Technical Leverage in The Python Ecosystem: Lessons Learned

Published in the Journal of Empirical Software Engineering (EMSE)
DOI: 10.1007/s10664-023-10355-2

Ranindya Paramitha · Fabio Massacci

UNIVERSITÀ DI TRENTO
Department of
Information Engineering and Computer Science

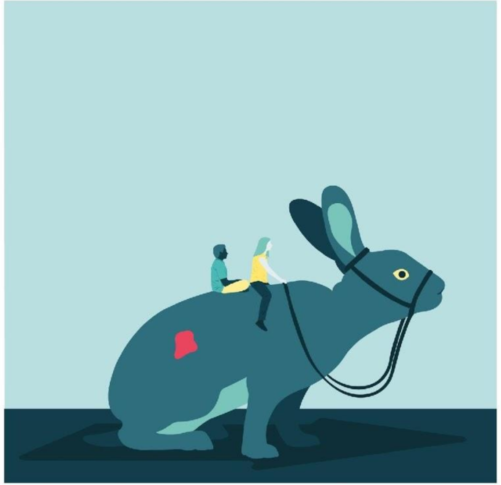
VU VRIJE UNIVERSITEIT AMSTERDAM

AssureMOSS

SEC4AI4SEC

SFSCon 2023 · Bolzano, November 10-11, 2023
The South Tyrol Free Software Conference 2023

Software, then and now



Software, what we think it is

Software, then and now

Software, what we think it is

how it really is

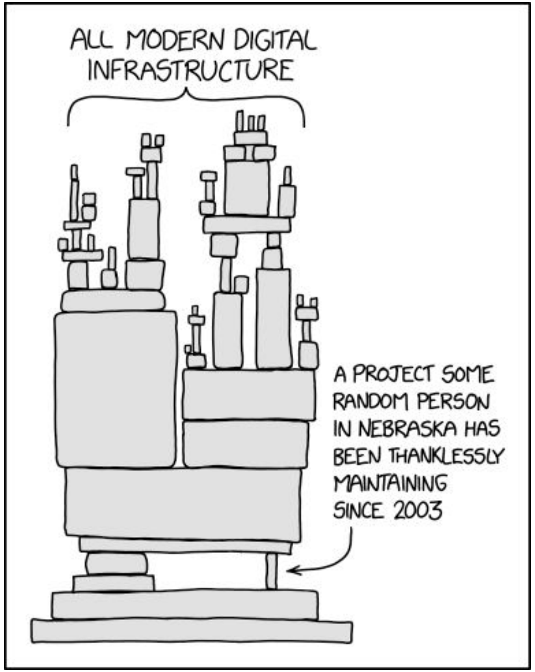
and the risks it brings

© Anna Formilan / annaformilan.com

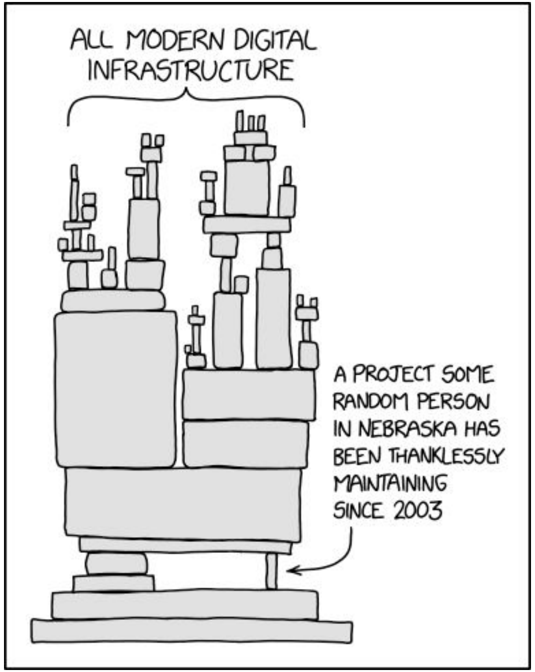
Nowadays:

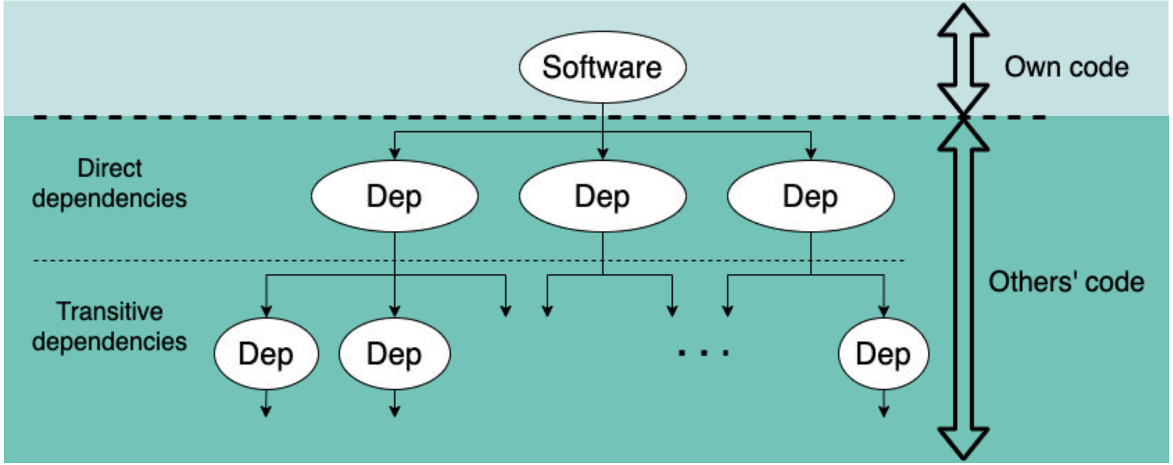
- Developers use FOSS (Free Open source software) as building blocks
- **Fraction of homegrown code as low as 5%** for industry software (Source SAP)
- **Dependency code = 4x own code** as industry average (Source BlackDuck)

