Statement Of Work for Software Composition Analysis

SUMMARY

Statement Of Work proposal for the plagiarism verification, license compliance certification and warranty for a software project. The document details the roadmap to verify and generate a detailed inventory for the third-party software components used inside a software project.

Agenda

Introduction

Software Composition Analysis is the activity for listing in detail what is used inside a software product. This includes listing the respective copyright holders and applicable legal terms, which can be either open source or under proprietary license terms.

On behalf of the customer, TripleCheck will proceed to:

1) Fully identify the third-party software components present on the product
2) Audit the license compatibility for the identified third-party components/snippets
3) Prioritize and solve license conflicts (if any) with the customer development team
4) Properly mark the source code that is IPR of the customer and the IPR of third parties
5) Prepare the software as a package for distribution to other parties
6) Provide an official compliance certificate by TripleCheck, as independent auditor
7) Issue a warranty for this product, valid for two years

About TripleCheck™

TripleCheck focuses on software originality (plagiarism) and licensing quality (compliance). The company was founded in Darmstadt, Germany in 2013. This company has been trusted over the years to verify the software products from customers on the telecom, aerospace, and transportation industries. TripleCheck is the only provider of technology for deep forensic identification of third-party items that is fully based in Europe, ranked by the NCC Group (UK) as one of the top 3 technologies for Software Composition Analysis[1].

Technology

The TripleCheck technology was first released to public in 2013 and is continuously improved with the feedback from customers. The main product is called Xray, which is currently on its second generation of releases.

TripleCheck as a forensic auditor, performs the following activities:

- Collecting an offline archive of publicly available software
- Detecting license terms on the software surface
- Detecting software plagiarism
- Generating software inventory lists (also known as Bill Of Materials)

[1] https://www.nccgroup.trust/globalassets/our-research/uk/whitepapers/2016/08/research-insights-volume-9-modern-security-vulnerabilty-discovery.pdf

The offline archive spans to about 2.5 petabytes of software collected from diverse public sources on the Internet over the years, namely github, sourceforge, bitbucket, stackoverflow, among others. This data is archived inside its own premises, from which are built the fingerprint databases that are later used for the offline matching of source code files and snippets.

In 2018, the TripleCheck Open Source archive amounts to 1,9 billion file fingerprints and 1,2 billion code snippets (methods, functions) from 55 programming languages. Technology-wise, only TripleCheck has the forensic capacity for identifying similar binary files in large scale, in addition to the exact fingerprint matching based on SHA1.

Audit team

For each customer is allocated a team that is adjusted to the specific context of the customer. This audit team works with the customer to complete the software verification tasks and consists of two critical members: a technical auditor (senior engineer) and a legal advisor (lawyer). The team is responsible for verifying and correcting the output from the tools, to investigate the origin of third-party items identified by the plagiarism database and to deliver accurate reports to the customer in regards to licensing compliance.

For software verifications on high-security environments (e.g. critical infrastructure), only EU/UK nationals are included on the audit team.

Security and privacy

TripleCheck has experience working on environments where the customer infrastructure requires a high standard for security and data privacy. TripleCheck includes the following characteristics to its service:

- **100% offline tooling and databases**. No network access required nor used for the third-party software identification
- **Auditable tools**. When necessary and by request of the customer, we provide the complete source code of the tools used within the customer secure environment
- **On-premise verifications**. When desired, we conduct the activities inside the customer premises, using the customer provided hardware.

Our staff is trained and experienced on the handling of sensitive information. TripleCheck will make the necessary adjustments in order to comply strictly with the customer security and privacy requirements.

Customer privacy

It is recommended that a Non-Disclosure Agreement is signed between both parties prior to any access to the software from the customer audit occurs. TripleCheck is available to promptly sign an NDA provided by the customer prior to starting the audit. In alternative, TripleCheck will provide a default NDA document to assure privacy in regards to the audit and respective deliverables.

The default approach to an audit starts with TripleCheck collecting the software from the customer through a hardware encrypted drive (provided and shipped by our company) or through any other mean preferred by the customer. The audit is performed on our infrastructure. In this scenario, the infrastructure is physically disconnected from public networks and the customer software is deleted from the TripleCheck equipment after the audit takes place.

When requested, TripleCheck will conduct the audit from the customer office. In such scenario it would be necessary a desk and computer for our expert to perform the audit, with an extra cost to account for travel expenses. This option assures our customers that the software being audited remains within his premises.

Data needed for price quote

The default data used for price quotes are the lines of code.

