Attributing Open-Source Contributions is Critical but Difficult: A Systematic Analysis of GitHub Practices and Their Impact on Software Supply Chain Security

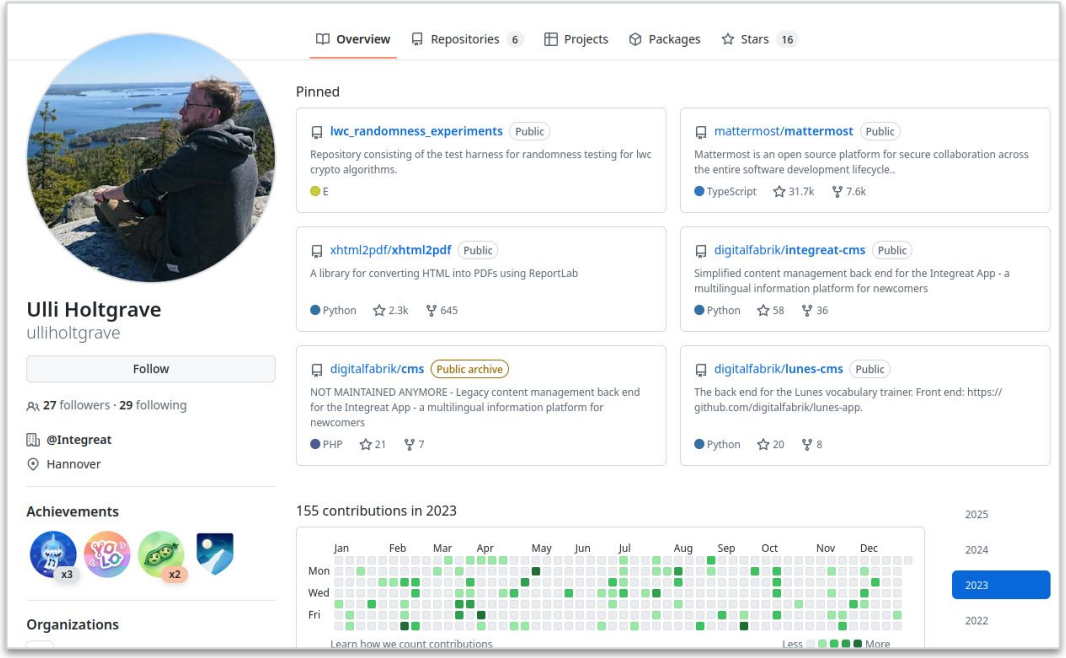
Jan-Ulrich Holtgrave*, Kay Friedrich*, Fabian Fischer*, Nicolas Huaman[⊥], Niklas Busch*, Jan H. Klemmer*, Marcel Fourné[‡], Oliver Wiese*, Dominik Wermke[§], Sascha Fahl*

*CISPA Helmholtz Center for Information Security, Germany
[⊥]Leibniz University Hannover, Germany
[‡]Paderborn University, Germany
[§]North Carolina State University, USA

Co-funded by the European Union

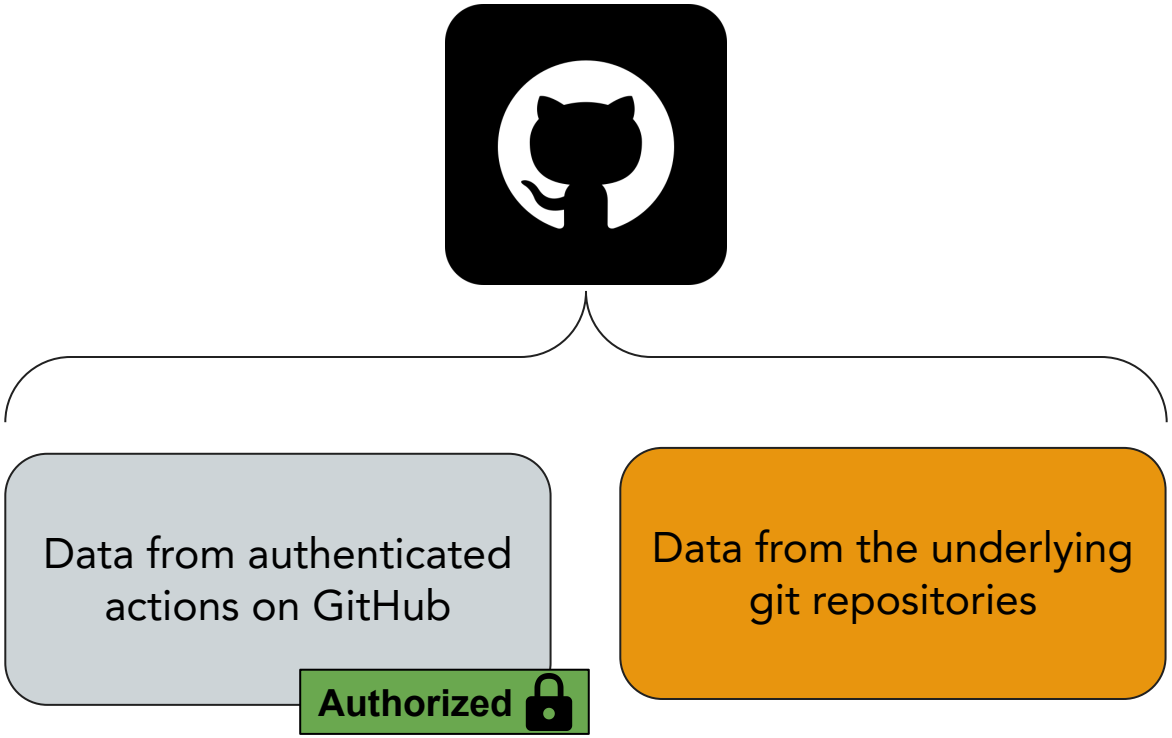
GitHub: A pillar of the OSS Supply Chain

