CE 815 – Secure Software Systems

ML-Based Vulnerability Detection Methods (Devign)

Mohammad Haddadian/Mehdi Kharrazi
Department of Computer Engineering
Sharif University of Technology

Acknowledgments: Some of the slides are fully or partially obtained from other sources. A reference is noted on the bottom of each slide, when the content is fully obtained from another source. Otherwise a full list of references is provided on the last slide.

Introduction

- As announced by the 2021 report, 98% of codebases contain open source components
- Meanwhile, 84% of code-bases have at least one open-source vulnerability
- 60% of them contain high-risk vulnerabilities
- By exploiting the OSS vulnerabilities reported in the vulnerability databases (e.g., NVD), attackers can perform “N-day” attacks against unpatched software systems

Problem

- A large volume of OSS security patches (e.g., GitHub commits fixing vulnerabilities) are silently released.

```
From 7f9822a48213dd2feca845dbbb6bcb8beb9550de
Subject: [PATCH] Add blinding to a DSA signature

This is based on side channel attacks demonstrated by (NCC Group)
for ECDSA which are likely to be able to be applied to DSA.
```

Not report to NVD

```
From 41bdc78544b8a93a9c6814b8bbbfef966272abbe
Subject: [PATCH] x86/tls: Validate TLS entries to protect espfix

Installing a 16-bit RW data segment into the GDT defeats espfix.
AFAICT this will not affect glibc, Wine, or dosemu at all.
```

Not provide explicit description

- Average users need to timely detect and apply security patches before being exploited by armored attackers.

Previous Solutions and Limitations

An OSS Patch

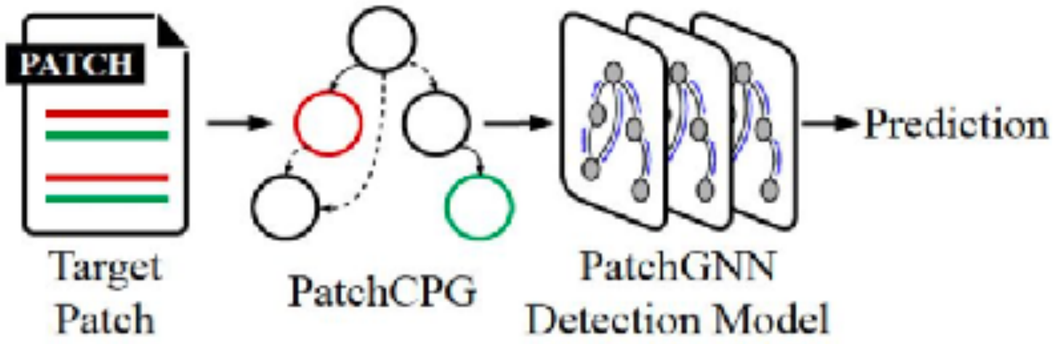
Natural Language Documentation
(Commit Message/Changelog)

- **Mining security keywords**
  Requiring well-maintained doc.

Source Code Changes

- **Regarding code as sequential data**
  Losing important semantics.
- **Our solution: representing code as graph**
  Retaining rich patch structural info.

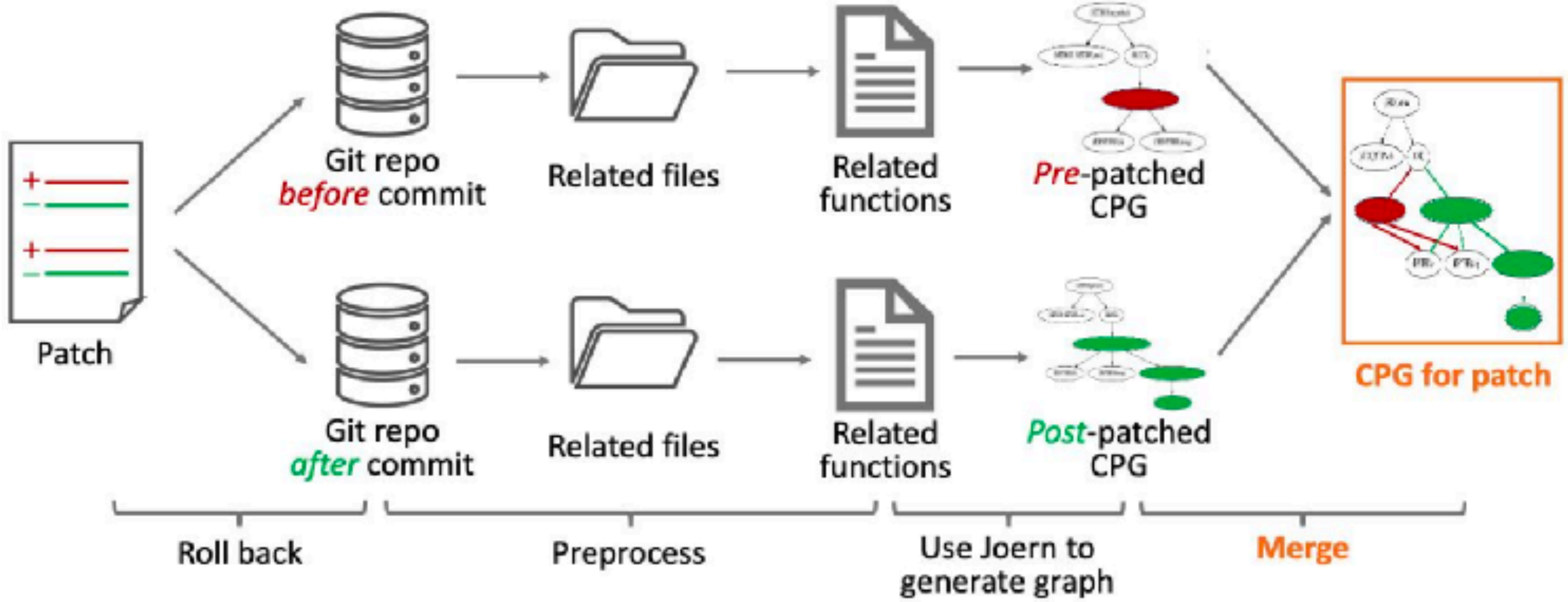
A Graph-Based Security Patch Detection System



