CHKPLUG: Checking GDPR Compliance of WordPress Plugins via Cross-language Code Property Graph

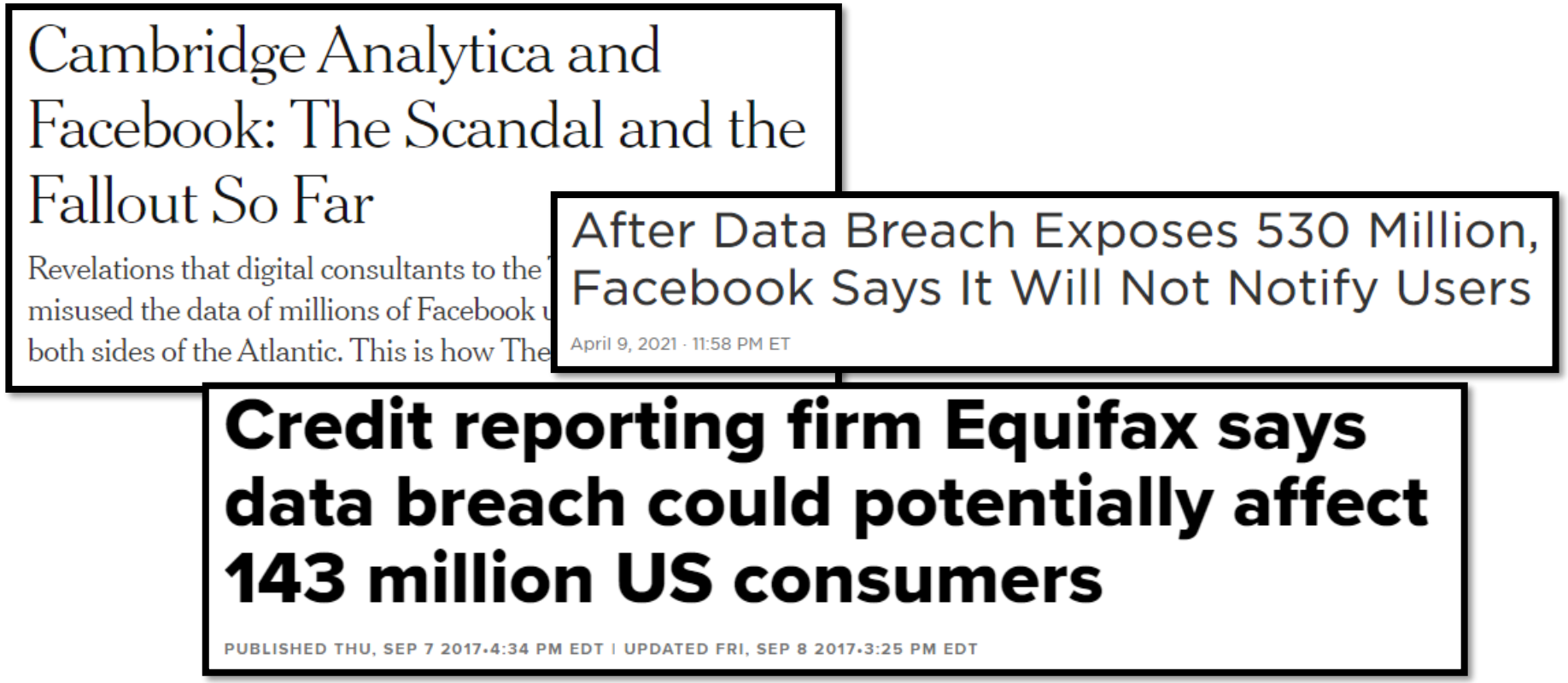
Faysal Hossain Shezan, Zihao Su, Mingqing Kang, Nicholas Phair,
Patrick William Thomas, Michelangelo van Dam, Yinzhi Cao, Yuan Tian

Recent data breach incidents

https://www.npr.org/2021/04/09/986005820/after-data-breach-exposes-530-million-facebook-says-it-will-not-notify-users
https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html
https://www.cnbc.com/2017/09/07/credit-reporting-firm-equifax-says-cybersecurity-incident-could-potentially-affect-143-million-us-consumers.html

Recent data breach incidents

https://www.npr.org/2021/04/09/986005820/after-data-breach-exposes-530-million-facebook-says-it-will-not-notify-users
https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html
https://www.cnbc.com/2017/09/07/credit-reporting-firm-equifax-says-cybersecurity-incident-could-potentially-affect-143-million-us-consumers.html



Recent data breach incidents

https://www.npr.org/2021/04/09/986005820/after-data-breach-exposes-530-million-facebook-says-it-will-not-notify-users
https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html
https://www.cnbc.com/2017/09/07/credit-reporting-firm-equifax-says-cybersecurity-incident-could-potentially-affect-143-million-us-consumers.html

Policy Enforcement

General Data Protection Regulation

Children's Online Privacy Protection Act

California Consumer Privacy Act

Virginia Consumer Data Protection Act

Colorado Privacy Act

Google continues to send data from EU websites to the US - despite two Court of Justice rulings. Austrian Data Protection Authority could fine Google up to €6 billion.

Fines for GDPR incompliance

https://www.skillcast.com/blog/20-biggest-gdpr-fines
https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/
https://noyb.eu/en/austrian-dpa-has-option-fine-google-eu6-billion

Following the introduction of GDPR in May 2018, initial reports showed that **data breach complaints increased by 160%**. This rate is alarming and indicates just how critical it is to ensure staff receive **comprehensive GDPR training**.

Top 20 GDPR fines so far

1. Amazon Europe - €746m fine (2021)
2. WhatsApp Ireland - €225m fine (2021)
3. Google Inc - €50m fine (2019)
4. H&M - €35.3m fine (2020)
5. TIM - €27.8m fine (2020)
6. British Airways - €22m fine (2020)

Google continues to send data from EU websites to the US - despite two Court of Justice rulings. Austrian Data Protection Authority could fine Google up to €6 billion.

Fines for GDPR incompliance

https://www.skillcast.com/blog/20-biggest-gdpr-fines
https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/
https://noyb.eu/en/austrian-dpa-has-option-fine-google-eu6-billion

Google continues to send data from EU w
US - despite two Court of Justice rulings.
Protection Authority could fine Google up

Following the introduction of GDPR in May 2018, initial reports showed that **data breach complaints increased by 160%**. This rate is alarming and indicates just how critical it is to ensure staff receive **comprehensive GDPR training**.

Top 20 GDPR fines so far

1. Amazon Europe - €746m fine (2021)
2. WhatsApp Ireland - €225m fine (2021)
3. Google Inc - €50m fine (2019)
4. H&M - €35.3m fine (2020)
5. TIM - €27.8m fine (2020)
6.

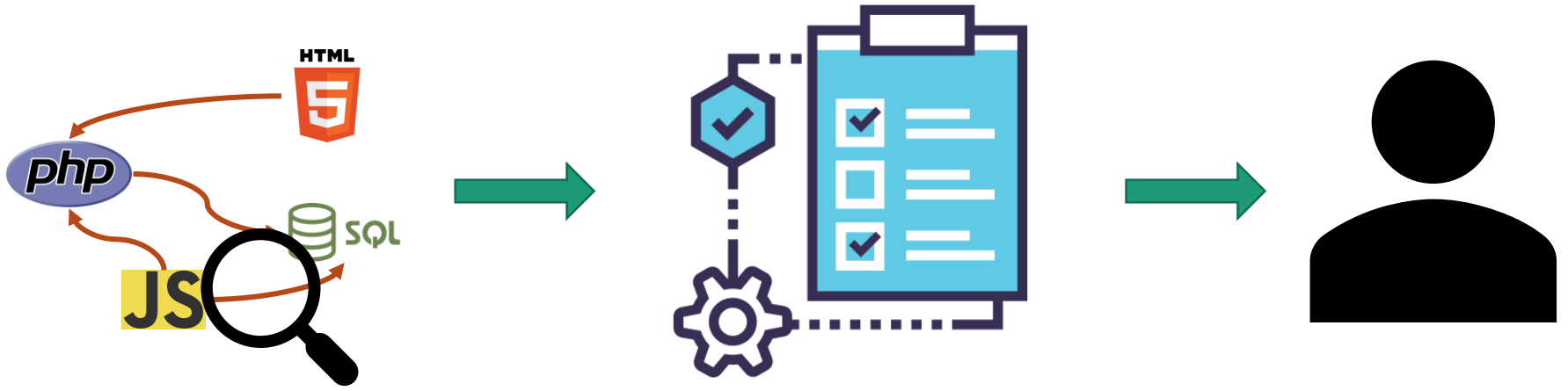
GDPR: Fines increased by 40% last year, and they're about to get a lot bigger

Non-compliant businesses, beware: analysts say that regulators are about to get much tougher with GDPR enforcement.