Technical leverage: learning from finance, now

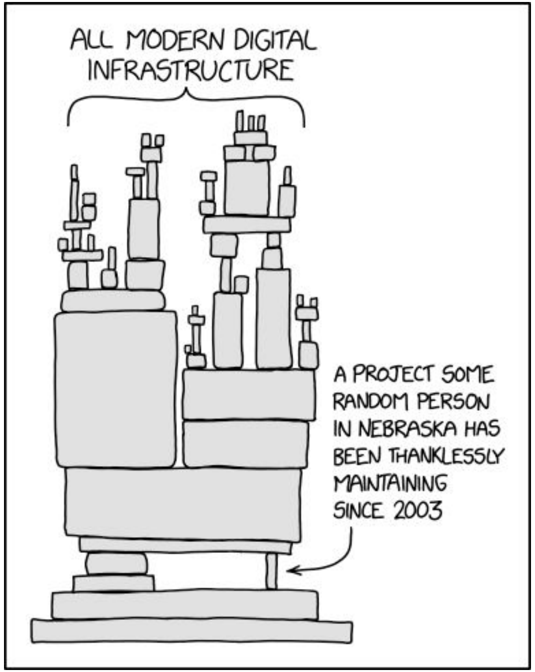

ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

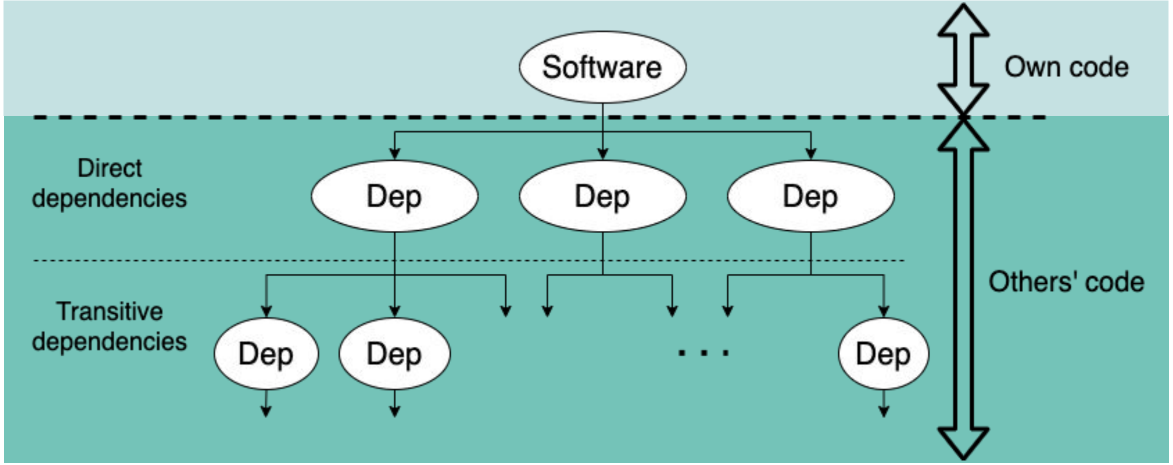
Technical leverage: learning from finance, now


ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003



Technical leverage: learning from finance, now

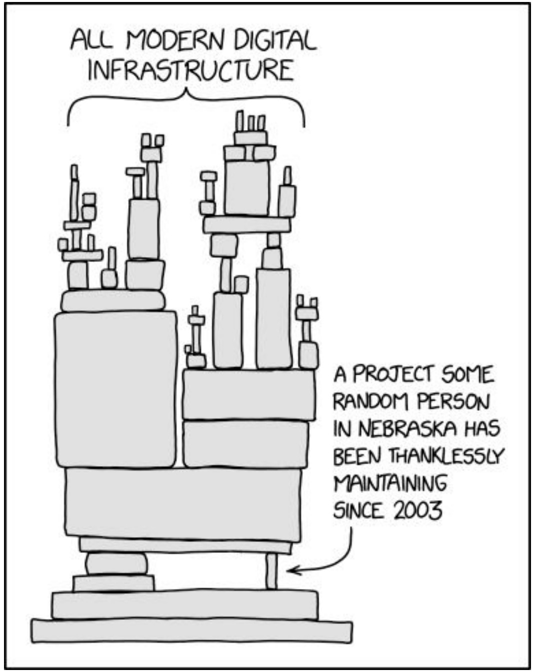

ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

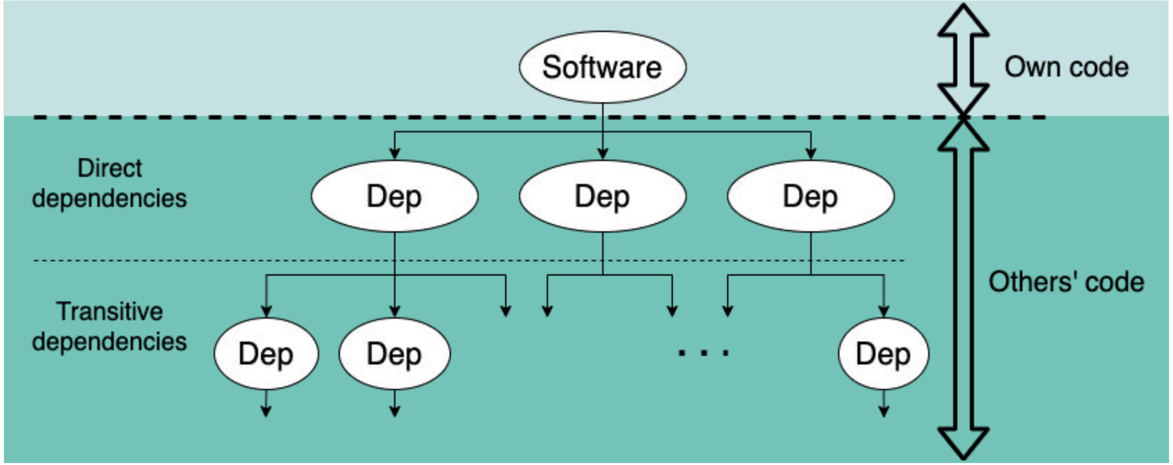


Finance

$$Leverage = \frac{Debt}{Equity}$$

Technical leverage: learning from finance, now


ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003



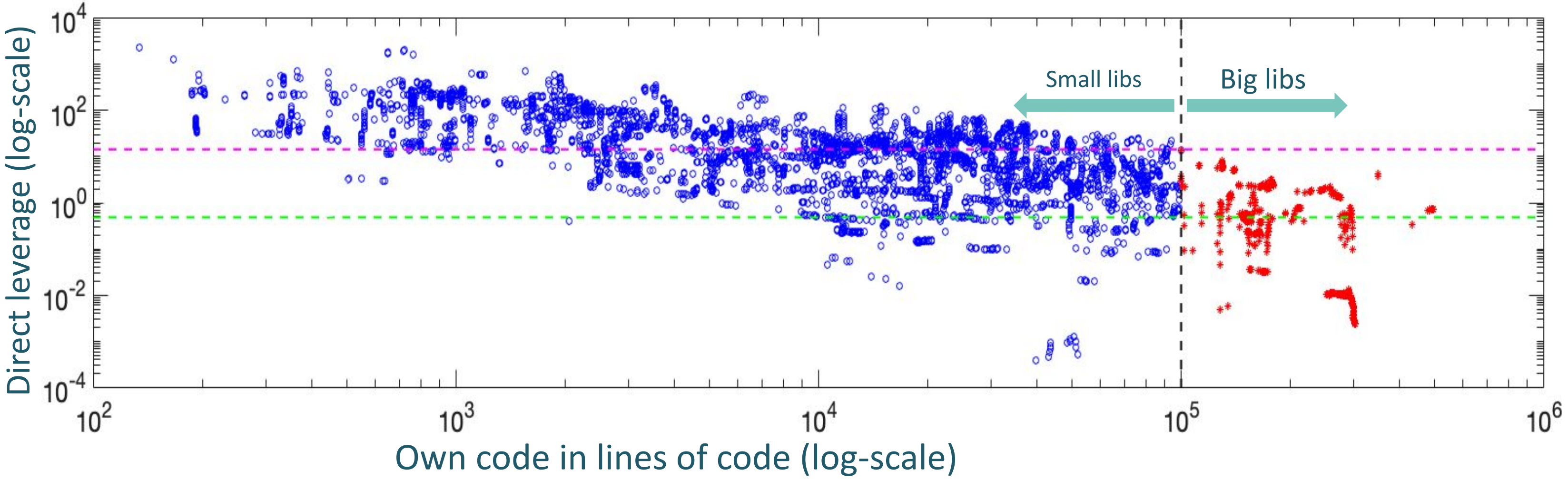
Finance

$$Leverage = \frac{Debt}{Equity}$$

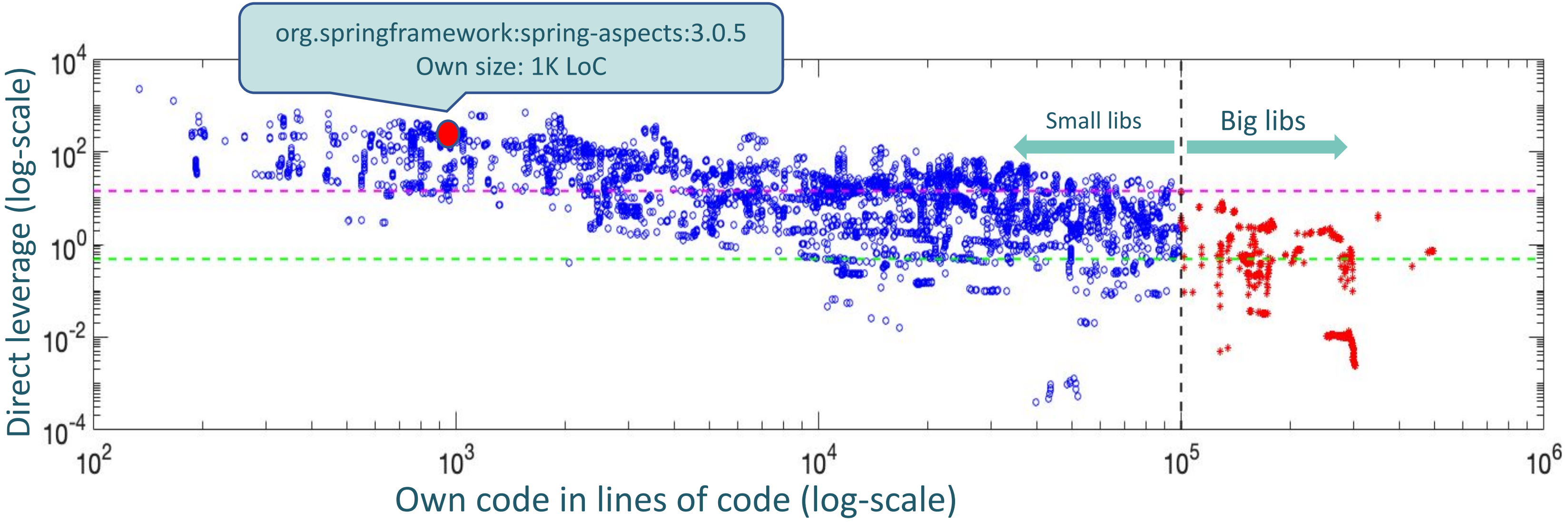
Software

$$Tech\ leverage = \frac{Others'\ code}{Own\ code}$$

Previous work in Java (Massacci & Pashchenko, ICSE’21)



Previous work in Java (Massacci & Pashchenko, ICSE’21)