- Most popular git hosting platform
- "Social Network" for open-source developers
- Lets you…
  - Host own software projects
  - Interact with other open-source projects via PRs or Issues
  - Follow the activities of other contributors
  - Investigate projects activities and contribution histories



Co-funded by the European Union

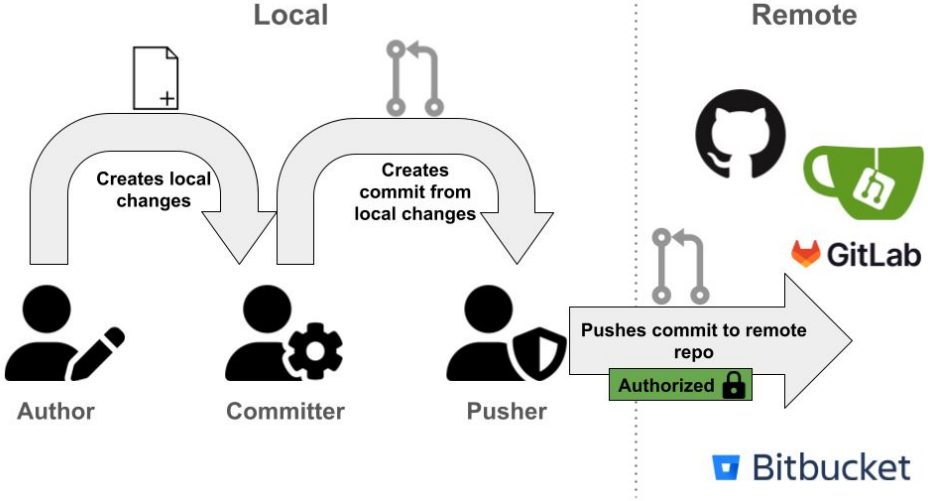
Problem: GitHub displays unverified data



Icons: Flaticon.com

Co-funded by the European Union

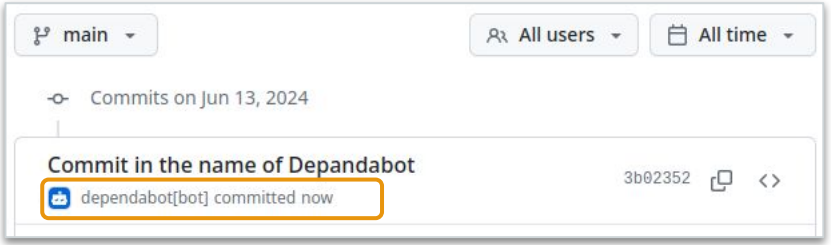
The git workflow




```
[user]
        email = dependabot[bot]@users.noreply.github.com
        name = dependatbot[bot]
```

```
tree fdd32538015c00a5eb392b9e11bf03516289887d
parent 01e9e8d1bf599b6296669053ecbfba014fcc7440
author dependabot[bot] <dependabot[bot]@users.noreply.github.com> 1718359303 +0200
committer dependabot[bot] <dependabot[bot]@users.noreply.github.com> 1718359303 +0200

Another commit in the name of Depandabot
```

