STIX Cyber Observable- Webpage Object

"Share webpage details"

What is it?

The Webpage Object is designed to allow an intel provider to record the following useful information:

- Record the javascript from an exploit redirection site
- Record excerpts from a conversation held within a web forum
- Record the redirect chain of multiple stages of exploit redirection to show how an attack was performed
- Record a web defacement
- Record changes to a webpage over time

When I've spoken with different dark web monitoring threat intelligence providers about STIX, inevitably the first question I get asked is 'Can we record content we're monitoring in crime forums in STIX v2?". And then the second question is can I record the multi-stage redirects that modern Exploit Kits redirect our clients through before exploiting them?

The answer today to both those questions is no. Unbelievably we have no mechanism for recording the content of webpages – no way of documenting interesting parts of the 1.8 billion websites that are estimated to currently exist. There is also no way for us to describe the multi-step redirection chain that clients get put through when being exploited.

Both of those subjects contain actionable threat intelligence, and its highly surprising that we don't yet have a way of codifying that information.

What benefits would it provide?

As mentioned earlier, the Webpage object gives intel providers a new way to share the sort of intel they want to share, and allows recipients to use that information to help detect maliciousness in their web traffic.

Creating a structure to recording webpage content allows web proxy vendors to use that information to detect badness.

By providing the ability to share text extracts from webpages, we allow threat intelligence providers to:

Send dark-web criminal forum extracts to their customers showing the threat actors discussing their upcoming plans.

Send extracts of ransomware .onion landing pages, helping your customers detect ransomware installs within your businesses network.

Record web-defacements.

Record website changes over time with a series of webpage objects.

This is currently a massive gap in our available objects, and is something I believe we must rectify in STIX 2.1.

How would it work?

I propose that we add a new STIX Cyber Observable (SCO) object to STIX v2.1 - the **Webpage object.** The Webpage object would allow an (x)HTML webpage to be modelled within STIX, allowing its information to be recorded. The Webpage object is flexible enough to allow only the important or interesting bits of the webpage to be recorded if desired.

I also propose that we add an **HTTP Response extension** for the **Network Object** to allow us to use the Webpage object to describe HTTP responses containing the webpage as well as the Webpage itself. This in turn will allow us to describe things like HTTP redirects, and to link HTTP requests with the webpage content that gets returned.

STIX Cyber Observable – Webpage object proposal

2.6. Webpage Object

Type Name: `webpage`

The Webpage Object represents an instance of a webpage, corresponding to the HTML W3C recommendations described at https://www.w3.org/TR/#tr_HTML. The webpage object represents the HTML content (and other content included in the HTML such as Javascript) sent by the webserver to the web client over HTTP or HTTPS. It does not include any server-side code required to generate the HTML.

All HTML MUST be escaped so that it can be represented within JSON as per RFC7159. A reminder that all quotation marks, reverse solidus, and the control characters (U+0000 through U+001F) MUST be escaped by preceding them with a reverse solidus (\) e.g. the HTML string `<link rel="stylesheet">` would be placed into the links list as `"<link rel=\"stylesheet\">` .

Any characters in the encoded value which cannot be decoded into Unicode MUST be replaced with the 'REPLACEMENT CHARACTER' (U+FFFD). If it is necessary to capture the raw HTML as observed, this can be achieved by referencing an `artifact` Object through the **webpage_data_ref** property.

2.6.1. Properties

Common Properties		
type, extensions		
Webpage Object Specific Properties		
url_ref, head, body, webpage_data_ref, redirects_to_ref, loads_content_from_refs		
Property Name	**Type**	**Description**
type (required)	`string`	The value of this property **MUST** be `webpage` .
url_ref (optional)	`object-ref`	Specifies the location of the webpage using a URL object. The object referenced in this property **MUST** be of type `url`.
head (optional)	`list` of type `string`	Specifies a `list` of type `string` where each list item contains HTML excerpts from the **head** element of the HTML content (as defined in the latest W3C HTML Recommendation) of the webpage. Each list item **MUST** be an individual HTML element and **MUST** have come from the head element of the webpage. The list item string **MUST** begin with the HTML element start tag and end with the HTML element end tag, e.g. "*<title>My Evil Site</title>*". All HTML **MUST** be escaped so that it can be represented within JSON as per RFC7159. List values **MUST** appear in the same order as present in the raw HTML of the webpage.
body (optional)	`list` of type `string`	Specifies a `list` of type `string` where each list item contains HTML excerpts from the **body** element of the HTML content (as defined in the latest W3C HTML Recommendation) of the webpage.