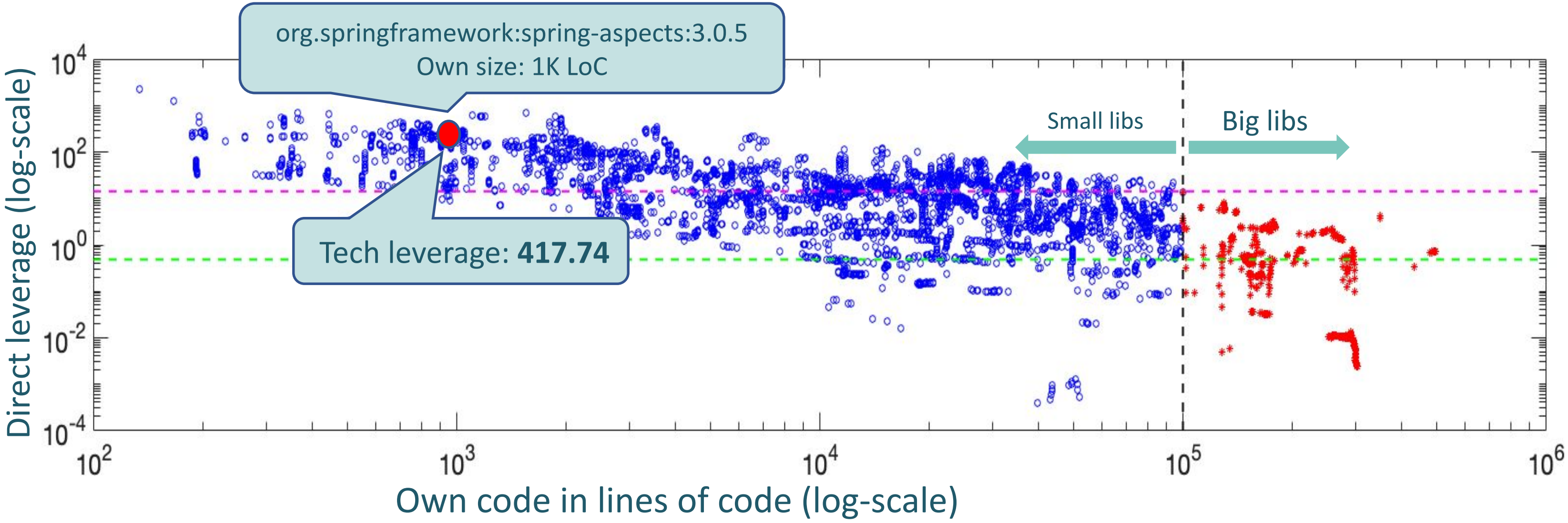


org.springframework:spring-aspects:3.0.5
Own size: 1K LoC

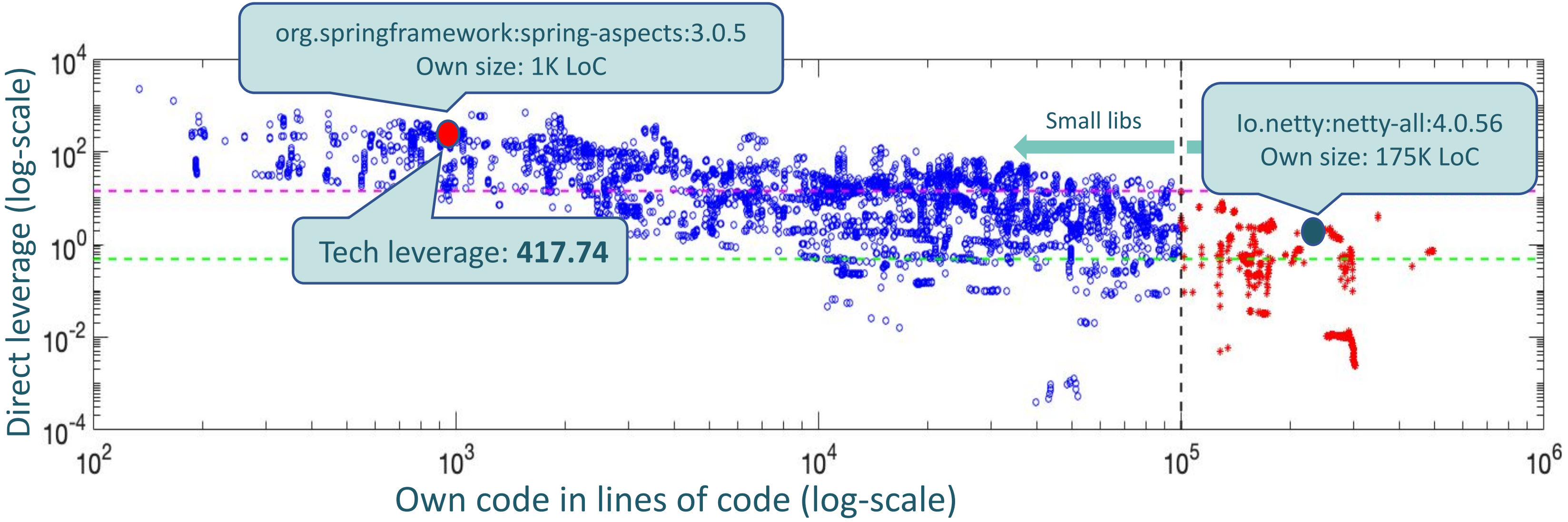
Small libs

Big libs

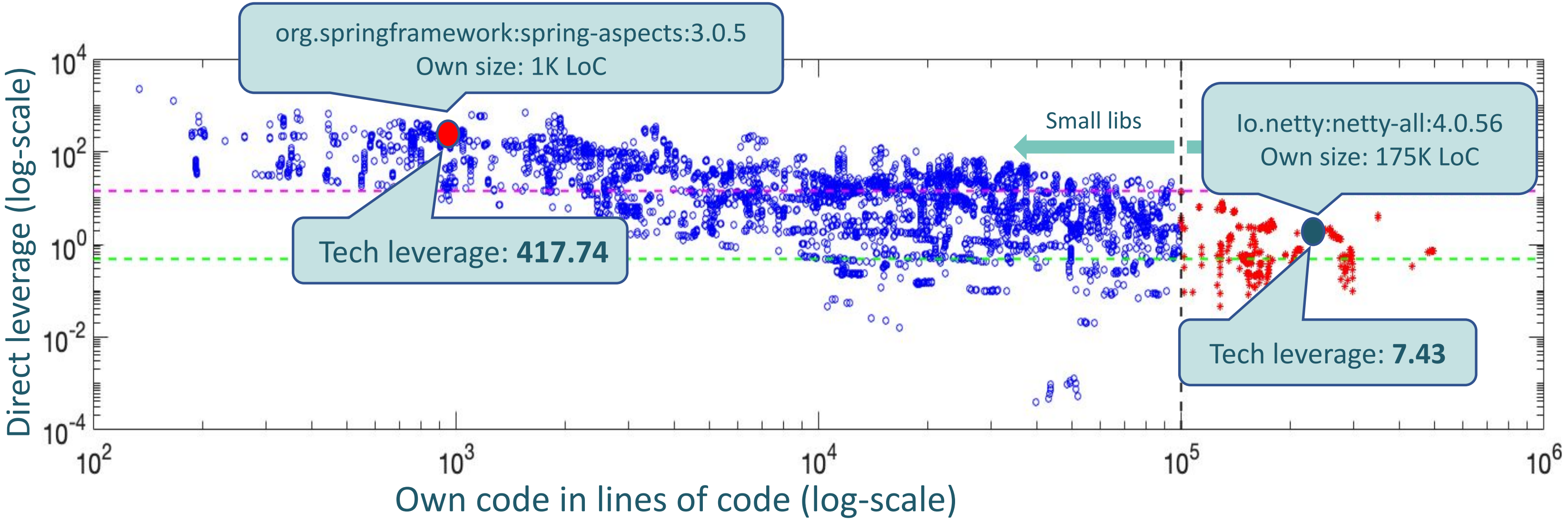
Previous work in Java (Massacci & Pashchenko, ICSE'21)



Previous work in Java (Massacci & Pashchenko, ICSE’21)



Previous work in Java (Massacci & Pashchenko, ICSE’21)



