

How Reliable is the Crowdsourced Knowledge of Security Implementation?

Presenter: Jordan Gillard



A little bit of background



A little bit of background

- Stack Overflow is a site for developers to help each other solve programming issues.



What is the problem?



What is the problem?

- Stack Overflow unintentionally propagates security vulnerabilities at an alarming rate.



What is the problem?

- Stack Overflow unintentionally propagates security vulnerabilities at an alarming rate.
- How much can we trust the security implementation suggestions on Stack Overflow?



What is the problem?

- Stack Overflow unintentionally propagates security vulnerabilities at an alarming rate.
- How much can we trust the security implementation suggestions on Stack Overflow?
- Can developers rely on the community to find vulnerabilities and identify secure counterparts?



Some more information on Stack Overflow



Some more information on Stack Overflow

- Relies on it's community to crowdsource knowledge.



Some more information on Stack Overflow

- Relies on it's community to crowdsource knowledge.
- Contains insecure answers - some of which are scored highly.



What technique is proposed?

What technique is proposed?

Category	Parameter	Insecure
SSL/TLS	HostnameVerifier	allow all hosts
	Trust Manager	trust all
	Version	<TLSv1.1
	Cipher Suite	RC4, 3DES, AES-CBC MD5, MD2
	OnReceivedSSLERror	proceed
Symmetric	Cipher/Mode	RC2, RC4, DES, 3DES, AES/ECB, Blowfish
	Key	static, bad derivation
	Initialization Vector (IV)	zeroed, static, bad derivation
	Password Based Encryption (PBE)	<1k iterations, <64-bit salt, static salt
Asymmetric	Key	RSA < 2,048 bit, ECC < 224 bit
Hash	PBKDF	<SHA224, MD2, MD5
	Digital Signature	SHA1, MD2, MD5
	Credentials	SHA1, MD2, MD5
Random	Type	Random
	Seeding	setSeed→nextBytes, setSeed with static values



Research Questions Raised



Research Questions Raised

- How prevalent are insecure coding suggestions on Stack Overflow?



Research Questions Raised

- How prevalent are insecure coding suggestions on Stack Overflow?
- Do community dynamics or Stack Overflow's reputation mechanism help developers choose secure answers over insecure ones?



Research Questions Raised

- How prevalent are insecure coding suggestions on Stack Overflow?
- Do community dynamics or Stack Overflow's reputation mechanism help developers choose secure answers over insecure ones?
- Do secure coding suggestions have more duplicates than insecure ones?



Research Questions Raised

- How prevalent are insecure coding suggestions on Stack Overflow?
- Do community dynamics or Stack Overflow's reputation mechanism help developers choose secure answers over insecure ones?
- Do secure coding suggestions have more duplicates than insecure ones?
- Why did users suggest duplicated secure or insecure answers on Stack Overflow?



How Security is Evaluated



How Security is Evaluated

- SSL/TLS

How Security is Evaluated

```
// Create a trust manager that does not validate certificate chains (trust all)
private TrustManager[] trustAllCerts = new TrustManager[] {
    new X509TrustManager() {
        public java.security.cert.X509Certificate[]
            getAcceptedIssuers() {return null;}
        public void checkClientTrusted(...) {}
        public void checkServerTrusted(...) {}    }};
public ServiceConnectionSE(String url) throws IOException {
    try {
        // Use the default TLSv1.0 protocol
        SSLContext sc = SSLContext.getInstance("TLS");
        // Install the trust-all trust manager
        sc.init(null, trustAllCerts, new java.security.
            SecureRandom()); ... } ...
    connection = (HttpsURLConnection) new URL(url).
        openConnection();
    // Use AllowAllHostnameVerifier that allows all hosts
    ((HttpsURLConnection) connection).setHostnameVerifier(new
        AllowAllHostnameVerifier());    }
```



How Security is Evaluated

- SSL/TLS
- Symmetric cyphers

How Security is Evaluated

```
// Declare a key parameter with a static value
private static byte[] key = "12345678".getBytes();
// Declare an IV parameter with a static value
private static byte[] iv = "12345678".getBytes();
public static String encrypt(String in) {
    String cypert = in;
    try {
        IvParameterSpec ivSpec = new IvParameterSpec(iv);
        // Create a secret key with the DES cipher
        SecretKeySpec k = new SecretKeySpec(key, "DES");
        // Declare a DES cipher
        Cipher c = Cipher.getInstance("DES/CBC/PKCS7Padding");
        c.init(Cipher.ENCRYPT_MODE, k, ivSpec);
        ... } }
```



