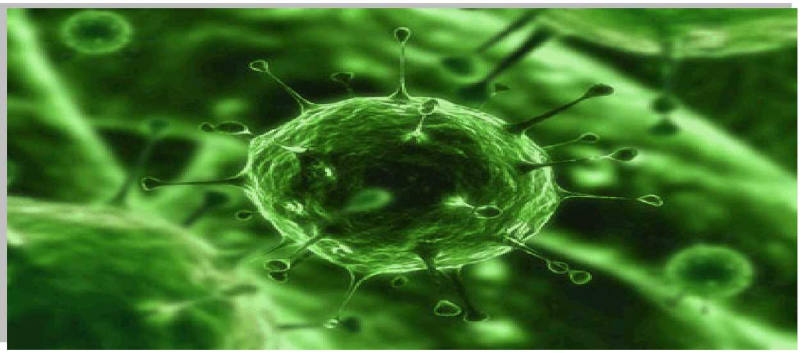
SAND2018-5915C



Malware Information Sharing by Federating Malware Repositories

Ken Chiang, Mike Bierma, Jon Crussell

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND Number:

What is FARM?

Sandia National Laboratories

- Forensic Analysis Repository for Malware
  - Repository of ~100 million analyzed and correlated malware.
  - Central capability @ Sandia National Labs (SNL) for .gov to use.
  - Dynamic, static, and correlation analysis.
  - Frontend and API access
  - Goals:
    - Triage suspicious software
    - Tie attack campaigns together

Inputs to FARM

Sandia National Laboratories

- Malware Samples from:
  - Integrated joint Cybersecurity Coordination Center (iJC3)
  - Email triage system for DOE labs
  - Incident response efforts
  - Public and private malware repositories
  - Network border exe triage across DOE labs and DHS
- Enrichment sources:
  - iJC3 RADAR
  - Yara Signatures from DOE/.gov partners and public
  - USCERT TLP green and white public reports

FARM outputs

Type	PE32 executable (GUI) Intel 80386, for MS Windows
Average Rating	Malicious
Submit Rating	

