



Attacking a Big Data Developer

Dr. Olaf Flebbe
of [ät oflebbe.de](mailto:of@flebbe.de)

ApacheCon Bigdata Europe
16.Nov.2016 Seville

About me

PhD in computational physics

Former projects: Minix68k (68k FP Emulation), Linux libm.so.5 (High Precision FP), perl and python for epoc, flightgear, msktutil...

PMC of Apache Bigtop

(Chief Software Architect at a European Software Integrator/Big Data)



Security

- ✦ The Internet is not a safe space any more
- ✦ Attackers are using increasingly complex attacks in order to penetrate enterprises
- ✦ There is no well established awareness for
 - ✦ Developers can be a attack vector!
 - ✦ Developers may create malicious artifacts by reusing insecure components.

Developer Attack Vector

- ✦ Any user of a software component which uses an insecure build process can be harmed and may create software artifacts which can penetrate its customer
- ✦ Method for investigation:
 - ✦ Compile a large code base
 - ✦ Looking for possible attack vectors

Method

- Catching complete network traffic when compiling a Big Data Distribution
- Create in depth package analysis of the traffic with an sophisticated network security monitor
- Store the representation in a NoSQL store
- Query

Toolset

Big Data Distro

```
:aass..
=XS22nai,>
=n-- +!!!"^^--
.vX>
.e<o;.
.v2`-{S>
..<de~..;)Sa,
.._aoX}::==>=-?Xo>,
.._aaoZe!`=><i=s+s;~*XXos,,
.....=iisaaoXXZY!"~.v(d=:nc-1s,~?SX#Xouass,,
=XXoXXXSXXXXXXXZUX21?!"^-.._au*`=u2` ]X>.+*a>,-"!Y1XSSX##ZXXXXXXXXXXXXoXXc
.{XXXXXX2*?!!"!^~---..._aa2!^-=dX(.+XXc.~!1nas,,----~^"!!"!!!?YSXXXXX2+
-"YSXXXXo=. _=sssaaav1!!~. _aXXe` )SXo>. ~"?Yooouass s,, vXXXXX2}~
-{XXZoai%*XXSSSX>.. <uXXX2~ {XXXXs,, .=dXXXZX2lii%uXXXe-
.<XXXXX%- <XXXXX1|==%vdXXXXXo;.. _vXXXXXXos_ =i|*XXXXX> -<XXXXX`
=SXXXXZc ..nXXX2> ---=2XXX2^-"|||}"---{ZXXX1-- :XXXXo; .)XXXX2`
=XXXXZc .nXXXX> =XXXXe.._s=> ..)XXXX1 .:SXXXXo; .)XXXX2..
<SXXXXc .nXXXX> =XXXXosummmBmma,)ZXXX1 :XXXX2; .)XXXXX.
<XXXXX( :nXXXXS; <XXXXXm#mmmWmmmmmoZXXX1 ..3XXXo; ..)XXXXX;
.nXXXXX; :XXXXX; =XXXXXmmmBmmWmB#XXXXX1 .nXXXX> :XXXXXc
=oZXXXXe; .<XXXX2` .)XXXXZmBmBmWmmW#2XXXX1 .vXXXXc vXXXXo;
+Y3S2Xz...vXXXXe .)ZXXXZmmWmBmmBBm#XXXXXo.. {XXXXz:..vSS2Y1=
---+"!"*!*Y1s|_ =uXSSSXZUXUXUXUXS2XX2n|_ =||%Y*?!!"!^^~---
```

```
.o.
.888.
.8"888.
.8'`888.
.880008888.
.8' `888.
o88o o8888o
0000
`888
888 .oo. .00000.
888 `888P"Y88b d88' `88b
888 888 .oP"888 888
888 888d8( 888 888 .o8 888 888 888 .o
888bod8P'`Y888"8o`Y8bod8P'o888o o888o`Y8bod8P'
o888bood8P' o888o`8o00000.
888 d" YD 888
o888o "Y88888P' o888o
```

<advertisement>

Bigtop



- Apache Bigtop is the „Debian“ of the Big Data Distributions
 - reused by Google for their Managed Hadoop Service
 - reused within Cloudera and Hortonworks
 - used by Canonicals Hadoop Offering
 - reused by the ODPI.org

Some components of Apache Bigtop



Components

- Compile Environment (based on docker)
- Convenience artifacts (i.e. repositories for Centos7, Centos6, Debian 8, Ubuntu 16.04, Ubuntu 14.04, Fedora 20, opensuse 42.1)
- Deployment Templates (puppet)
- Orchestration with Juju Charms
- Automatic Testing Environment
- And ... non intel architectures (ppc64le, aarch64)

