Anomaly Detection on User-agents

Peter van Bolhuis

Overview

- Introduction
- Research Question
- User-agents
- Scoring host anomalies
- Verification
- Conclusion

Introduction

- Methods
  - Statistical
  - Knowledge based
  - Machine learning

Research Question

What is the effectiveness of statistical anomaly detection when applied to user-agent strings?

User-agents

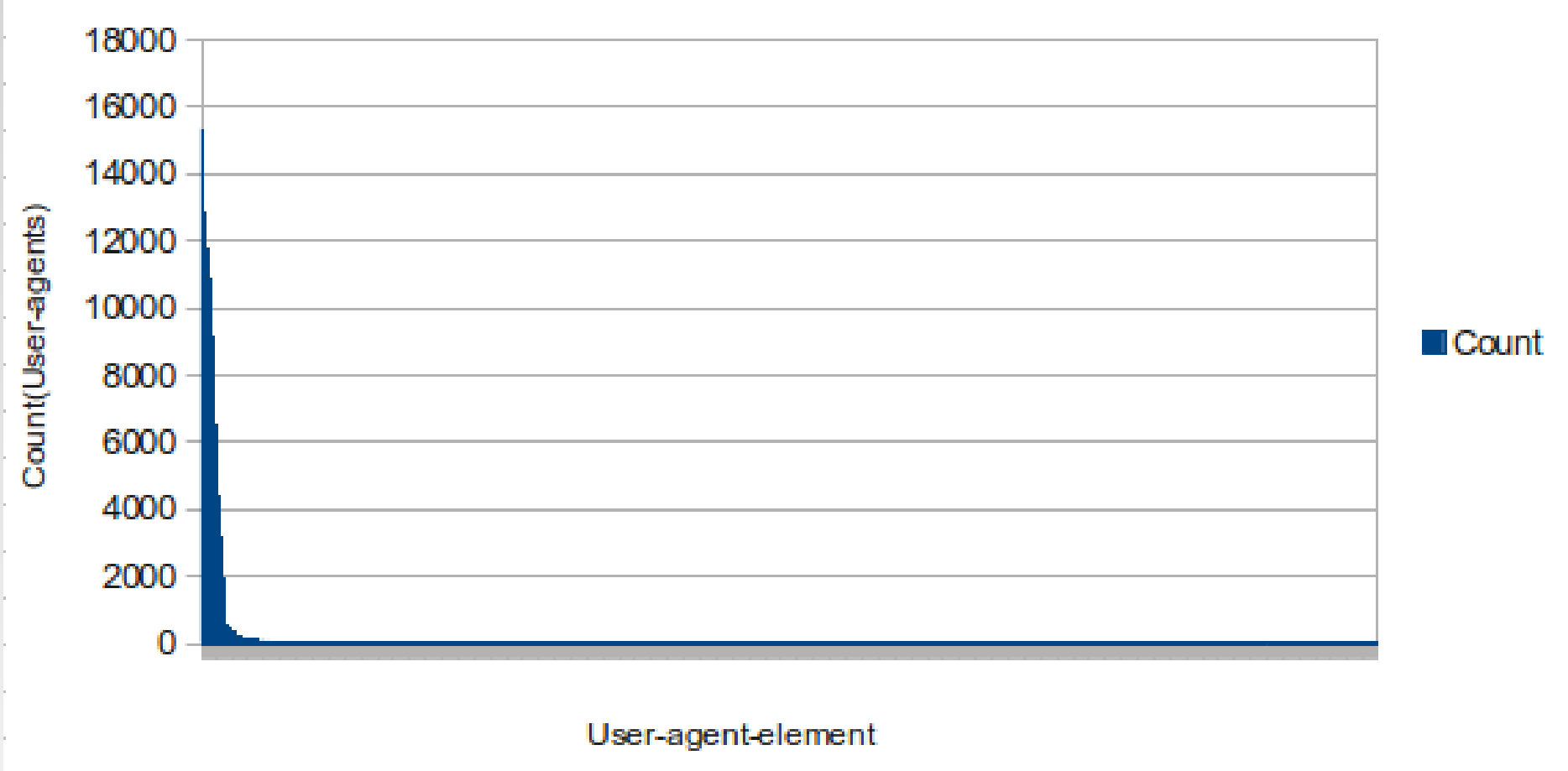
- Programs that "*act on behalf of a user*"
- Identify themselves with a string
  - Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/7.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; eSobiSubscriber 2.0.4.16; BRI/1; MAAR; .NET4.0C; AskTbORJ/5.15.9.29495; .NET4.0E; BRI/2) Funshion/1.0.0.1