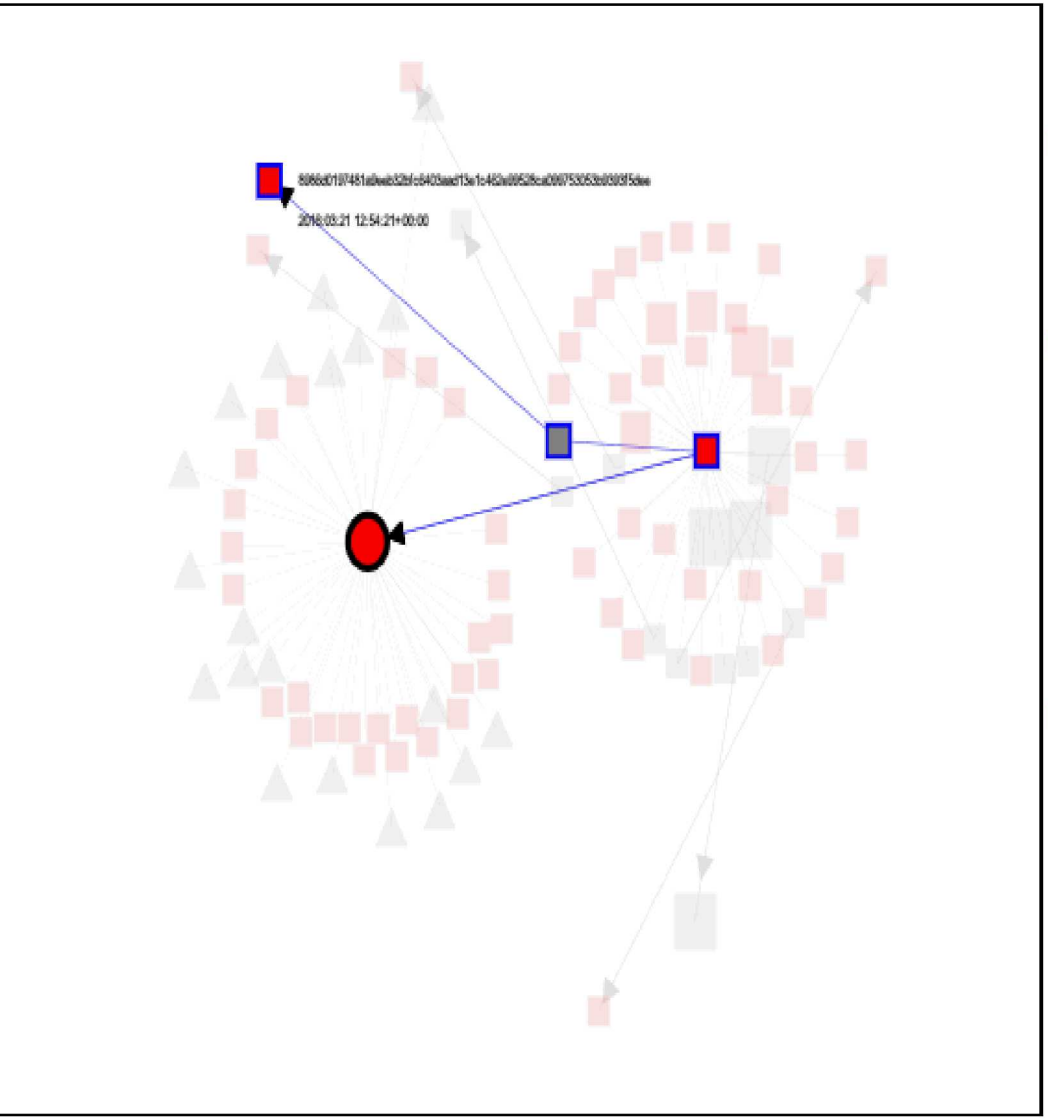
Submission Info [6] | Tool Results [25] | Comments [4] | Related [53] | Intel [14] | Rerun Tools | Links | Download

Sample Intel:

Code	Indicator	Threat Level	Confidence
YARA-1	Yara APT signature match	10	1.0
AV-1	Matches Anti-virus signature	8	0.33
AVATAR-1	Classified as malware by Avatar	8	0.9537035
WXP-2	Calls out to unknown domain in Windows XP running on baremetal hardware	8	1.0
WXP-8	Generates traffic to unknown IPs in Windows XP running on baremetal hardware	8	1.0
VWXP-2	Calls out to unknown domain in Windows XP running in a virtual machine	8	1.0
VWXP-8	Generates traffic to unknown IPs in Windows XP running in a virtual machine	8	1.0
VW7-2	Calls out to unknown domain in Windows 7 32bit running in a virtual machine	8	1.0
VW7-8	Generates traffic to unknown IPs in Windows 7 32bit running in a virtual machine	8	1.0
VW7-64-2	Calls out to unknown domain in Windows 7 64bit running in a virtual machine	8	1.0
VW7-64-8	Generates traffic to unknown IPs in Windows 7 64bit running in a virtual machine	8	1.0
HASHNINJA-1	Has been rated as malicious	6	0.9
HERDWARE-1	Related to malware	6	0.753292261905
VERIFYSIGS-2	Non-existent Embedded Certificate	4	0.7

- **Intelligence scores for triage**
- Related samples based on unpacking, memory analysis, behavior, and code similarity
- Searchable results

FARM outputs



- Intelligence scores for triage
- **Related samples based on unpacking, memory analysis, behavior, and code similarity**
- Searchable results

FARM outputs

- Intelligence scores for triage
- Related samples based on unpacking, memory analysis, behavior, and code similarity
- **Searchable results**

FARM is great, so why federate?