How Security is Evaluated

- SSL/TLS
- Symmetric cyphers
- Asymmetric cyphers



How Security is Evaluated

- SSL/TLS
- Symmetric cyphers
- Asymmetric cyphers
- Hash



How Security is Evaluated

- SSL/TLS
- Symmetric cyphers
- Asymmetric cyphers
- Hash
- Random



How Security is Evaluated

```
byte [] keyStart = "encryption key".getBytes();  
SecureRandom sr = SecureRandom.getInstance("SHA1PRNG");  
sr.setSeed(keyStart);
```



Methodology



Methodology

- Extracted relevant code snippets from Stack Overflow



Code Extraction



Code Extraction

- Check question tags and answer posts



Code Extraction

- Check question tags and answer posts
- Filtering by question tags and security APIs



Clone Detection



Clone Detection

- 25,855 code snippets from 23,224 posts



Clone Detection

- 25,855 code snippets from 23,224 posts
- 2,657 clone groups that contained 8,690 code snippets



Code Labeling

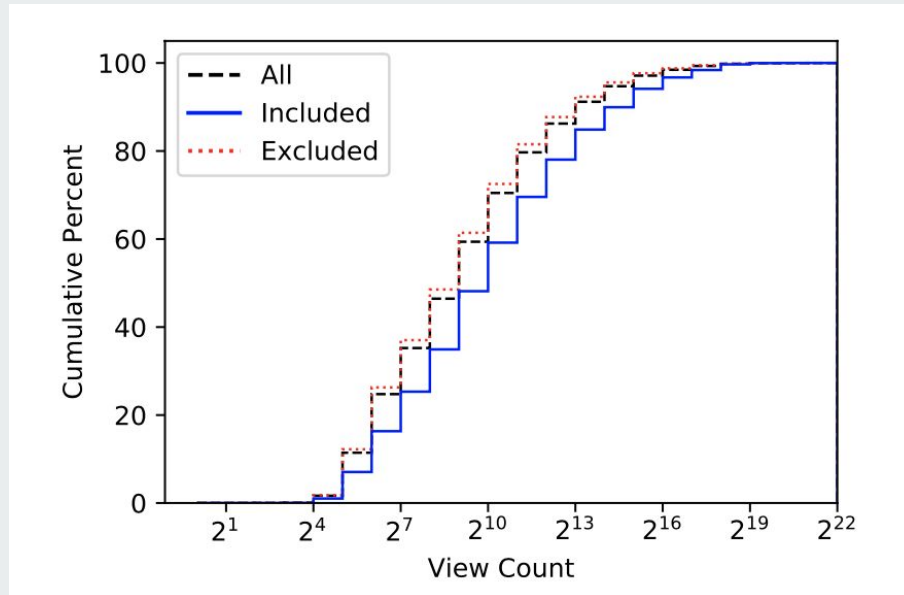
- 8,690 code snippets
- Labelled secure, insecure, or irrelevant



Code Labeling

	Secure	Insecure	Mixed	Irrelevant	Total
# of clone groups	587	326	40	1,704	2,657
# of snippets	1,802	1,319	0	5,569	8,690
# of answer posts	785	644	0	2,133	3,562