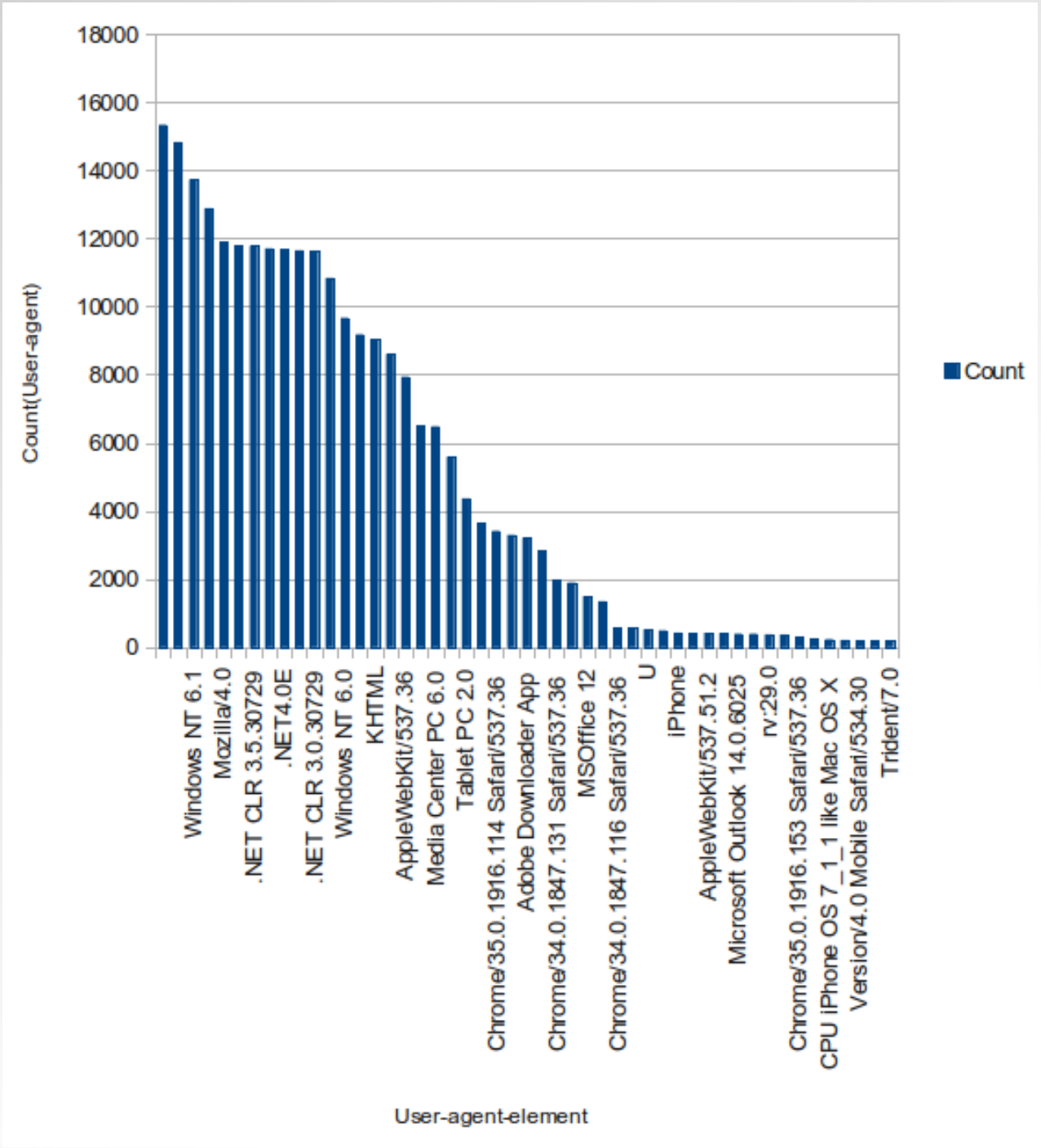
User-agents (2)

