ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

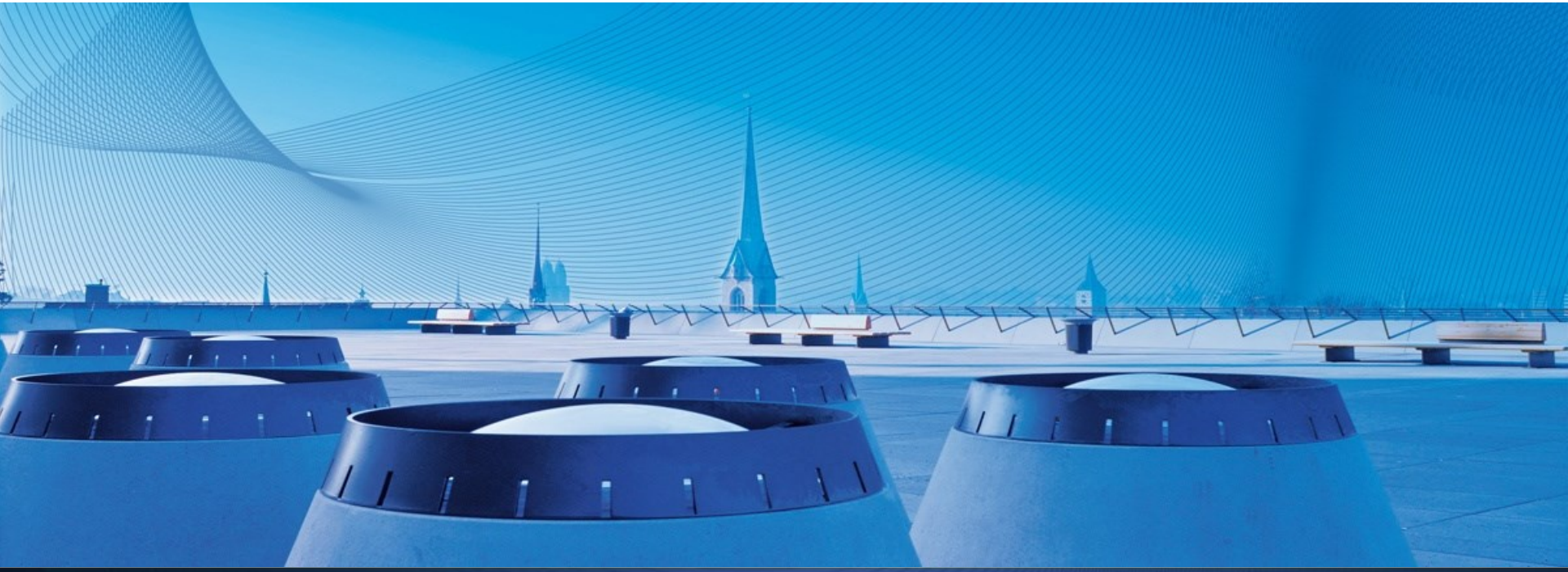
Institut für
Technische Informatik und
Kommunikationsnetze

DFRWS EU 2015

Hviz: HTTP(S) Traffic Aggregation and Visualization for Network Forensics

David Gugelmann, Fabian Gasser, Bernhard Ager (ETH Zurich, Switzerland)
Vincent Lenders (armasuisse, Thun, Switzerland)

Date: 24. March 2015 Location: Dublin



INTRODUCTION

Motivation and problem statement

- HTTP(S) traffic is important for digital forensics:
  - Many organizations allow Web browsing
  - Main protocol in corporate networks
  - Used by malware as C&C-channel

- Nowadays Web sites are quite complex:
  - Loading a single Web site can cause dozens to hundreds of HTTP(S) requests
  - Content is loaded from many different servers

➔ **Difficult to manually reconstruct, identify and analyze suspicious Web activity**

Contributions

- Hviz - HTTP(S) traffic visualizer:
  - Grouping, aggregation and correlation of HTTP events
    - Number of events reduced by nearly a factor of 20
    - ➔Much easier for an investigator to spot anomalies
  - Interactive graph visualization of HTTP(S) activity of a workstation
    - Represent event timeline
    - ➔Visualize “what a user/malware did”
- Evaluation using synthetic and real-world HTTP traces

DESIGN GOALS AND DATA PROCESSING

Design goals

I. Visualize the timeline of Web browsing, i.e., which sites a user visited

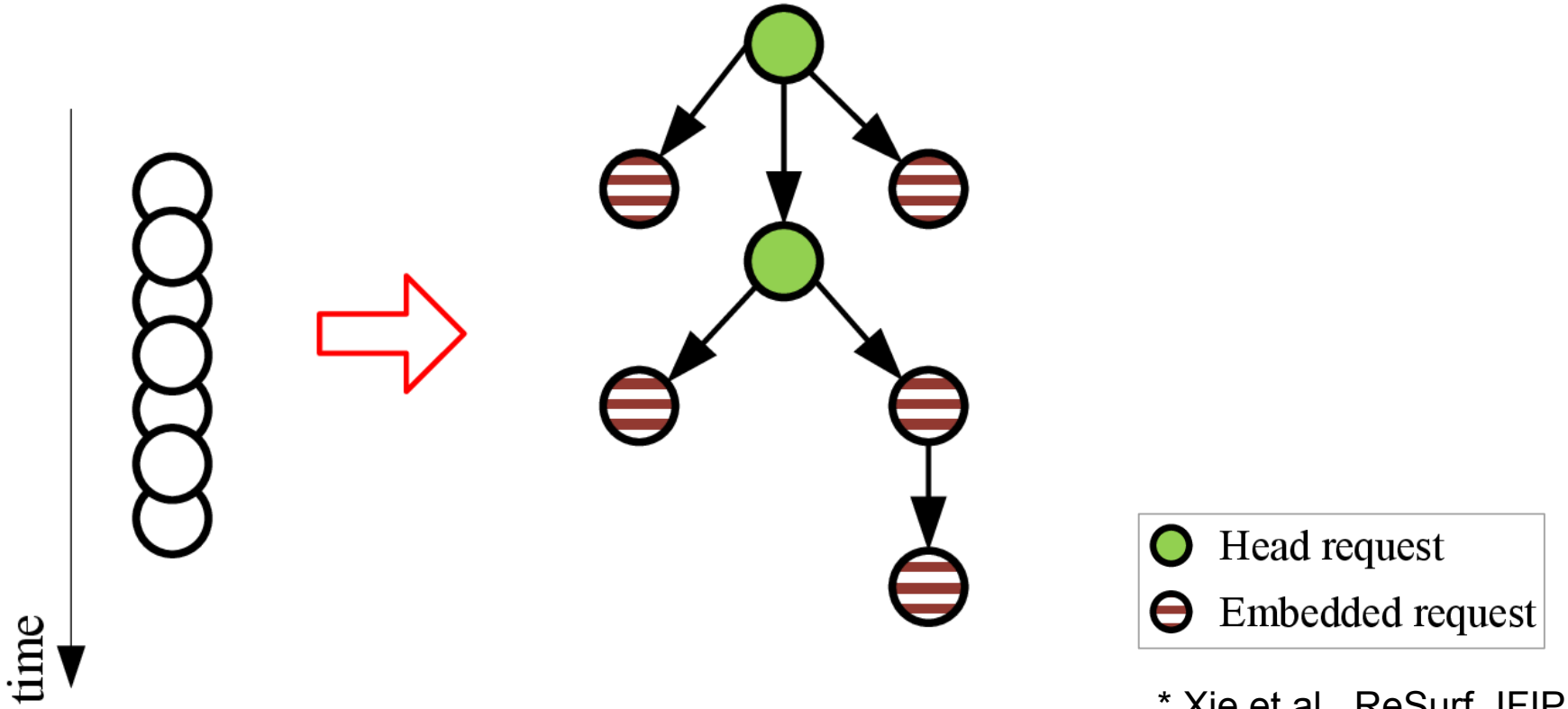
II. Support an investigator to understand why a request happened:
- Result of regular Web browsing
- Malware activity
- …

