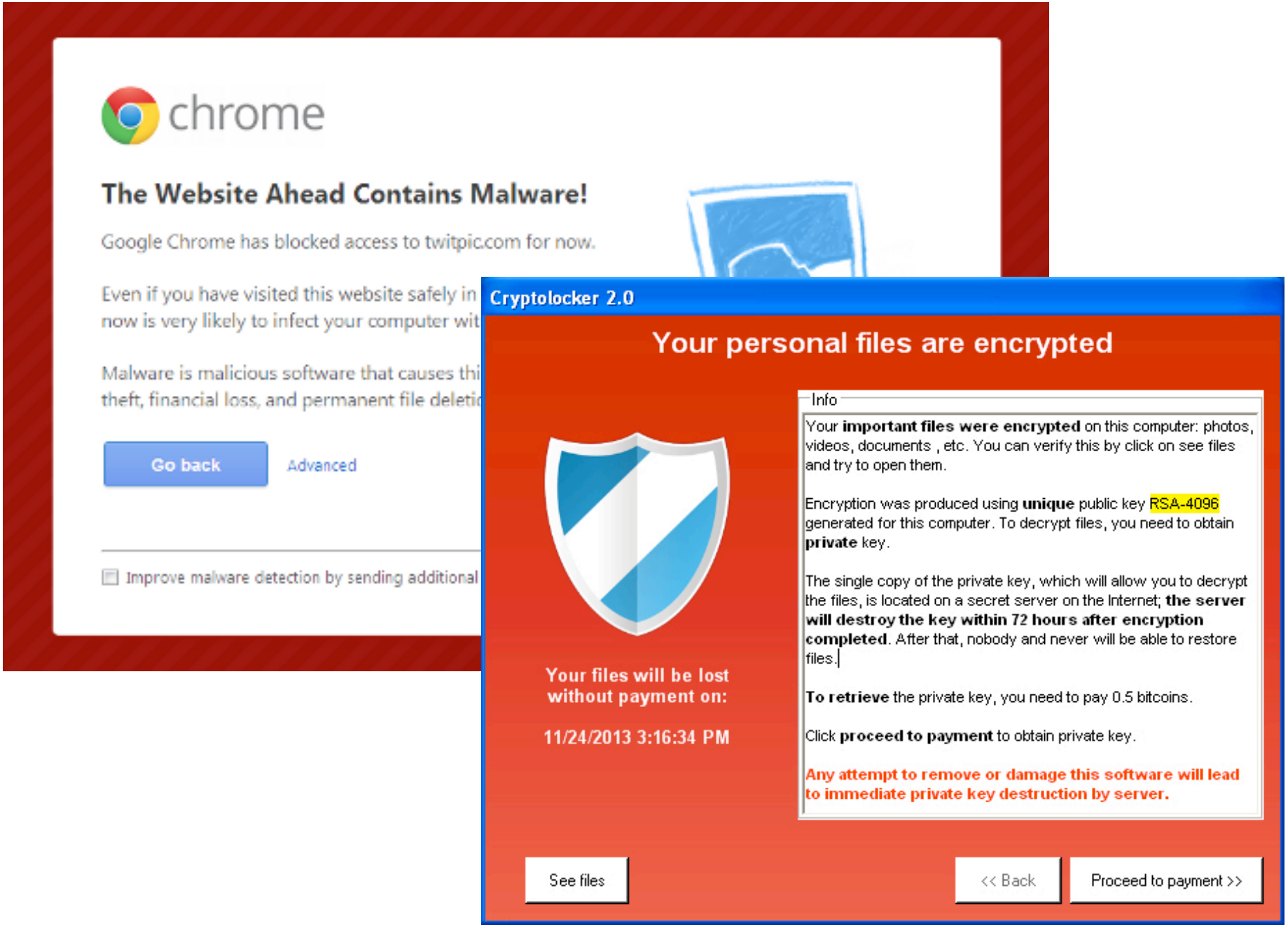
# Large-scale drive-by download detection: visit$^{n}$. process. analyse. report.

Adriaan Dens
Martijn Bogaard

Students Master of System and Network Engineering

Under supervision of:

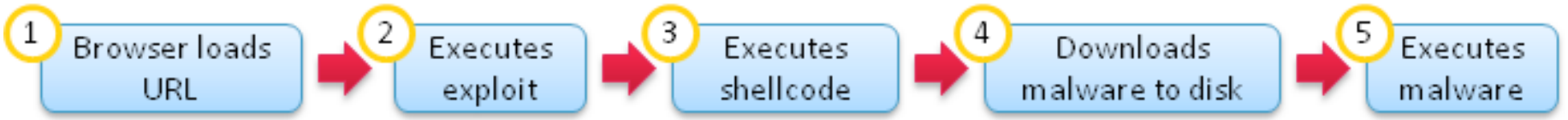
Nationaal Cyber Security Centrum
*Ministerie van Veiligheid en Justitie*




Source: http://www.bleepstatic.com/fhost/uploads/2/cryptolocker-2.0.jpg

# Drive-by downloads

- Vulnerable browsers and plugins
  - Remote code execution
- Malvertising
- Hard to detect




Source: http://blog.armorize.com/2011/04/newest-adobe-flash-0-day-used-in-new.html (modified)

Hacking News | Malwares | Cyber Attack | Vulnerabilities | Hacking Groups | Spy

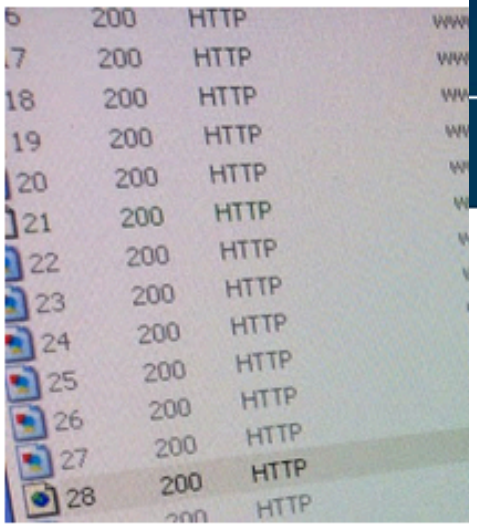
# The Hacker News™

Security in a serious way

## Dutch News site spread Malware on 100000 Computers

Friday, March 16, 2012 Mohit Kumar

Dutch News site spread **Malware on 100000 Computers**




malware) to users of IE. Nu.nl has

provide visitors to the news site wi

exploit kit was placed.