		Each list item **MUST** be an individual HTML element and **MUST** have come from the body element of the webpage. The list item string **MUST** begin with the HTML element start tag and end with the HTML element end tag, e.g. "*Click here*". All HTML **MUST** be escaped so that it can be represented within JSON as per RFC7159. List values **MUST** appear in the same order as present in the raw HTML of the webpage.
webpage_data_ref (optional)	`object-ref`	Specifies the complete raw binary contents of the webpage, including both the headers and body, as a reference to an `artifact` object. The object referenced in this property **MUST** be of type `artifact` **Note: This property does NOT include the HTTP headers. HTTP Headers are specified in a HTTP Response extension of a Network Object.**
redirects_to_ref (optional)	`object-ref`	Specifies the next URL that this webpage redirects the client to, as a reference to a `url` object. The object referenced in this property **MUST** be of type `url`. This property **MAY** contain a URL extracted from anywhere on the webpage, or a URL generated dynamically from scripts included in the webpage running when viewed by the web client.

loads_content_from_refs (optional)	`list` of type `object-ref`	Specifies a list of URLs that this webpage instructs the web client to load content from, with each list entry as a reference to a different `url` object. The objects referenced in this property **MUST** all be of type `url`. This list **MAY** contain URLs extracted from anywhere on the webpage, or URLs generated dynamically from scripts included in the webpage running when viewed by the web client.

2.6.3. Examples

Hacked Website redirecting to exploit site using Javascript

```
{
  "0": {
    "type": "url",
    "value": "https://mymainnews.com/news/index.html"
  },
  "1": {
    "type": "webpage",
    "url_ref" "0",
    "body": [
      "<script src=\"https://cdnjs.cloudflare.com/ajax/libs/jquery/3.1.1/jquery.js\_"
type=\"text/javascript\"></script>",
      "<script src=\"https://myhackedsite.com/files/uploads/d.js\_"
type=\"text/javascript\"></script>"
    ]
  }
}
```

Multiple chained exploit site redirects

```
{
  "0": {
    "type": "url",
    "value": "https://wayneindustries.com/research/index.html"
  },
  "1": {
    "type": "webpage",
    "url_ref" "0",
    "body": [
      "<iframe src=\"http://exssredaf.intsite.info/content/lfw?327xas\">"],
    ],
    "redirects_to_ref": "2"
  },
  "2": {
    "type": "url",
    "value": "http://exssredaf.intsite.info/content/lfw?327xas"
  },
  "3": {
    "type": "webpage",
    "url_ref" "2",
    "body": [
      "<script src=\"http://389553.teaparty2.biz/uploads/j.jpg\">"],
    ],
```

```
    "loads_content_from_refs": ["4"]
  },
  "4": {
    "type": "url",
    "value": "http://389553.teaparty2.biz/uploads/j.jpg"
  },
  "5": {
    "type": "webpage",
    "url_ref" "2",
    "body": [
      "<script> eval(function(p,a,c,k,e,d){e=function(c){return
c};if(!''.replace(/^/,String)){while(c--){d[c]=k[c]||c}k=[function(e){return
d[e]}];e=function(){return'\\w+'};c=1};while(c--){if(k[c]){p=p.replace(new
RegExp('\\b'+e(c)+'\\b','g'),k[c])}}return p}('1 0="2
3!";4(0);',5,5,'thing|var|Hello|World|alert'.split('|'),0,{})) </script>"],
    ]
  }
}
```

Recording Forum Posts on Website (including the HTTP request)

```
{
  "0": {
    "type": "domain-name",
    "value": "wayneindustries.com"
  },
  "1": {
    "type": "network-traffic",
    "dst_ref": "0",
    "protocols": [
      "tcp",
      "http"
    ],
    "extensions": {
      "http-request-ext": {
        "request_method": "get",
        "request_value": "/research/index.html",
        "request_version": "http/1.1",
        "request_header": {
          "Accept-Encoding": "gzip,deflate",
          "User-Agent": "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.6)
Gecko/20040113",
          "Host": "wayneindustries.com"
        },
        "message_webpage_ref": "2"
      }
    }
  },
  "2": {
    "type": "webpage",
    "url_ref" "3",
    "body": [
      "<div class=\"highlight\">I hack3d the Internetz and no 1 will get me eva.</div>",
      "<div class=\"highlight\">I've managed to get into the CIA man!</div>",
    ]
  },
  "3": {
    "type": "url",
    "value": "https://wayneindustries.com/research/index.html"
  }
}
```

2.12 Network Traffic Object

The following HTTP Response Extension object is an extension of the Network Traffic object.

2.12.2 HTTP Response Extension

Type Name: `http-response-ext`

The HTTP response extension specifies a default extension for capturing network traffic properties specific to HTTP responses. The key for this extension when used in the **`extensions`** dictionary **MUST** be `http-response-ext`.