Fines for GDPR incompliance

https://www.skillcast.com/blog/20-biggest-gdpr-fines
https://www.zdnet.com/article/gdpr-fines-increased-by-40-last-year-and-theyre-about-to-get-a-lot-bigger/
https://noyb.eu/en/austrian-dpa-has-option-fine-google-eu6-billion

Automatic Compliance Check is Needed



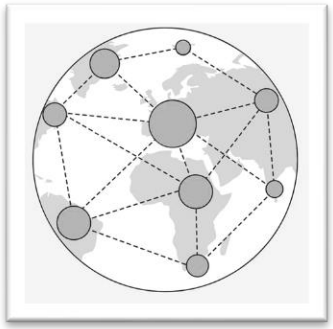
Complex Cross-Language Interaction

Privacy Regulations

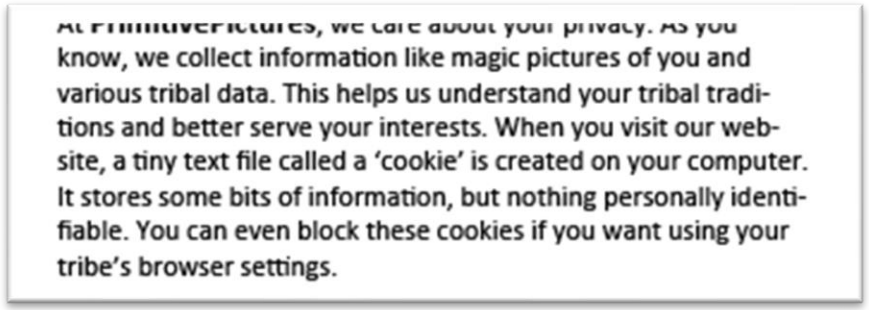
User

We need to identify how the app violates GDPR automatically and provide possible solution to fix that

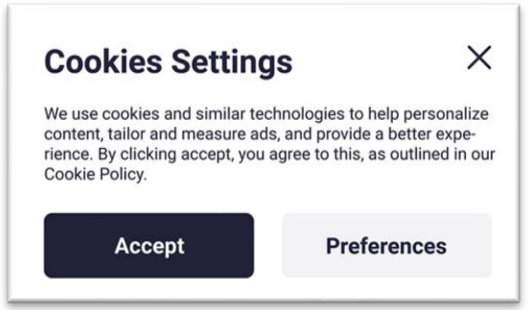
Previous Work Will Not Work



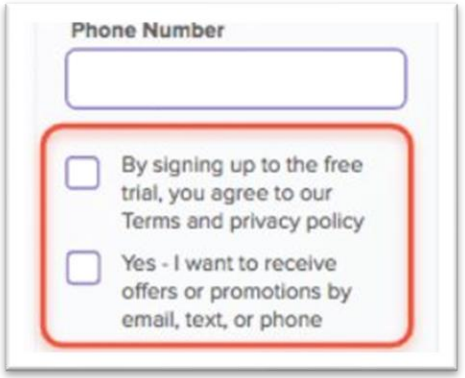
Network Traffic



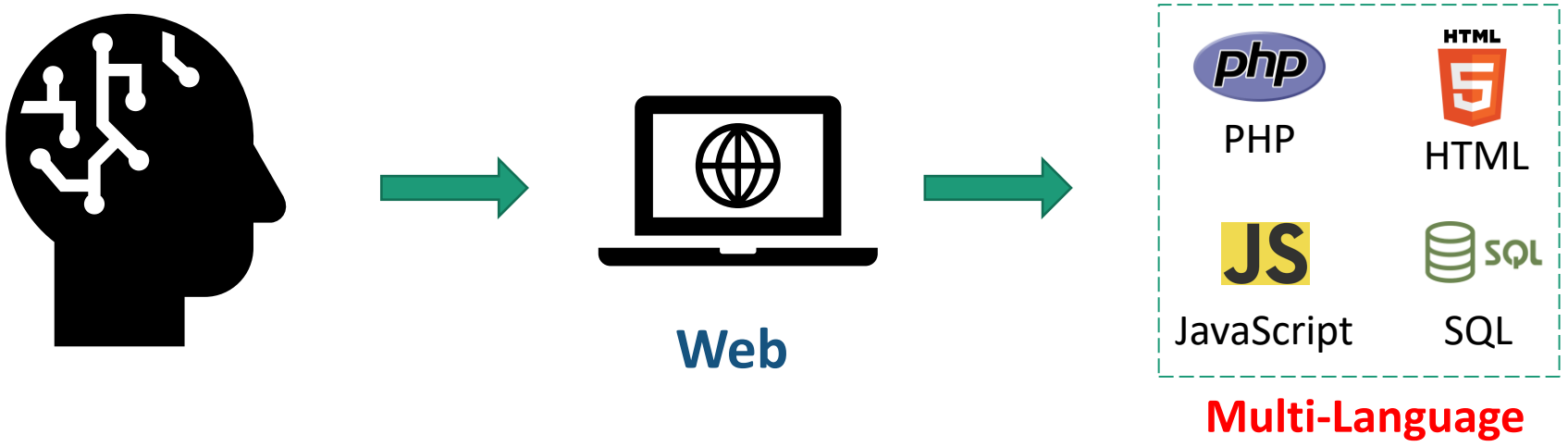
Privacy Policy



Cookie



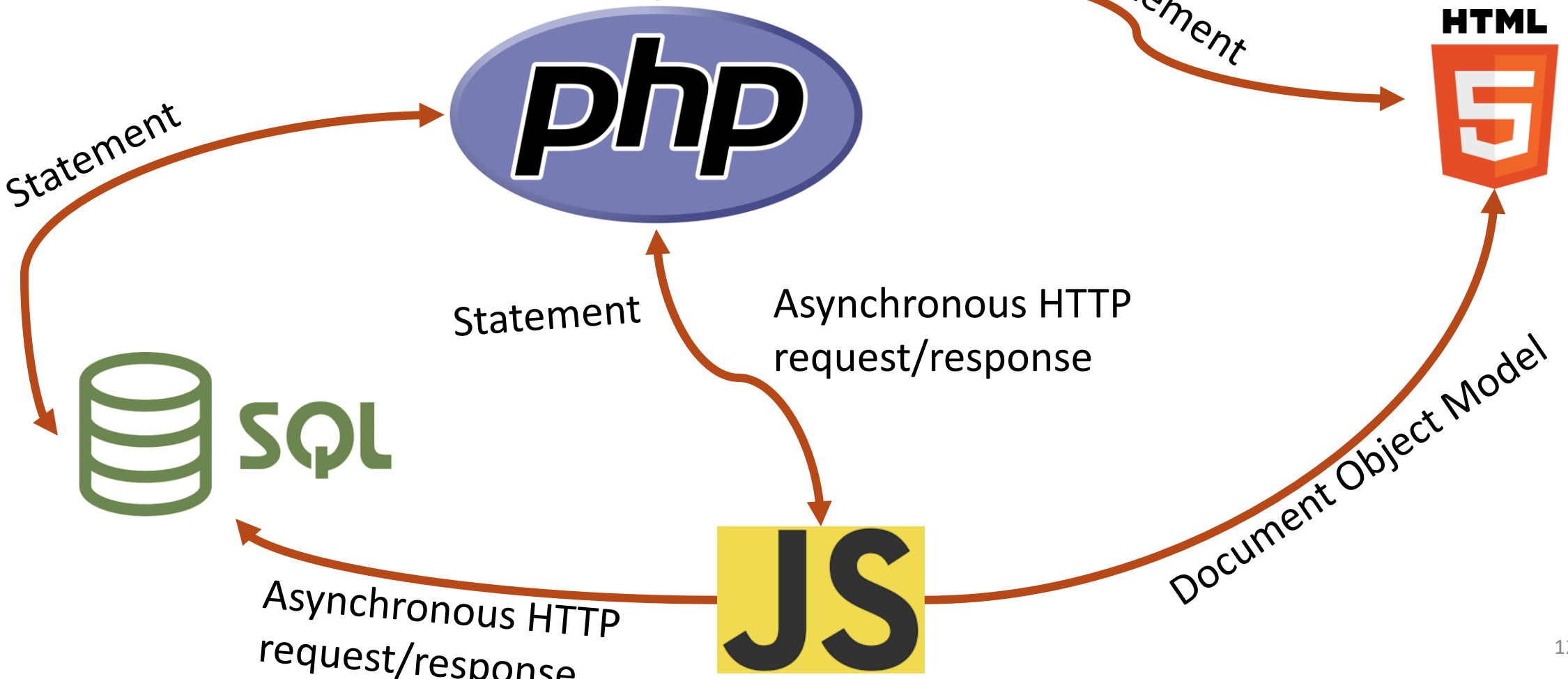
User Consent



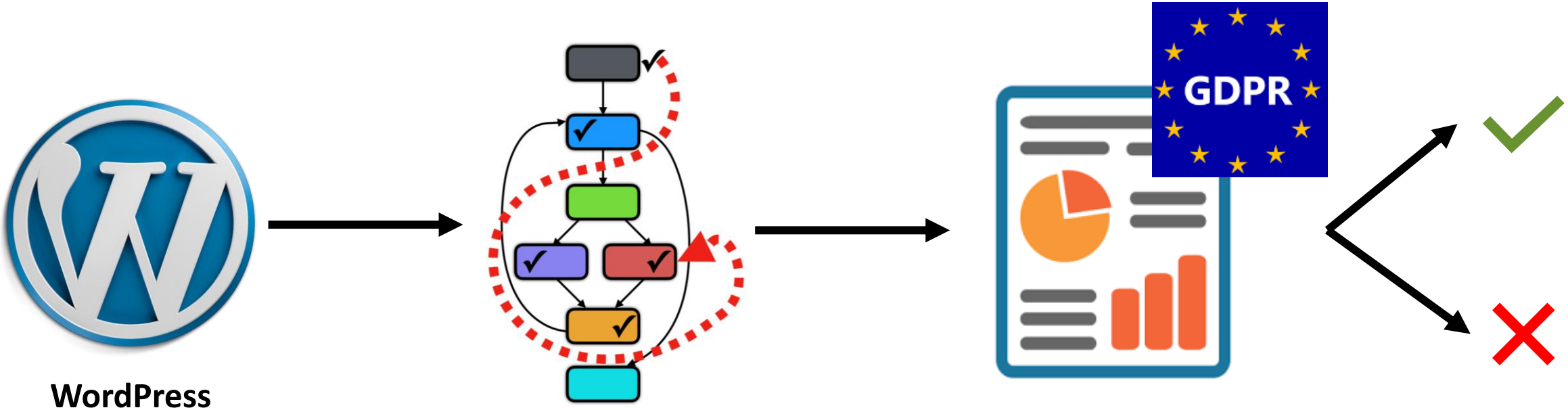
Existing information flow traversal limited to single programming language