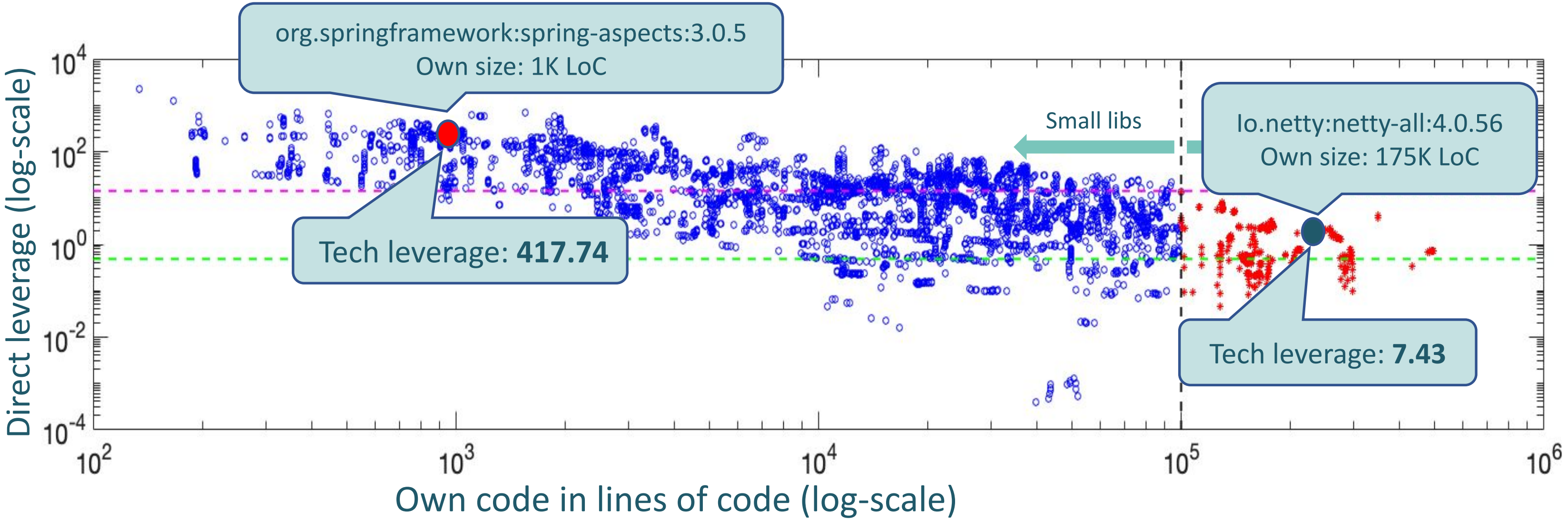
org.springframework:spring-aspects:3.0.5
Own size: 1K LoC

Tech leverage: **417.74**

Io.netty:netty-all:4.0.56
Own size: 175K LoC

Tech leverage: **7.43**

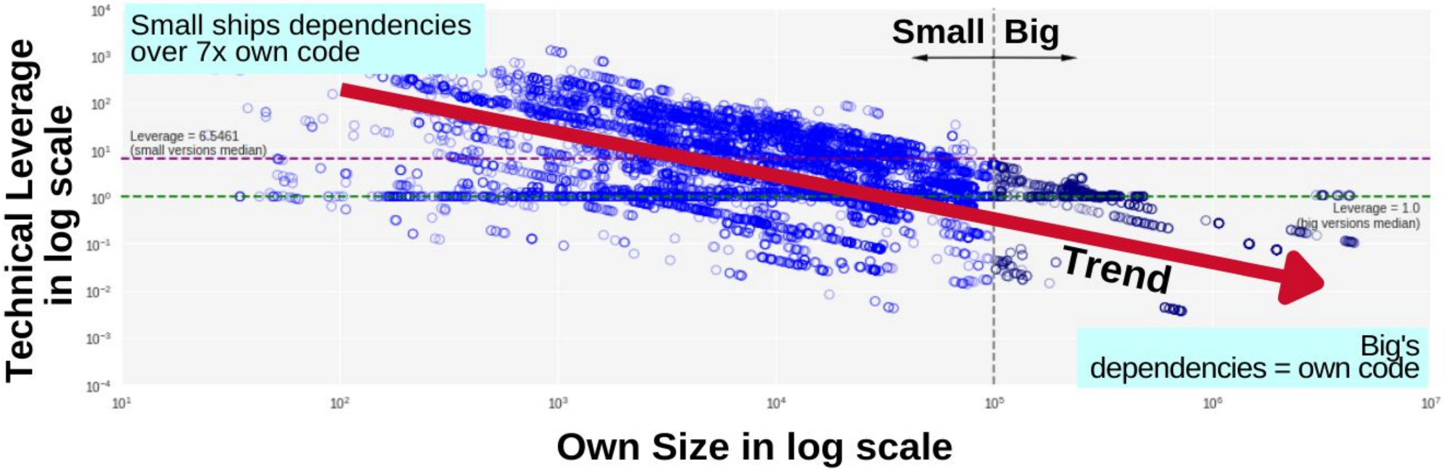
Previous work in Java (Massacci & Pashchenko, ICSE'21)



How about in another ecosystem?

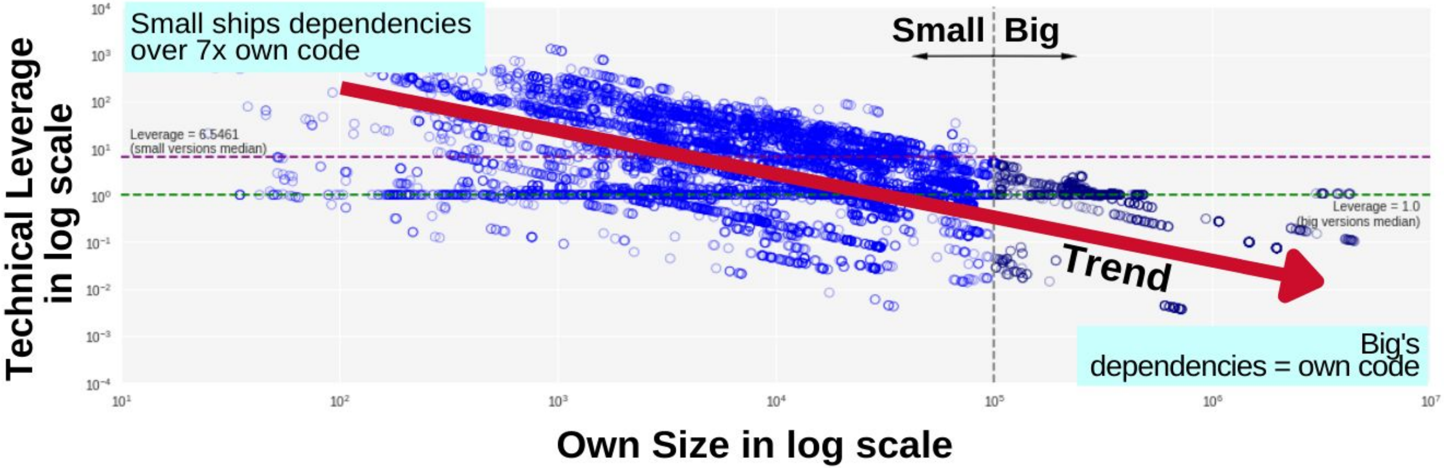
How is the evolution of technical leverage?