III. Reduce the number of displayed events
- Allow to quickly grasp the big picture

IV. Prevent HTTP activity from getting lost in the shuffle
- E.g., malware activity should be visible despite the large numbers of requests caused by regular Web browsing

Step I: Detecting user clicks

- Request graph and request classification*



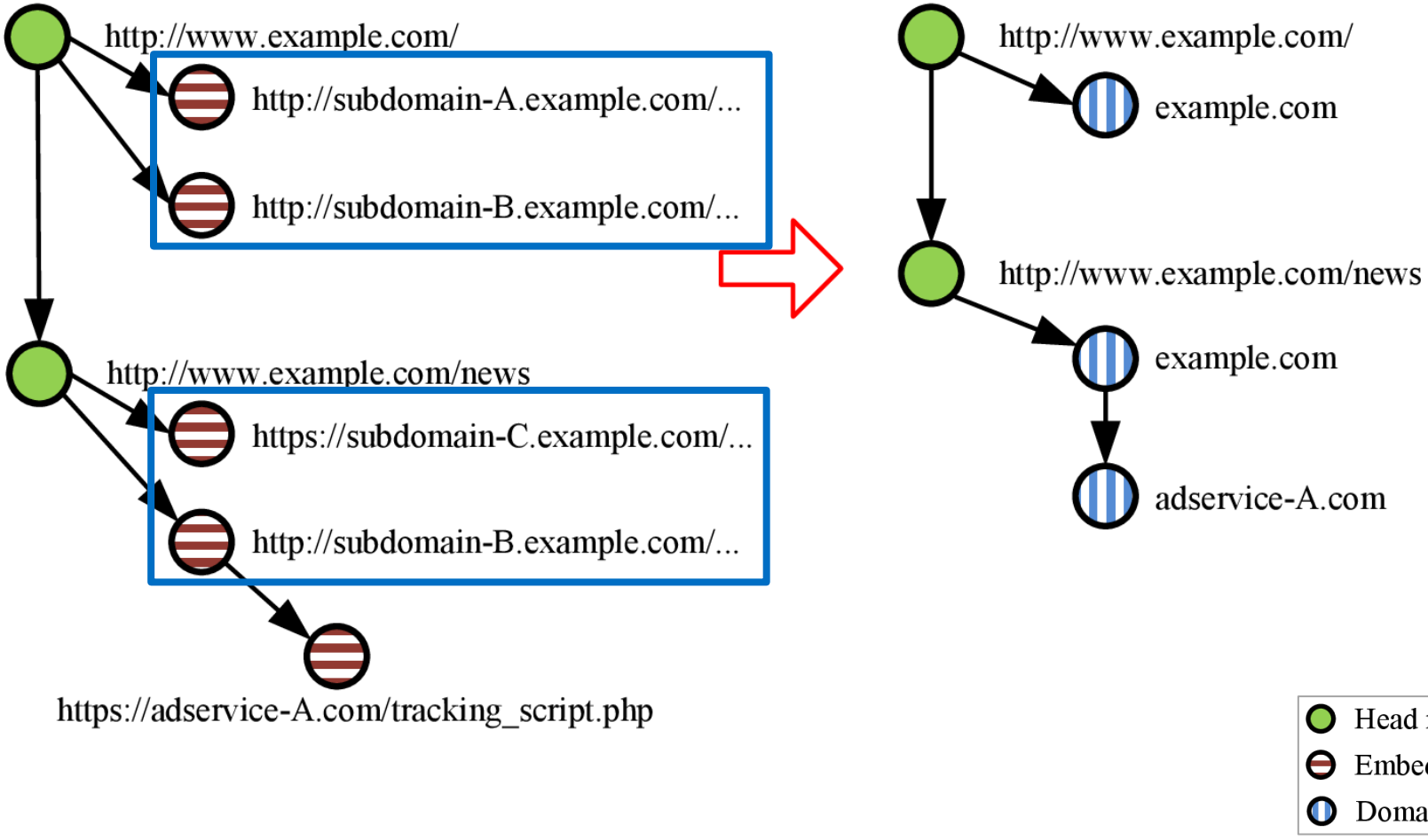
* Xie et al., ReSurf, IFIP Networking, 2013

➔ Head requests represent “big picture”