</advertisement>

Bro



- ✦ Bro: The Network Security Monitor
- ✦ www.bro.org
- ✦ Flexible, High performance, Stateful in depth Analysis
- ✦ Analyse HTTP, HTTPS Certificate Chains, Fingerprinting of Downloads, Analyse DNS Requests and Answers

Elastic Search, Kibana

- The ELK Stack, built on Apache Lucene
- Simple NoSQL RESTful Database with a powerful Analysis Tool

Setup

Docker Container
Apache Bigtop

eth0

tcpdump -i eth0

Network
Trace

Internet

Analytic Toolchain

- ✦ github:

- ✦ dockerhub:

[danielguerra69/bro-debian-elasticsearch](#)

(pull request pending, regarding checksums)

Docker
compose

Docker Container
Bro



Docker Container
Elastic Search

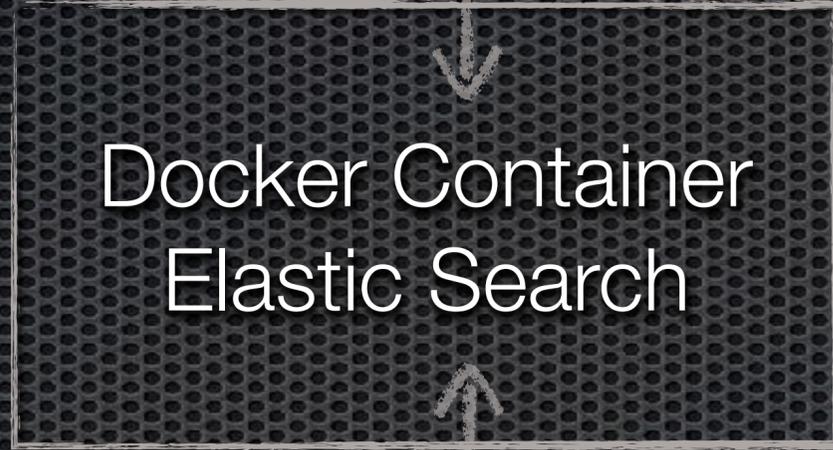
Docker Container
Index Config

Docker Container
Kibana

Docker Container
Kibana Config



1969



5601



Docker / Docker Compose

- Orchestration on a single node of
 - Bro
 - Elastic Search (Cluster)
 - Kibana
 - Index Generation in Elastic
 - Dashboard and Query generation in Kibana
- Many thanks to [danielguerra/bro-debian-elasticsearch](#) on [github/dockerhub](#)!

Workflow

- Compile in Docker container
- bigtop/slaves:trunk-debian-8
 - add tcpdump
 - tcpdump -i eth0 -s 0 -w FILE &
 - ./gradlew pkg
- See <https://cwiki.apache.org/confluence/display/BIGTOP/How+to+build+Bigtop-trunk>

Recapulate <http://> vs <https://>

https://

- ✦ Use of TLS for establish a secure channel
- ✦ Authentication of connection
- ✦ Need to check the certificate chain back to a trusted „root“ cert.
- ✦ Everything needed integrated into maven 3.3.x
(Upgrade!)

http://

- ✦ Data may be modified in between
- ✦ Data are not authenticated
- ✦ Data may be sent from a different server
- ✦ contraproductive to add <http://repo.maven.org> to `<repositories/>`!



Use of TLS Version (Sidetrack)

- Only TLS 1.2 is considered secure
 - services.gradle.org on TLS 1.1
 - Many TLS 1.1 connections

Abandoned Projects

- DNS NXDOMAIN Answer

This visualization is linked to a saved search: **NXDOMAIN**

query: Descending

nexus.codehaus.org	1
nexus.codehaus.org.local	1
snapshots.repository.codehaus.org	1
snapshots.repository.codehaus.org.local	1
maven-repository.dev.java.net	6
maven-repository.dev.java.net.local	6
maven.jamon.org	6
maven.jamon.org.local	6
repository.codehaus.org	6
repository.codehaus.org.local	6

Count

metrics



Advanced

sub-buckets

Abandoned Projects

- ✦ Code trying to download from a non resolving address
- ✦ java.net (Oracle)
- ✦ codehaus.org (Individual)
- ✦ What if a malicious guy is allocating these domains ?
 - ✦ Asking the WHOIS entry of codehaus.org for comment

WHOIS Owner of codehaus.org

Hi,

Yes it is a risk I am aware of - at this stage I'll be keeping hold of the domain names indefinitely. If that position ever changes I'll keep Apache in mind as a potential benevolent owner.

Cheers,

Ben Walding

Apache Mission Statement:

TPKDTNFY!