2.12.2.1 Properties

Property Name	Type	Description
`response_url_ref` (optional)	`object-ref`	Specifies the URL object that specifies the network location that the HTTP Response was received from. The object referenced in this property **MUST** be of type `url`.
`response_status` (optional)	`integer`	Specifies the HTTP status code for the HTTP response, as an integer.
`response_version` (optional)	`string`	Specifies the HTTP version portion of the HTTP response line, as a lowercase string.
`response_content_type` (required)	`string`	Specifies the HTTP content_type header portion of the HTTP response line, as a lowercase string.
`response_header` (optional)	`dictionary`	Specifies all of the HTTP header fields that may be found in the HTTP server response, as a dictionary. All HTTP header fields must be specified in the order they are shown in the HTTP response if the **response_header** property is used. Each key in the dictionary **MUST** be the name of the header field and **SHOULD** preserve case, e.g., `User-Agent`. The corresponding value for each dictionary key **MUST** be a `string`.
`response_request_ref` (optional)	`object-ref`	Specifies the HTTP Request object that resulted in this HTTP Response being generated. The object referenced in this property **MUST** be of type `network-traffic` with a `http-request-ext` extension.
`message_body_length` (optional)	`integer`	Specifies the length of the HTTP message body, if included, in bytes.
`message_body_data_ref` (required)	`object-ref`	Specifies the data contained in the HTTP message body. The object referenced in this property **MUST** be of type `artifact` or of type `webpage`.
`redirects_to_ref` (optional)	`object-ref`	Specifies the URL object that specifies the network location that this HTTP Response is redirecting us to. This property is used to relate the URL that the HTTP client will be redirected to if it receives a HTTP 3xx Redirection Status code. The object referenced in this property **MUST** be of type `url`.

Examples

HTTP Request using Webpage object and URL object

```
{
  "0": {
    "type": "ipv4-addr",
    "value": "198.51.100.53"
  },
  "1": {
    "type": "network-traffic",
    "src_ref": "0",
    "protocols": [
      "tcp",
      "http"
    ],
    "extensions": {
      "http-response-ext": {
        "response_url_ref": "3",
        "response_status": 200,
        "response_version": "http/1.1",
        "response_content_type": "text/html",
        "response_header": {
          "Date": "Mon, 27 Jul 2017 12:28:53 GMT ",
          "Server": "Apache/2.2.14 (Win32)",
          "Last-Modified": "Wed, 22 Jul 2009 19:15:56 GMT",
          "Content-Length": "88",
          "Content-Type": "text/html",
          "Connection": "Closed"
        },
        "message_body_length": 88,
        "message_body_data_ref": "2",
      }
    }
  },
  "2": {
    "type": "webpage",
    "url_ref" "3",
    "body": [
      "<div class=\"highlight\">I hack3d the Internetz and no 1 will get me eva.</div>",
      "<div class=\"highlight\">I've managed to get into the CIA man!</div>",
    ]
  },
  "3": {
    "type":"url",
    "value":"http://www.happydaze1988.com"
  }
}
```

HTTP Request, and HTTP Response using Webpage object and URL object

```
{
  "0": {
    "type": "domain-name",
    "value": "wayneindustries.com"
  },
  "1": {
    "type": "network-traffic",
    "dst_ref": "0",
    "protocols": [
      "tcp",
      "http"
    ],
    "extensions": {
      "http-request-ext": {
        "request_method": "get",
        "request_value": "/research/index.html",
        "request_version": "http/1.1",
        "request_header": {
          "Accept-Encoding": "gzip,deflate",
```

```
          "User-Agent": "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.6)
Gecko/20040113",
          "Host": "wayneindustries.com"
        },
        "message_webpage_ref": "2"
      }
    }
  },
  "2": {
    "type": "network-traffic",
    "src_ref": "0",
    "protocols": [
      "tcp",
      "http"
    ],
    "extensions": {
      "http-response-ext": {
        "response_url_ref": "4",
        "response_status": 200,
        "response_version": "http/1.1",
        "response_content_type": "text/html",
        "response_header": {
          "Date": "Mon, 27 Dec 2017 12:28:53 GMT ",
          "Server": "Apache/2.2.14 (Win32)",
          "Last-Modified": "Wed, 12 Jan 2017 19:15:56 GMT",
          "Content-Length": "3558",
          "Content-Type": "text/html",
          "Connection": "Closed"
        },
        "response_request_ref": "2",
        "message_body_length": 88,
        "message_body_data_ref": "3",
      }
    }
  },
  "3": {
    "type": "webpage",
    "url_ref" "4",
    "body": [
      "<div class=\"highlight\">I hack3d the Internetz and no 1 will get me eva.</div>",
      "<div class=\"highlight\">I've managed to get into the CIA man!</div>",
    ]
  },
  "4": {
    "type":"url",
    "value":"http://wayneindustries.com/research/index.html"
  }
}
```

HTTP Request redirecting to new URL (HTTP 302 Redirection)

```
{
  "0": {
    "type": "ipv4-addr",
    "value": "198.51.100.53"
  },
  "1": {
    "type": "network-traffic",
    "src_ref": "0",
    "protocols": [
      "tcp",
      "http"
    ],
    "extensions": {
      "http-response-ext": {
        "response_url_ref": "3",
        "response_status": 302,
        "response_version": "http/1.1"
      }
    }
  },
  "2": {
```

```
    "type":"url",
    "value":"http://www.happydaze1988.com"
  },
  "3": {
    "type":"url",
    "value":"http://www.happydaze1988.nl"
  }
}
```