Challenges

Cross-Language Dataflows



Key Insight



WordPress Plugin

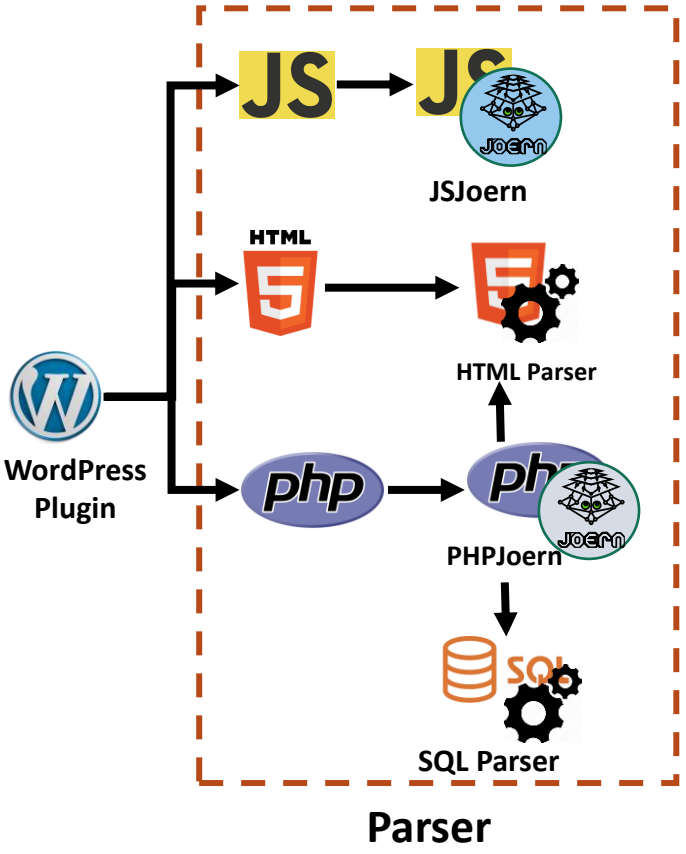
Dataflow Traversal

GDPR Mapping

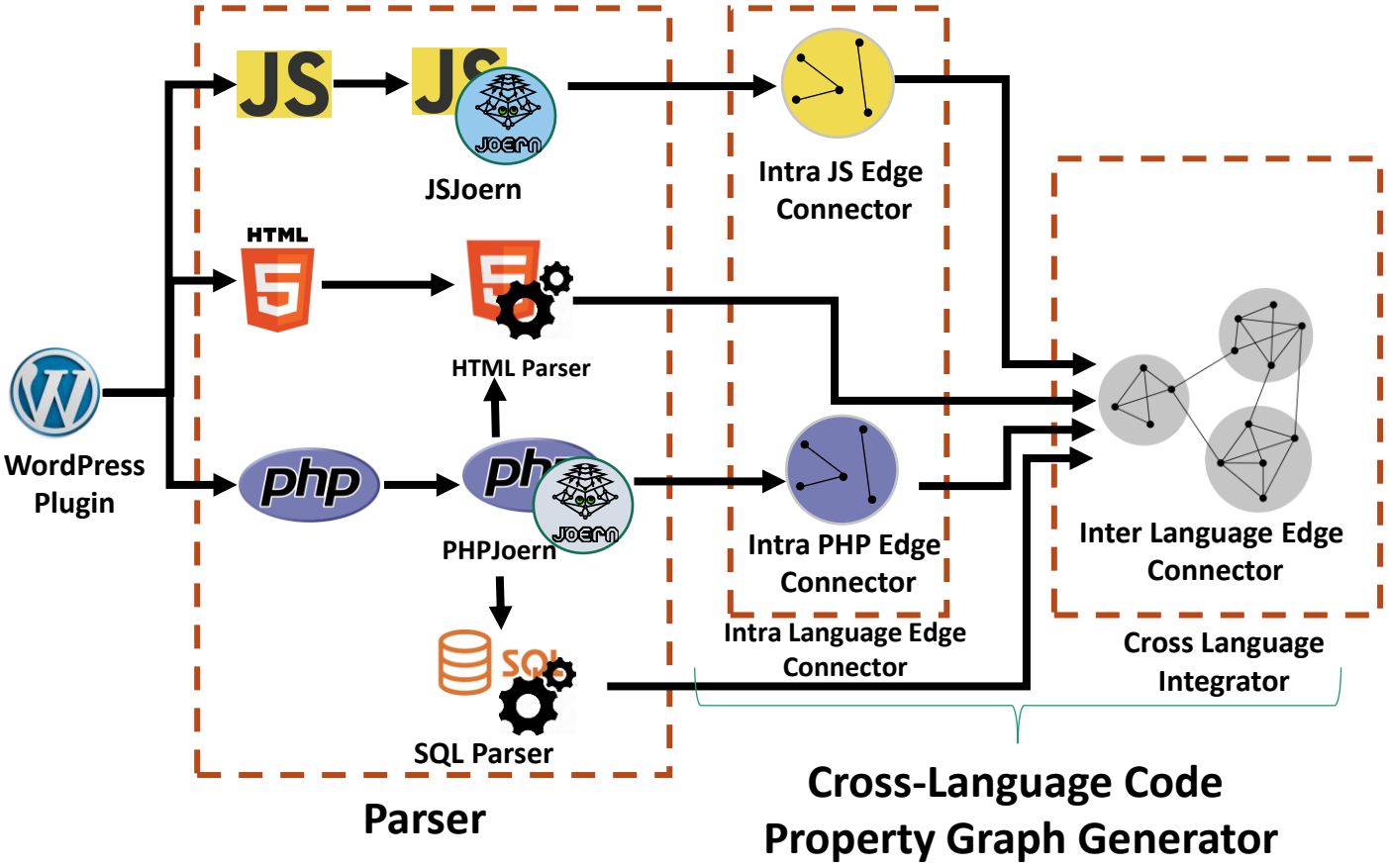
Goal

Check Compliance of GDPR Policies

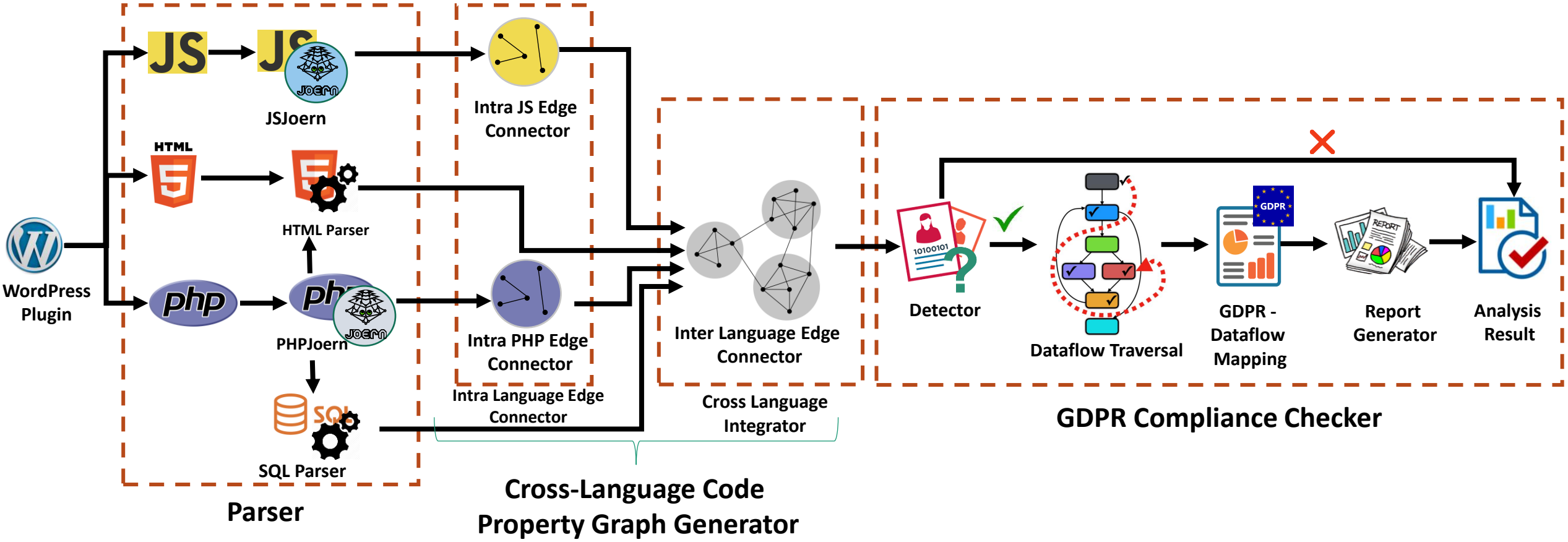
System Overview of CHKPLUG



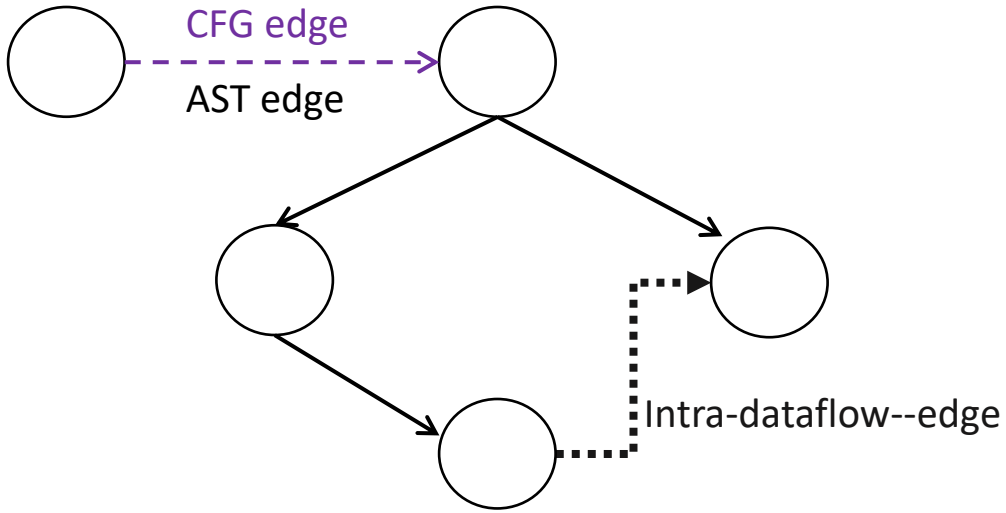
System Overview of CHKPLUG



System Overview of CHKPLUG