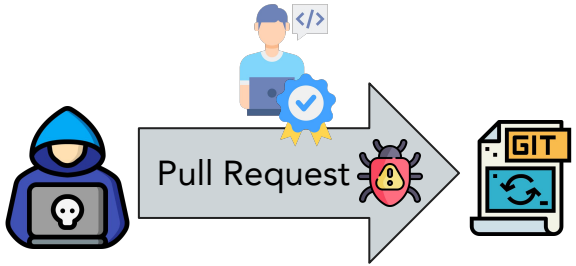


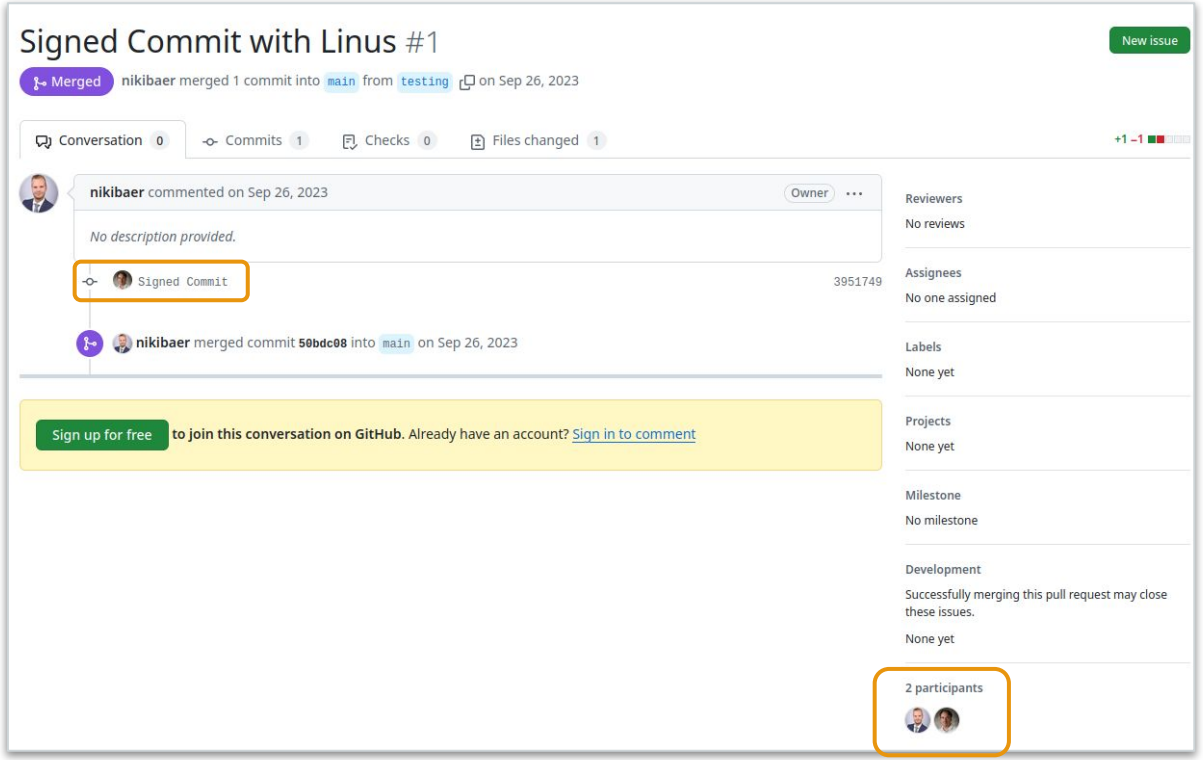
Co-funded by the European Union

Attackers can fake participants

Contributor Spoofing

Add established OSS contributors as (co-)authors to attackers' malicious PRs and leverage social trust for more lenient code reviews. Goal: Inject malicious code to an OSS project.