➔ Request graph shows how user arrived at a Web page

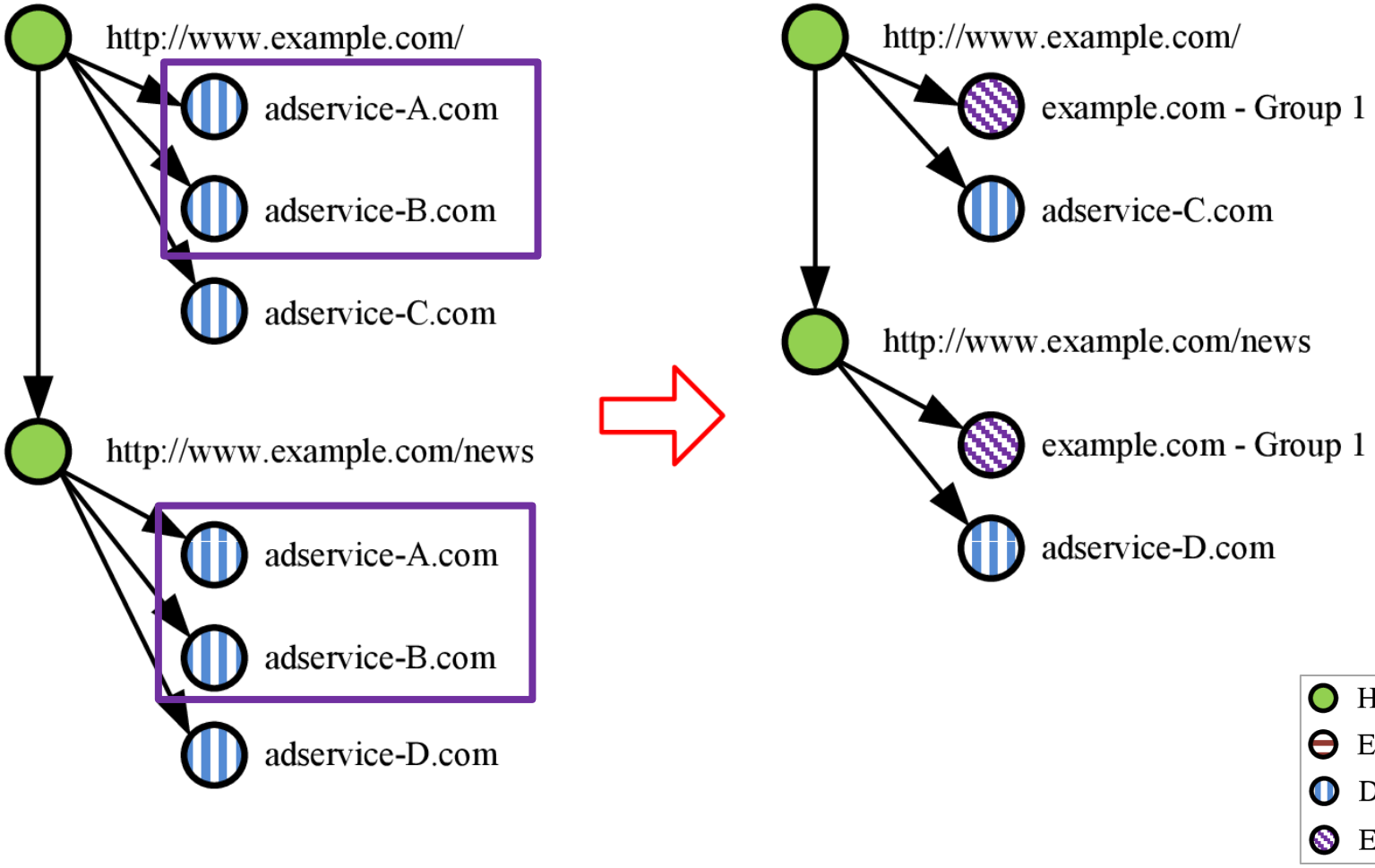
Step II.a: Domain aggregation

- Aggregate embedded requests to domain events



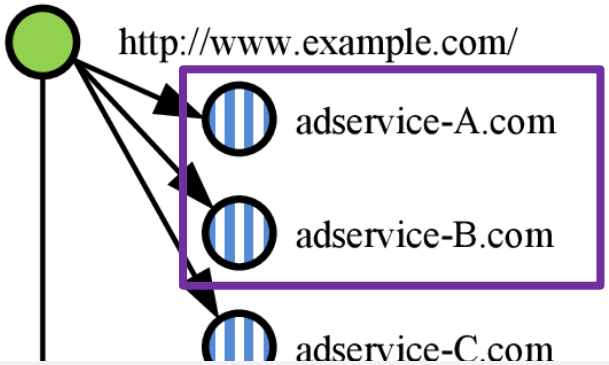
Step II.b: FIM aggregation

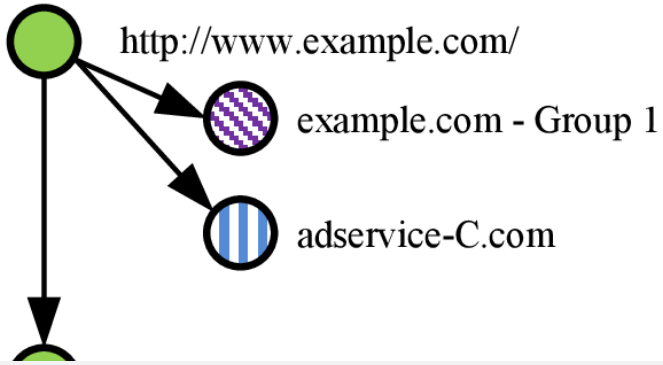
- Aggregate domain events using frequent itemset mining (FIM)



Step II.b: FIM aggregation

- Aggregate domain events using frequent itemset mining (FIM)