RQ1: How is technical leverage distribution in the Python ecosystem?



As in Java, Python developers also tend to ship a lot of other people's code (RQ1)

RQ1: How is technical leverage distribution in the Python ecosystem?

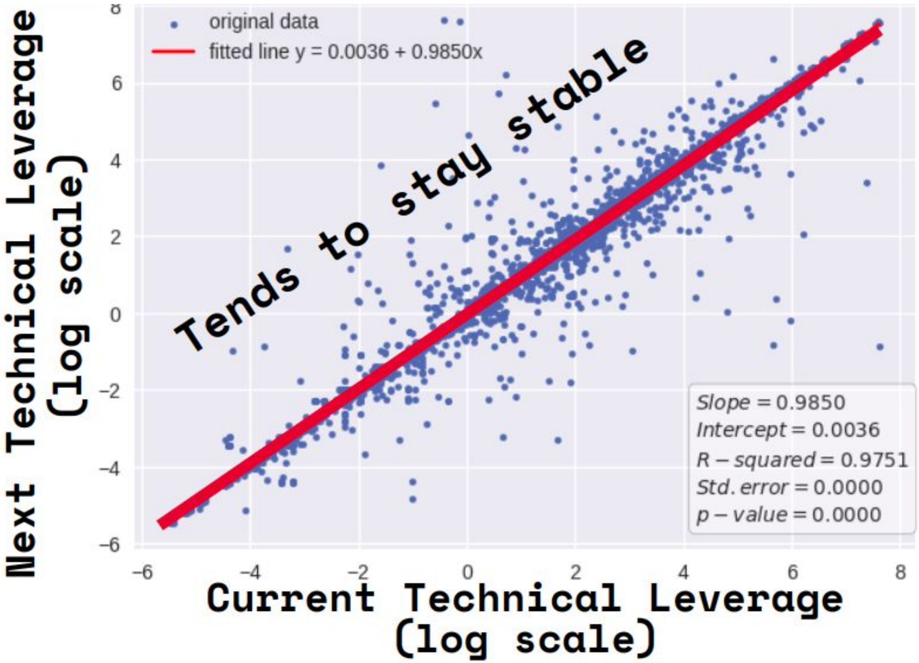


As in Java, Python developers also tend to ship a lot of other people's code (RQ1)

LESSON LEARNED

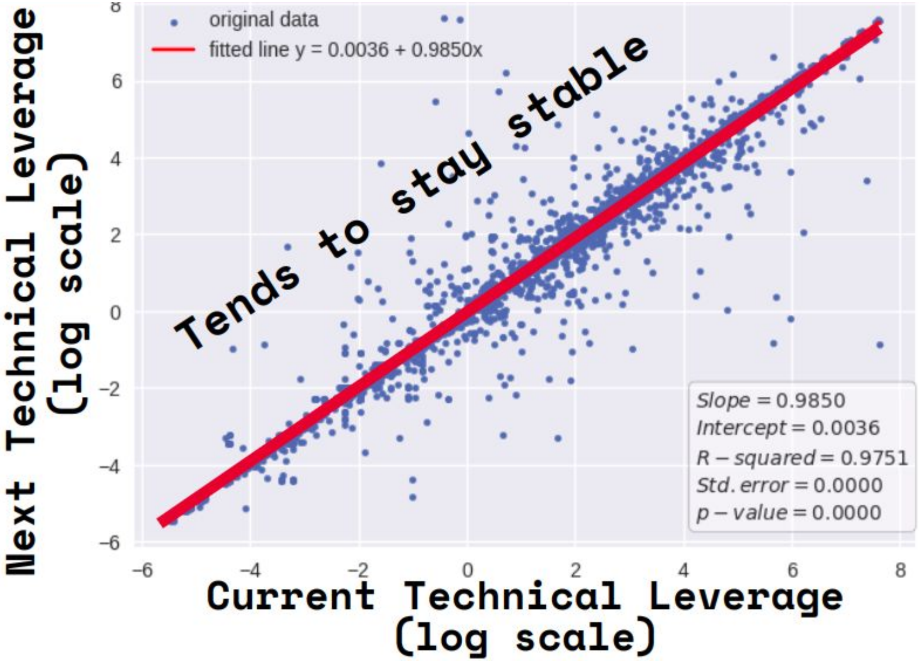
Understanding what else will come along when a package is adopted is *key* to assessing its level of uncontrollable risk.

RQ2: How does the technical leverage metric change when we move to newer versions in a package?



If you are highly leveraged, you will stay so.

RQ2: How does the technical leverage metric change when we move to newer versions in a package?

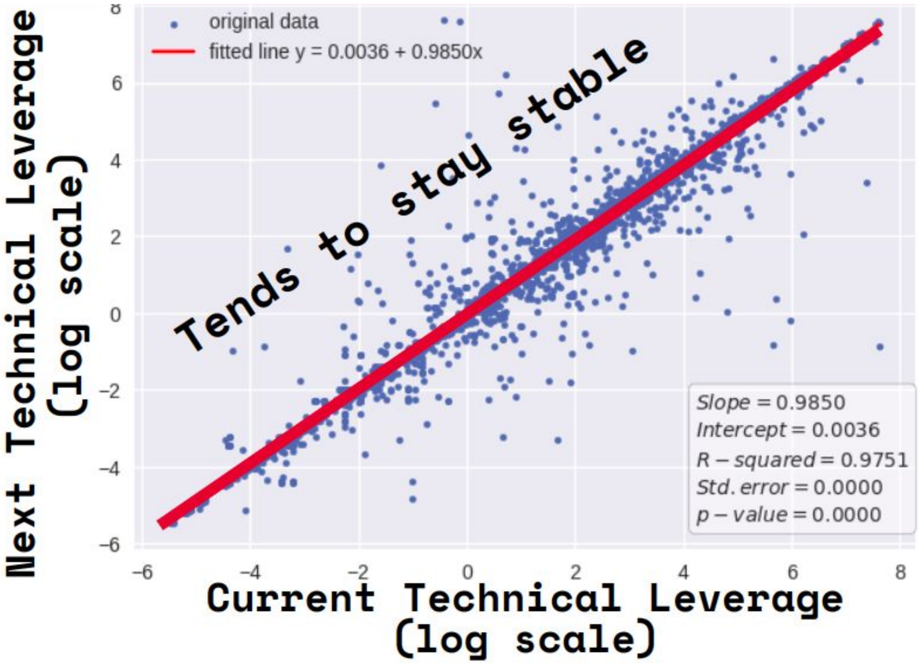


If you are highly leveraged, you will stay so.