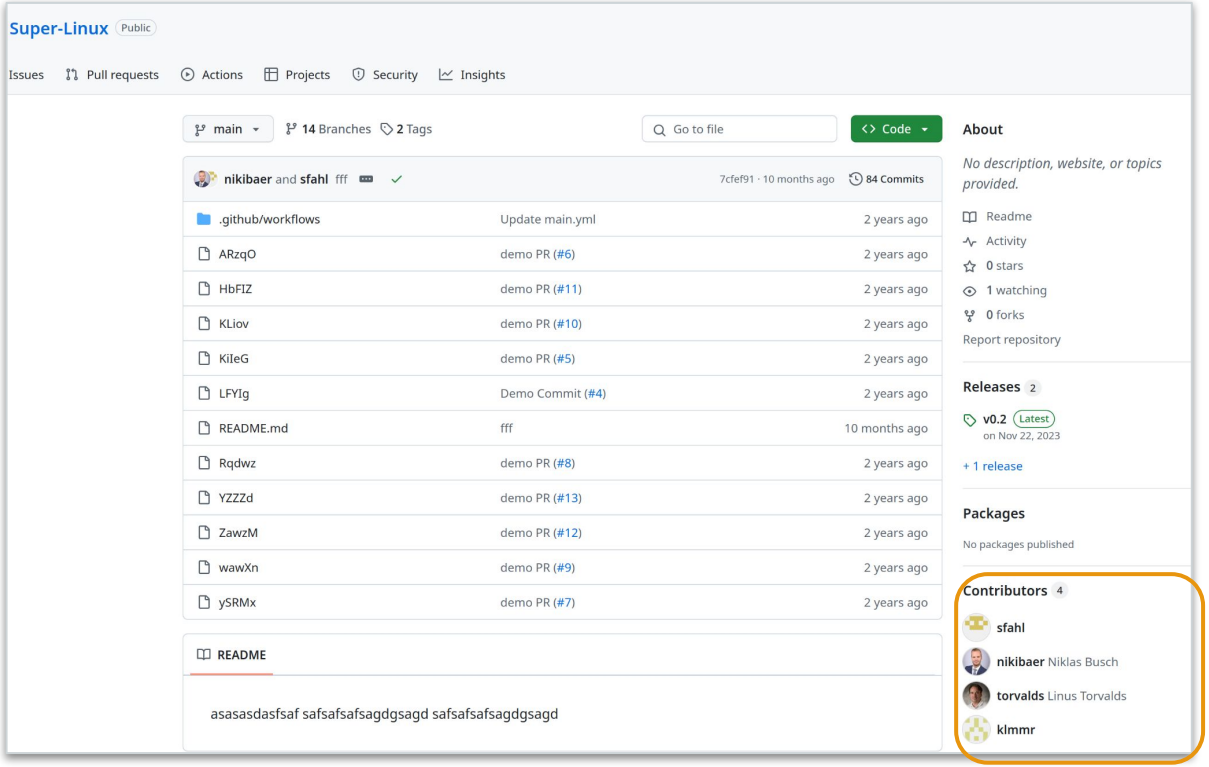
Icons: Flaticon.com

Co-funded by the European Union

Attackers can fake contributors

Reputation Hijacking

Create fake commits to malicious repository in the name of established OSS contributors and thus push its reputation.
Goal: Increase reputation of a project, e.g., for repository confusion attacks.



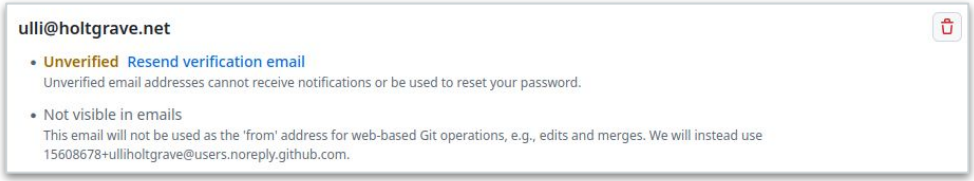
Icons: Flaticon.com

Co-funded by the European Union

Attackers can hijack unlinked commits

Contribution Hijacking

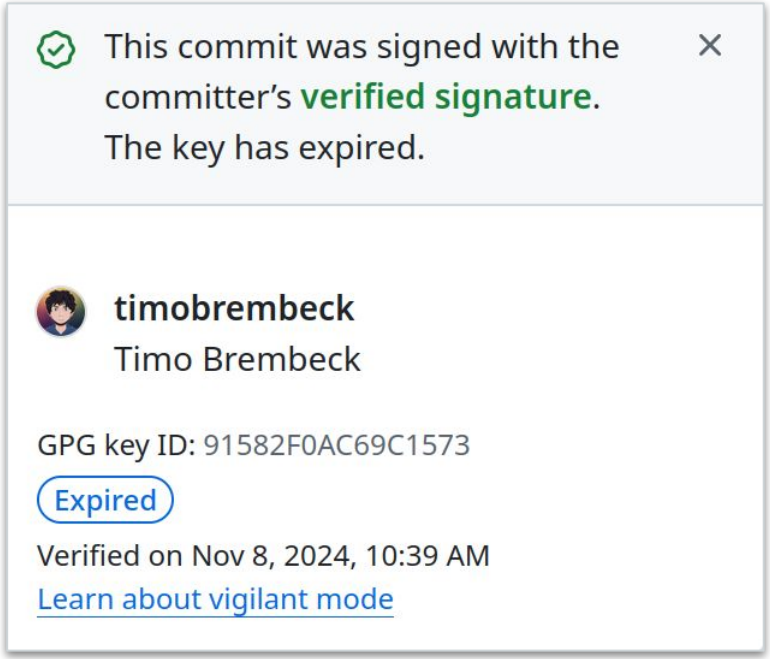
Claim unassigned commits in large OSS projects by adding the email to your account. Use Domain Hijacking for full attribution. Goal: Create fake legend, e.g., for infiltrating other OSS projects (see XZ-utils).