FOX IT

Home About Back to fox-it.com

# Malicious advertisements served via Yahoo

Posted on January 3, 2014 by joostbijl

# Challenges current systems

- Slow
- Single website at a time
- Hard to maintain



# Challenges current systems

- Slow
- Single website at a time
- Hard to maintain



- Have we already said they're slow?

# Challenges current systems

- Slow
- Single website at a time
- Hard to maintain



- Have we already said they're slow?
  - But that's not really their fault, it was never designed for this purpose

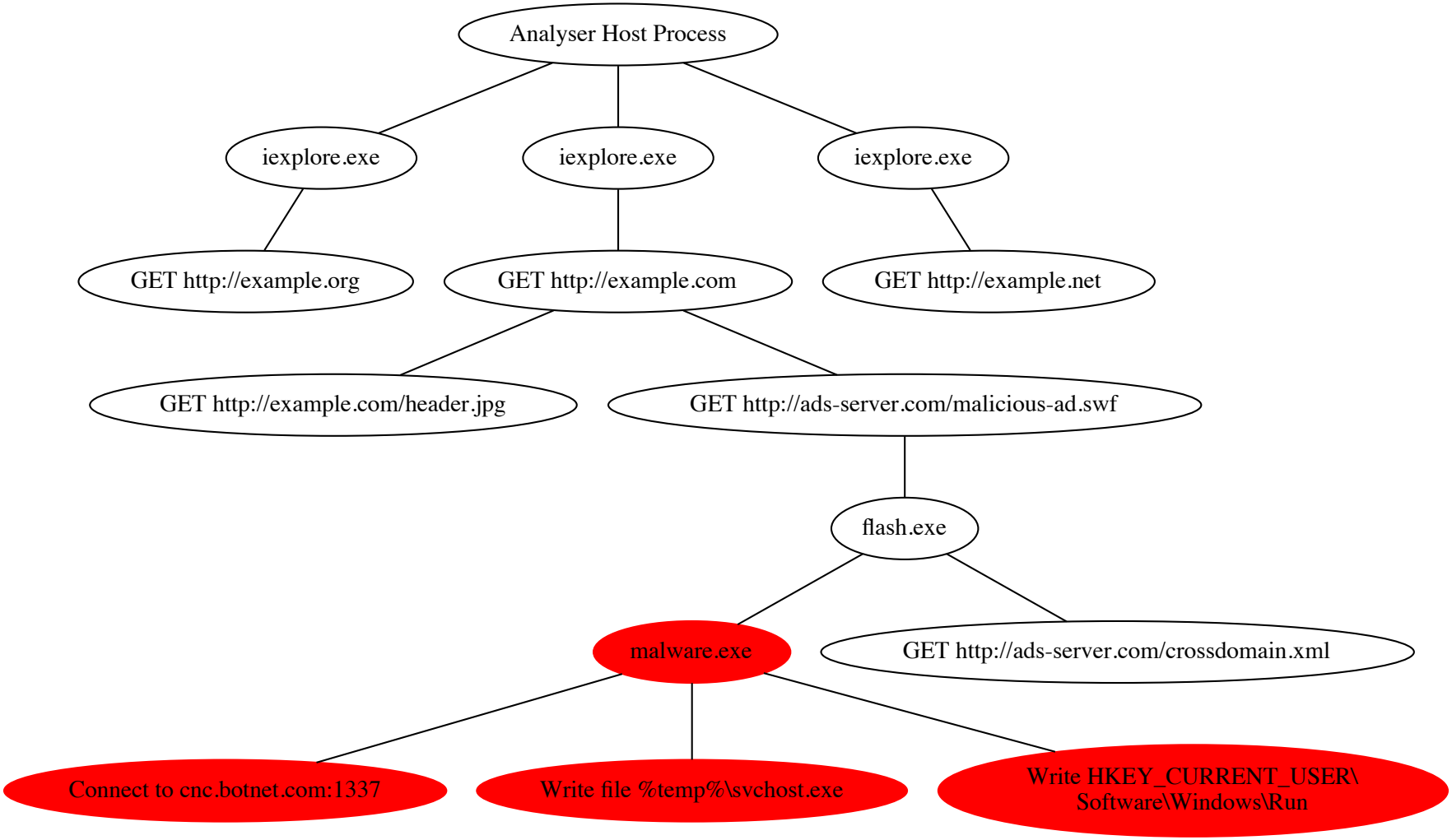
# Research question

How can we concurrently visit multiple URLs and still be able to determine which URL was responsible for malicious activities?

# Research question

How can we **concurrently** visit multiple URLs and still be able to determine **which URL** was responsible for malicious activities?

# The Goal


Analyser Host Process
iexplore.exe
iexplore.exe
iexplore.exe
GET http://example.org
GET http://example.com
GET http://example.net
GET http://example.com/header.jpg
GET http://ads-server.com/malicious-ad.swf
flash.exe
malware.exe
GET http://ads-server.com/crossdomain.xml
Connect to cnc.botnet.com:1337
Write file %temp%\svchost.exe
Write HKEY_CURRENT_USER\
Software\Windows\Run