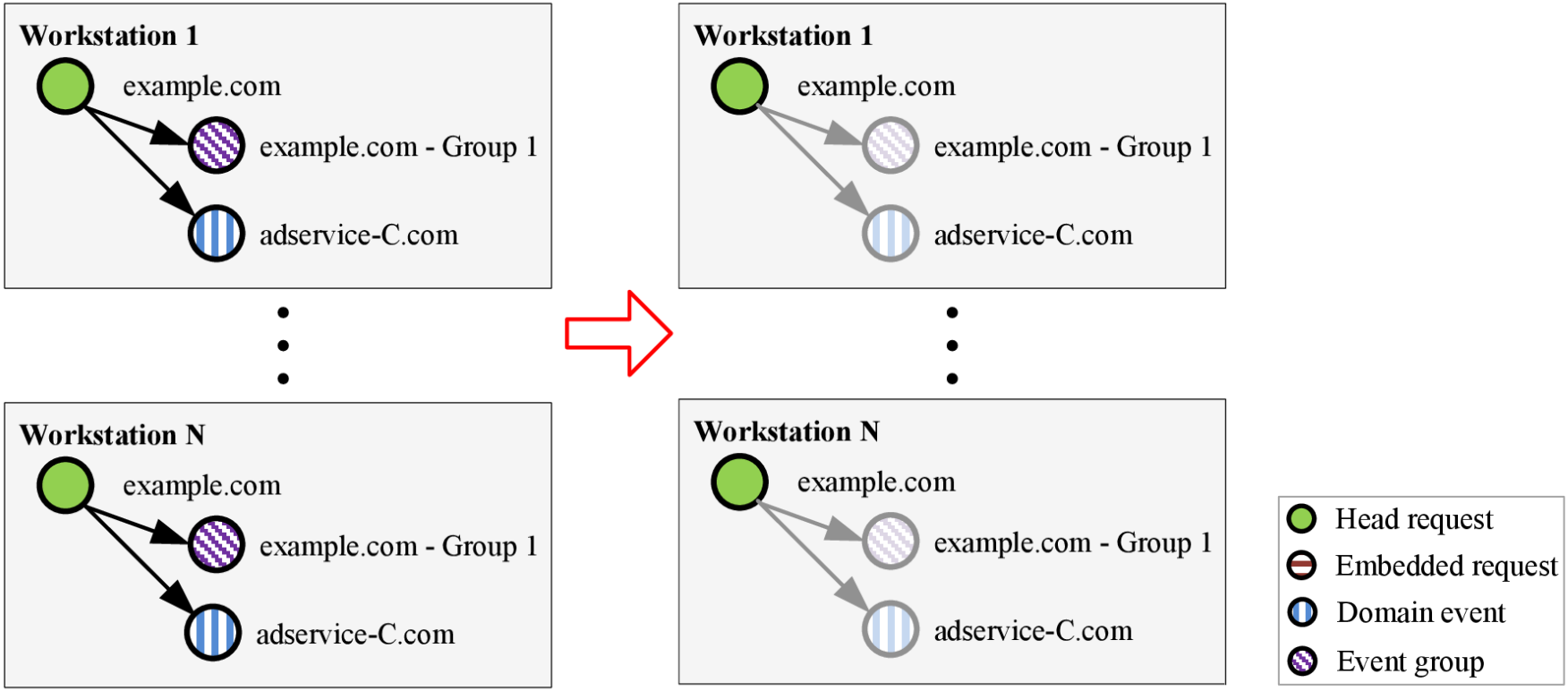


➔ Advantages of aggregation over only displaying head events:

- Requests that are not related to Web browsing (e.g. malware) are visible
- Easier to identify and handle misclassified nodes
  - Attackers could intentionally cause misclassifications