Verifying the Prevalence of Sampled Posts



Major Findings

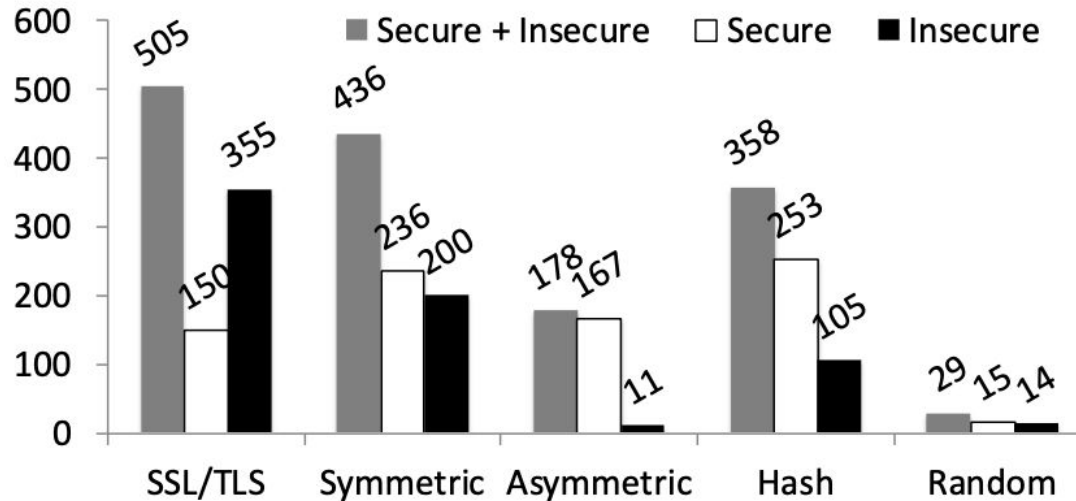
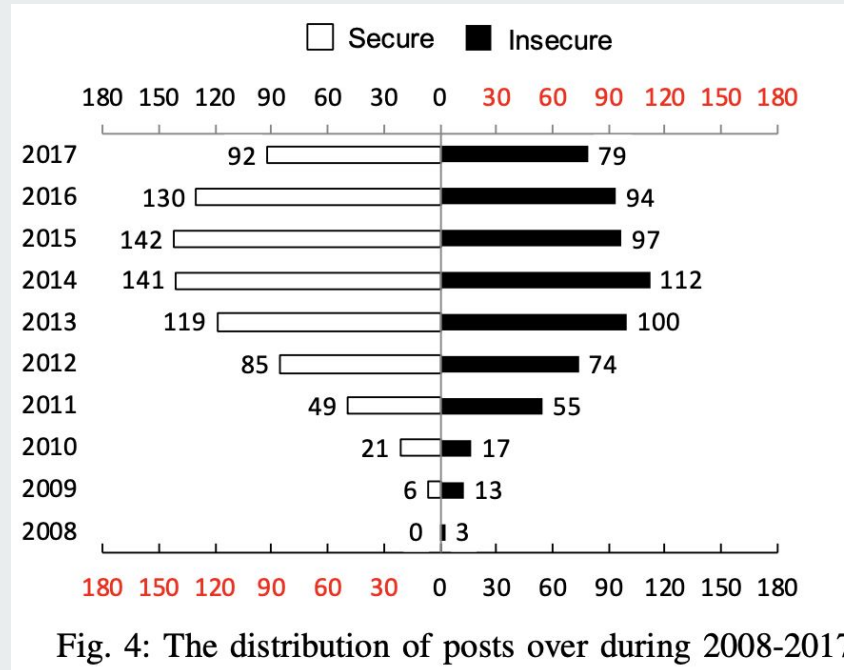


Fig. 3: The distribution of posts among different categories

Major Findings





Major Findings



Major Findings

- Vulnerabilities found *after* the insecure code was posted.



Major Findings

- Vulnerabilities found *after* the insecure code was posted.
- Only few secure answers correct outdated insecure ones.



Major Findings



Major Findings

- Secure posts have higher rep, except for SSL/TLS posts



Major Findings

- Secure posts have higher rep, except for SSL/TLS posts
- Insecure posts have higher scores, more comments, favorites, and views.



Why do insecure posts have higher scores?



Why do insecure posts have higher scores?