Sandia National Laboratories

- Cross repository boundary discovery of potentially sensitive attack documents that are similar.

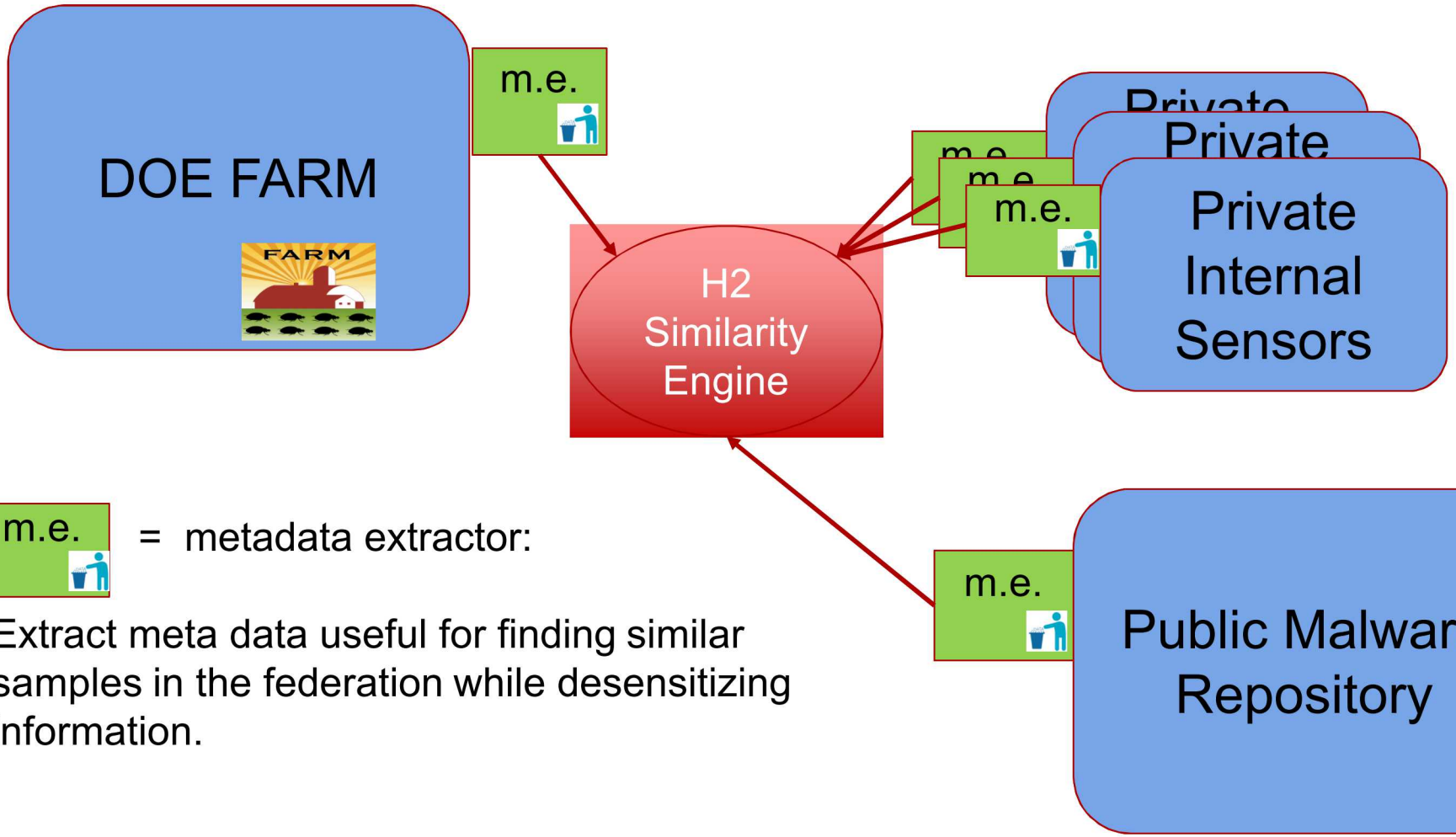
HERDWARE– scalable similarity engine

Sandia National Laboratories

- Swiss Army Knife for identifying nearest K-neighbors
  - Parsers:
    - assembly, javap, dexdump, javascript, …
  - Featurers:
    - n-grams, abstract syntax trees (AST), …
  - Similarity measures:
    - jaccard, minhash, …
  - Storers:
    - SQLite, PostgreSQL, …