- PatchCPG: a new graph representation of inherent code change structures.
    - Syntax and semantics: AST + control & data dependency graph.
    - Changes and relations with context: pre-patch + post-patch graph.
- PatchGNN: a tailored GNN model to capture diverse patch structural information

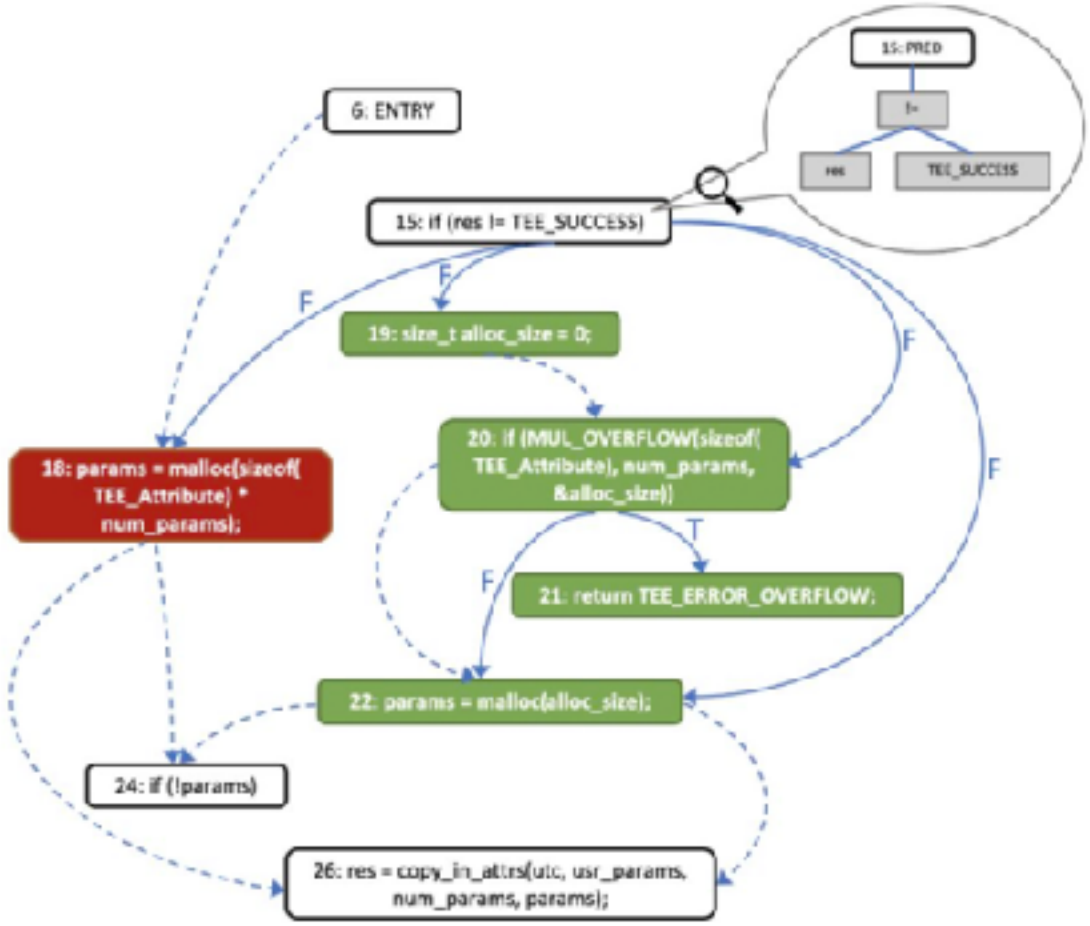
PatchCPG: From Patch to Graph

- Challenge: how to construct PatchCPG?



Patch Code Property Graph (PatchCPG)

