





JSReg 0.1

- Started to create a JavaScript parser
- Recreating JavaScript within JavaScript
- Why do you need to sandbox JavaScript?
- There's got to be a better way?



- Since we want to execute JavaScript in JavaScript why not use the engine itself?
- Instead of parsing, why not rewrite instead!
- Char by char seems longwinded especially when we have regex



- You don't know a value until it's executed
- E.g. `x=func();obj[x];` // what is x?
- RegEx isn't good for recursive values like square bracket notion in JavaScript
- Regex is slow



- I want to share JavaScript safely
- Browsers don't provide the tools to do it (ES5 is getting there)
- SOP has expired, we need something else. That something doesn't exist





The design cont.

Main regexp is constructed using all the others



- Main regex is run in global mode without start or end anchor: /(…)|…/g
- The regex starts from the next valid match
- Skips stuff that isn't matched
- Regex lastIndex keeps track of position



The design cont.

Matching string is either rewritten or returned literally for performance



- The rewrite can be called recursively if required (but it gets complicated) considering the left context of the match
- JavaScript has no lookbehind!
- Hard to know what context the code your matching is in. E.g. `{}[1,2,3]` is an array
- `1,{}[1]` is a Object literal



- Code is converted from variable to \$variable\$
- Square bracket notion is rewritten from obj[x] to \$obj\$[JSREG_FUNC.gp(\$x\$)]
- We are forcing JavaScript into a whitelist of allowed commands



How to rewrite object literals?

- 1,{'a':123} is rewritten to 1,{'\$a\$':123};
- 1,{'\
- a':123} normalized to 1,{'\$a\$':123};
- 1,{'\x61':123}; rewritten to 1,{'\$\x61\$':123};
- The strict nature of object property names makes it easier



It can't be that easy?



```
counter--;
leftContext += "\u009a-fa-F){4)}(?[w$ _]\u009a-fa-";
return '\u009a-fa-F){4)}';
} else {
    counter--;
    leftContext += "\u009a-fa-F){4)}(?[w$ _]\u009a-fa-";
    return '\u009a-fa-F){4)}';
}
```



“Str\

ing” to “String”

```
RegExp ("(?: (?: ['] (?: \\ \\ {2} | \\ \\ \\  
['] | [^'] ) * ['] ) | (?: [\" ] (?: \\ \\ {2}  
} | \\ \\ \\ [\" ] | [^\" ] ) * [\" ] ) ) ")
```



Matching regexes

- Normalize

/reg\
ex/ to /regex/

- RegExpr ("(?:[\\/] (?:\\\[(?:\\\\[\\
\\])]+\\]|\\\\\\\\[\\/]|[^\\/*]) (?:\\
\\[(?:\\\\\\\\[\\]|[^\\])]+|^[\\])]+)+|\\\\\\\\[\\
/]|^[\\/]) *? [\\/] (?:[a-zA-
Z]*)) ")



Matching regexes cont.

- `regexpsLeft = new RegExp ( '?' + endStatement.source + '|' + operators.source + '|[(]+)' + spaces.source )`

- Need to know the context
- Difference between 1/1/1 and regex







How to match HTML

- `allowedTags =`
`/ (?:form|optgroup|button|legend|fieldset|label...`
- `allowedAttributes =`
`/ (?:type|accesskey|align|alink|alt...`
- `attributeValues =`
`RegExp ("(?:\"[^\"]\"{0,"+attributeLength+"}\\\"|\"[^\s'\\\"`>]{1,"+attributeLength+"}\"|'[^']*{0,"+attributeLength+"}'|'')")`

```
RegExp ("(?:\"[^\"]\"{0,\"+attributeLength+\"}\\\"|\"[^\\"'\\s'\\\"`>]{1,\"+attributeLength+\"}|'[^']*\"{0,\"+attributeLength+\"}'")
```

attributeLength+"} ') ") ,



How to match HTML cont.

Restrictions placed on length

Linked regexes again

- Very easy compared to JavaScript
- ```
RegExp (' ('+styleTag.source+') | (<\\\\\\\\/?[a-z0-9]{1,10} (? : '+'attributes.source+') {0,'+maxAttributes+' } (? : \\\\/?>) | ('+text.source+') | ('+invalidTags.source+') ', 'ig')
```



# How to lockdown HTML

- ID/Names attributes are unique to the application
- Image requests are proxied
- Using the DOM to decode and place HTML in the document