- Key Benefit:
  - Allows us to quickly find similar samples in large repository of millions of samples within a few minutes

Federation Concept



m.e. = metadata extractor:

Extract meta data useful for finding similar samples in the federation while desensitizing information.

Concept of operations

Sandia National Laboratories

- Central meta data similarity engine (Herdware2)
- Distribute meta data extraction client.
- Participants of federation choose to send any information about its own files with its own unique id.
- Share to receive
  - Share meta data to allow ***discovery*** of similar malware across the federation.
  - Subsequent sharing of sensitive attack context controlled by the participants.

Similarity engines available

- Current
  - Windows executable 

  - PDF 

- Future plans
  - Office docs 

Exe metadata process description

- Disassemble
- Extract opcodes and throw away operands and data

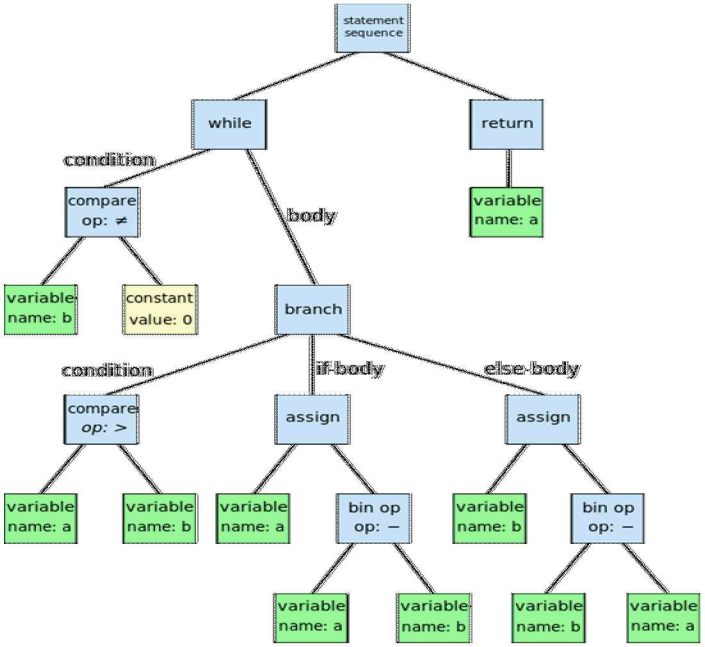
  - i.e. mov, push, pop, ret
- Group into n-grams, basic blocks, and functions.