Icons: Flaticon.com

Signing commits adds cryptographic proof

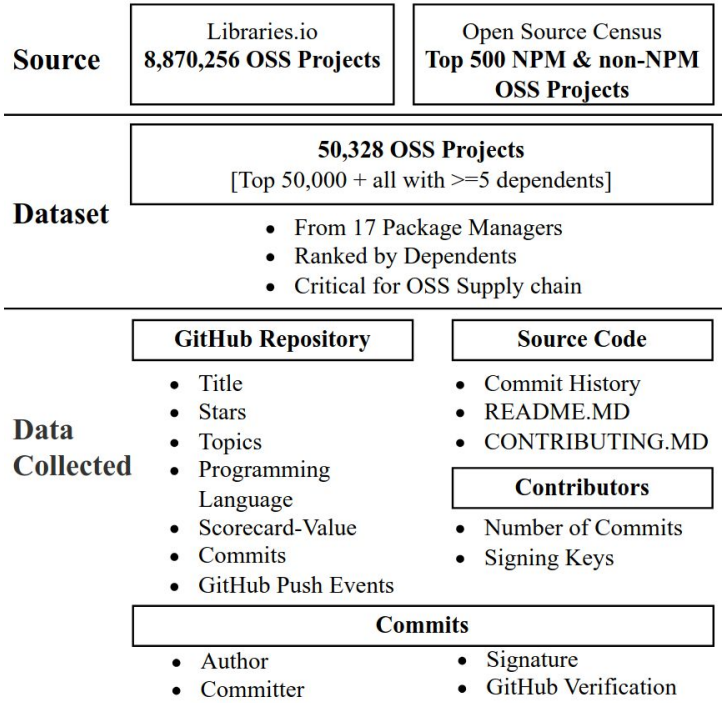
- Git offers committer signatures for commits and adds a layer of authenticity to commit objects
- GitHub supports PGP, SSH, and S/MIME and awards it with a “verified” badge



Measuring the attack surface for OSS

Methodology

- Top 50,328 open-source packages by dependent count hosted on GitHub
- 26,170,564 commits
- We investigated:
  - Disparity of git roles within a commit
  - Prevalence of claimable commits
  - Prevalence of commit signing on GitHub and signature validation status

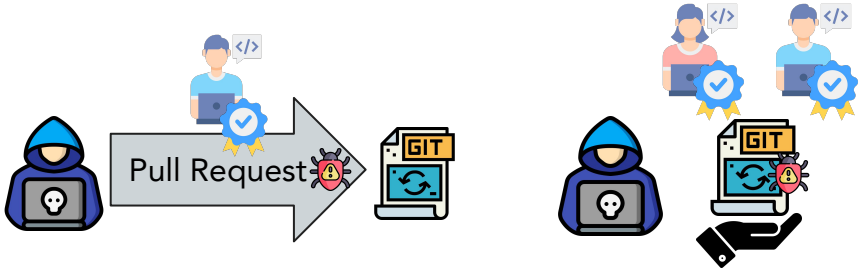


Co-funded by the European Union

Scenarios are feasible and difficult to detect

Contributor Spoofing & Reputation Hijacking

- 85.9% of the projects contain at least one commit with diverging roles
- 2,368 repositories contain no commit with parity (e.g., mirrors)



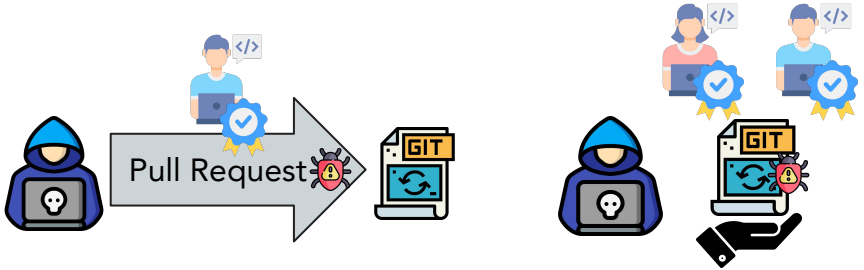
Icons: Flaticon.com

Co-funded by the European Union

Scenarios are feasible and difficult to detect

Contributor Spoofing & Reputation Hijacking

- 85.9% of the projects contain at least one commit with diverging roles
- 2,368 repositories contain no commit with parity (e.g., mirrors)



Contribution Hijacking

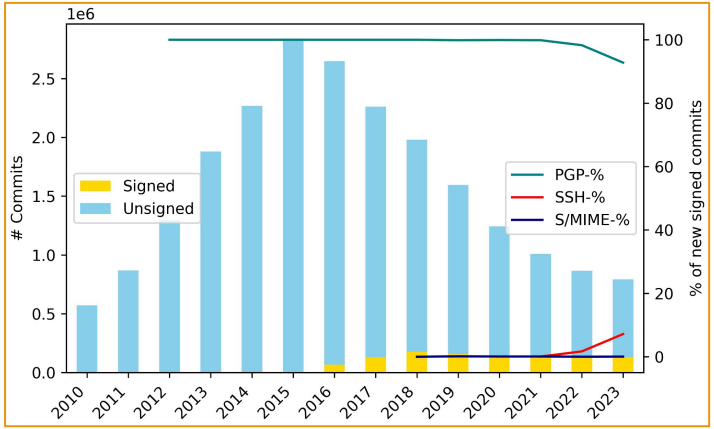
- 3,013,817 unlinked commits
- 573,043 contained valid email addresses
- We found 4,107 available domains that can be taken over via *Domain Hijacking*