# Proxied images prevent CSRF

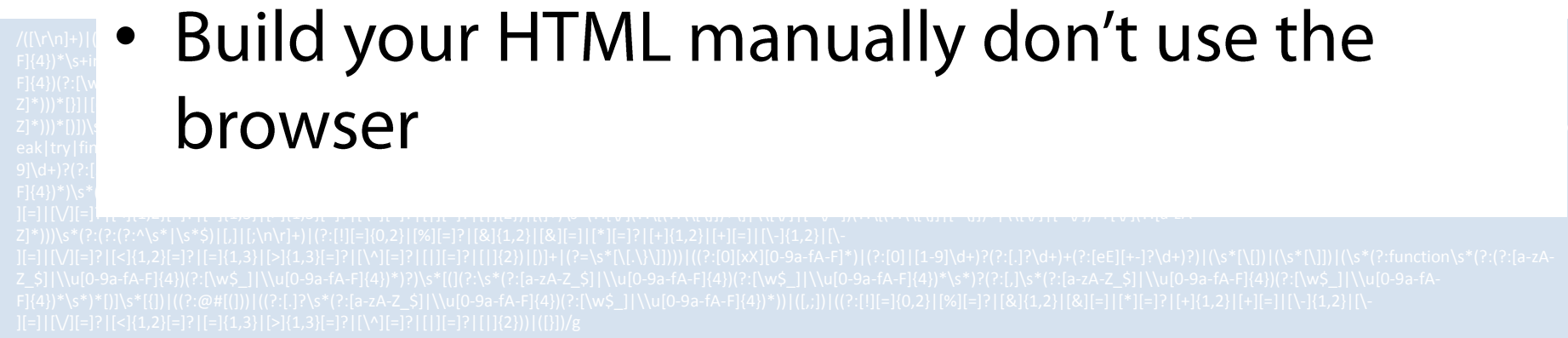
- ``
- ``
- You don't want sandboxed content escaping to the outside world and conducting CSRF
- Through the proxy no cookies are sent originating from the client computer

- Through the proxy no cookies are sent originating from the client computer



# Using the DOM

- InnerHTML sucks. Doesn't represent a true rendering of the HTML source
- Style is HTML decoded and manipulated on IE
- Solution is to use undefined attributes `sandbox-style=`
- Build your HTML manually don't use the browser





- Whitelist all properties and values
- Only positive match discard everything else
- Hex escape urls with spaces after the encoded character
- Proxy image requests



# How to match CSS cont.

Restrictive selectors

```
selectorStart = new
```

```
RegExp (' ((? : (? : [. #] \\ \\ w { 1 , 20 } | for
m | optgroup . . .
```

Whitelisted values

```
units = new
```

```
RegExp (' (? : (? : normal | auto | (? : [+ -
] ? [\\ \\ / . \\ \\ d] { 1 , 8 } \\ \\ s *) { 1 , 4 } (? : px
| % | pt | pc | em | mm | ex | in | cm) ?)) ')
```

Hex encode with space & always  
quote the value

```
<div style="background:
url('http://www.gmodules.com/ig/proxy?url=http\3a
/\3c \3e ') repeat scroll 0% 0%
transparent;">test</div>
```



# Putting it all together

- JSReg whitelists the code
- HTMLReg handles CSS with CSSReg
- Extend the window or global object inside JSReg
- A separate DOM API can then be inserted inside the sandbox, even provide ES5 methods to sandboxed code using ES3 browsers





## Advanced use case

- [Hackvertor.co.uk](http://Hackvertor.co.uk)
- Allows researchers to share sandboxed JavaScript
- Code is extended automatically to reuse code
- Yahoo pipes used to get external sites



input

%61

## a

```
{“HTML”:
 “unicode info”}
```

## JSON

## Decode and sandbox

# Sandboxed HTML

# Advanced use case cont.

a-fA-

typeof|throw|br  
1-



# Alternatives

- Facebook JS
- Caja
- Microsoft Sandbox

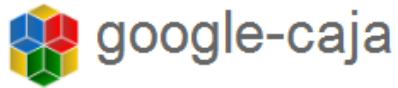


## facebook

- `<div style=background-image:url('http://&quot;;xss/**/&#x3a;expression(alert(1));+&quot;')!important;></div>`
- Now fixed

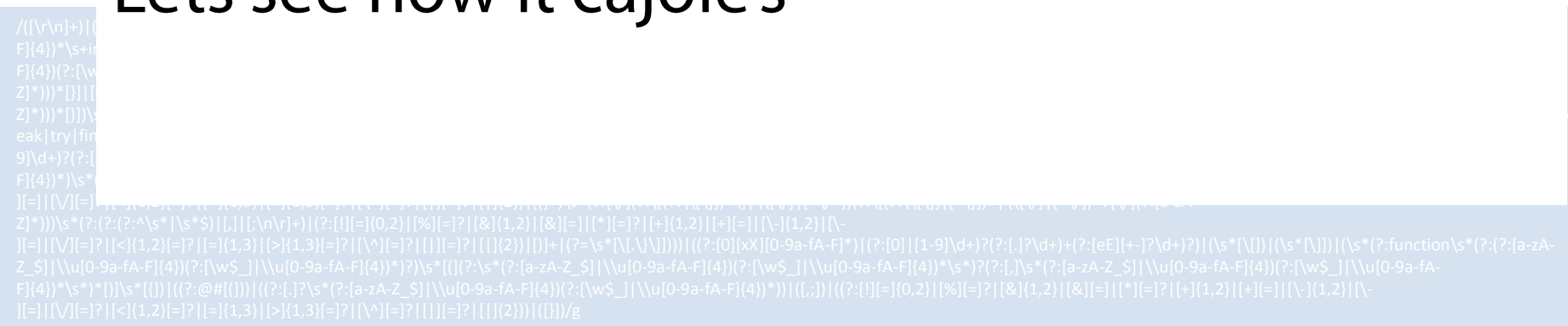


# Alternatives cont.



- `<script>`  
    `Array(4294967295).join(Array(4294967295));`  
    `</script>`

Lets see how it cajole's



ow|br

[illegible]



- ## Alternatives cont.



\* Credits Soroush Dalili