{"mnemonics": ["push", "push", "push", "mov", "call", "mov", "mov", "pop", "retn"], "complexity": 1, "name": "sub_55DC46", "leaders": [0]}
{"mnemonics": ["retn"], "complexity": 1, "name": "nullsub_1", "leaders": [0]}
{"mnemonics": ["push", "mov", "push", "push", "push", "push", "call", "mov", "add", "pop", "retn"], "complexity": 1, "name": "sub_55DCA9", "leaders": [0]}
{"mnemonics": ["push", "mov", "push", "push", "push", "push", "call", "mov", "add", "pop", "retn"], "complexity": 1, "name": "sub_55DCC5", "leaders": [0]}
{"mnemonics": ["push", "push", "mov", "mov", "push", "push", "call", "push", "push", "call", "add", "pop", "pop", "retn"], "complexity": 1, "name": "sub_55DCE1", "leaders": [0]}
{"mnemonics": ["push", "mov", "call", "mov", "mov", "call", "and", "push", "lea", "mov", "call", "mov", "call", "retn"], "complexity": 1, "name": "sub_55DD01", "leaders": [0]}
{"mnemonics": ["push", "push", "mov", "call", "mov", "mov", "pop", "retn"], "complexity": 1, "name": "sub_55DD9C", "leaders": [0]}
{"mnemonics": ["push", "push", "mov", "call", "mov", "mov", "pop", "retn"], "complexity": 1, "name": "sub_55DDE1", "leaders": [0]}
{"mnemonics": ["push", "mov", "call", "mov", "mov", "call", "and", "push", "lea", "mov", "call", "mov", "call", "retn"], "complexity": 1, "name": "sub_55DDF9", "leaders": [0]}
{"mnemonics": ["push", "mov", "call", "mov", "mov", "push", "call", "and", "lea", "call", "lea", "call", "lea", "call", "lea", "call", "push", "mov", "push", "call", "pop", "pop", "mov", "call", "retn"], "complexity": 1, "name": "sub_55DE75", "leaders": [0]}
{"mnemonics": ["push", "mov", "call", "mov", "mov", "push", "mov", "call", "pop", "push", "push", "lea", "call", "push", "push", "lea", "call", "push", "push", "lea", "call", "push", "push", "lea", "call", "or", "mov", "call", "call", "retn"], "complexity": 1, "name": "sub_55DECA", "leaders": [0]}