Step III: Filtering based on correlation

- Fade out navigation paths that are common to many computers

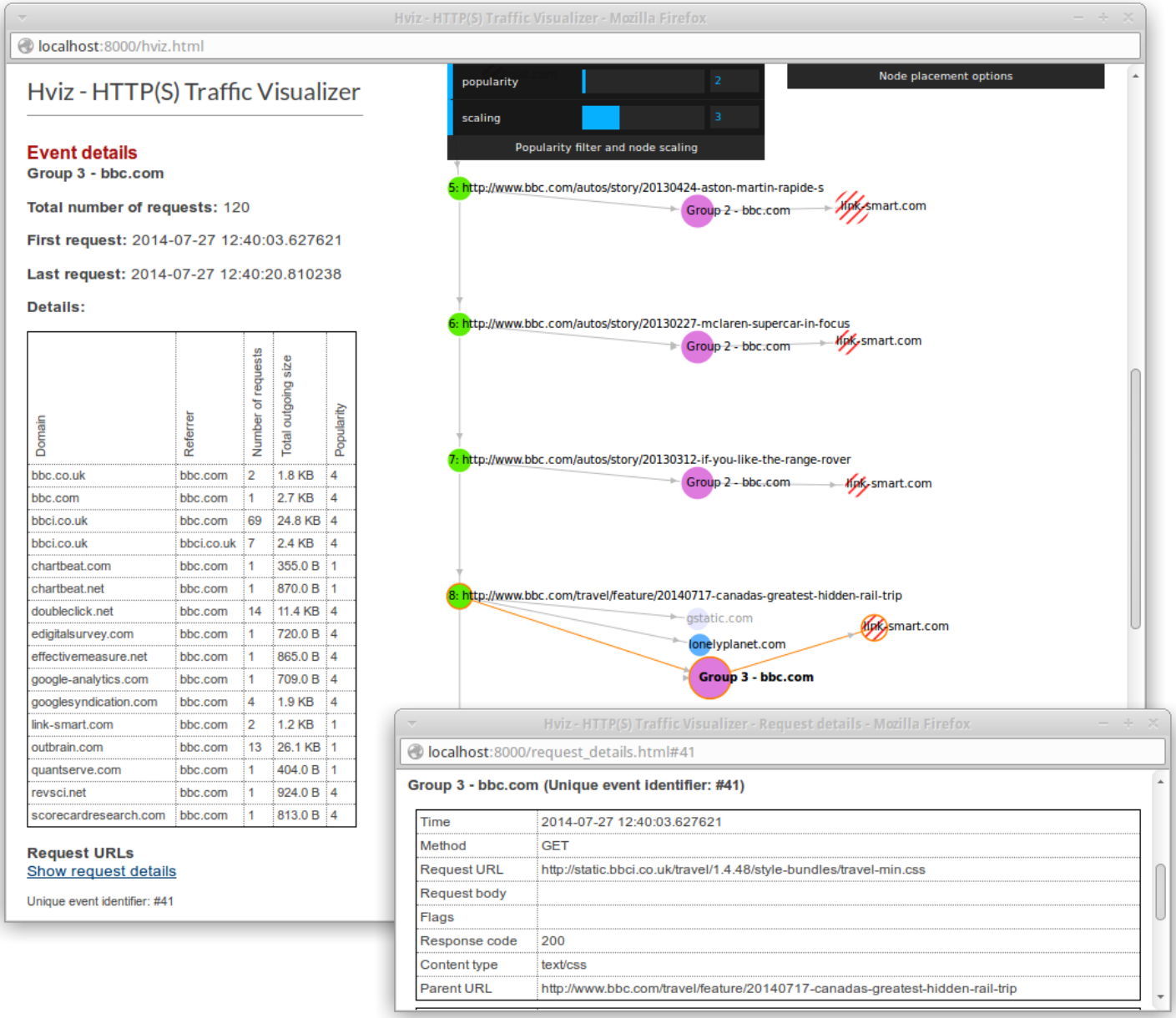


➔ Focus on a workstation's singular traffic

IMPLEMENTATION

Implementation

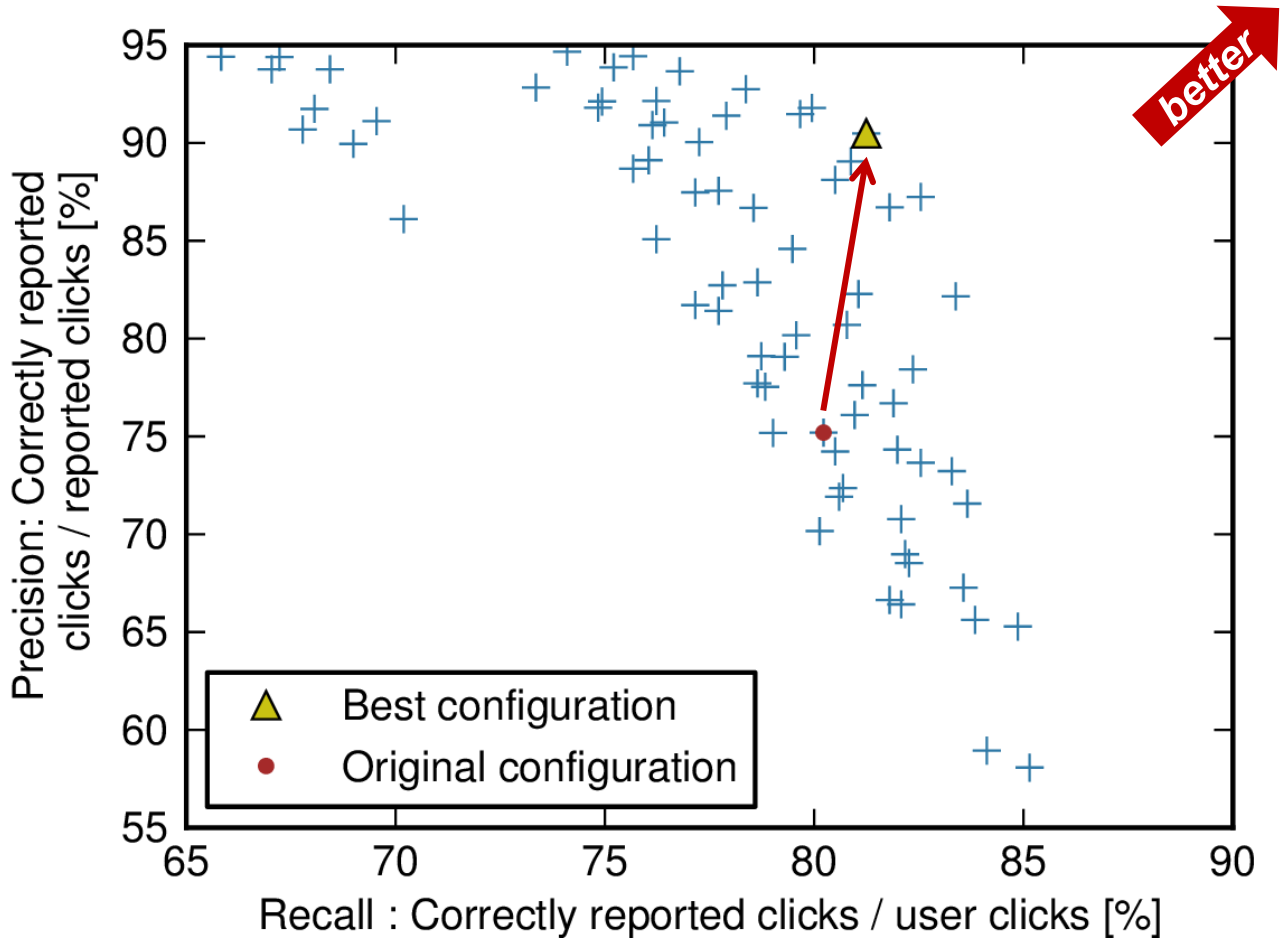
- Backend processing
  - Bro IDS to parse libpcap files
    - HTTP activity
  - Mitmproxy scripting API for mitmdump logs
    - HTTP and HTTPS activity
  - Python program
    - NetworkX
    - PyFIM (Frequent Item Set Mining for Python)
- Frontend
  - Running in Web browser
  - 3D.js


Hviz - HTTP(S) Traffic Visualizer - Mozilla Firefox
localhost:8000/hviz.html
Hviz - HTTP(S) Traffic Visualizer
popularity
2
scaling
3
Popularity filter and node scaling
Node placement options
5: http://www.bbc.com/autos/story/20130424-aston-martin-rapide-s
Group 2 - bbc.com
link-smart.com
6: http://www.bbc.com/autos/story/20130227-mclaren-supercar-in-focus
Group 2 - bbc.com
link-smart.com
7: http://www.bbc.com/autos/story/20130312-if-you-like-the-range-rover
Group 2 - bbc.com
link-smart.com
8: http://www.bbc.com/travel/feature/20140717-canadas-greatest-hidden-rail-trip
gstatic.com
lonelyplanet.com
link-smart.com
Group 3 - bbc.com
Event details
Group 3 - bbc.com
Total number of requests: 120
First request: 2014-07-27 12:40:03.627621
Last request: 2014-07-27 12:40:20.810238
Details:
Domain | Referrer | Number of requests | Total outgoing size | Popularity
bbc.co.uk | bbc.com | 2 | 1.8 KB | 4
bbc.com | bbc.com | 1 | 2.7 KB | 4
bbci.co.uk | bbc.com | 69 | 24.8 KB | 4
bbci.co.uk | bbci.co.uk | 7 | 2.4 KB | 4
chartbeat.com | bbc.com | 1 | 355.0 B | 1
chartbeat.net | bbc.com | 1 | 870.0 B | 1
doubleclick.net | bbc.com | 14 | 11.4 KB | 4
edigitalsurvey.com | bbc.com | 1 | 720.0 B | 4
effectivemeasure.net | bbc.com | 1 | 865.0 B | 4
google-analytics.com | bbc.com | 1 | 709.0 B | 4
googlesyndication.com | bbc.com | 4 | 1.9 KB | 4
link-smart.com | bbc.com | 2 | 1.2 KB | 1
outbrain.com | bbc.com | 13 | 26.1 KB | 1
quantserve.com | bbc.com | 1 | 404.0 B | 1
revsci.net | bbc.com | 1 | 924.0 B | 4
scorecardresearch.com | bbc.com | 1 | 813.0 B | 4
Request URLs
Show request details
Unique event identifier: #41
Hviz - HTTP(S) Traffic Visualizer - Request details - Mozilla Firefox
localhost:8000/request_details.html#41
Group 3 - bbc.com (Unique event identifier: #41)
Time | 2014-07-27 12:40:03.627621
Method | GET
Request URL | http://static.bbci.co.uk/travel/1.4.48/style-bundles/travel-min.css
Request body |
Flags |
Response code | 200
Content type | text/css
Parent URL | http://www.bbc.com/travel/feature/20140717-canadas-greatest-hidden-rail-trip