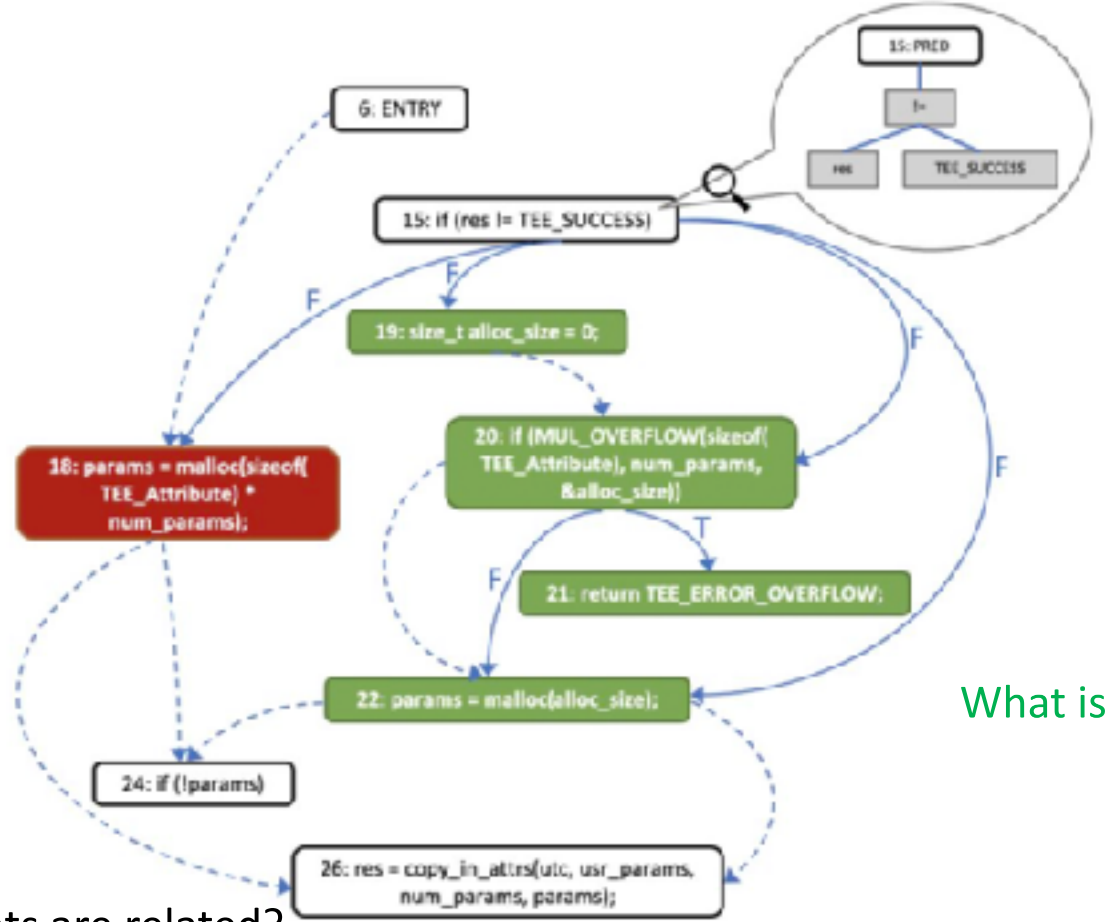
- A joint graph encodes rich patch structural information.



Patch Code Property Graph (PatchCPG)

- A joint graph encodes rich patch structural information.

What is deleted?



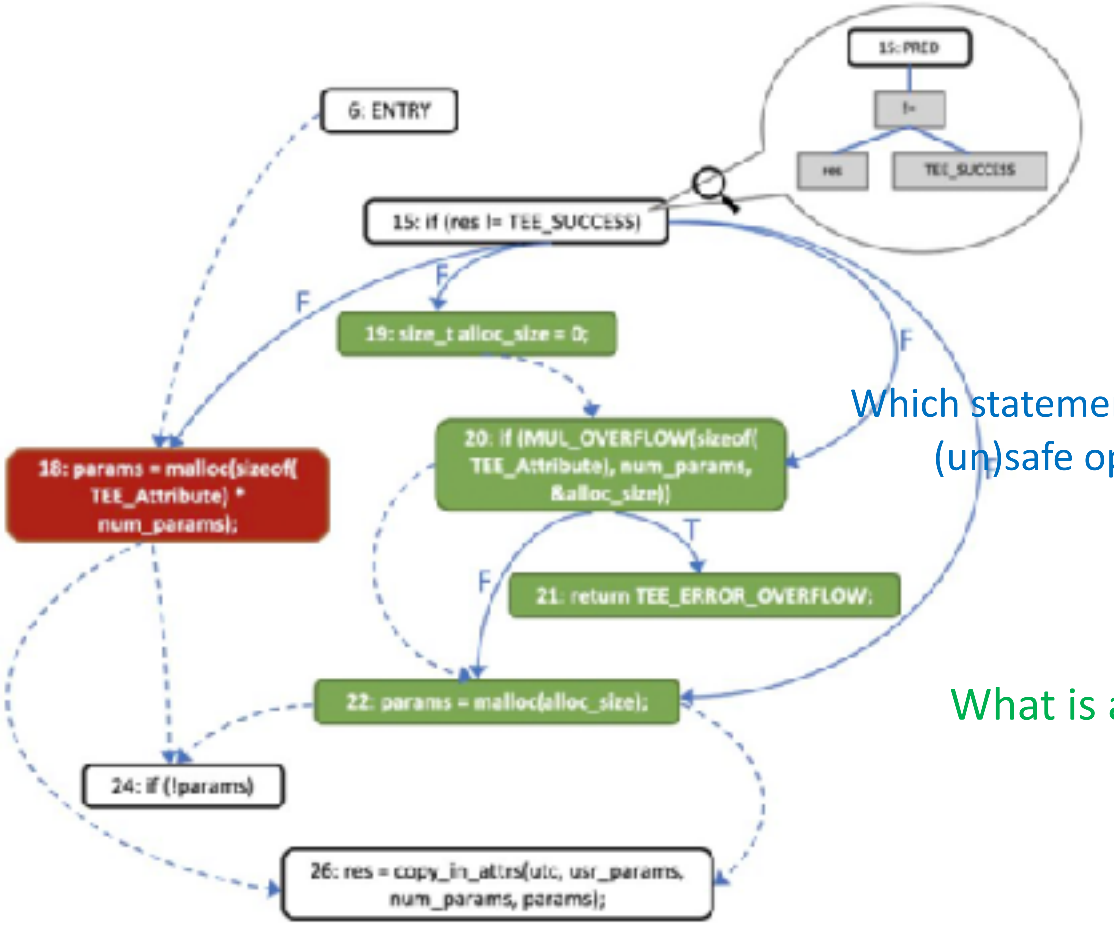
What is added?

What context statements are related?

Patch Code Property Graph (PatchCPG)

- A joint graph encodes rich patch structural information.

What is deleted?