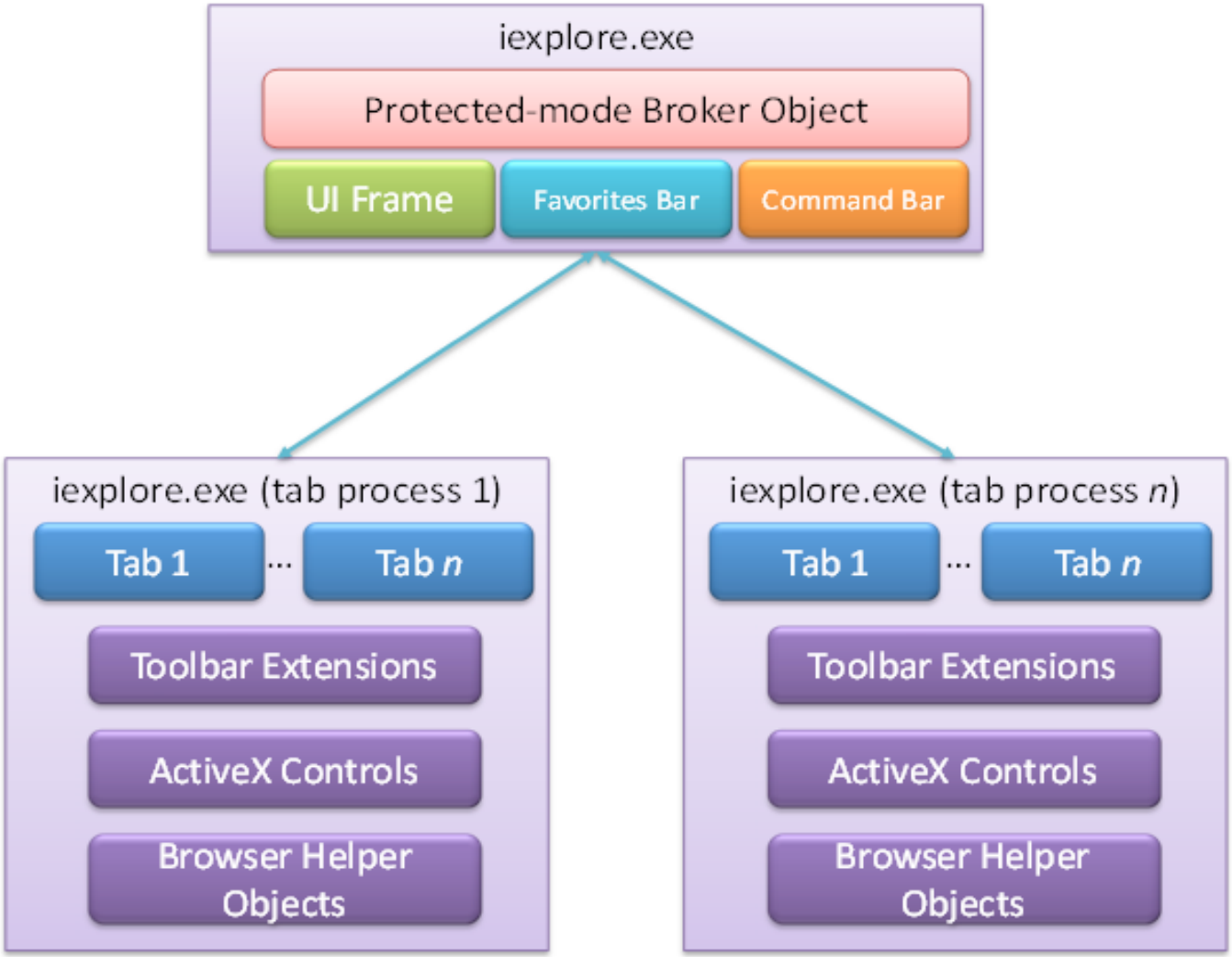
# Algorithm

- Platform independent
  - PoC for Windows 7
- Browser independent
  - PoC for Internet Explorer 8
- Fast(er)
- Limited maintenance needed

# Determine malware origin

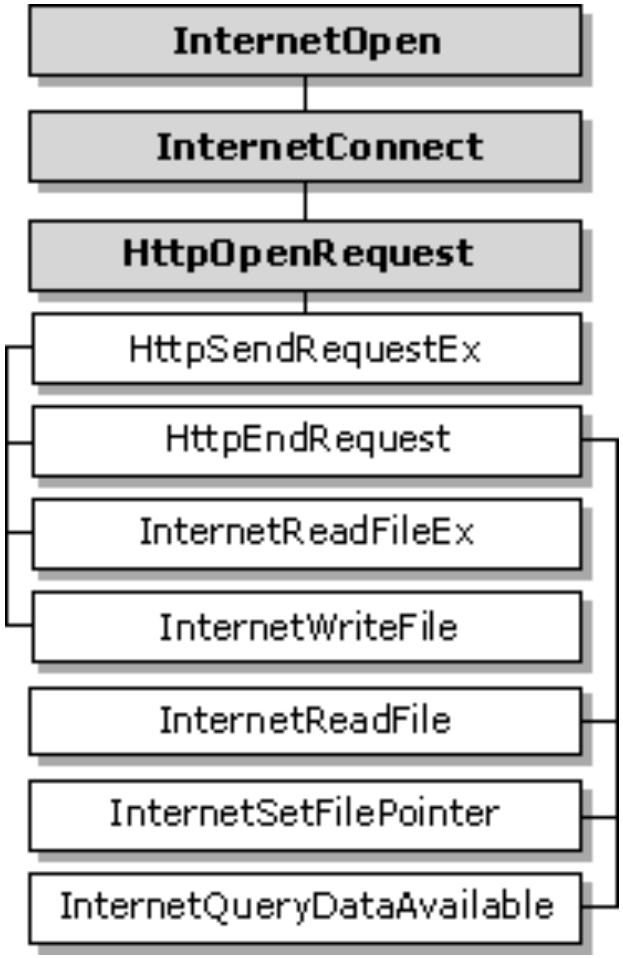
- One website
  - Many HTTP requests
- Easy solutions:
  - Modify web browser and add logging
  - Monitor network
- Scalable solution:
  - Observing API calls
  - Includes thread and process context

# Web browser process model




Source: http://blogs.msdn.com/b/ie/archive/2008/03/11/ie8-and-loosely-coupled-ie-lcie.aspx