EVALUATION

Evaluation – Detecting user clicks

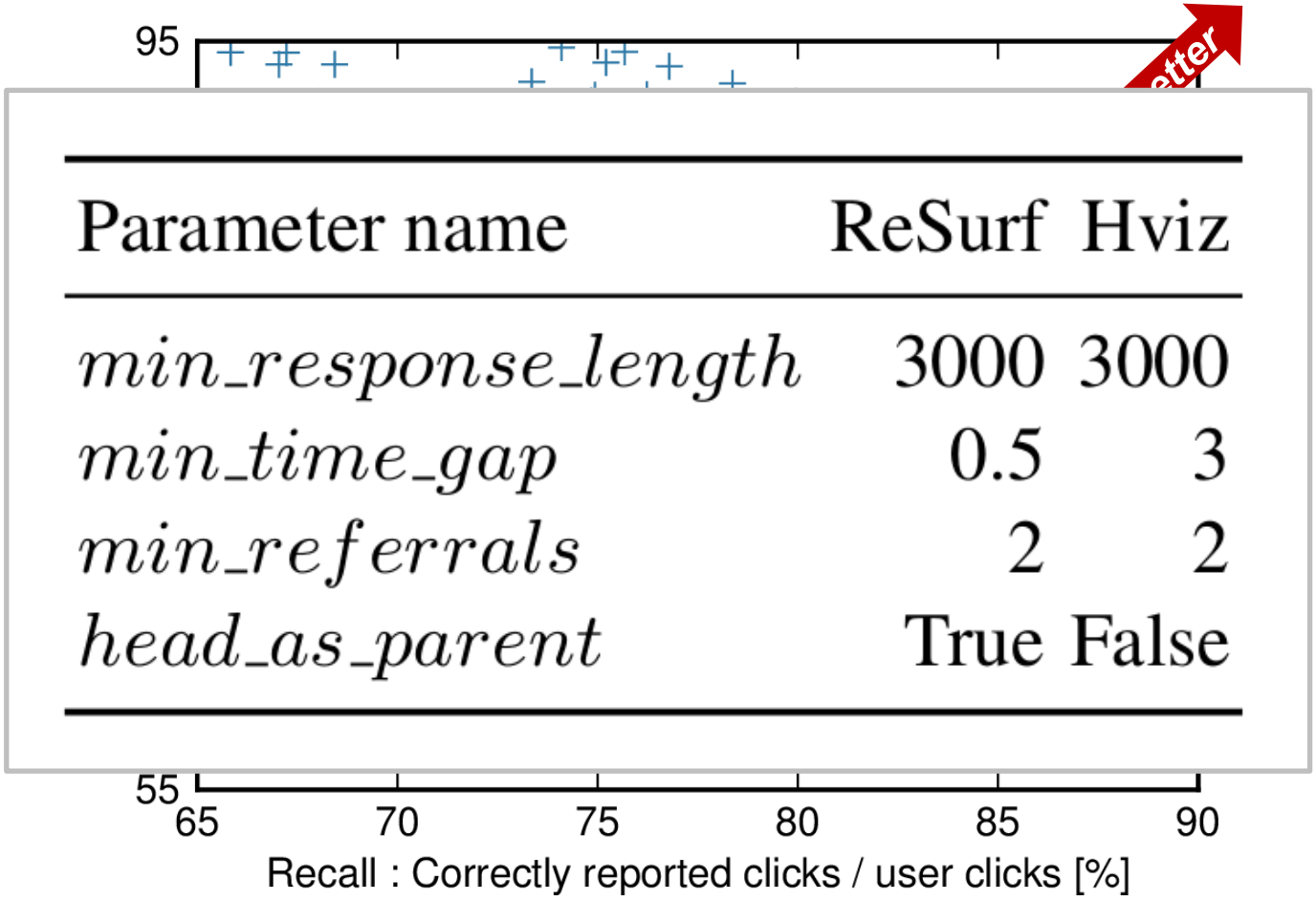
- Evaluation dataset: automated Web browsing on top 300 Alexa sites



➔ Parameters improved over original ReSurf algorithm

Evaluation – Detecting user clicks

- Evaluation dataset: automated Web browsing on top 300 Alexa sites

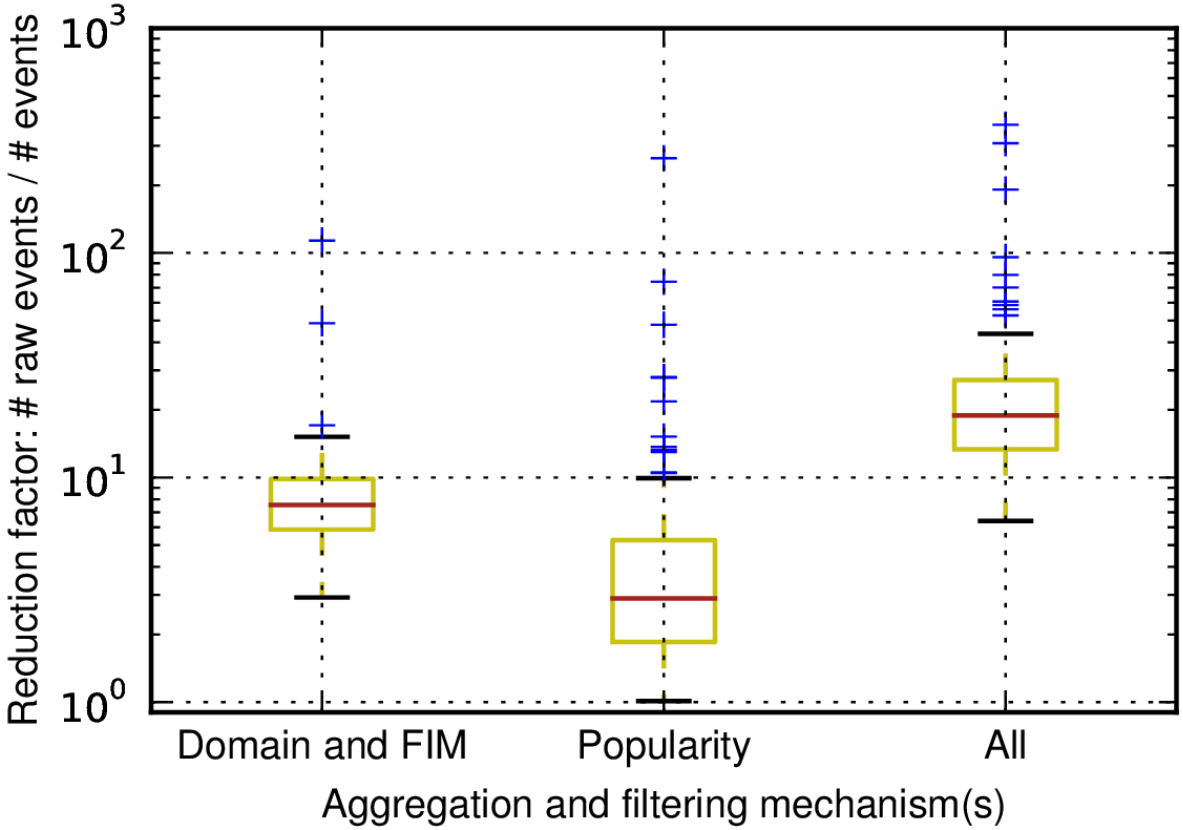


Parameter name	ReSurf	Hviz
min_response_length	3000	3000
min_time_gap	0.5	3
min_referrals	2	2
head_as_parent	True	False

➔ Parameters improved over original ReSurf algorithm

Evaluation – Aggregation and filtering

- Evaluation dataset: HTTP traffic from a university network, 24h, 1.8k clients, 5.7M requests