LESSON LEARNED

Once a package has been adopted, the level of uncontrollable risk is unlikely to change over time

RQ2: How does the technical leverage metric change when we move to newer versions in a package?



If you are highly leveraged, you will stay so.

LESSON LEARNED

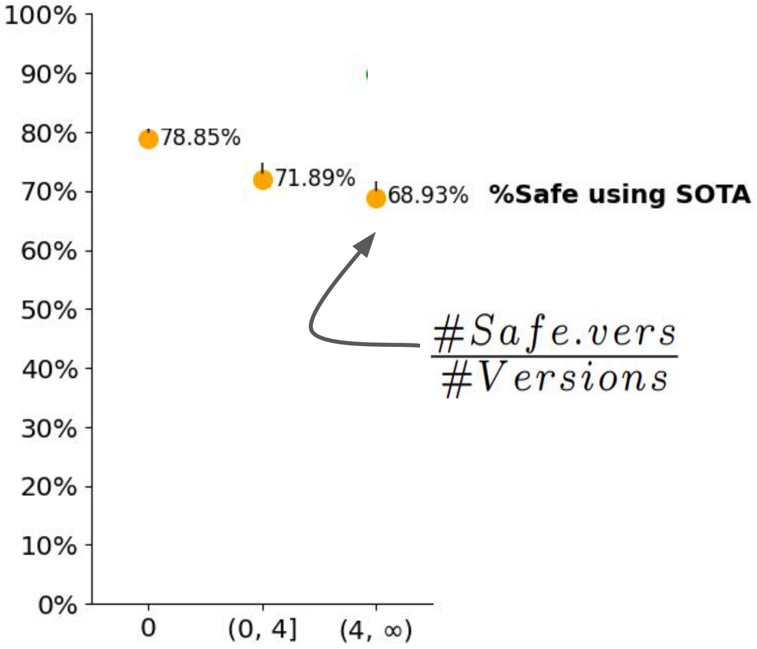
Once a package has been adopted, the level of uncontrollable risk is unlikely to change over time

Good News

If the risk was consciously accepted

Otherwise → ***Bad News***

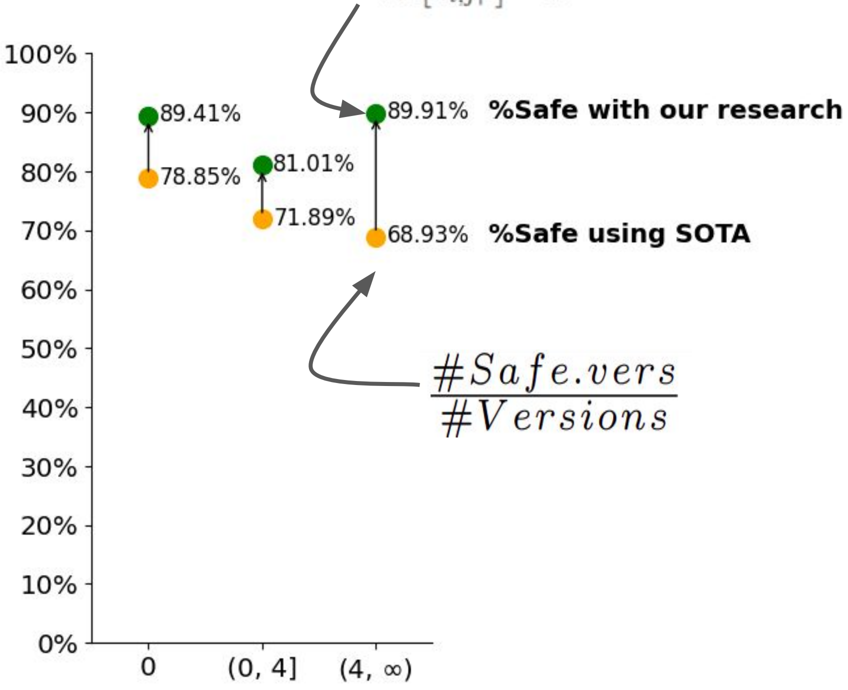
RQ3: Does technical leverage capture the risk of having vulns?



$$\frac{\#Safe.vers}{\#Versions}$$

RQ3: Does technical leverage capture the risk of having vulns?

$$Pr[safe\ vers|\lambda] = \frac{1}{N_{\Lambda=\lambda}} \sum_{\substack{\text{package } i \\ \wedge \mathbb{E}[\lambda_{i,j}|i] = \lambda}} \left(1 - \frac{v_i}{n_i}\right) \qquad (7)$$

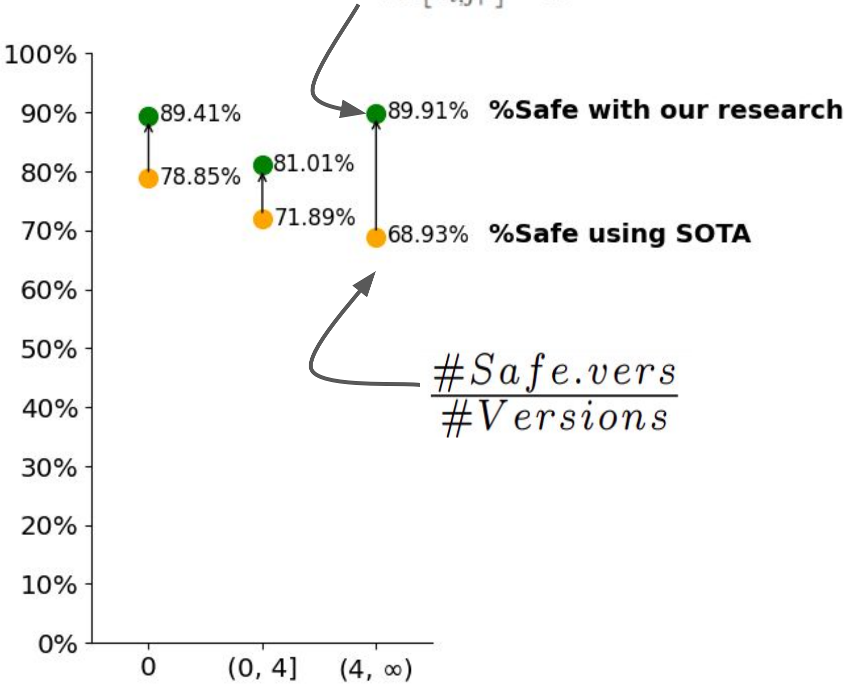


$$\frac{\#Safe.vers}{\#Versions}$$

The probability of choosing a safe package version is higher than the proportion in the ecosystem.