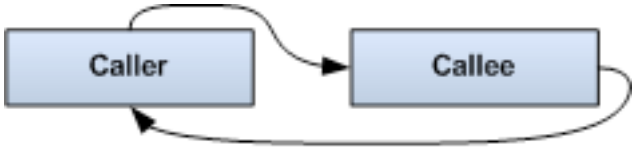
# Web browser network stack



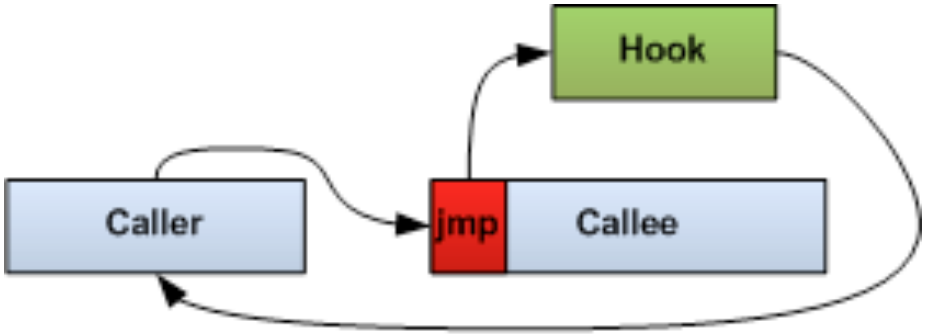

Source: https://msdn.microsoft.com/en-us/library/windows/desktop/aa383766%28v=vs.85%29.aspx

Source: http://newgre.net/node/5

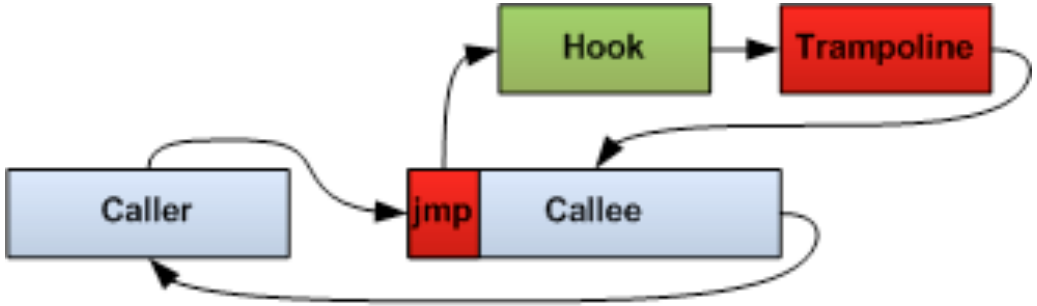
# API hooking




Situation without Hooking




Hook function is called without calling original function




Hook function is called, which in turn calls original function

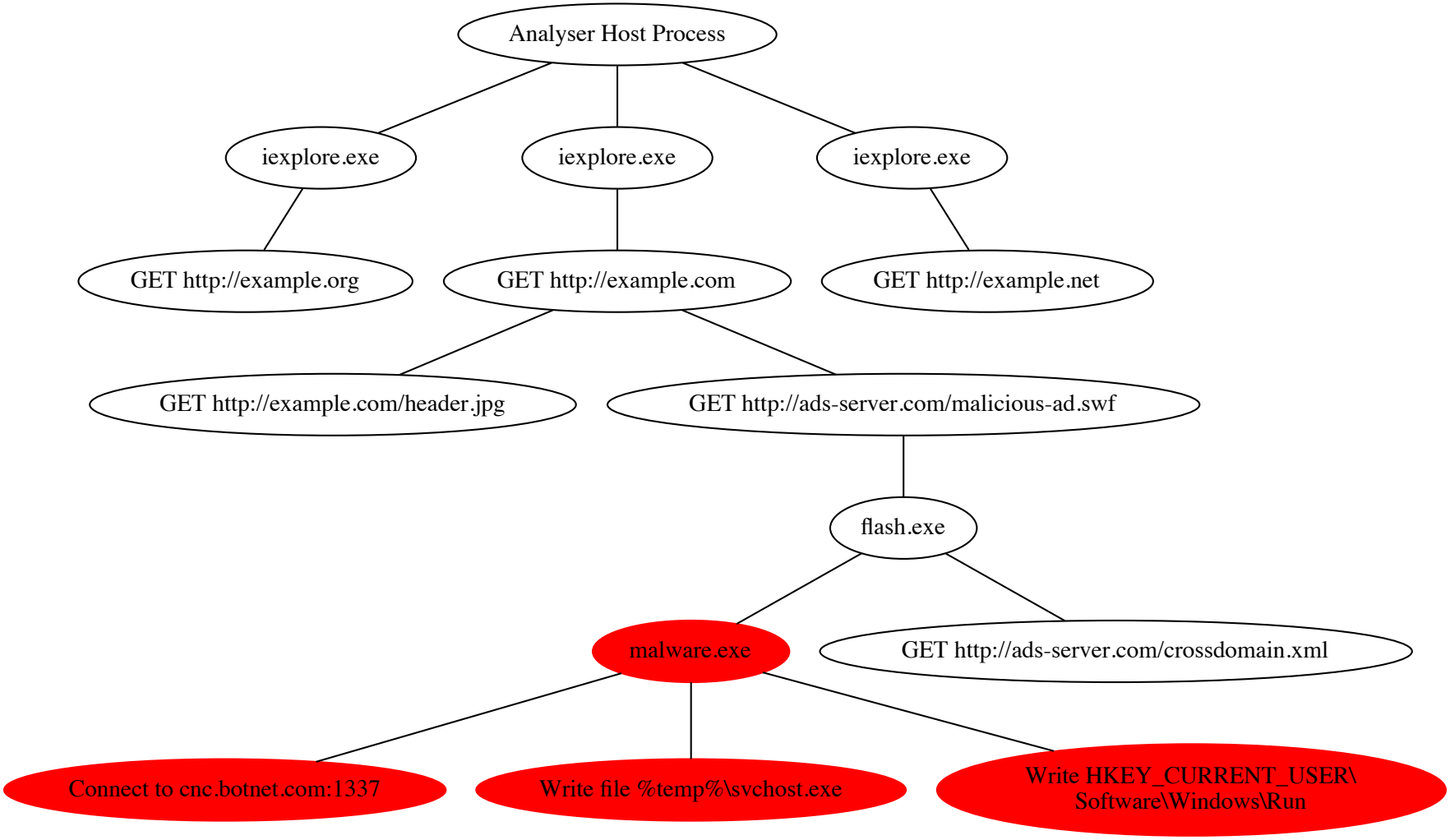
# Connecting the dots…

| | | | |
|---|---|---|---|
| InternetOpenW | ProxyBypass =><br>AccessType => 0x00000000<br>Agent => Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E)<br>Flags => 0x10000000<br>ProxyName => | SUCCESS | 0x00cc0004 |
| InternetConnectW | Username =><br>Service => 3<br>InternetHandle => 0x00cc0004<br>ServerName => www.example.com<br>Flags => 0x00000000<br>ServerPort => 80<br>Password => | SUCCESS | 0x00cc0008 |
| HttpOpenRequestW | Version =><br>InternetHandle => 0x00cc0008<br>Flags => 4194816<br>Verb => GET<br>Referer =><br>Path => / | SUCCESS | 0x00cc000c |
| HttpAddRequestHeadersW | Headers => Accept-Language: en-us User-Agent:Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E) Accept-Encoding: | SUCCESS | 0x00000001 |