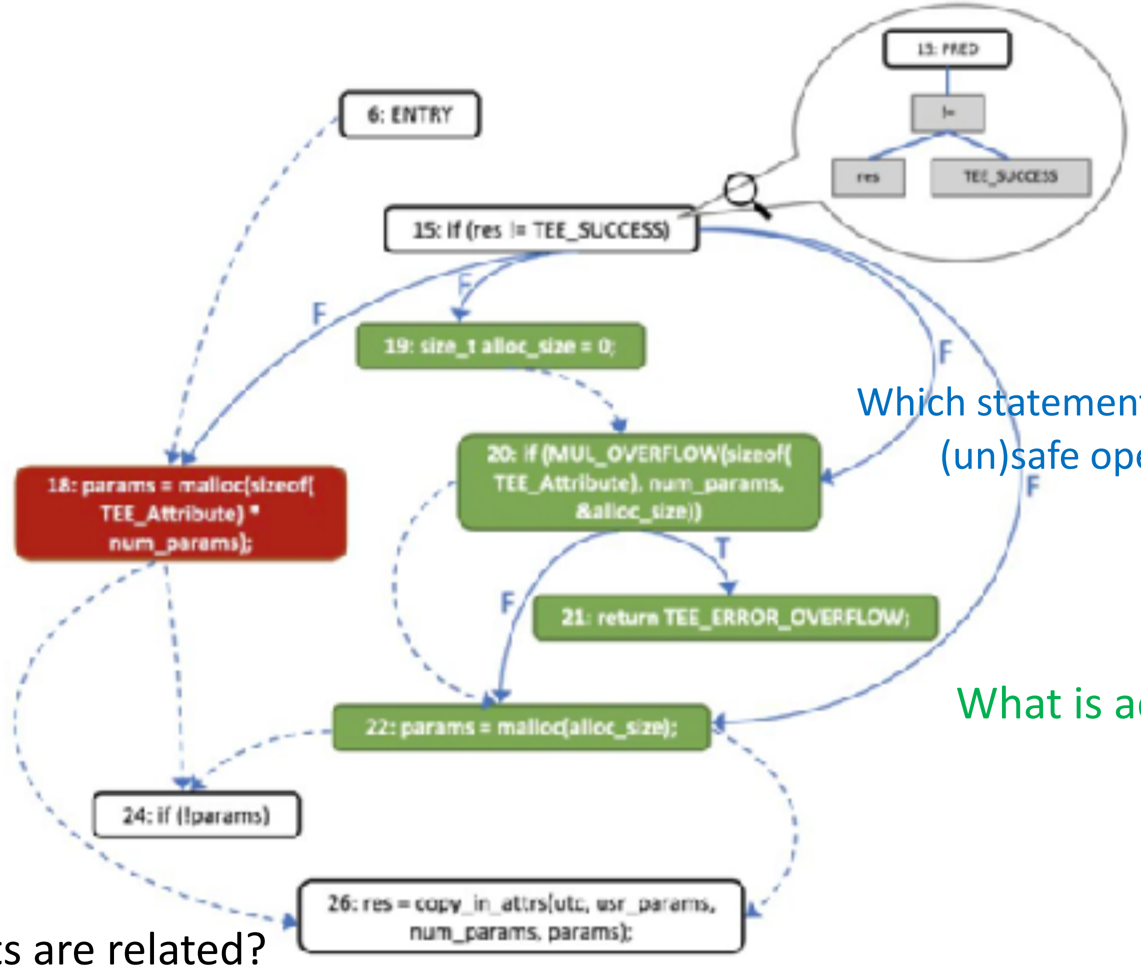
Which statements decide the (un)safe operation?

What is added?

What context statements are related?

Patch Code Property Graph (PatchCPG)

- A joint graph encodes rich patch structural information.



Which statements decide the (un)safe operation?

What is deleted?

What is added?

Where the value comes from?

What context statements are related?

Patch Code Property Graph (PatchCPG)

- A joint graph encodes rich patch structural information.

How each statement looks like?