Shady sites

- personal home pages



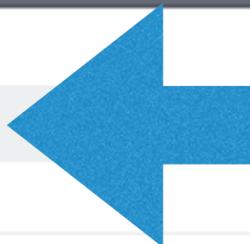
Count

Add metrics

query: Descending

add sub-buckets

people.apache.org	58
archive.cloudera.com	52
maven.twtr.com	40
apache.osuosl.org	36
www.apache.org	36
bower.herokuapp.com	26
files.couchbase.com	20
maven.glassfish.org	20
maven.java.net	18
nexus.codehaus.org	18
nexus.codehaus.org.local	18
snapshots.repository.codehaus.org	18
snapshots.repository.codehaus.org.local	18
repo2.maven.org	16
maven.restlet.com	12
maven.restlet.org	12
maven-repository.dev.java.net	10
repository.jboss.com	10
maven-repository.dev.java.net.local	6
maven.atlassian.com	6
maven.jamon.org	6



Shady sites

- ✦ HBase used people.apache.org
- ✦ Rescue: Has been cleaned up in current master, without my intervention. THANKS!

Shady resources

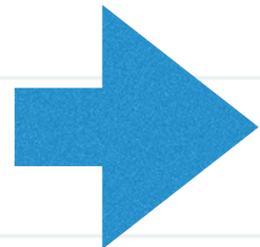
- Things not to download by a compile job. Never, ever!

01-01 2013-01-01 2014-01-01 2015-01-01 2016-01-01

ts per month



	query	host	method	uri
4th 2016, 21:36:16.246	www.pyx4me.com	-	-	-
4th 2016, 21:36:16.246	www.pyx4me.com	-	-	-
4th 2016, 21:36:17.054	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.pom
4th 2016, 21:36:17.366	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.pom.sha1
4th 2016, 21:36:17.649	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.pom.md5
4th 2016, 21:36:18.668	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.exe
4th 2016, 21:36:19.360	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.exe.sha1
4th 2016, 21:36:19.640	-	www.pyx4me.com	GET	/maven2/sysinternals/ junction/1.04/junction-1.04.exe.md5



MAVEN DOWNLOADING EXE ?



SAY IT AGAIN

Shady resources

- ✦ The „official“ Maven Junction plugin is downloading junction.exe (a copy of a non free tool from sysinternals now microsoft)
- ✦ It is supposed to create a symlink in NTFS (Windows Filesystem)
- ✦ Doing „ln -s“ on unix
- ✦ WTF ?

Company Headquarter



HTTPS to the rescue?

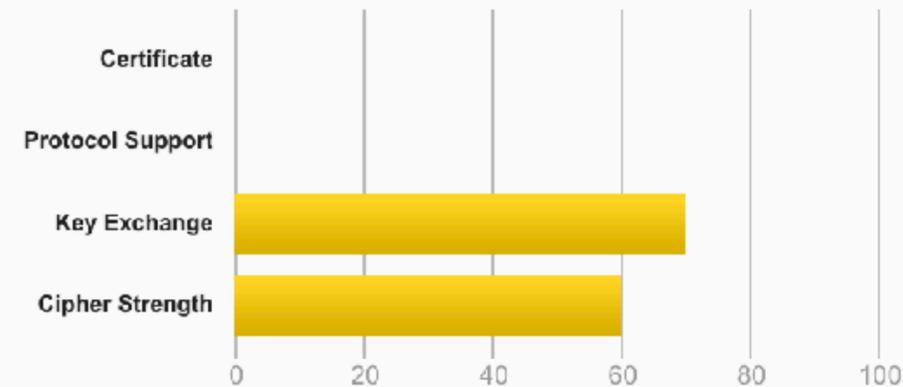
SSL Report: pyx4me.com (99.231.144.170)

Assessed on: Tue, 06 Sep 2016 19:55:19 UTC | **HIDDEN** | [Clear cache](#)

[Scan Another »](#)

Summary

Overall Rating



Visit our [documentation page](#) for more information, configuration guides, and books. Known issues are documented [here](#).

This server's certificate is not trusted, see [below](#) for details.

This server supports SSL 2, which is obsolete and insecure, and can be used against TLS (DROWN attack). Grade set to F. [MORE INFO »](#)

This server is vulnerable to the POODLE attack. If possible, disable SSL 3 to mitigate. Grade capped to C. [MORE INFO »](#)

This server supports weak Diffie-Hellman (DH) key exchange parameters. Grade capped to B. [MORE INFO »](#)

This server does not mitigate the [CRIME attack](#). Grade capped to C.

A real threat ?

- Apache Flink < 1.2
 - Contacted Flink PMC on 11th Sep
 - FLINK-4732 Adresses this issue
 - New Apache Flink release fixes this issue
 - Special thanks to the whole Apache Flink PMC!