# Algorithm in 4 simple steps

- Visit (the URLs)
- Process (the data)
- Analyse (the graph)
- Report (the findings)

# Analyse graph



# Proof of Concept

- Cuckoo
  - Support for analysis of single website
  - Cuckoomon (API hooking)
- Windows 7 + IE 8
- Actual detection is out of scope
  - And up to the user in existing solutions

# Running it…

# Proof of Concept

```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```

# Proof of Concept

```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```

# Proof of Concept

```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```

# Proof of Concept

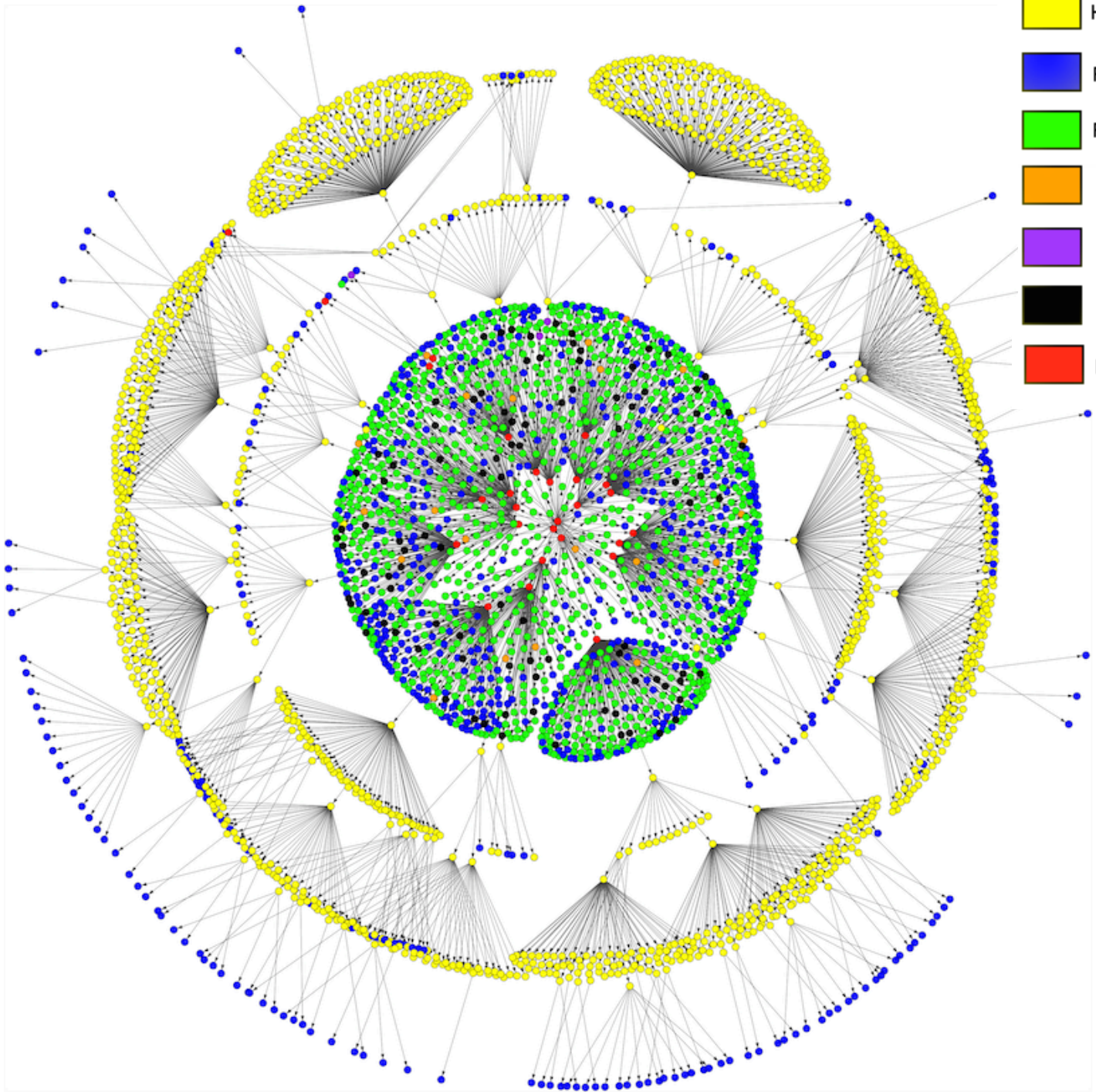
```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```