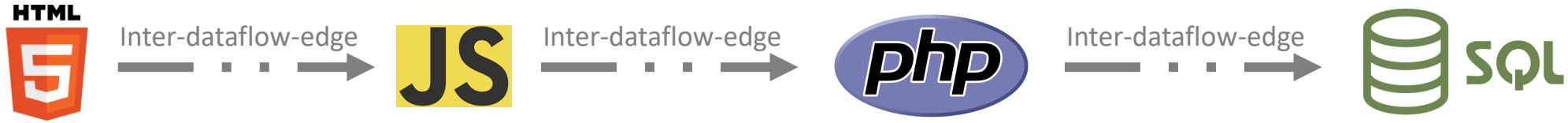


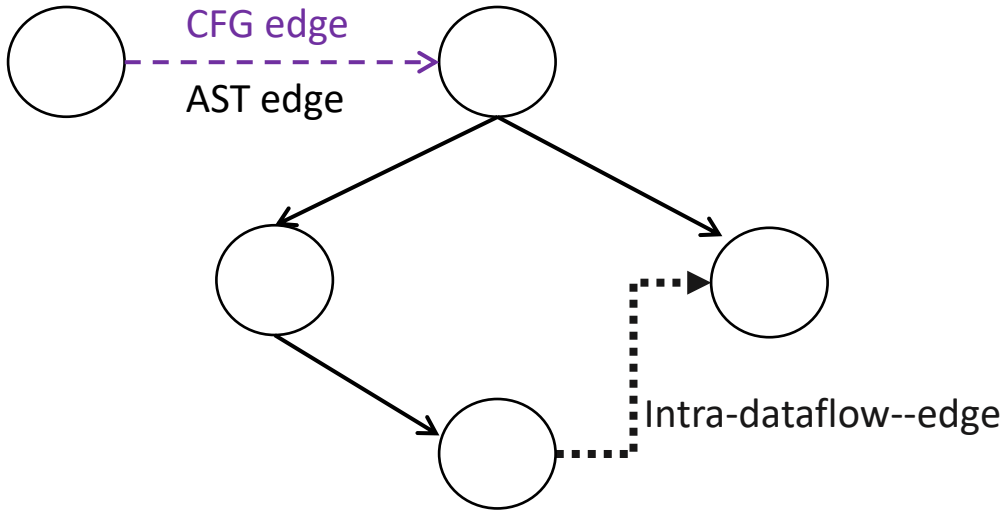
Procedure



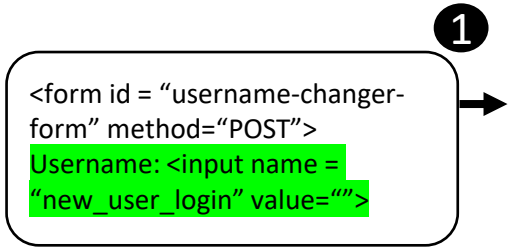
Create Code Property Graph for all the four programming languages

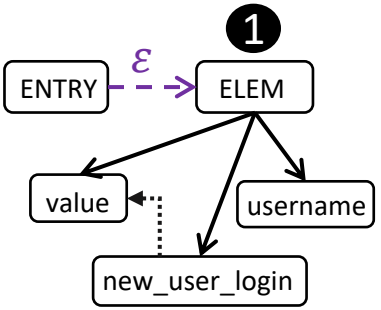
Procedure

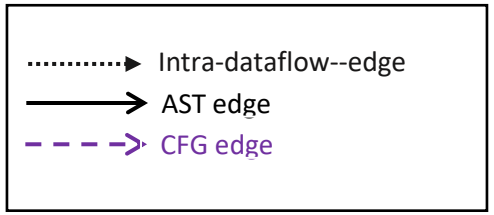




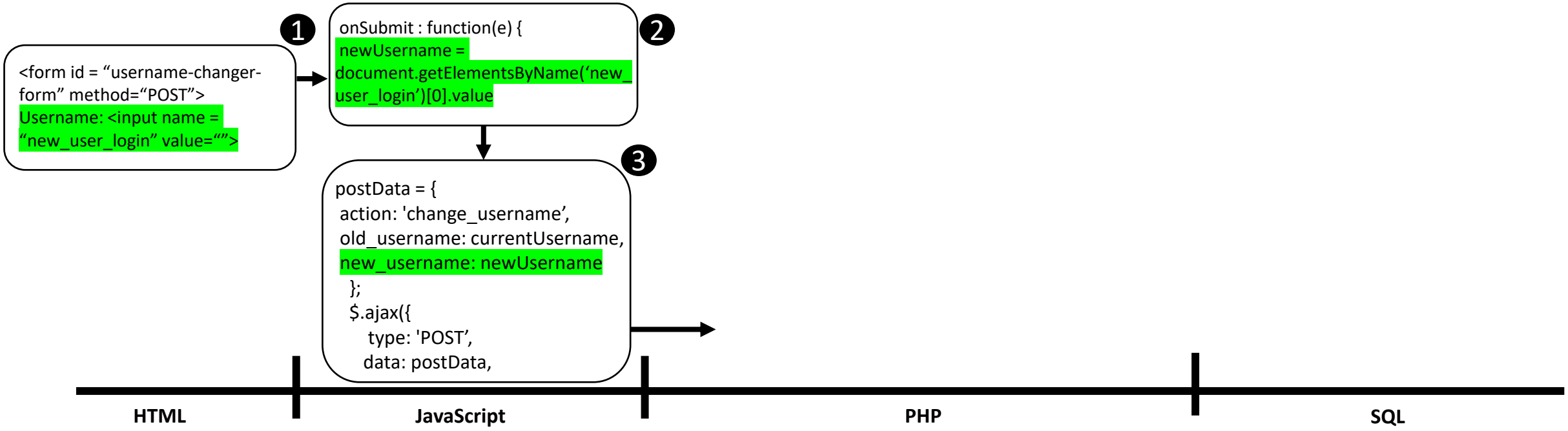
Connect nodes from different programming languages to create Cross-Language Code Property Graph