PDF metadata process description

- Extract javascript from PDF, throw away content

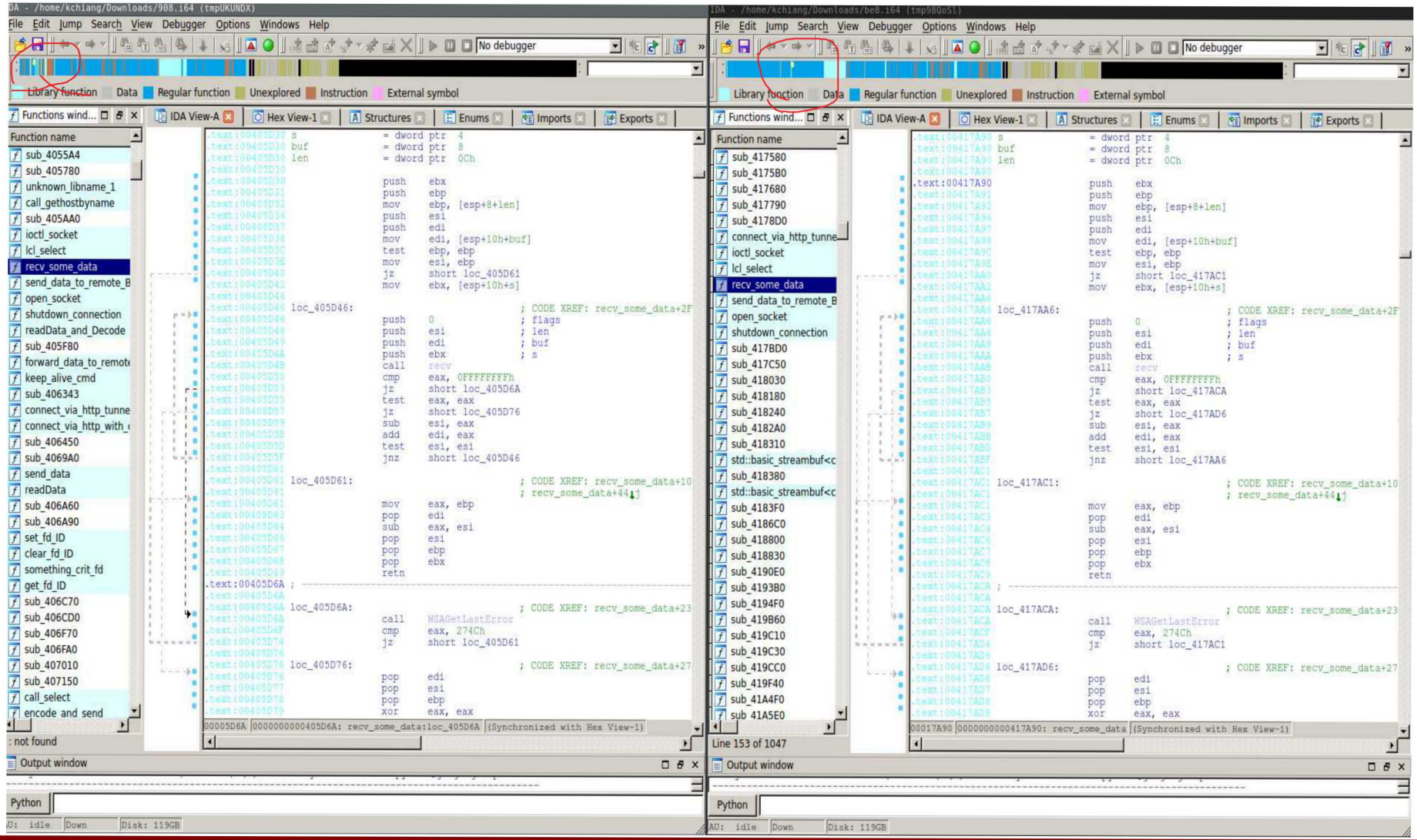
- Convert javascript to AST



{u'body': [{u'end': 224, u'alternate': None, u'consequent': {u'end': 224, u'alternate': None, u'consequent': {u'body': [{u'start': 165, u'end': 174, u'type': u'ExpressionStatement', u'expression': {u'right': {u'start': 169, u'end': 173, u'type': u'Literal', u'value': u'el', u'raw': u'"el"'}, u'end': 173, u'start': 165, u'operator': u'=', u'type': u'AssignmentExpression', u'left': {u'start': 165, u'end': 166, u'type': u'Identifier', u'name': u'x'}}}, {u'start': 183, u'end': 192, u'type': u'ExpressionStatement', u'expression': {u'right': {u'start': 188, u'end': 191, u'type': u'Literal', u'value': u'y', u'raw': u'"y"'}, u'end': 191, u'start': 183, u'operator': u'=', u'type': u'AssignmentExpression', u'left': {u'start': 183, u'end': 185, u'type': u'Identifier', u'name': u'zz'}}}, {u'start': 201, u'end': 218, u'type': u'ExpressionStatement', u'expression': {u'right': {u'end': 217, u'computed': False, u'object': {u'start': 205, u'end': 210, u'type': u'Identifier', u'name': u'event'}, u'start': 205, u'property': {u'start': 211, u'end': 217, u'type': u'Identifier', u'name': u'target'}, u'type': u'MemberExpression'}, u'end': 217, u'start': 201, u'operator': u'=', u'type': u'AssignmentExpression', u'left': {u'start': 201, u'end': 202, u'type': u'Identifier', u'name': u'z'}}}], u'start': 155, u'end': 224, u'type': u'BlockStatement'}, u'start': 110, u'test': {u'right': {u'start': 152, u'end': 153, u'type': u'Literal', u'value': 0, u'raw': u'0'}, u'end': 153, u'start': 114, u'operator': u'===', u'type': u'BinaryExpression', u'left': {u'start': 114, u'end': 147, u'type': u'CallExpression', u'arguments': [{u'start': 130, u'end': 133, u'type': u'Literal', u'value': 321, u'raw': u'321'}, {u'start': 135, u'end': 141, u'type': u'Literal', u'value': 513613, u'raw': u'513613'}, {u'start': 143, u'end': 146, u'type': u'Literal', u'value': u'a', u'raw': u'"a"'}], u'callee': {u'end': 129, u'computed': False, u'object': {u'start': 114, u'end': 125, u'type': u'Identifier', u'name': u'ImageField1'}, u'start': 114, u'property': {u'start': 126, u'end': 129, u'type': u'Identifier', u'name': u'ZZA'}, u'type': u'MemberExpression'}}}, u'type': u'IfStatement'}, u'start': 70, u'test': {u'right': {u'start': 100, u'end': 104, u'type': u'Literal', u'value': None, u'raw': u'null'}, u'end': 104, u'start': 74, u'operator': u'===', u'type': u'BinaryExpression', u'left': {u'start': 74, u'end': 95, u'type': u'CallExpression', u'arguments': [], u'callee': {u'end': 93, u'computed': False, u'object': {u'start': 74, u'end': 78, u'type': u'ThisExpression'}, u'start': 74, u'property': {u'start': 79, u'end': 93, u'type': u'Identifier', u'name': u'execInitialize'}, u'type': u'MemberExpression'}}}, u'type': u'IfStatement'}, {u'start': 225,