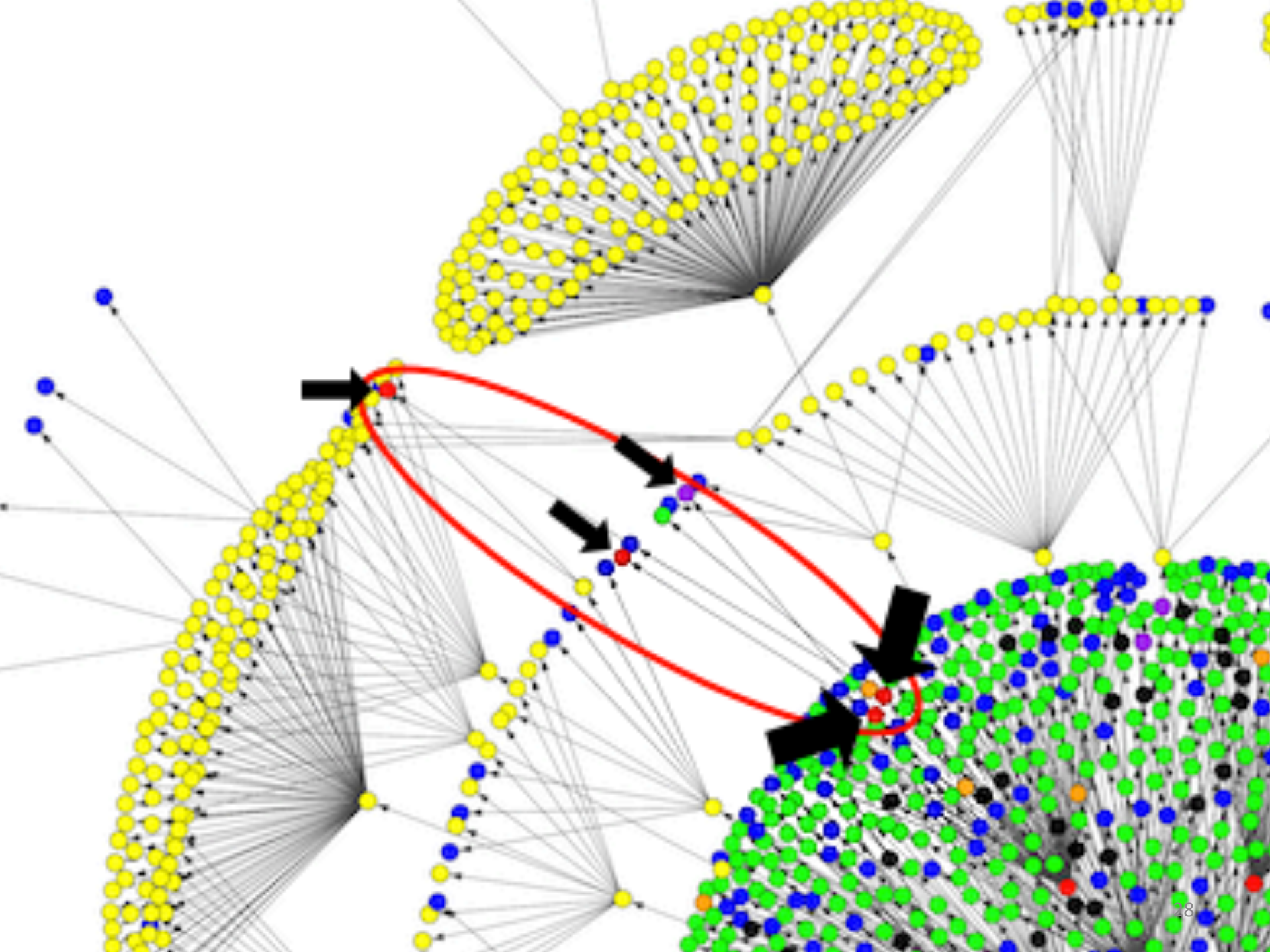
# Proof of Concept

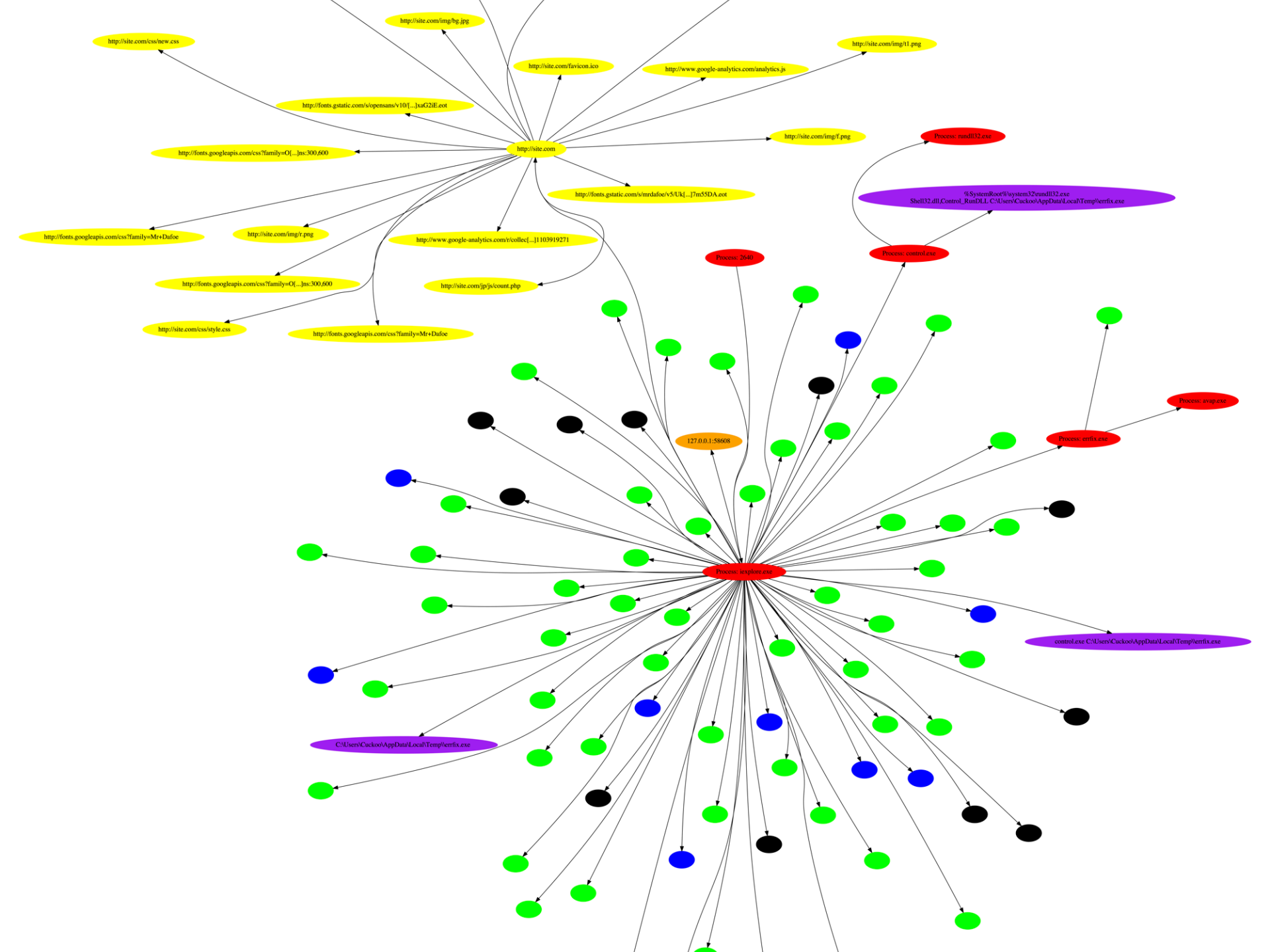