Icons: Flaticon.com

Co-funded by the European Union

Only 5% of commits are signed

- Recent trend towards more SSH signing



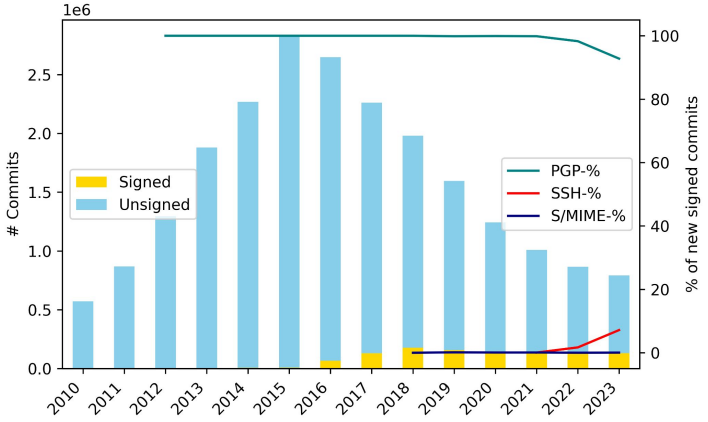
Result	PGP		SSH		S/MIME		GitHub PGP	
Valid	890,171	82%	10,660	90%	2	0%	2,960,651	100%
Unknown key	104,760	10%	810	7%	–	–	1	0%
No user	53,973	5%	119	1%	157	29%	0	0%
Bad email	19,918	2%	–	–	0	0%	0	0%
Unverified email	19,690	2%	273	2%	0	0%	0	0%
Invalid	212	0%	2	0%	1	0%	941	0%
GPG error	384	0%	–	–	–	–	111	0%
Bad certificate†	–	–	–	–	363	67%	–	–
No signing key	162	0%	–	–	–	–	0	0%
OCSP revoked*	–	–	–	–	15	3%	–	–

We also found 5 instances of unknown signature types we could not process.
* Entry in the certificate chain was revoked. † Certificate could not be verified.

Co-funded by the European Union

Only 5% of commits are signed

- Recent trend towards more SSH signing
- 98.9% used PGP and 1.1% used SSH for signing



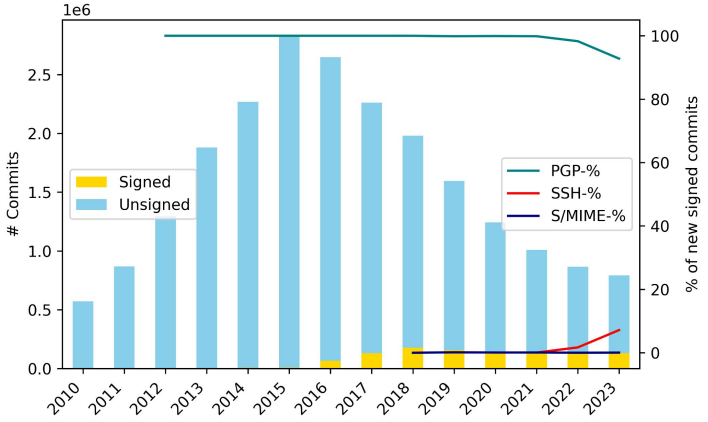
Result	PGP		SSH		S/MIME		GitHub PGP	
Valid	890,171	82%	10,660	90%	2	0%	2,960,651	100%
Unknown key	104,760	10%	810	7%	–	–	1	0%
No user	53,973	5%	119	1%	157	29%	0	0%
Bad email	19,918	2%	–	–	0	0%	0	0%
Unverified email	19,690	2%	273	2%	0	0%	0	0%
Invalid	212	0%	2	0%	1	0%	941	0%
GPG error	384	0%	–	–	–	–	111	0%
Bad certificate†	–	–	–	–	363	67%	–	–
No signing key	162	0%	–	–	–	–	0	0%
OCSP revoked*	–	–	–	–	15	3%	–	–

We also found 5 instances of unknown signature types we could not process.
* Entry in the certificate chain was revoked. † Certificate could not be verified.