On top of this estimation may be added other factors such as license complexity, urgency of delivering results, or extra effort required for the audit. The detailed pricing list is provided on the next section. For calculating the LOC (lines of code) and code complexity, TripleCheck provides an offline tool that can be downloaded from:

> http://triplecheck.tech/estimator

This tool does not modify any files on your disk, nor uses any network connection: it runs offline. The usage instructions are provided on the webpage, it will generate a single line of text that contains the information that is necessary for a complete quote.

The text signature is a short snippet that will look similar to:

```
1v0i1-iz-24fj-bc-7m
```

For the sake of transparency, each value within “-” is respective to:

1) total file size
2) number of files
3) number of LOC
4) number of license references with higher risk (type A, e.g. copyleft licenses)
5) number of license references with lower risk (type B, e.g. permissive licenses)

Each value represents a number. This number is represented on clear text inside the tool log and then converted to a different representation so that its content is kept minimally private while transiting through email.

On this example, the value "`iz`" represents the number of `683` files. This can be verified through an online tool such as `https://www.dcode.fr/base-n-convert` where the radix number should be set as 36, then typing "`iz`" on the "numbers to convert" box and finally clicking on the "convert numbers" button. Using the same method for `24fj` we obtain `99055` LOC and so forth.

When possible, TripleCheck recommends that the software to be verified is placed inside a single folder, where the tool can then generate a text signature for that folder on a single run.

Understanding that the software to be verified from the customer might be large and that combining the folders might be a difficult task, it is possible to simply provide a list with the text signatures and then TripleCheck will combine them together.

Once the data is gathered, an exact quote can be provided, using the pricing list detailed on the next section.

Pricing list

Price for software verification

Prices on this section are following the values from the TripleCheck 2018 pricing list. Using the LOC measure provided by the customer, each block of software up to **500 thousand LOC to be verified has a standard fixed price of 6800 EUR** (VAT not included).

Based on data from previous audits, in average each block of 500 thousand LOC will contain up to 200 different third-party libraries and up to 170 000 third-party code matches that will need to be processed by the auditors.

For each block of code, the total verification time is estimated between **2 to 6 calendar weeks** depending on the audit conditions. For example: if working remotely or on-premises, availability of Internet for auditors to research information from lesser known libraries found, and then the status of existing licensing quality of customer code (e.g. properly identified as authored or third-party). The TripleCheck Xray **tooling requires between 1 day to one week** for completing the plagiarism scan for each block, depending on the hardware used for performing the scan (customer provided or TripleCheck internal infrastructure).

When necessary, TripleCheck will suggest the necessary adjustments to reduce the necessary plagiarism scan time. For example, specific details to avoid the deep scanning of well-known and large third-party software items such as Wordpress, Linux and similar, since these are identifiable without need for a deep scan.

Price for the tooling

On the cases when the **customer software cannot leave the customer premises** and that several scans will need to be performed across the year, the TripleCheck tooling will need to be installed and available on the customer side. The license for installing and using the **tooling on-premises has a standard fixed price of 28 200 EUR per year** (VAT not included).

This tooling license has no limitation in terms of files, CPUs or products to be scanned. It is a fixed price with full support and updates, without added costs. The license limits the tool usage and distribution to the customer premises (e.g. excluding contractors and customers). The output from the tool can be redistributed and exists the possibility of purchasing a license renewal for additional years when desired by the customer.

Travel costs

Audit teams working from customer premises in Germany have a fixed price of 193 EUR per day, or 341 EUR per day when working from premises within the EU and UK (VAT not included). Prices are per team, not per person and these already include all travel costs.

Travel costs are only performed with permission from the customer. TripleCheck recommends that certain activities such as the provenance investigation of missing copyright or license terms is performed off-site from the TripleCheck premises, possibly other parts of the activity will be performed from Darmstadt, if and when permitted by the customer.

Other points

Once the metrics become available, TripleCheck proposes a time line and final price quote for appreciation by the customer prior to place an order. For each verified block of code, the customer receives a two-year warranty that permits the customer development teams to run routine verifications after the initial audit, assuring that non-contamination of customer product is preserved as the code gets further developed.

By request of the customer, the price quotes can be provided as GBP or USD using the current exchange rate.

The quote is valid for two months after being issued. Payment is due on net-45 payment period, counted from the day when the order is placed to TripleCheck by the customer.