Attacking

- Men in the middle (MITM) Attack
- Intercepting http traffic
- Demo with ettercap:
 - ARP Poisoning
 - DNS Attack
 - (SSL Forging)

Demo of Apache Flink Exploit for Windows

- Forge maven to download and run calc.exe rather junction.exe

Attack details

- ✦ Need priviledged network position (for instance in the same subnet as victim)
- ✦ Prepare webserver for offering attacking packages, configuring DNS forgery to point to attacking machine. (Disabling off SSL forgery)
- ✦ Starting ettercap, create ARP Spoofing, default router is host1 dev host2
- ✦ profit.

A statement of the authors:

Hi Olaf

The project is actually abandoned and no-longer supported.

BTW today there is a better way todo all this directly in java.

Files.createSymbolicLink(newLink, target);

Your suggestions ?

Vlad Skarzhevsky

Even „normal“ maven plugins are dangerous:

- ✦ Hacking maven-compile or plexus-compile
- ✦ For instance flume (Update: current flume is fixed and upstream to Apache Bigtop)

Fixing zookeeper

- ✦ ant/ivy based source
- ✦ Contacted via security@zookeeper.apache.org
- ✦ Fixed in ZOOKEEPER-2594
- ✦ Was using abandoned repositories and non TLS-Sources
- ✦ Special thanks to Patrick Hunt!

Trying to fix tomcat

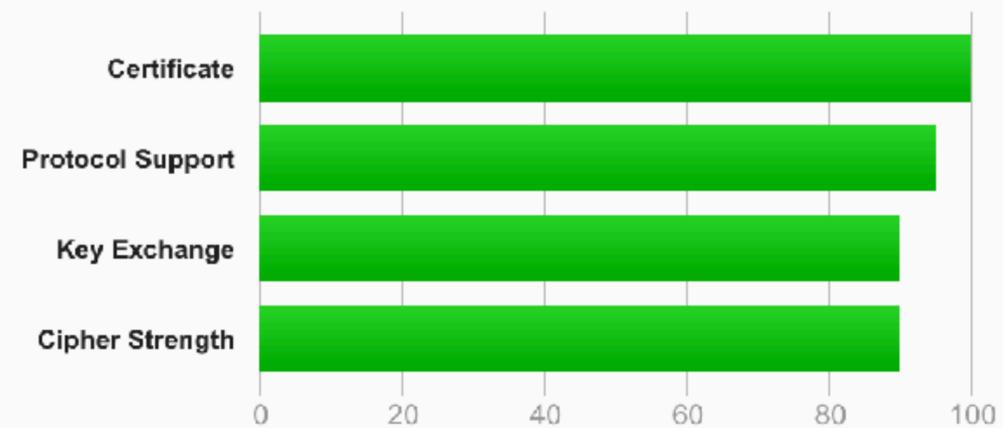
- ✦ NSIS (Windows Installer) sourceforge.net only supports non TLS downloads.
- ✦ Sidetracked: <http://www.apache.org/dyn/closer.lua>
- ✦ Only a few of the mirrors support TLS
- ✦ How to automatically prove the trust?

You are here: [Home](#) > [Projects](#) > [SSL Server Test](#) > [mirror.klaus-uwe.me](#) > 2a02:c205:0:5156:0:0:0:1

SSL Report: [mirror.klaus-uwe.me](#) (2a02:c205:0:5156:0:0:0:1)

Summary

Overall Rating



Visit our [documentation page](#) for more information, configuration guides, and books. Known issues are documented [here](#).

HTTP Strict Transport Security (HSTS) with long duration deployed on this server. [MORE INFO »](#)

Do No Trust ... Verify

- ✦ Either bundle GPG Keys or Checksums.
- ✦ TBD

Best practices for Devs

- Migrate at least to Maven 3.3.x
 - It uses and validates TLS Connections!
 - TLS Connection to repo.maven.org is built in
- Check KEYS or Checksums with official website

Best practices for maven

- Look out for `<repository>` tags in pom.xml
 - Download only from trusted sites
 - Use https://
 - Do not enable snapshot repositories
- If you need snapshot features:
 - Use maven profiles and enable `<repositories>` in `<profile>` section

Best practices for ivy

- Same as maven
- Use `https://` repositories.

Best practices for downloading from apache.org

- ✦ INFRA does not like/guarantee downloads from apache.org
For instance <https://www-us.apache.org>
- ✦ Validate with checksums (for instance sha1) within source
- ✦ Or validate GPG Keys supplied with source
- ✦ But that's tough ..

Unsolved Problems

- ✦ Who is security@ for maven plugins at maven central ?
(for instance maven junction)
- ✦ How do we transport trust for artifacts at dist.a.o /
archive.a.o ?
 - ✦ IMHO keys of individual dev's are suboptimal
 - ✦ Maybe reuse maven repo ?

Questions?

- ✦ Contact me at
- ✦ of ät oflebbe.de