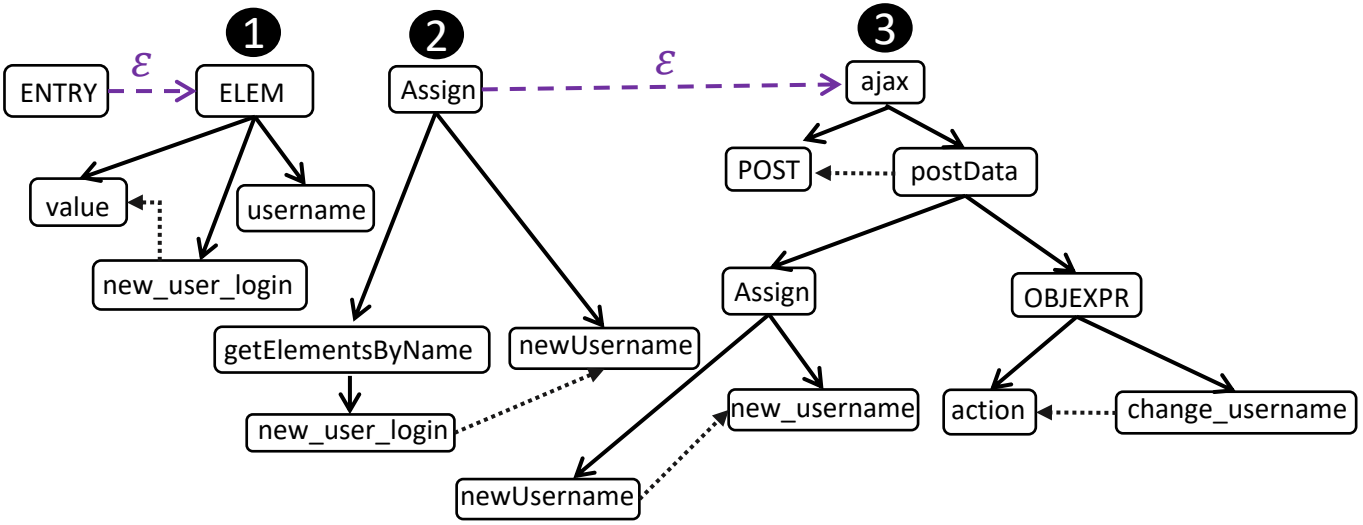


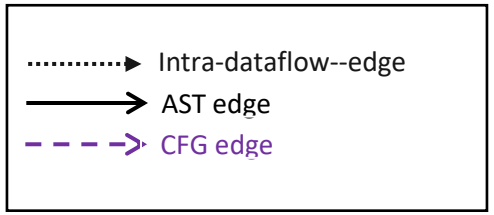




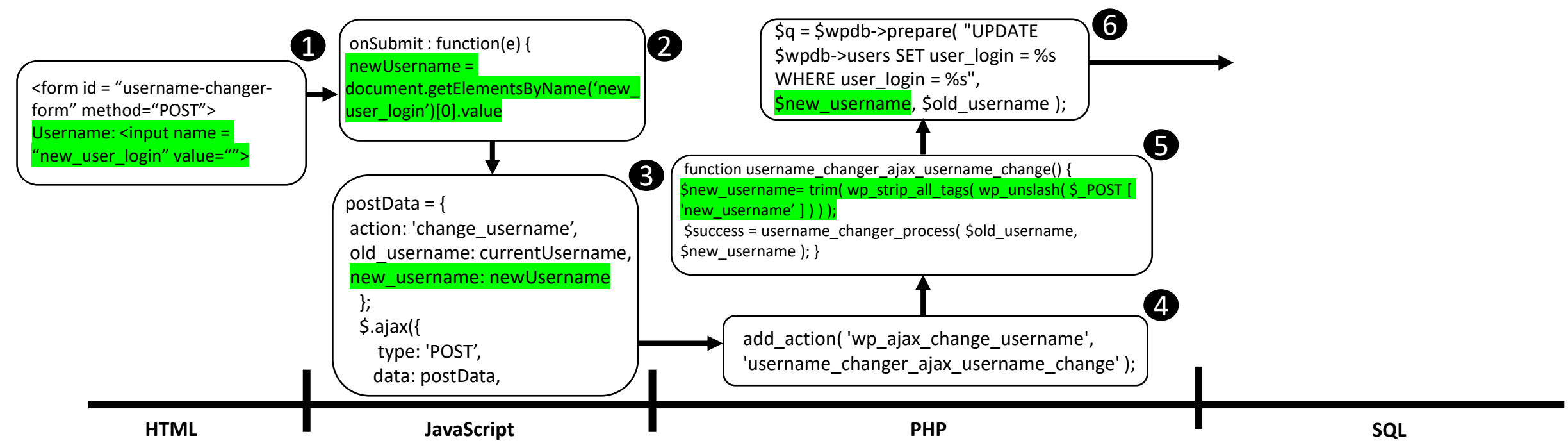
Cross-Language Code Property Graph

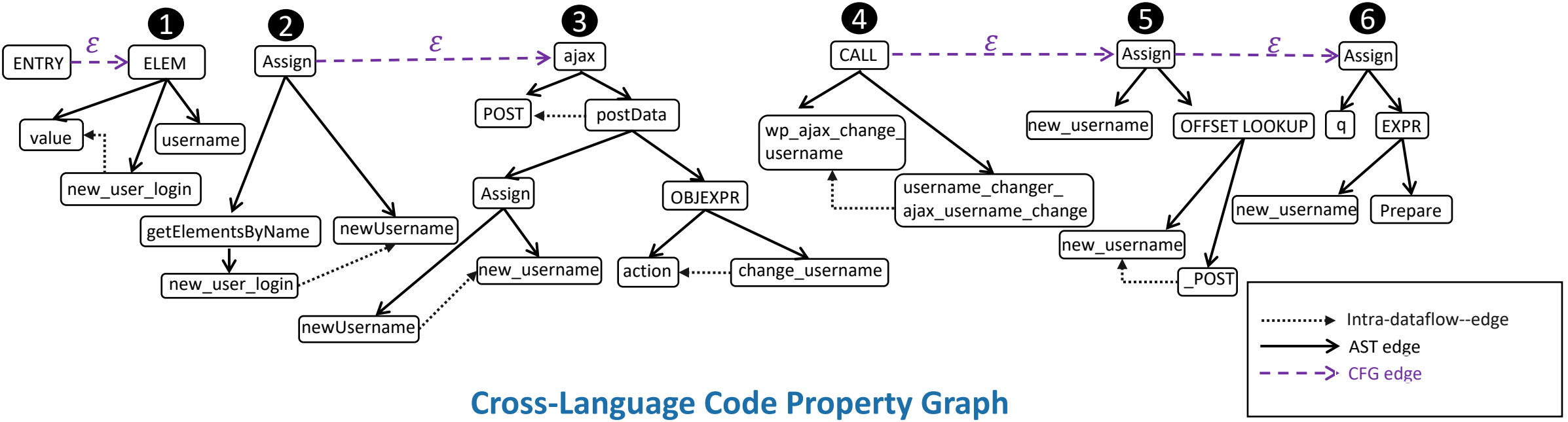




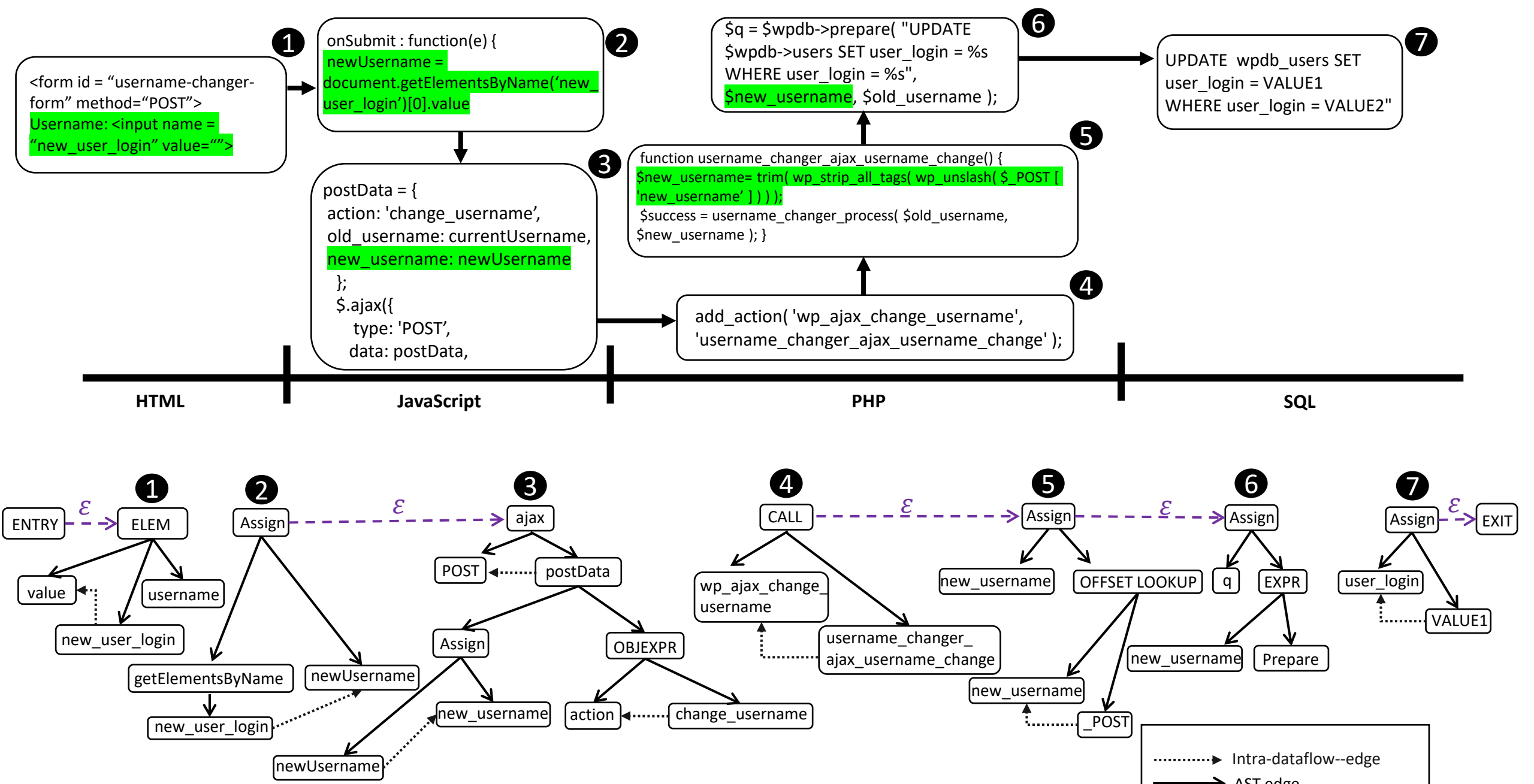


Cross-Language Code Property Graph

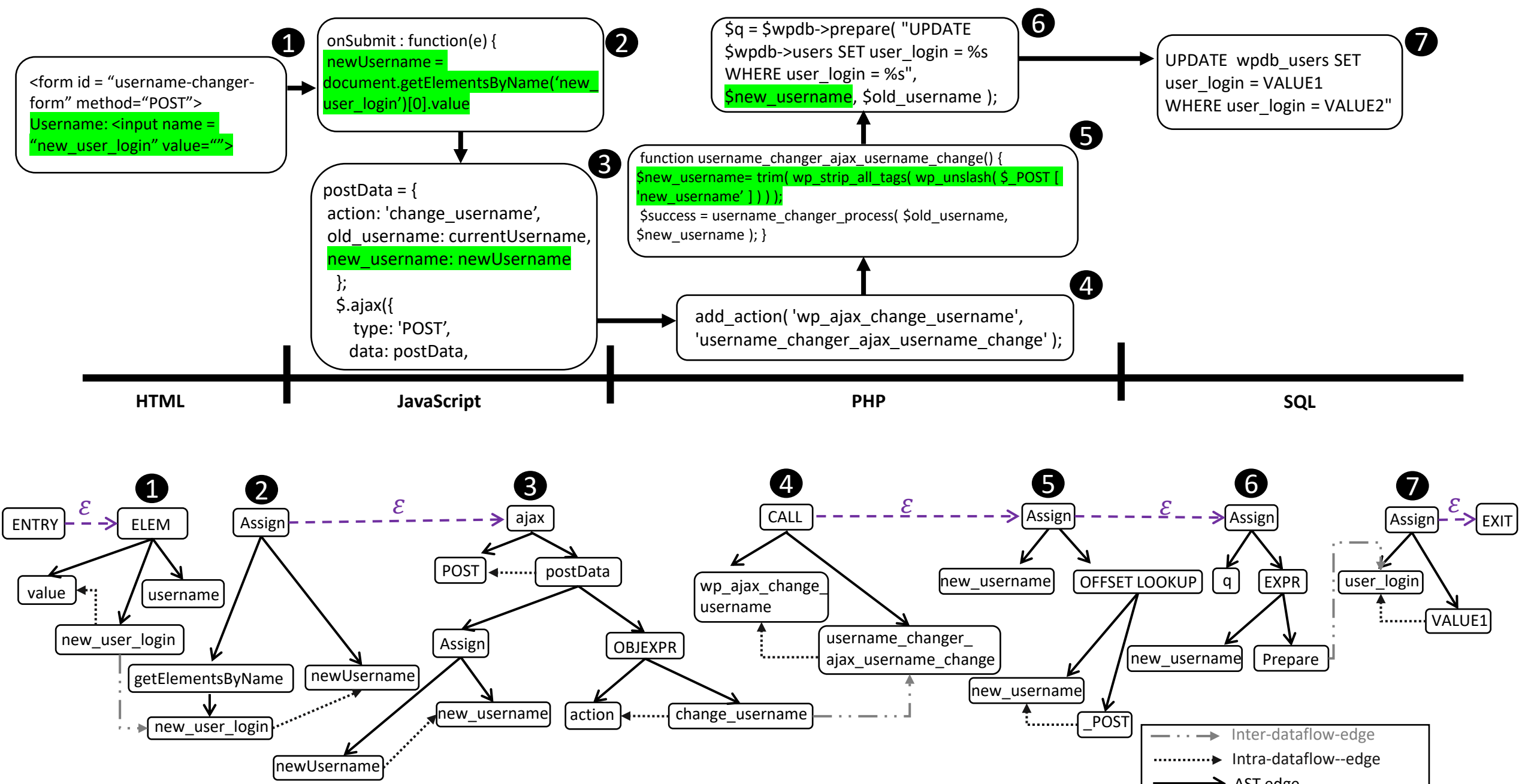




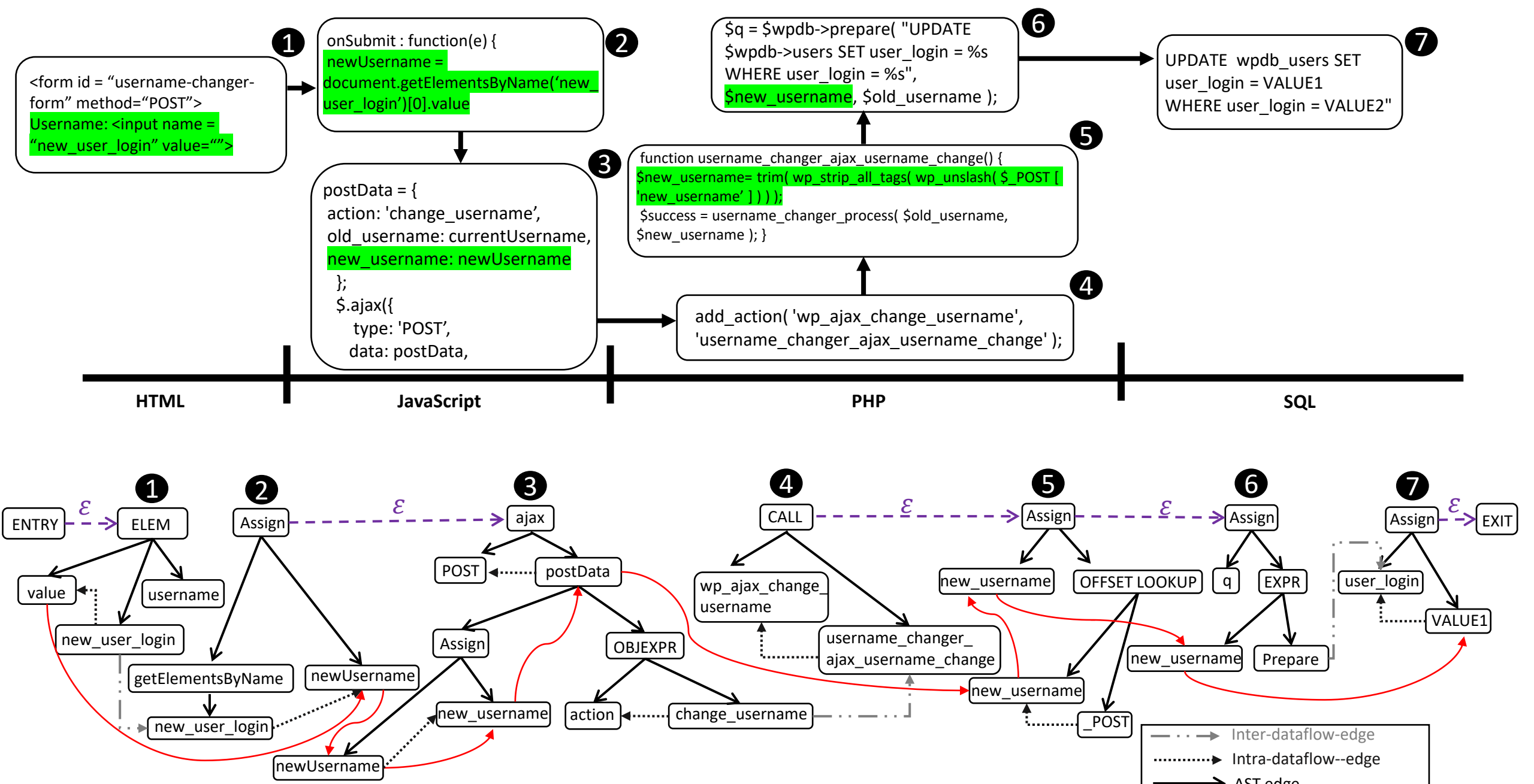
Cross-Language Code Property Graph



Cross-Language Code Property Graph



Cross-Language Code Property Graph



Cross-Language Code Property Graph

Queries for Checking Compliance

ID	Action	Flow
P_{access}	Store	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow storage_{DB}(PII)$
	Export	$counter_{retrieve\ DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow export_{interface}(PII)$
P_{delete}	Store	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow storage_{DB}(PII)$
	Delete	$counter_{retrieve\ DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow delete_{interface}(PII)$
P_{share}	Send	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow remote_{request}(PII)$
	Disclosure	$doesExist(privacy\text{-}policy)\ \&\&\ disclose\ (sent_{PII}, remote_{URL})$
$P_{security}$	Send	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow secure_{node} \rightarrow remote_{request}(PII)$