Summary of the pricing list

Item	Price
Software audit (warranty included)	6800 EUR per 500k LOC
TripleCheck Xray	28 200 EUR/year
On-premises audit	DE: 193 EUR/day. EU/UK: 341 EUR/day

Initial deliverables

On its first stage, TripleCheck creates a detailed inventory of the third-party items.

During this first stage verification are often found items that need correction (e.g. item with a problematic license). When generating the inventory and discovering such item, TripleCheck will inform the customer as early as possible about the item and propose a solution to address the issue. This permits the customer to have an early warning and enough time for deciding on the course of action.

The final output from the first stage a set of reports that detail each of the third-party items inside the verified products. These reports are made available in HTML, Excel and text formats.

If the verified software is compliant with the software reuse list at this stage, TripleCheck produces an official statement (certificate) to prove the compliance.

When the software has items that need further work, this is called as second stage where TripleCheck is supporting the development teams with licensing advice (included on the warranty) and re-evaluating the output from newer scans. Products reaching this stage are typically compliant after the development team completes the corrective changes, which tend to take between 2 to 4 calendar-months, the official statement is then produced by TripleCheck.

The third-stage is maintenance, meaning the schedule of periodic verifications after each 6 or 12 months to assure that non-contamination is preserved. This is included with the two-year warranty, which can be extended by the customer.

Items that need correction

The customer typically applies corrective changes on the items considered as problematic, or in this case that are not compliant with the software reuse list. The ideal end result is when TripleCheck can consider the licensing compliance level as satisfactory and produce a certificate.

Still, code changes may be necessary, for example, due to incompatible licensing terms such as strong copyleft, commercial licenses that need to be acquired, icons that need to be changed or because of third-party software that has uncertain provenance.

It might also happen that it won’t be possible to address all items due to budget and effort considerations on the customer side. The customer decides which changes should take place and will apply them according to his plan.

Documenting the source code

During the verification, TripleCheck will be marking in detail which code belongs to the customer, and which code belongs to other parties (e.g. OSS libraries, proprietary code). For that purpose, TripleCheck creates a special folder called “inventory” where the necessary legal documentation is stored. Details about what is generated and stored on this folder can be found on the following location:

http://triplecheck.tech/inventory/

This folder is what permits customers to prove their compliance with the third-party license requirements for their products. An example of the inventory is provided with the document.

Certifying the audited software

In the quality of independent auditor, TripleCheck will review the licensing compliance status of the software when the changes have been applied and exists a pre-indication that the software gathers the sufficient conditions to be considered as compliant, according to the industry standards as monitored by TripleCheck.

A certificate of compliance is issued up to one year after the audit has started. This document details the relevant aspects to be considered in regards to the license terms of the adopted third-party software, along with an expert opinion that justifying their usage as correct, based on the experience of the auditor.

The external audit and respective certificate are a proof that the customer has engaged in rigorous care to follow the respective license requirements of the third-party software. However, this certificate should not be understood as a legal shield, nor the opinion of TripleCheck should be understood as legal advice. TripleCheck is the technical expert that is focused on identifying with specialised tools the possible items of plagiarism and copyright infringement present inside the customer software. When engaging with TripleCheck, the customer agrees that the auditor cannot be held liable for legal consequences regarding the IPR contained inside the audited software. TripleCheck recommends that the customer includes his legal team for advice on the output of this audit as early as possible.

Warranty scope

TripleCheck is the only provider including a two-year warranty with each software verification.

This warranty protects the audited software up to two years, it is applicable up to 30% of modifications from the original code to permit customers to continue the product development.

The customer will receive at no extra charge the technical support for the next two years, even as new third-party software items continue to be added by the customer development team.

The following features are included:

1) Expert email/phone support to clarify open source conflicts/questions (reasonable use)
2) Routine scans at periodic time intervals (typically 6 months) to verify compliance
3) Warranty remains valid until a maximum of 30% changes on the source code files

Conclusion

After the audit is concluded, the following items are typically delivered to the customer:

a) Excel file listing of the third-party software and expert evaluation on their compliance
b) HTML report for distribution to other parties, including the compliance quality score
c) Compliance Warranty and Certificate valid for two years
d) Email/Phone support with a dedicated expert to answer questions about s/w licenses
e) (optional) Zip file of the audited software ready for distribution to other parties

For any further questions, please refer to Mr. Brito, Managing Director of TripleCheck GmbH that will be available to assist with any detail of the requirements or any other topic related to software composition analysis.

Direct contact:

email: brito@triplecheck.tech
phone: +49 162 327 2929

With best regards,

Your TripleCheck team