```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```

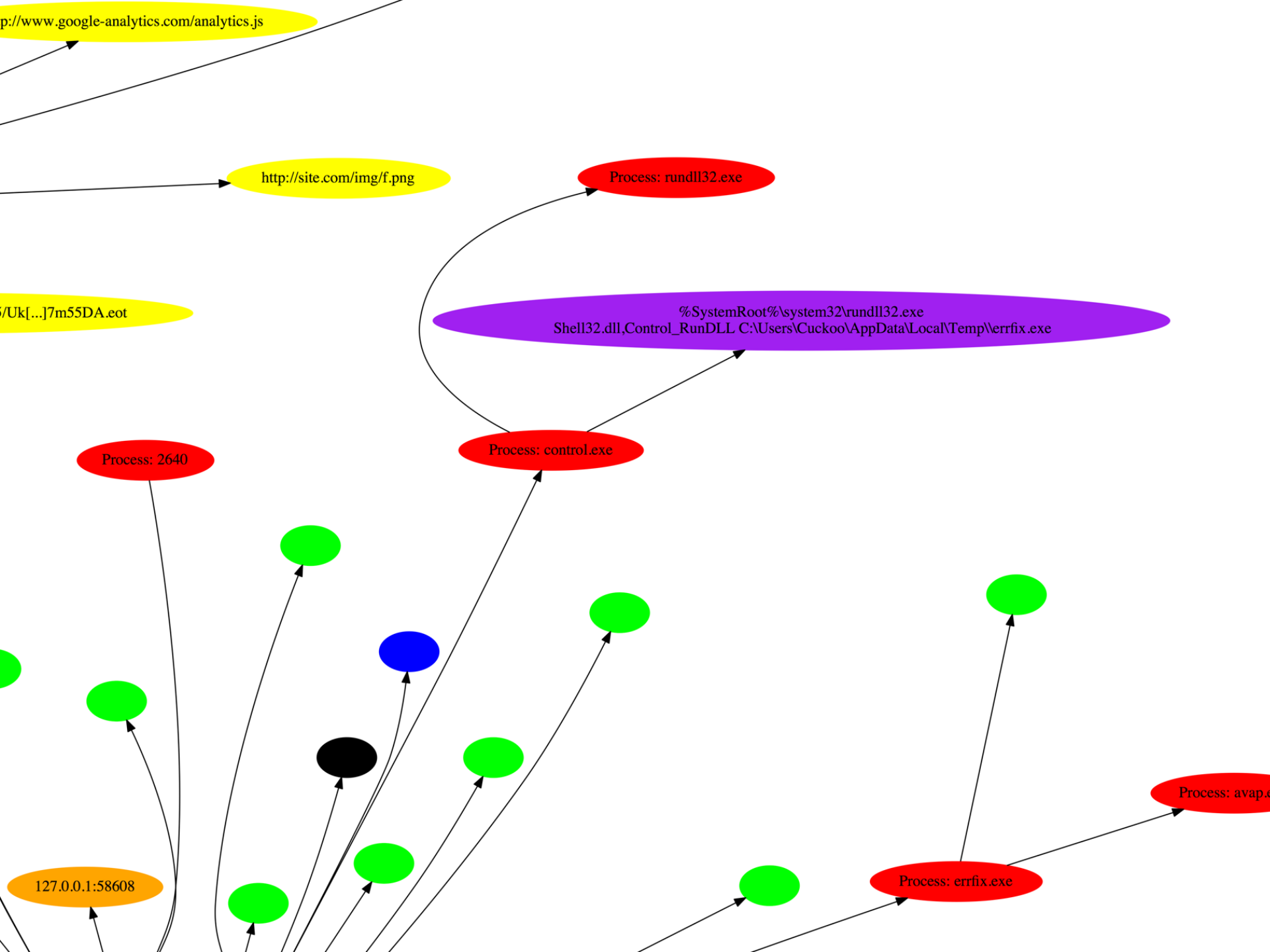
# Proof of Concept