- Problem:
  - Mozilla/4.0 (compatible; version 1.33.7)
  - Mozilla/4.0 (compatible; version 1.33.8)
  - Dalvik/1.4.2 (AskTbORJ/5.15.9.29495)

User-agents (3)

- Splitting on elements
  - Mozilla/4.0
  - Dalvik/1.4.2
  - Compatible
  - Version


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Count(User-agents)
User-agent-element
Count


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Count(User-agent)
Count
Windows NT 6.1
Mozilla/4.0
.NET CLR 3.5.30729
.NET4.0E
.NET CLR 3.0.30729
Windows NT 6.0
KHTML
AppleWebKit/537.36
Media Center PC 6.0
Tablet PC 2.0
Chrome/35.0.1916.114 Safari/537.36
Adobe Downloader App
Chrome/34.0.1847.131 Safari/537.36
MSOffice 12
Chrome/34.0.1847.116 Safari/537.36
U
iPhone
AppleWebKit/537.51.2
Microsoft Outlook 14.0.6025
rv:29.0
Chrome/35.0.1916.153 Safari/537.36
CPU iPhone OS 7_1_1 like Mac OS X
Version/4.0 Mobile Safari/534.30
Trident/7.0
User-agent-element

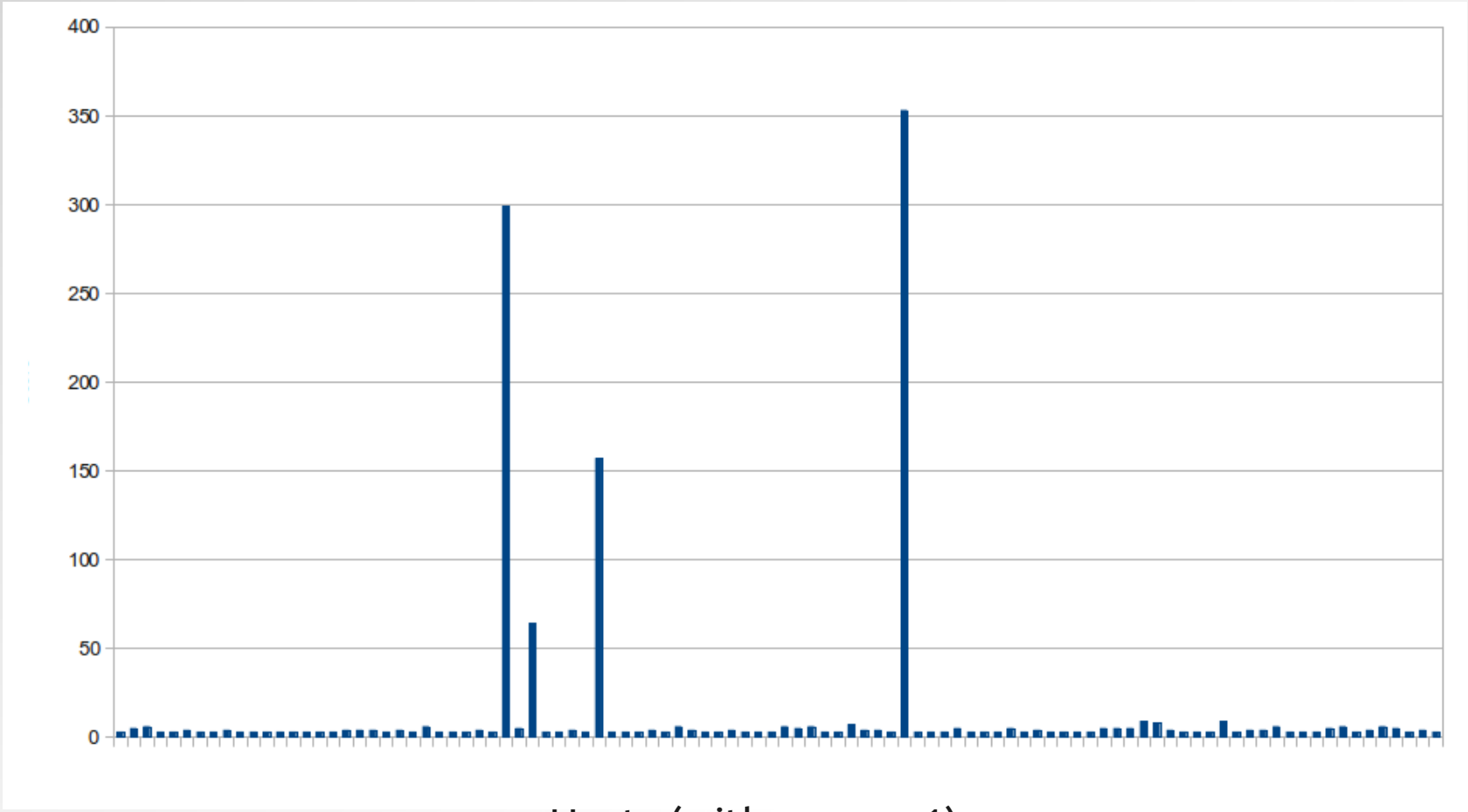
Scoring host anomalies

- Elements with the lowest *n* occurrences give a host a +1

User-agent element	#Occurrences	Increases score
Mozilla/4.0	100	No
AppleWebKit/537.36	20	No
Dalvik/1.4.0	10	Yes
AppWorld/5.0	2	Yes
Q10/10.2.1.3175	2	Yes
zh-cn	2	Yes
4012FREE	1	Yes

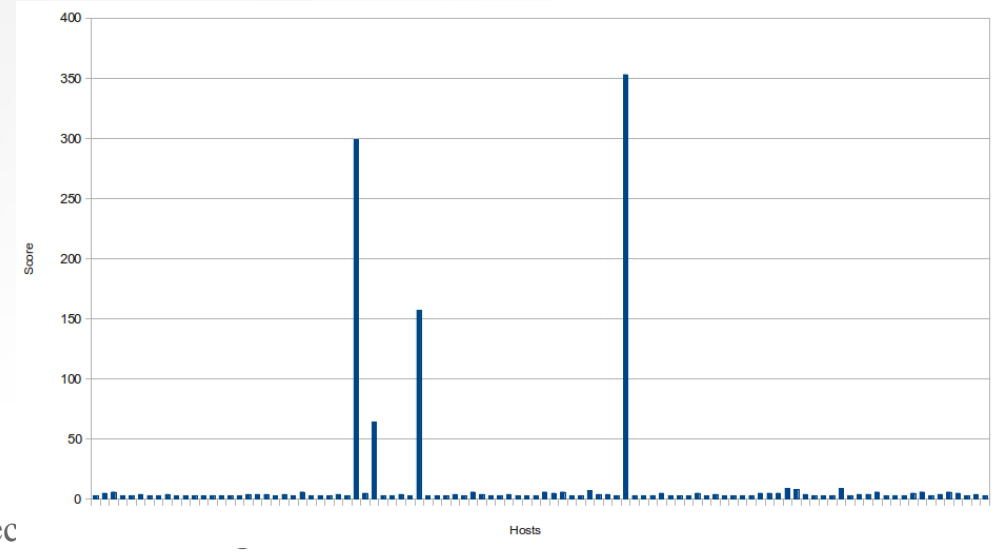
Table for *n* = 3

Scoring host anomalies (2)



Verification

Host	Score	Result of verification
A	299	Host was a phone: Compliance incident
B	64	Host infected with Conduit Browser Hijacker
C	157	Host was a proxy
D	353	Host was a proxy



Conclusion

- User-agent strings can be used for anomaly detection
  - Best results on uniform networks
  - Anomalies are not necessarily infections, but rather installed software packages

Demo

Thank you