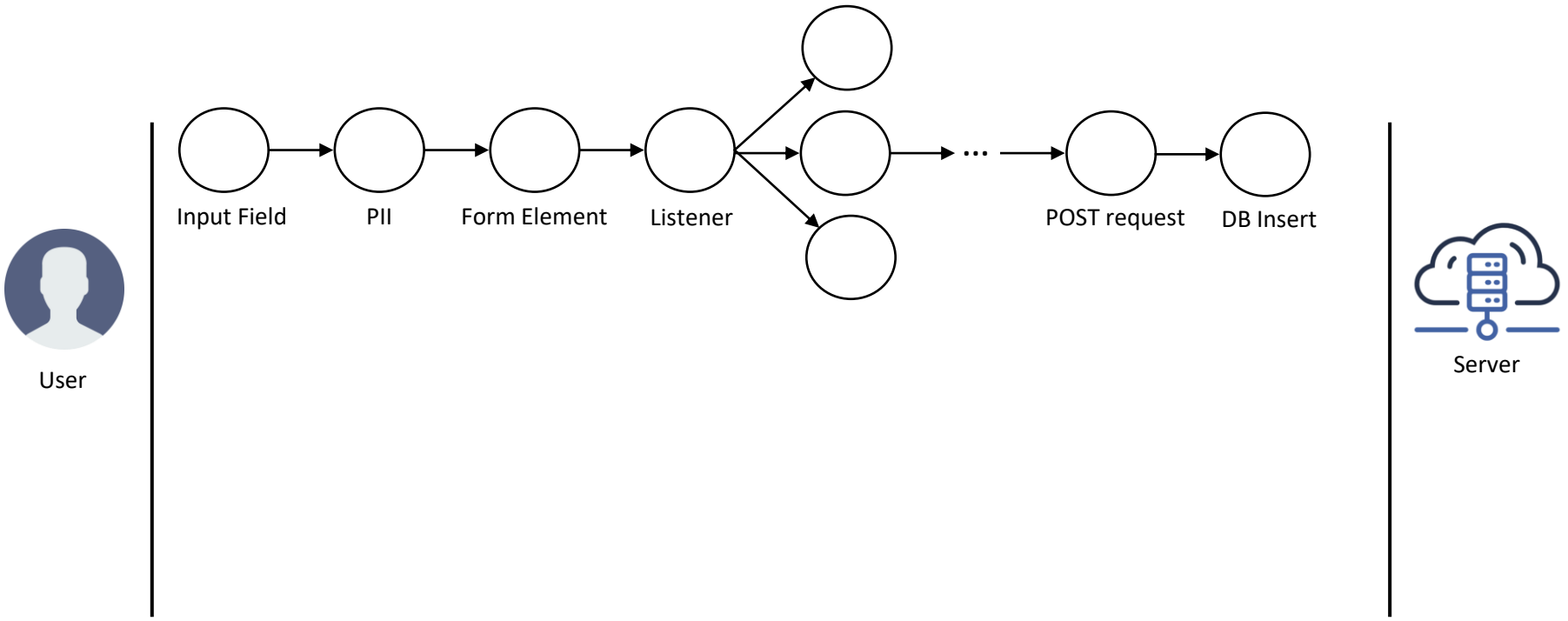
GDPR policies covered by CHKPLUG

Queries for Checking Compliance

ID	Action	Flow
P_{access}	Store	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow storage_{DB}(PII)$
	Export	$counter_{retrieve\ DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow export_{interface}(PII)$
P_{delete}	Store	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow storage_{DB}(PII)$
	Delete	$counter_{retrieve\ DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow delete_{interface}(PII)$
P_{share}	Send	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow intermediate_{node\_1,\ node\_2,....,node\_N} \rightarrow remote_{request}(PII)$
	Disclosure	$doesExist(privacy\text{-}policy)\ \&\&\ disclose\ (sent_{PII}, remote_{URL})$
$P_{security}$	Send	$collect_{HTML}(PII) \mid\mid retrieve_{DB}(PII) \rightarrow secure_{node} \rightarrow remote_{request}(PII)$

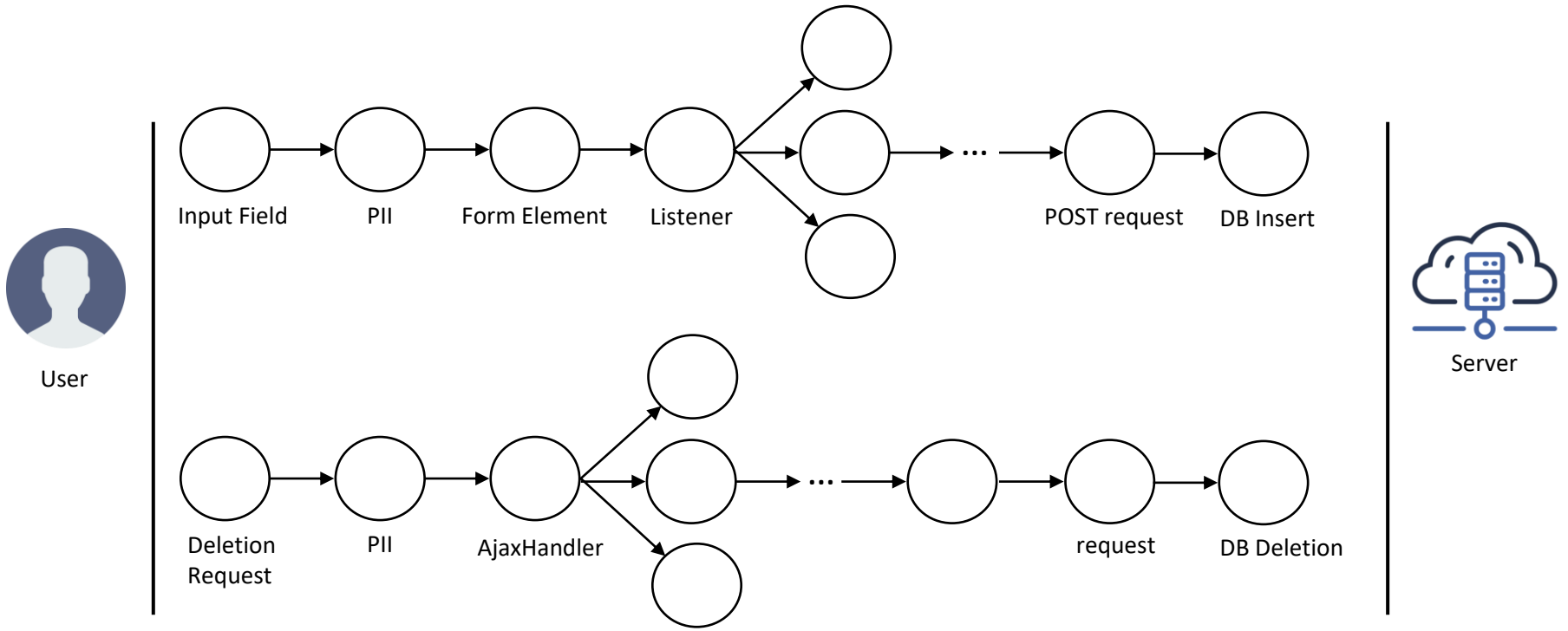
GDPR policies covered by CHKPLUG

Example Query



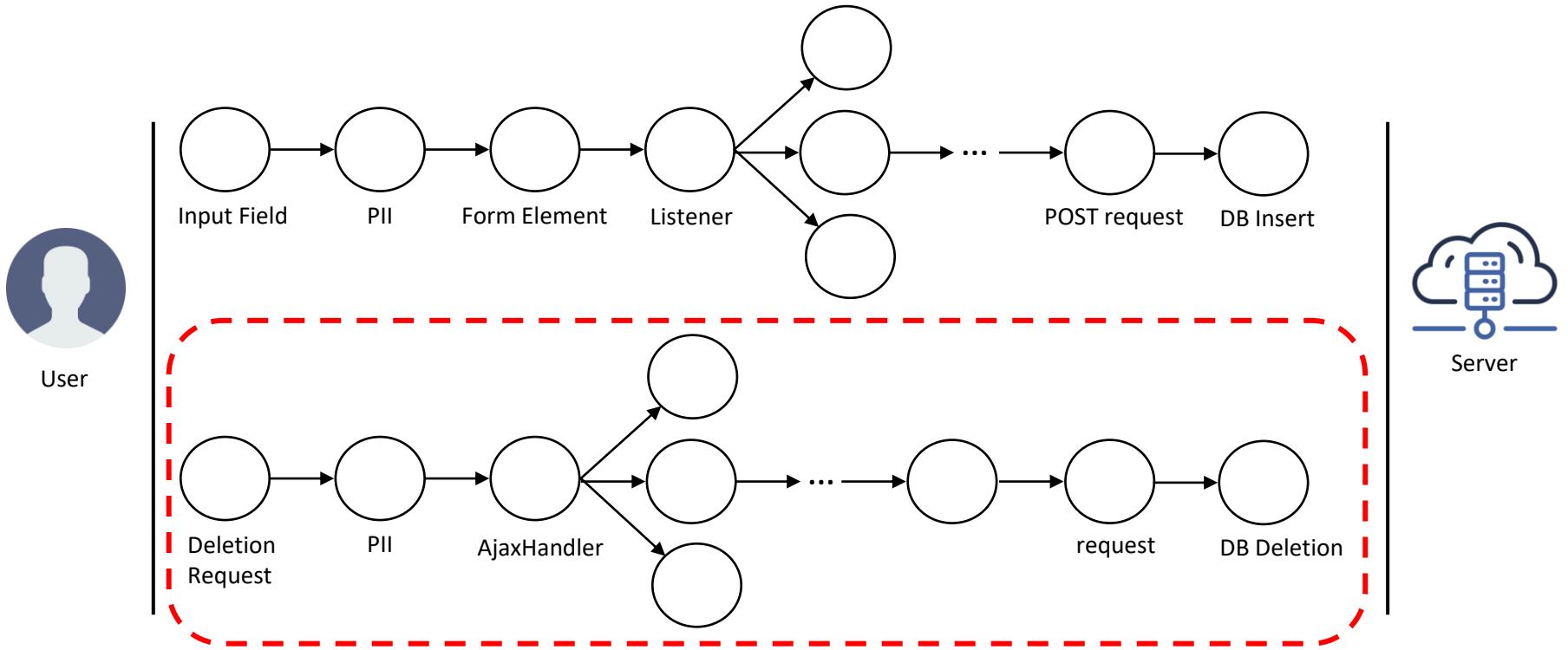
Queries used for identifying violation

Example Query



Queries used for identifying violation

Example Query



Missing Deletion Graph -> Data Deletion Violation

Evaluation

#Compliant	$\#P_{access}$	$\#P_{delete}$	$\#P_{share}$	$\#P_{security}$	#Plugin
✓	200	131	190	190	124
✗	0	69	10	10	76