RQ3: Does technical leverage capture the risk of having vulns?

$$Pr[safe\ vers|\lambda] = \frac{1}{N_{\Lambda=\lambda}} \sum_{\substack{\text{package } i \\ \wedge \mathbb{E}[\lambda_{i,j}|i] = \lambda}} \left(1 - \frac{v_i}{n_i}\right) \qquad (7)$$



$$\frac{\#Safe.vers}{\#Versions}$$

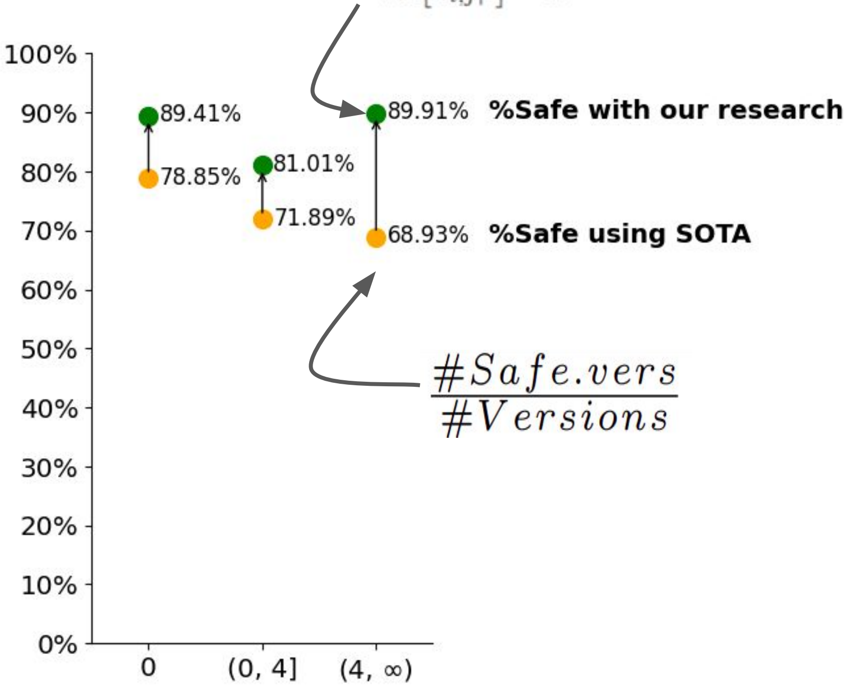
The probability of choosing a safe package version is higher than the proportion in the ecosystem.

LESSON LEARNED

Making statistics based on versions is good for claiming to have done a 'large case study' but is not representative of the reality on the field.

RQ3: Does technical leverage capture the risk of having vulns?

$$Pr[safe\ vers|\lambda] = \frac{1}{N_{\Lambda=\lambda}} \sum_{\substack{\text{package } i \\ \wedge \mathbb{E}[\lambda_{i,j}|i] = \lambda}} \left(1 - \frac{v_i}{n_i}\right) \qquad (7)$$



$$\frac{\#Safe.vers}{\#Versions}$$

The probability of choosing a safe package version is higher than the proportion in the ecosystem.

LESSON LEARNED

Making statistics based on versions is good for claiming to have done a ‘large case study’ but is not representative of the reality on the field.

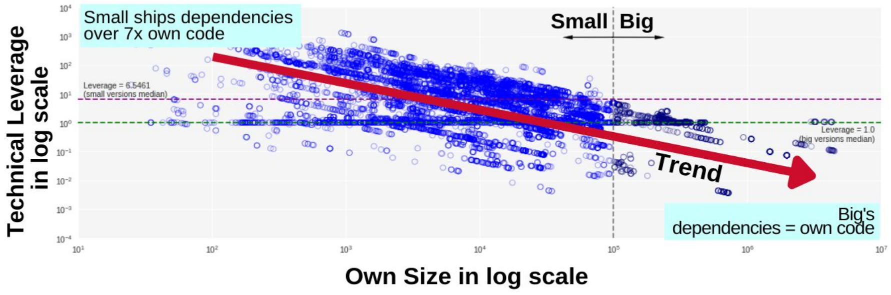
Good News

Life can be better than researchers depict it.

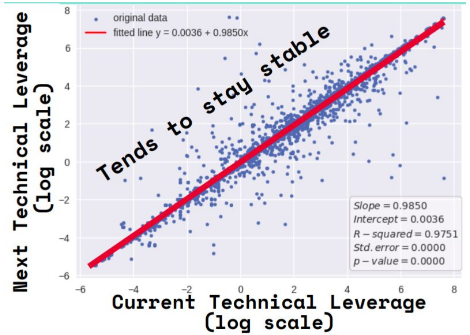
Technical Leverage in The Python Ecosystem: Lessons Learned

Ranindya Paramitha · *Fabio Massacci*

RQ1: How is technical leverage distribution in the Python ecosystem?

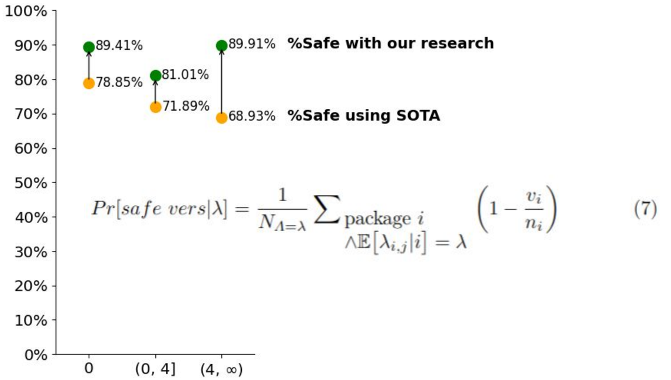


RQ2: How does the technical leverage metric change when we move to newer versions in a package?



MY WEB

RQ3: Does technical leverage capture the risk of having vulns?



$$Pr[safe\ vers|\lambda] = \frac{1}{N_{\Lambda=\lambda}} \sum_{\substack{\text{package } i\\ \wedge\mathbb{E}[\lambda_{i,j}|i] = \lambda}} \left(1 - \frac{v_i}{n_i}\right) \qquad (7)$$

PARTICIPATE IN OUR EXPERIMENT!

“Do you think it is an Ethical Decision?”

Some questions on an example from the ACM Code of Ethics and professional conduct

- Your participation is **voluntary**.
- The QR code will redirect you to an online survey.
- The **aim** of the study is to find out how people think about **ethical decisions in security.**
- The data is collected **anonymously** and will be used for research purposes.
- It will take you approximately **5-10 minutes**