Co-funded by the European Union

Only 5% of commits are signed

- Recent trend towards more SSH signing
- 98.9% used PGP and 1.1% used SSH for signing
- GitHub supports S/MIME signatures, but next to no prevalence in our dataset



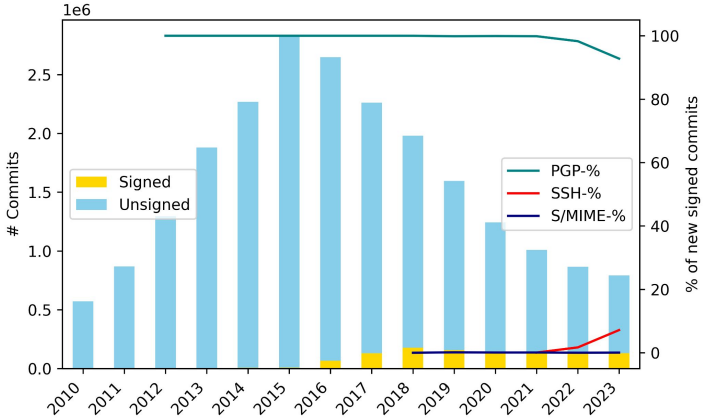
Result	PGP		SSH		S/MIME		GitHub PGP	
Valid	890,171	82%	10,660	90%	2	0%	2,960,651	100%
Unknown key	104,760	10%	810	7%	–	–	1	0%
No user	53,973	5%	119	1%	157	29%	0	0%
Bad email	19,918	2%	–	–	0	0%	0	0%
Unverified email	19,690	2%	273	2%	0	0%	0	0%
Invalid	212	0%	2	0%	1	0%	941	0%
GPG error	384	0%	–	–	–	–	111	0%
Bad certificate†	–	–	–	–	363	67%	–	–
No signing key	162	0%	–	–	–	–	0	0%
OCSP revoked*	–	–	–	–	15	3%	–	–

We also found 5 instances of unknown signature types we could not process.
* Entry in the certificate chain was revoked. † Certificate could not be verified.

Co-funded by the European Union

Only 5% of commits are signed

- Recent trend towards more SSH signing
- 98.9% used PGP and 1.1% used SSH for signing
- GitHub supports S/MIME signatures, but next to no prevalence in our dataset
- Noticeable Problem: About 10% of commits cannot be verified due to no known signing key



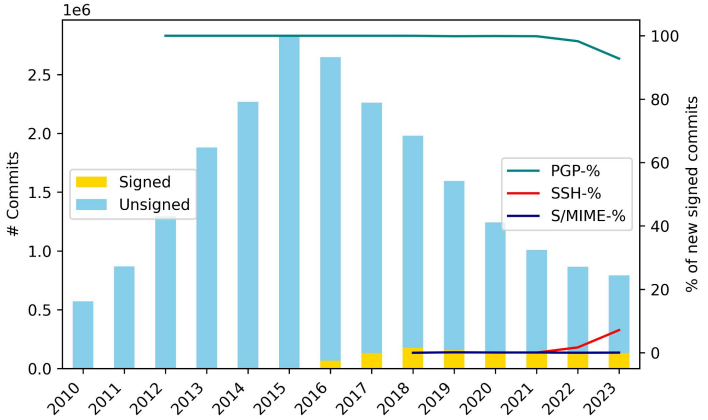
Result	PGP		SSH		S/MIME		GitHub PGP	
Valid	890,171	82%	10,660	90%	2	0%	2,960,651	100%
Unknown key	104,760	10%	810	7%	–	–	1	0%
No user	53,973	5%	119	1%	157	29%	0	0%
Bad email	19,918	2%	–	–	0	0%	0	0%
Unverified email	19,690	2%	273	2%	0	0%	0	0%
Invalid	212	0%	2	0%	1	0%	941	0%
GPG error	384	0%	–	–	–	–	111	0%
Bad certificate†	–	–	–	–	363	67%	–	–
No signing key	162	0%	–	–	–	–	0	0%
OCSP revoked*	–	–	–	–	15	3%	–	–

We also found 5 instances of unknown signature types we could not process.
* Entry in the certificate chain was revoked. † Certificate could not be verified.

Co-funded by the European Union

Only 5% of commits are signed

- Recent trend towards more SSH signing
- 98.9% used PGP and 1.1% used SSH for signing
- GitHub supports S/MIME signatures, but next to no prevalence in our dataset
- Noticeable Problem: About 10% of commits cannot be verified due to no known signing key



Result	PGP		SSH		S/MIME		GitHub PGP	
Valid	890,171	82%	10,660	90%	2	0%	2,960,651	100%
Unknown key	104,760	10%	810	7%	–	–	1	0%
No user	53,973	5%	119	1%	157	29%	0	0%
Bad email	19,918	2%	–	–	0	0%	0	0%
Unverified email	19,690	2%	273	2%	0	0%	0	0%
Invalid	212	0%	2	0%	1	0%	941	0%
GPG error	384	0%	–	–	–	–	111	0%
Bad certificate†	–	–	–	–	363	67%	–	–
No signing key	162	0%	–	–	–	–	0	0%
OCSP revoked*	–	–	–	–	15	3%	–	–

We also found 5 instances of unknown signature types we could not process.
* Entry in the certificate chain was revoked. † Certificate could not be verified.