Performance: 75% similar exe samples

Sandia National Laboratories



Structurally similar code in PDF (90%)

Sandia National Laboratories

```
function t9HcD4XJVoH4() {
    var ymSiMtdF2urF = unescape;
    return ymSiMtdF2urF;
}



eval("var xx" + "xx" + "cx = ev" + "al;")

var vPZTDqX6SV56 = "";
xxxxcx("rs6eTpTAcvsz = th" + "is.in" + "fo.p" + "ro" + "du" + "cer;");
var pcn = "%";

function bWErQ77PVftj(yQoywt4qm9UF, wxPatIZjooaF) {
    var xx = yQoywt4qm9UF.replace(/kol2/g, wxPatIZjooaF);
    return xx;
}
vPZTDqX6SV56 = bWErQ77PVftj(rs6eTpTAcvsz, pcn);

function theRNYAtyzsu(fwle5Ft4DJe8) {
    var vvvvc = eval;
    vvvvc(fwle5Ft4DJe8);
}



var pZRrZJmRFX7K = t9HcD4XJVoH4();


var u1GcKg0DGa5p = pZRrZJmRFX7K(vPZTDqX6SV56);
theRNYAtyzsu(u1GcKg0DGa5p);
```

```
function dr83Wmq6yqTl4q() {
    var x1H0sOBPB3UAty = unescape;
    return x1H0sOBPB3UAty;
}



eval("var xx" + "xx" + "cx = ev" + "al;")

var uwoWXDzzBPrWok = "";
xxxxcx("bDTJElS8AeWLh1 = th" + "is.in" + "fo.p" + "ro" + "du" + "cer;");
var pcn = "%";

function laySXnz51z5COv(WDfcQ0wfvPm64I, QkJxLHQs6Px702) {
    var xx = WDfcQ0wfvPm64I.replace(/kol2/g, QkJxLHQs6Px702);
    return xx;
}
uwoWXDzzBPrWok = laySXnz51z5COv(bDTJElS8AeWLh1, pcn);

function pnx31cfQmb7h1s(z5BEpVBAnb3s1z) {
    var vvvvc = eval;
    vvvvc(z5BEpVBAnb3s1z);
}



var zccsgo8VJIg7ui = dr83Wmq6yqTl4q();


var uD5mpgfWUm3n5i = zccsgo8VJIg7ui(uwoWXDzzBPrWok);
pnx31cfQmb7h1s(uD5mpgfWUm3n5i);
```

Interested in sharing?

Sandia National Laboratories

- If your organization is interested in getting FARM accounts OR

- If you want to participate in the federated concept, we will be piloting a pdf document similarity study to test out concept, email:
- farm [at] sandia.gov

Demo video

- Triaging a sample that is not in public repositories using the federation concept to gain intelligence about the unknown sample.

Questions?

Sandia National Laboratories