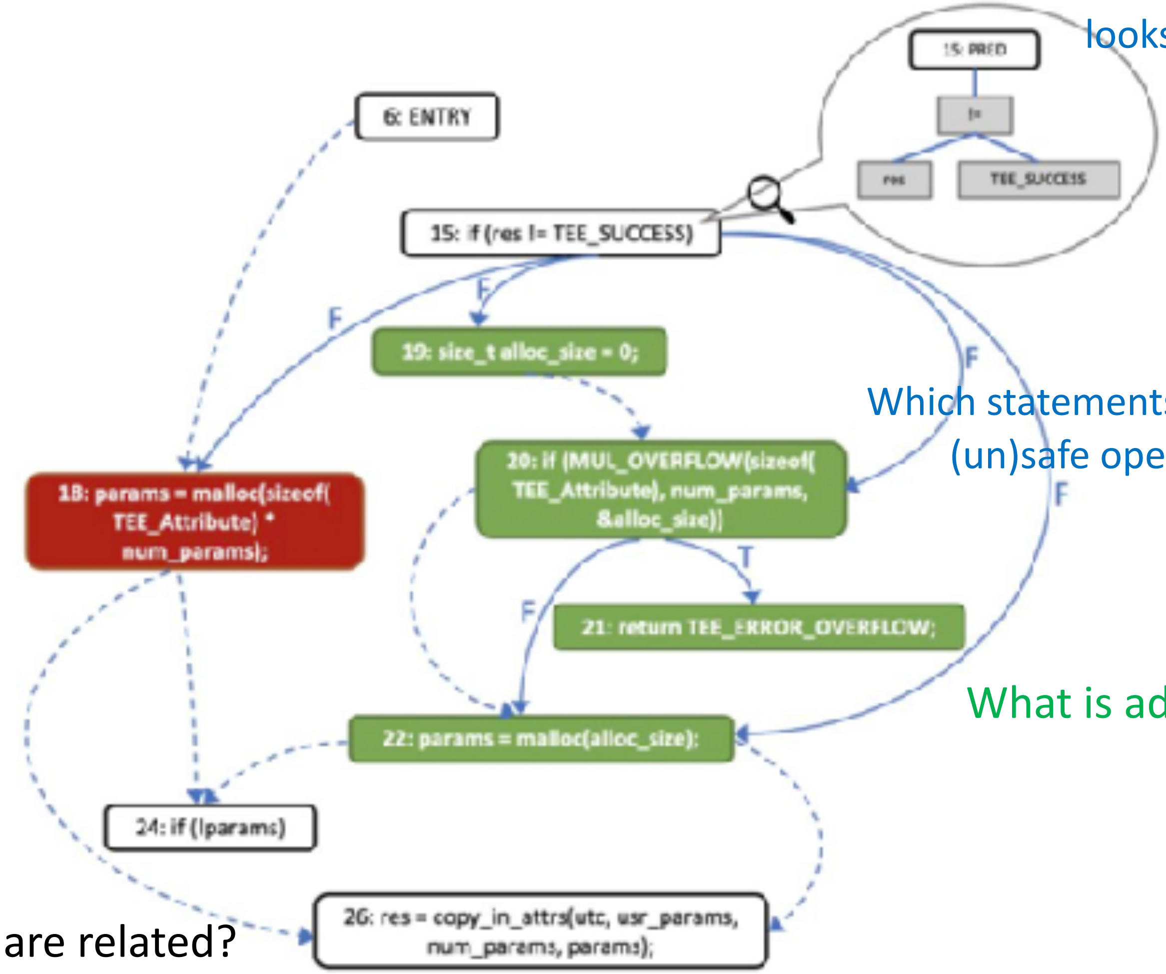
Which statements decide the (un)safe operation?

What is deleted?

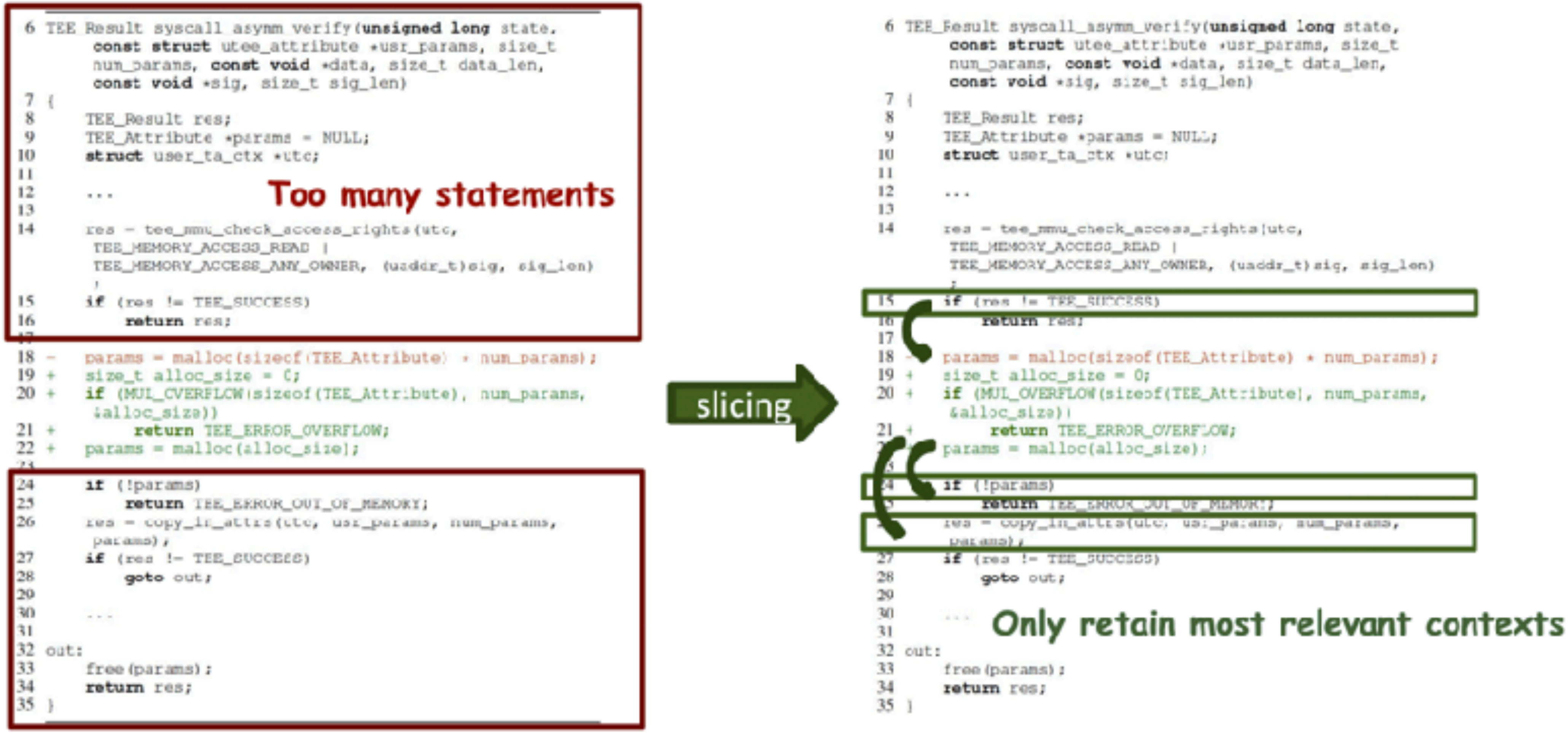
What is added?