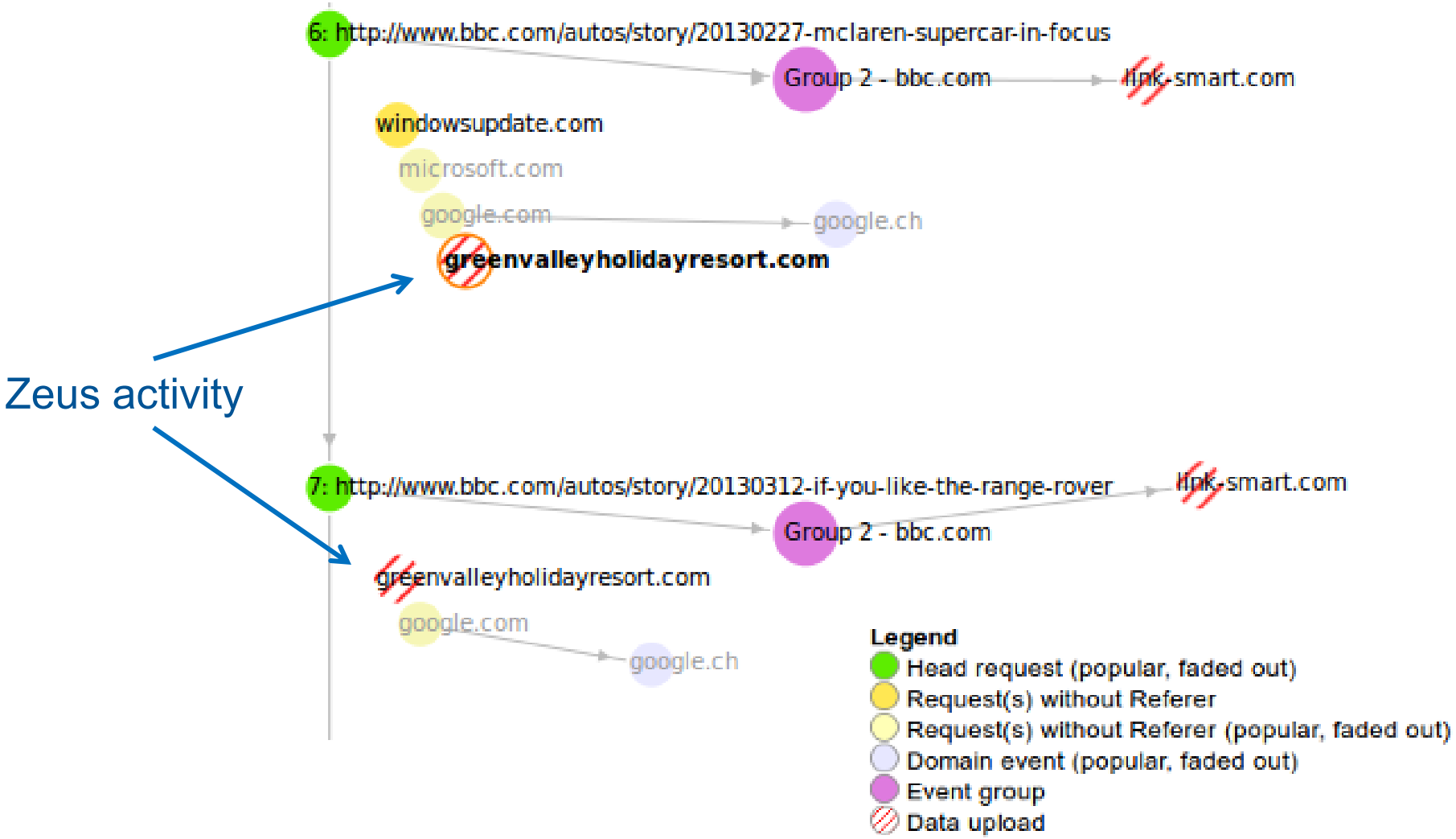
Event reduction factors:

- Domain and FIM grouping: 7.5
- Popularity-filter (threshold 10/1.8k): 2.9

➔ Overall reduction factor: **19**

USAGE SCENARIOS

Zeus malware activity during regular Web browsing

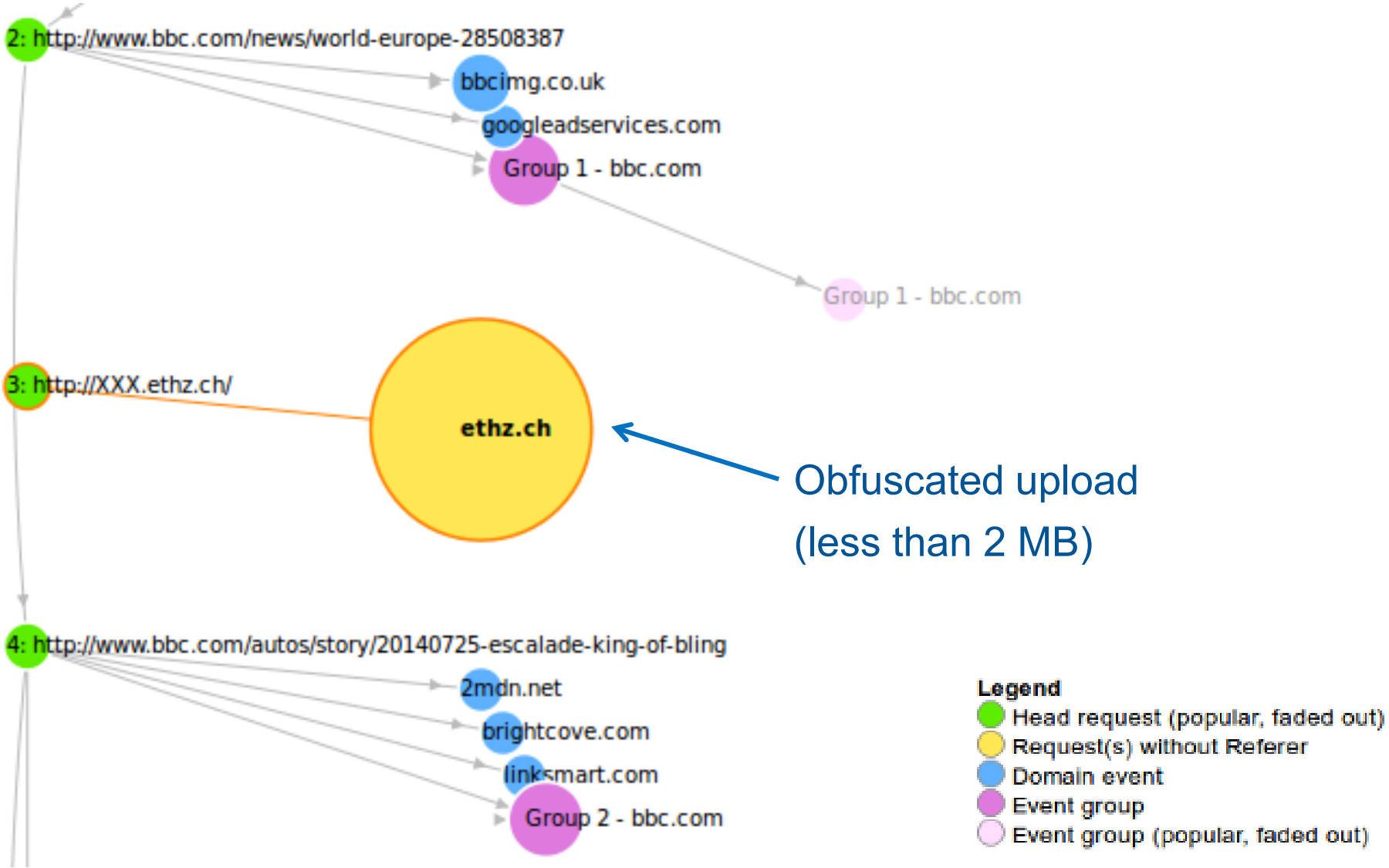


greenvalleyholidayresort.com (Unique event identifier: #41)