Ground Truth Labelled Data of 200 plugins

Evaluation

#Compliant	$\#P_{access}$	$\#P_{delete}$	$\#P_{share}$	$\#P_{security}$	#Plugin
✓	200	131	190	190	124
✗	0	69	10	10	76

Ground Truth Labelled Data of 200 plugins

Policy	TP	TN	FP	FN	TPR	TNR
P_{access}	0	200	0	0	100%	100%
P_{delete}	66	127	5	2	97%	96.2%
P_{share}	7	189	1	3	70%	99.5%
$P_{security}$	9	189	1	1	90%	99.5%
Total	**82**	**705**	**7**	**6**	**89.3%**	**98.8%**

Performance of CHKPLUG on 200 plugins

How many plugins violate GDPR as reported by CHKPLUG?

Plugins Violating GDPR

Policy	#Violation	Percentage
P_{access}	0	0%
P_{delete}	368	13.52%
P_{share}	19	0.7%
$P_{security}$	36	1.3%

Measurement analysis on 2,722 plugins

Personal Information	P_{delete}		P_{share}		$P_{security}$	
	WP	Non-WP	WP	Non-WP	WP	Non-WP
USERNAME	310	11	15	1	11	3
EMAIL	29	3	2	0	5	4
PASSWORD	18	1	0	0	0	0
ADDRESS	12	1	2	0	0	0
FIRST_NAME	12	0	2	0	3	3
LAST_NAME	12	1	2	0	3	3
IP	7	2	1	0	5	4
STATE	8	0	2	0	1	0
COUNTRY	7	0	1	0	1	0
PHONE	6	0	0	0	0	1
POSTCODE	4	0	1	0	1	0
CITY	4	0	1	0	1	0
BIRTHDAY	1	0	0	0	0	0

Breakdown of PIIs in GDPR violations

Conclusion

- Cross-programming language analysis
- Generic framework which automatically detects GDPR violations

Introducing GDPR compliance checker in cross-platform programming languages (*e.g. PHP, JavaScript*)

CHKPLUG

Challenge:

Interaction among multiple programming language

Solution:

Cross-language code property graph to identify the information flow path

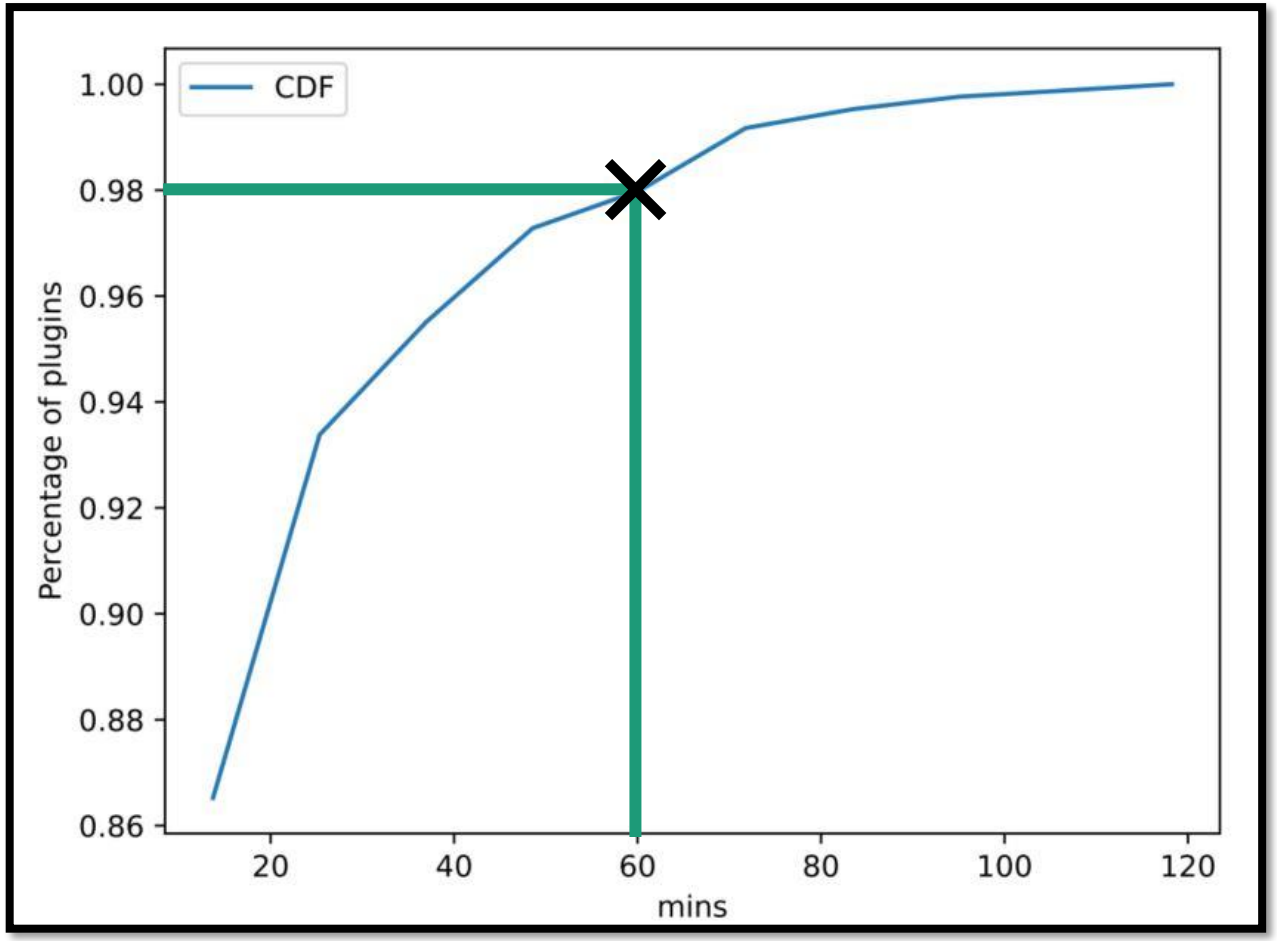
Thank You for listening!

fs5ve@virginia.edu

https://github.com/faysalhossain2007/CHKPLUG

Back Up Slide

Computation Overhead of CHKPLUG



98% of plugins finish analysis within an hour

Why WordPress Plugin?

Popular among website builders

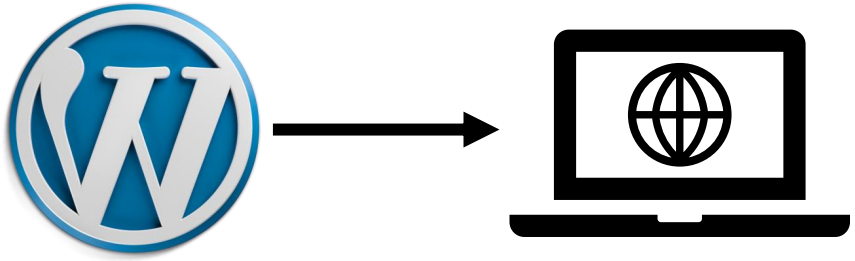
```
    //add loading spinner to home product card
    jQuery('.add-to-cart-in-image a').on('click', function(){
        jQuery(this).text('Adding...');
    });

});

document.addEventListener( 'DOMContentLoaded', function () {
    var elms = document.getElementsByClassName( 'splide' );
    var options = {
        type   : 'loop',
        perPage: 4,
        perMove: 4,
        pagination: false,
        gap: '1em',
        breakpoints: {
            768: {
                perPage: 2,
                perMove: 2,
            },
        }
    };
    for ( var i = 0, len = elms.length; i < len; i++ ) {
        new Splide( elms[ i ], options ).mount();
    }

} );
```

Availability of source code



Website owners integrate WordPress plugin without modification

Key Insights

Match WordPress plugin behaviors represented as **data- and control-flows** and extracted via **cross-language** static analysis against a set of predefined rules of **GDPR** articles

Previous Work Will Not Work

Previous Work	PHP	JavaScript	HTML	SQL
NAVEX [A]	✓	✗	✗	✗
ODGen [B]	✗	✓	✗	✗

A. NAVEX: Precise and scalable exploit generation for dynamic web applications. Alhuzali et al. (USENIX SEC 2018)
B. Mining Node.js Vulnerabilities via Object Dependence Graph and Query. Li et al. (USENIX SEC 2022)