Where the value comes from?

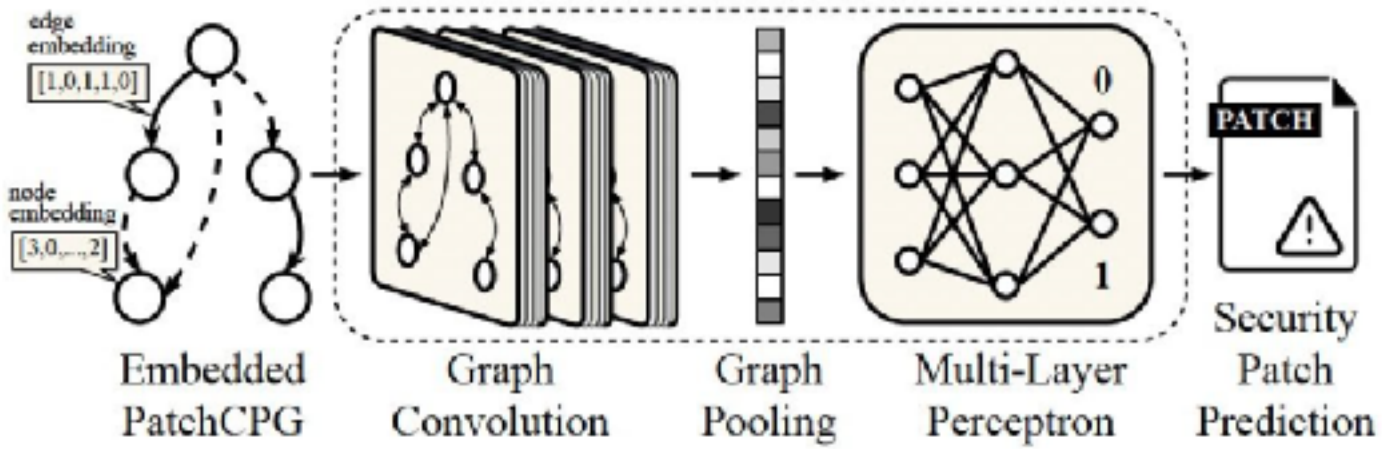
What context statements are related?



Reducing Noisy Information by Slicing



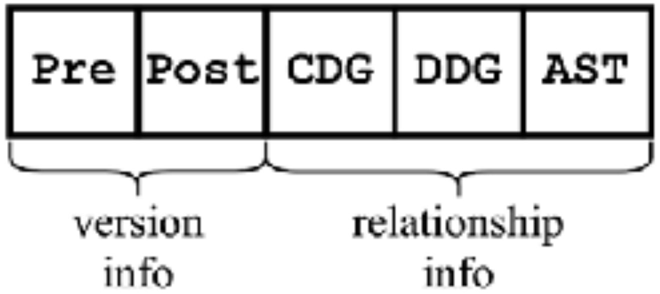
PatchGNN: Detect Security Patches from PatchCPGs



- Challenge 1: how to embed the PatchCPGs?
- Challenge 2: how to learn multiple attributes (CDG/DDG/AST/pre/post)?

PatchCPG Embeddings

- Node Embedding
  - 20-dimensional vulnerability features.
    - code snippet metadata
    - identifier and literal features
    - control flow features
    - operator features
    - API features
- Edge Embedding
  - 5-dimensional binary vector.



e.g., [1,1,0,1,0] means the edge is a context edge of data dependency.