Time	2014-07-27 12:39:18.780000
Method	POST
Request URL	http://www.greenvalleyholidayresort.com/bytecode/gate.php
Request body	0x25 0x6A 0xBD 0x9B 0x44 0x66 0x97 0xCF 0x9F 0x96 0x99 0xB9 0x10 0x88 0x4D 0xE4 0x70 0x97 0xDF 0x28 0xE8 0xF1 0xD4 0xDD 0x7D 0x70 0x0B 0x34 0xDE 0x09 0x45 0x75 0x45 0x36 0x79 0xEB 0x26 0x88 0x3B 0xC7 0x90 0xC2 0x17 0x09 0x25 0x90 0xC6 0xF8 0x7B 0x41 0xDB 0x7D 0x62 0x54 0x82 0x81 0x22 0x79 0x96 0x48 0x70 0xD7 0xAB 0xB5 0x29 0x7D 0xFF 0x32 0x88 0x90 0x17 0x9F 0x9C 0xA8 0x11 0x27 0x76 0xC8 0x4E 0xCA 0x30 0x0E 0x02 0x2C 0x1D 0x75 0x50 0x50 0x8B 0xB9 0xC9 0x62 0x39 0xC5 0xBD 0xCE 0x33 0x61 0x57 0x7A 0x52 0x8D 0x73 0x92 0x97 0xED 0xC1 0x9A 0x04 0x11 0xAE 0xD9 0x7A 0xF0 0xBF 0x97 0xE3 0xCB 0x0E 0xD2 0xC2 0x31 0x6D 0x99 0x83 0xDA 0x00 0x42 0x85 0x80 0x73 0x4A 0x7B 0x03 0xED 0x31 0xF3 0x48 0xFB 0x40 0x43 0x94 0x92 0xD1 0xB4 0x1B 0xA5 0x9B 0xA8 0xDC 0x20 0x22 0xCC 0x5C 0x11 0xDE 0x82 0xB6 0xFF 0x28 0xBA 0x66 0xB1 0x0B 0x04 0x0B 0xEB 0x09 0x71 0xB7 0xFB 0x6D 0x9E 0x0A 0xC0 0x4D 0xDD 0xB7 0x26 0xAA 0x95 0x74 0x6C 0x4E 0xD6 0x4B 0xA2 0xB4 0x7E 0xD6 0x6D 0x73 0x19 0x75 0xCD 0x61 0x11 0x1A 0x2B 0x4B 0xD9 0xD7 0xFA 0xC6 0xDC 0x4E 0x86 0x01 0x5A 0x3A 0x41 0xF3 0xE4 0xA9 0x17 0x6E 0x89 0xD9 0x4C 0xFC 0xA7 0xCD 0xA6 0x7D 0xE9 0xEF 0x9A 0xA5 0x68 0x0C 0xCA 0x47 0xB1 0x12 0x19 0xA2 0x7F 0x71 0x1F 0xDC 0x3C 0x17 0xEB 0x67 0x22 0x17 0xF5 0x2F 0xDF 0x1D 0xEC 0xA7 0xC2 0x00 0x7C 0x88 0x25 0x2C 0xF9 0x0B 0xF2 0x1B 0xF8 0xDC 0xA8 0x03 0x89 0xD2 0x62 0xDF 0xD3 0xB9 0xA4 0x87 0xBA 0xA3 0xB6 0x72 0x1F 0xC9 0x65 0x88 0x27 0xE0 0x4A 0x28 0x1F 0x65 0x7A 0x0C 0xEB 0xE0 0xA5 0xA1 0x7B 0x05 0xD5 0xB2 0xBA 0x03 0xC0 0x25 0xFB 0x08 0x24 0x5B 0x6D 0xEB 0x5B 0xFB 0xFE 0xFB 0xAF 0x32 0xF5 0xE2 0x4D 0xB5 0x83 0x0E 0x56 0x25 0xF3 0x47 0xEC 0xA0 0x4D 0x43 0x19 0xCA 0xD7 0xDC 0x5F 0xF0 0x40 0x5C 0xC8 0xE3 0xE7 0xA7 0x63 0xC8 0xC9 0x3C 0x1E 0x42 0x6D 0x24 0xD6 0x7B 0xB5 0xEF 0x07 0x17 0xED 0x4D 0xBB 0xCB 0xA3 0x99 0xD1 0x0B 0x49 0x16 0x3C 0x66 0xA7 0xFC 0x77 0x1B 0x64 0xA5 0x12 0xB2 0xF6 0x6F 0x99 0x87 0x12 0xD3 0xF5 0x91 0x2E 0xB6 0x52 0x44 0x43
Flags	REQUEST_BODY_PRESENT; NO_PARENT_FOUND
Response code	200
Content type	text/html
Parent URL	None

Time	2014-07-27 12:39:26.270000
Method	POST
Request URL	http://www.greenvalleyholidayresort.com/bytecode/gate.php
	0x30 0xBE 0xD2 0xB2 0x57 0x23 0xF1 0x15 0x72 0x16 0xC4 0x1E 0x5E 0x7E 0x11 0xE5 0xEE 0xA4 0x78 0x2

Data exfiltration



DFRWS 2009 Challenge

- Part of DFRWS 2009 forensics challenge:
  - Illegal Mardi Gras images have been shared
  - A suspect denies being responsible for any shared images

➔ Hviz shows at a glance that corresponding Web pages have been searched for and accessed
(which does not proof that the suspect indeed shared these images, but it is an indication that the system should be analyzed)

SUMMARY

Summary

- Hviz visualizes Web browsing activity in a graph:
  - Number of active events reduced by a factor of 19 by grouping, aggregation and correlation
- An investigator can interactively filter and explore Web activity:
  - Understand the “big picture”
  - Zeus malware activity and obfuscated uploads as small as a few MB clearly stand out
- Live demonstration:

  http://hviz.gugelmann.com

Thank you for your attention!
Questions?