```
$ python cuckoo.py &
$ python utils/mass-analyse.py url_list.txt
Warning: Task with ID 22 is not yet completed; Waiting...
INFO:root:Parse log....
[...]
PID 2876 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.com/
PID 3656 'iexplore.exe' spawned from parent PID 2860
Visiting: http://unsuspicious.com/
PID 2108 'iexplore.exe' spawned from parent PID 2860
Visiting: http://google.nl/
PID 3064 'iexplore.exe' spawned from parent PID 2860
Visiting: http://imdb.com/
PID 1012 'iexplore.exe' spawned from parent PID 2860
Visiting: http://facebook.com/
PID 3728 'control.exe' spawned from parent PID 3656
PID 2848 'repfix.exe' spawned from parent PID 3656
PID 1944 'rundll32.exe' spawned from parent PID 3728
PID 3780 'ynuni.exe' spawned from parent PID 2848
[...]
Analyser 'Subprocess_from_tab': The URL 'http://unsuspicious.com' spawns
    a process called 'control.exe', 'repfix.exe', 'rundll32.exe' and
    'ynuni.exe'.
```


HTTP Request
File operations
Registry operations
Socket operations
Shell command
Anomalities
Process created



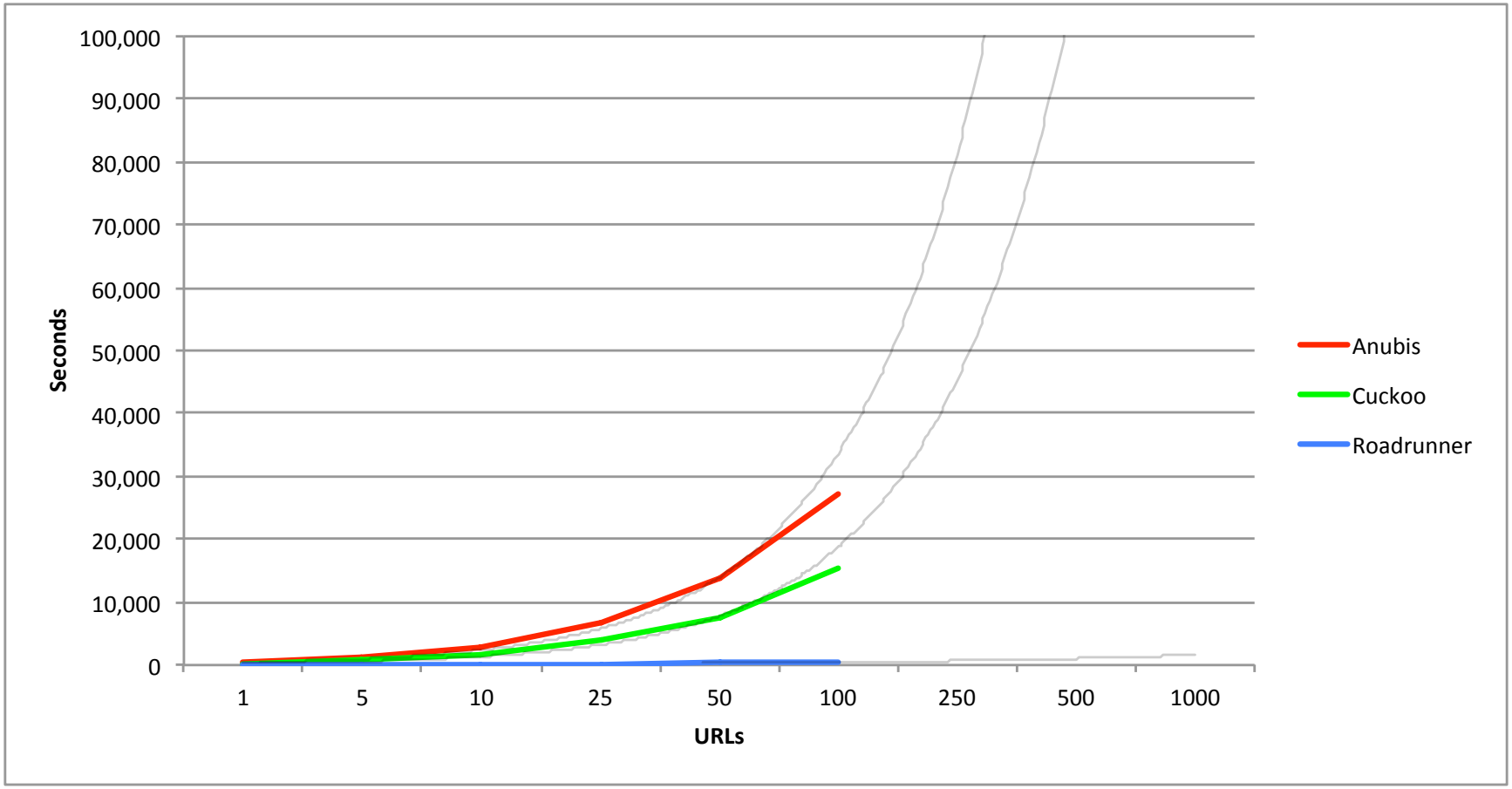

http://site.com/img/bg.jpg
http://site.com/css/new.css
http://site.com/img/t1.png
http://site.com/favicon.ico
http://www.google-analytics.com/analytics.js
http://fonts.gstatic.com/s/opensans/v10/[...]xaG2iE.eot
http://site.com/img/f.png
Process: rundll32.exe
http://fonts.googleapis.com/css?family=O[...]ns:300,600
http://site.com
%SystemRoot%\system32\rundll32.exe
Shell32.dll,Control_RunDLL C:\Users\Cuckoo\AppData\Local\Temp\errfix.exe
http://fonts.gstatic.com/s/mrdafoe/v5/Uk[...]7m55DA.eot
http://fonts.googleapis.com/css?family=Mr+Dafoe
http://site.com/img/r.png
http://www.google-analytics.com/r/collec[...]110391271
Process: control.exe
Process: 2640
http://fonts.googleapis.com/css?family=O[...]ns:300,600
http://site.com/jp/js/count.php
http://site.com/css/style.css
http://fonts.googleapis.com/css?family=Mr+Dafoe
Process: avap.exe
Process: errfix.exe
127.0.0.1:58608
Process: iexplore.exe
control.exe C:\Users\Cuckoo\AppData\Local\Temp\errfix.exe
C:\Users\Cuckoo\AppData\Local\Temp\errfix.exe


p://www.google-analytics.com/analytics.js
http://site.com/img/f.png
Process: rundll32.exe
/Uk[...]7m55DA.eot
%SystemRoot%\system32\rundll32.exe
Shell32.dll,Control_RunDLL C:\Users\Cuckoo\AppData\Local\Temp\\errfix.exe
Process: control.exe
Process: 2640
127.0.0.1:58608
Process: errfix.exe
Process: avap

But is it also faster?

# Benchmarks

| | 1 URL | 10 URLs | 100 URLs |
|---|---|---|---|
| Anubis | 273,8s | 2810s | ~27500s |
| Cuckoo | 152,8s | 1507s | ~15000s |
| Roadrunner | 48,8s | 102,4s | 450,9s |
| Improvement | 3-6x | 14-27x | 33-60x |




Comparison of benchmark results

# Research question

How can we **concurrently** visit multiple URLs and still be able to determine **which URL** was responsible for malicious activities?

- Use tabbed browsing
- Make use of modern browser architectures
- In-depth process monitoring (API Hooking)
- Graph based analysis

# Future Work

- Finishing touch PoC
  - Stability
  - Known false positives
  - Correctly creating the graph is hard
    - But all data is available for manual analysis like before
      - And now you know where to start looking
- Better analysis
  - Machine learning?

# Thanks!

- Our great supervisors from the NCSC
  - Jop van der Lelie & Wouter Katz
- The Cuckoo developers
  - Especially Jurriaan Bremer for helping us in understanding Cuckoo and solving our own introduced bugs

# Questions?

Adriaan Dens
adriaan.dens@os3.nl

Martijn Bogaard
martijn.bogaard@os3.nl

Check it out now:
https://github.com/MartijnB/cuckoo/tree/multi-url