→ Signed commits are rarely used and the current adoption offers only limited protection and authenticity

Co-funded by the European Union

How can we make future open-source contributions harder to fake?

- **Email Validation:** We urge GitHub to rethink their email verification to not rely on unverified emails for contributions.
- Improve GitHub UI:
  - Distinguish GH signatures
  - Display the pusher of a commit
- Add author signatures for git commit objects: Improve commit authenticity.

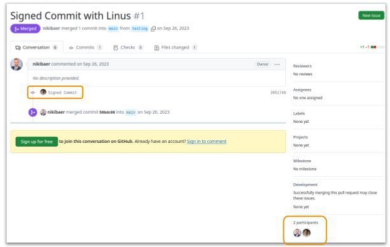
We disclosed our findings to GitHub and they could reproduce our attacks. They consider making attributions more strict in the future.

Summary

Attackers can fake participants

Contributor Spoofing

Add established OSS contributors as (co-)authors to attackers' malicious PRs and leverage social trust for more lenient code reviews.
Goal: Inject malicious code to an OSS project.

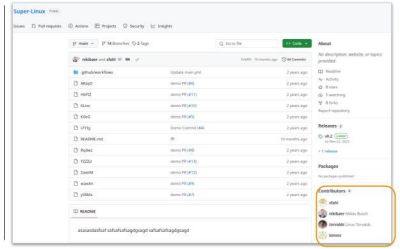


Icons: Flaticon.com

Attackers can fake contributors

Reputation Hijacking

Create fake commits to malicious repository in the name of established OSS contributors and thus push its reputation.
Goal: Increase reputation of a project, e.g., for repository confusion attacks.

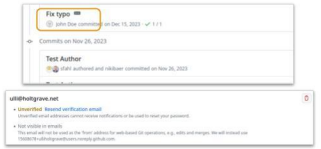


Icons: Flaticon.com

Attackers can hijack unlinked commits

Contribution Hijacking

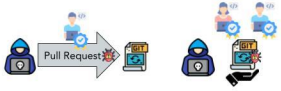
Claim unassigned commits in large OSS projects by adding the email to your account. Use Domain Hijacking for full attribution.
Goal: Create fake legend, e.g., for infiltrating other OSS projects (see XZ-utils).



Icons: Flaticon.com

Scenarios are feasible and difficult to detect

Contributor Spoofing & Reputation Hijacking
- 85.9% of the projects contain at least one commit with diverging roles
- 2,368 repositories contain no commit with parity (e.g., mirrors)



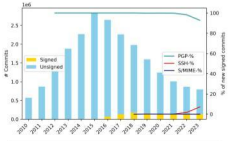
Contribution Hijacking
- 3,013,817 unlinked commits
- 573,043 contained valid email addresses
- We found 4,107 available domains that can be taken over via *Domain Hijacking*

Icons: Flaticon.com

Only 5% of commits are signed

- Recent trend towards more SSH signing
- 98.9% used PGP and 1.1% used SSH for signing
- GitHub supports S/MIME signatures, but next to no prevalence in our dataset
- Noticeable Problem: About 10% of commits cannot be verified due to no known signing key

→ Signed commits are rarely used and the current adoption offers only limited protection and authenticity



Result	PGP	SSH	S/MIME	GitHub PGP
Valid	890,171 82%	10,660 90%	2 0%	2,960,651 100%
Unknown key	104,760 10%	810 7%	1 0%	1 0%
No user	53,873 5%	119 1%	157 29%	0 0%
Bad email	19,918 2%	- -	0 0%	0 0%
Unverified email	19,690 2%	273 2%	0 0%	0 0%
Invalid	212 0%	2 0%	1 0%	941 0%
GPG error	384 0%	- -	- -	111 0%
Bad certificate[1]	- -	- -	363 67%	0 0%
No signing key	162 0%	- -	- -	0 0%
OCSP revoked[2]	- -	- -	15 3%	0 0%

We also found 5 instances of unknown signature types we could not process.
[1] Entry in the certificate chain was revoked. [2] Certificate could not be verified.

How can we make future open-source contributions harder to fake?

- Email Validation: We urge GitHub to rethink their email verification to not rely on unverified emails for contributions.
- Improve GitHub UI:
  - Distinguish GH signatures
  - Display the pusher of a commit
- Add author signatures for git commit objects: Improve commit authenticity.

We disclosed our findings to GitHub and they could reproduce our attacks. They consider making attributions more strict in the future.

Contact:
Jan-Ulrich Holtgrave (jan-ulrich.holtgrave@cispa.de)