- Software engineers need a quick fix



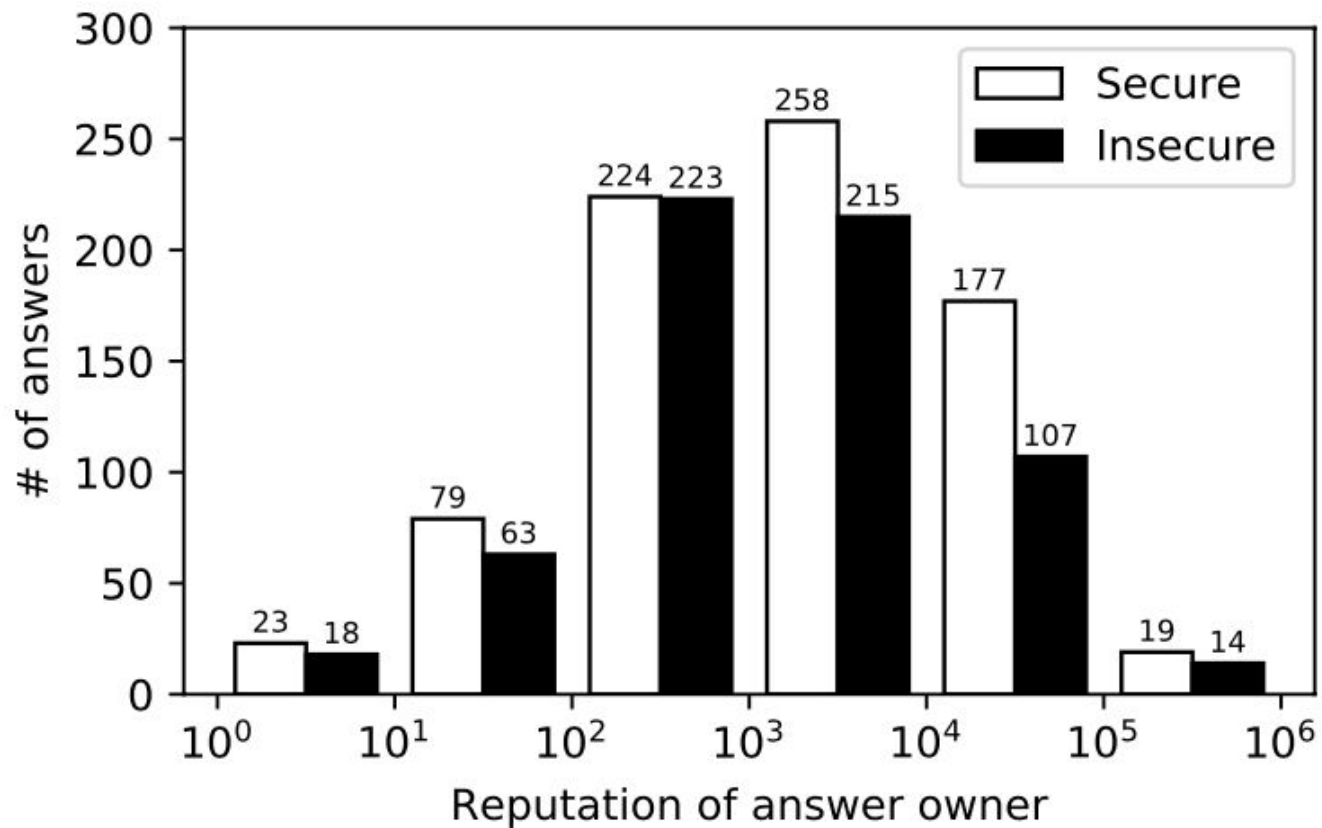
Why do insecure posts have higher scores?

- Software engineers need a quick fix
- Insecure algorithms are still supported by Java libraries



Why do insecure posts have higher scores?

- Software engineers need a quick fix
- Insecure algorithms are still supported by Java libraries
- Insecure posts have had more time to accumulate points






Why do high rep users post insecure code?

- Users earned scores for being an expert in areas other than security.




Why do high rep users post insecure code?

- Users earned scores for being an expert in areas other than security.
- Users cannot rely on Stack Overflow's reputation mechanism to identify secure answers.



Comparison of Accepted Answers



Comparison of Accepted Answers

- An accepted answer is not a good indication that it is secure



Code Duplication



Code Duplication

- Repeated code does not indicate it is secure code



Why do user's copy/paste code?



Why do user's copy/paste code?

- Users suggest general “best-practices” to try and get points



Why do user's copy/paste code?

- Users suggest general “best-practices” to try and get points
- Users did not post insecure code to mislead others



Related work



Related work

- Security API misuses



Related work

- Security API misuses
- Developer studies



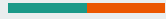
Related work

- Security API misuses
- Developer studies
- Empirical studies on Stack Overflow



Related work

- Security API misuses
- Developer studies
- Empirical studies on Stack Overflow
- Duplication Detection Related to Stack Overflow or Vulnerabilities



Concluding remarks & recommendations



Concluding remarks & recommendations

- Tool builders



Concluding remarks & recommendations

- Tool builders
- Stack Overflow developers



Concluding remarks & recommendations

- Tool builders
- Stack Overflow developers
- Designers of crowdsourcing platforms



In conclusion



In conclusion

- Secure and insecure advice is about even



In conclusion

- Secure and insecure advice is about even
- The reputation mechanism isn't good for promoting secure suggestions



In conclusion

- Secure and insecure advice is about even
- The reputation mechanism isn't good for promoting secure suggestions
- The reputation mechanism encourages users to provide duplicated code



Discussion