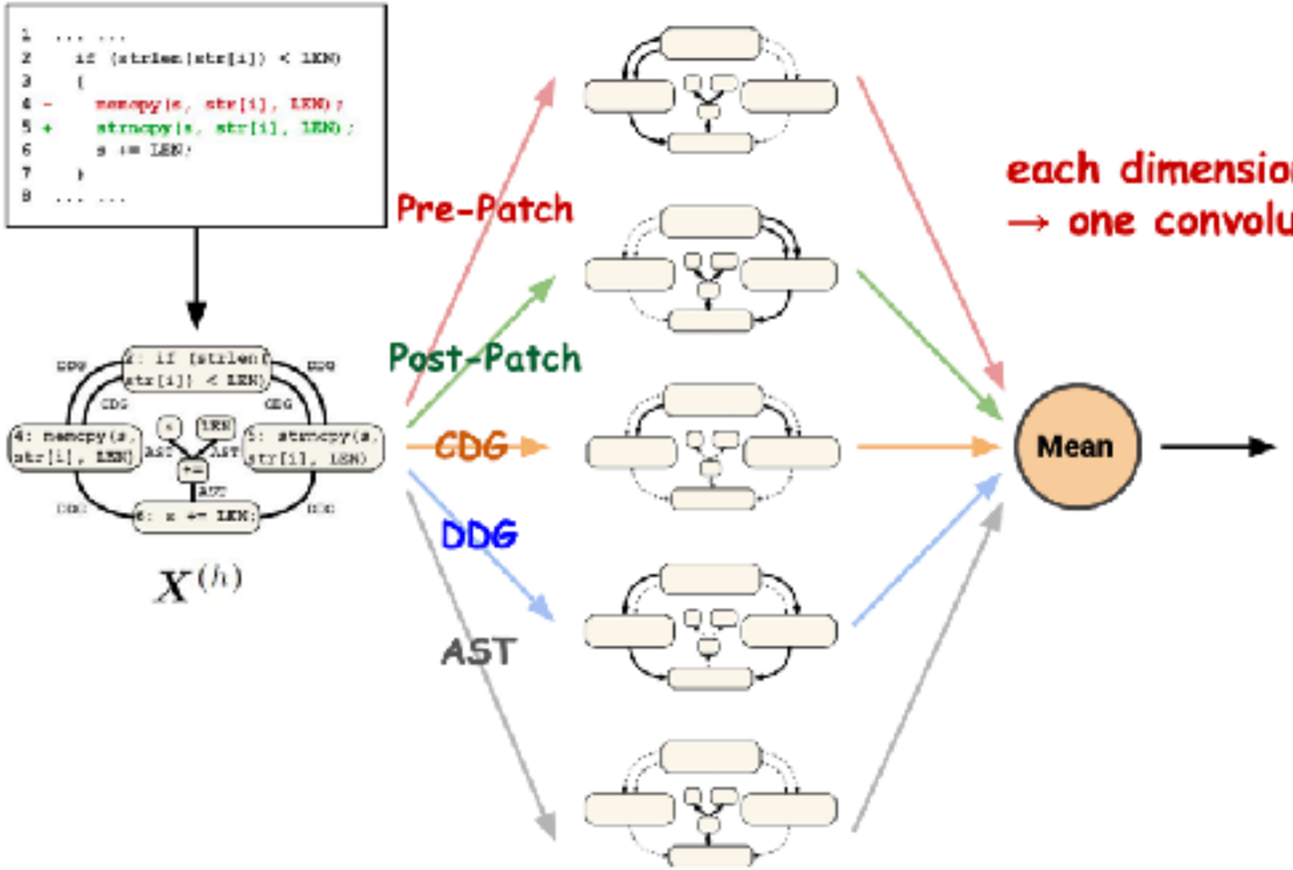
PatchCPG Embeddings

TABLE I: The involved tokens or sub-tokens of the control flow features, the operator features, and the API features.

Features	Matched Tokens or Sub-tokens
condition	if, switch
loop	for, while
jump	return, break, continue, goto, throw, assert
arithmetic[†]	++, --, +, -, *, /, %, =, +=, -=, *=, /=, %=
relational	==, !=, >=, <=, >, <
logical	&&, ||, !, and, or, not
bitwise[*]	&, |, <<, >>, ~, ∧, bitand, bitor, oxr
memory API	alloc, free, mem, copy, new, open, close, delete, create, release, sizeof, remove, clear, dequene, enquene, detach, attach
string API	str, string
lock API	lock, mutex, spin
system API	init, register, disable, enable, put, get, up, down, inc, dec, add, sub, set, map, stop, start, prepare, suspend, resume, connect,

[†] Operator * is determined as dereference operator or arithmetic operator.
[*] Operator & is determined as address-of operator or bitwise operator.

PatchGNN with Multi-Attribute Graph Convolution



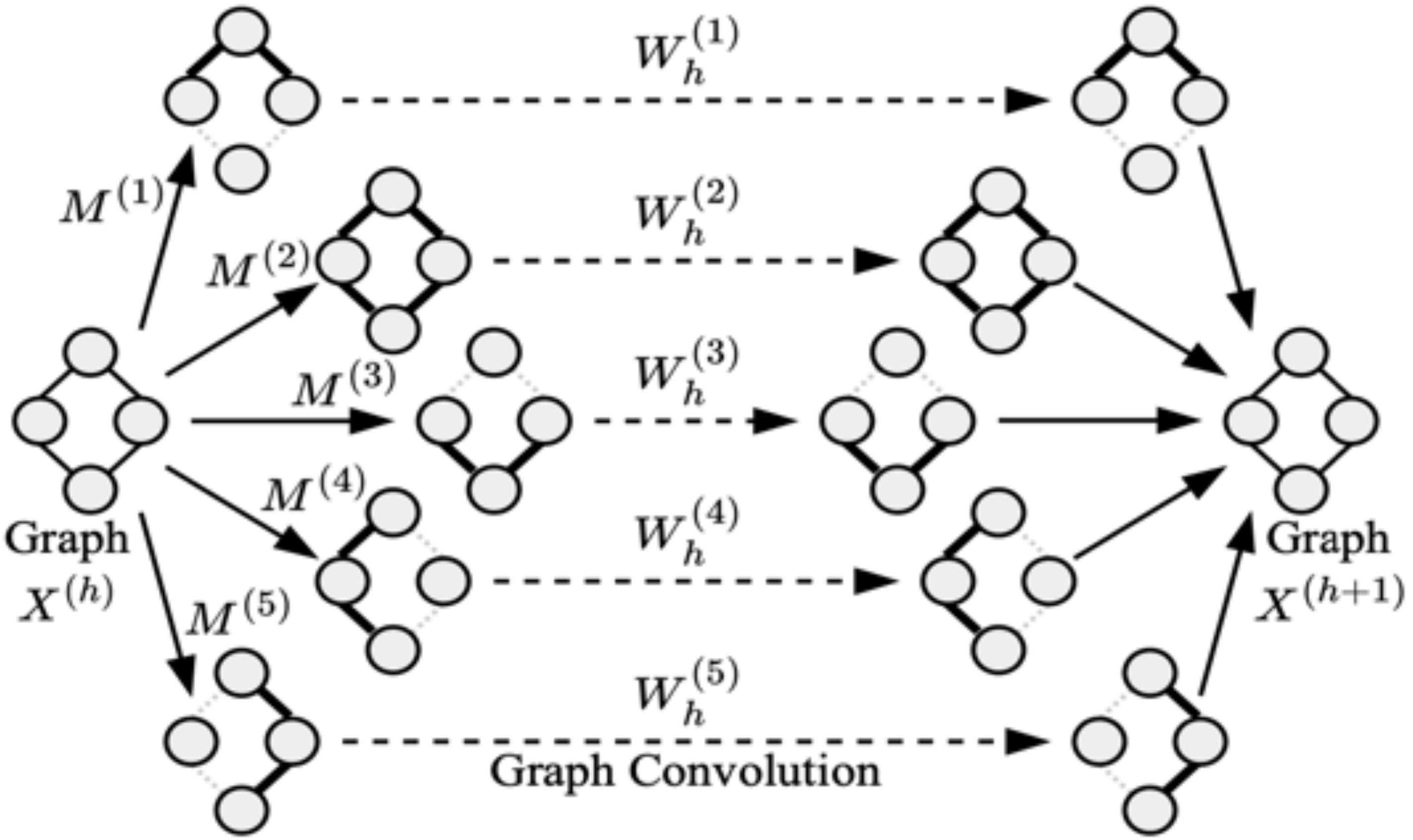
each dimension of edge embeddings
→ one convolution channel

$$X^{(h+1)} = \frac{1}{5}\sum_{k=1}^{5} \sigma(\tilde{L}_k \cdot X^{(h)} \cdot W_k^{(h)})$$

where

$$\tilde{L}_k = \tilde{D}_k^{-1/2} \cdot \tilde{A}_k \cdot \tilde{D}_k^{-1/2}$$
$$\tilde{A}_k = A \odot M^{(k)} + I$$
$$\tilde{D}_k = diag(\sum_j A_{:j})$$

PatchGNN with Multi-Attribute Graph Convolution



Implementation & Evaluation

Implementation

- **5K new LoC** in Scala and Python on top of *Joern* parser and *PyTorch* library.

Datasets:

- PatchDB: 12K security patches from 300+ GitHub repos.
- SPI-DB: 10K security patches from FFmpeg and QEMU.

Evaluation:

- Compared with sequential-based patch detector.
- Compared with vulnerability detection methods.
- Case study on four popular OSS repos.

Compared with Sequential-based Solution

- Accuracy 10.8%↑
- F-1 score: 0.096↑
- Precision: 28.82%↑
- False Positive Rate: 14.62%↓

Method	Dataset	General Metrics		Special Metrics	
		Accuracy	F1-score	Precision	FP Rate
TwinRNN [1][2]	PatchDB	69.60%	0.461	48.45%	19.67%
	SPI-DB	56.37%	0.512	49.07%	41.57%
GraphSPD	PatchDB	80.39%	0.557	77.27%	5.05%
	SPI-DB	63.04%	0.503	63.96%	19.16%

[1] PatchRNN: A Deep Learning-Based System for Security Patch Identification.
[2] SPI: Automated Identification of Security Patches via Commits.

Compared with Vulnerability Detection Solutions

- **2.5 - 50x** detection rate of vulnerability detectors.

Method	# $\text{Vul}_{\text{pre-patch}}$	# $\text{Vul}_{\text{post-patch}}$	# $\text{Patch}_{\text{security}}$	TP Rate
Cppcheck[3]	3	1	2	0.54%
flawfinder[4]	109	108	1	0.27%
ReDeBug[5]	29	29	0	0.00%
VUDDY[6]	22	16	21	5.71%
VulDeePecker[7]	3	0	3	0.82%
GraphSPD	-	-	53	14.40%

[3] Cppcheck. https://cppcheck.sourceforge.io.
[4] flawfinder. https://dwheeler.com/flawfinder/.
[5] Redebug: finding unpatched code clones in entire os distributions.
[6] VUDDY: A scalable approach for vulnerable code clone discovery.
[7] VulDeePecker: A deep learning- based system for vulnerability detection.

Case Study on False Negetives

```
commit 247d30a7dba6684ccce4508424f35fd58465e535
if (!s1->current_frame.data[0]
    ||s->width != s1->width
    ||s->height!= s1->height) {
    if (s != s1)
-       copy_fields(s, s1, golden_frame, current_frame);
+       copy_fields(s, s1, golden_frame, keyframe);
    return -1;
}
```

Listing 4: Security patch for a double free (CVE-2011-3934).

```
commit 360e95d45ac4123255a4c796db96337f332160ad
if (priv->cac_id_len) {
    serial->len=MIN(priv->cac_id_len, SC_MAX_SERIALNR);
-   memcpy(serial->val,priv->cac_id,priv->cac_id_len);
+   memcpy(serial->val,priv->cac_id,serial->len);
    SC_RETURN(card->ctx,SC_DEBUG_NORMAL,SC_SUCCESS);
}
```

Listing 5: A patch with similar patterns (CVE-2018-16393).

Case Study on OSS Repos

- **NGINX**: detect 21 security patches (Precision: 78%).

Changes w/	CVE	Total Commits	Valid Commits	Detected S.P.	Confirmed S.P.	Precision
1.19.x	3	180	127	7	6	86%
1.17.x	3	134	82	4	3	75%
1.15.x	1	203	120	7	4	57%
1.13.x	1	270	157	9	8	89%
Sum.	8	787	486	27	21	78%

- **Xen**: detect 29 security patches (Precision: 55%).
- **OpenSSL**: detect 45 security patches (Precision: 66%).
- **ImageMagick**: detect 6 security patches (Precision: 46.2%).

Conclusion

- Silent security patches can be leveraged by attackers to launch N-day attacks.
- GraphSPD presents patches as graphs and identifies security patches with graph learning, achieving higher accuracy and fewer false alarms.
- GraphSPD can be extended to other programming languages.

Acknowledgments

- [GraphSPD] GraphSPD: Graph-Based Security Patch Detection with Enriched Code Semantics, S. Wang, X. Wang, K. Sun, S. Jajodia, H. Wang, Q. Li, IEEE S&P 2023.