

## Form Scraping – Wrap-Up

This Proof of Concept (PoC) was supposed to query all Microsoft Internet Explorer windows and, if a form were present, get the contents. The functionality would be achieved using the Microsoft Component Object Model (COM) Application Programming Interface (API).

After approximately three days of research, Blackbird believes this method to be outdated and esoteric. Additionally, Blackbird found that all references to the Sinowal Trojan exhibiting this functionality come from a single Virus Bulletin report (June 2014).

The dated code was evident during implementation of the PoC. To enumerate all Internet Explorer windows prior to monitoring them for web forms, one used the SHDocVw class. This class should have been available after manually importing (#import) shdocvw.dll. While this library exists on systems, multiple problems were observed during the import in the IDE. The inability to successfully import the required libraries prevents class declaration, instantiation, and also prevented the enumeration of all Internet Explorer windows.

Further investigation yields almost no articles dated more recently than 2004. Most articles assume Administrator privileges and Internet Explorer 5.0. References were found to Windows Vista and Internet Explorer 8.0, but they were sparsely documented and were viewed as the most recent versions at the time.

Although the initial Virus Bulletin report appeared to describe the process in great detail, we found these details to be both esoteric and misapplied. Intermixing COM and C++ terminology within the same sentence gives the impression of a thorough description, but ends up being very misleading. For instance, consider the following statement:

“The Invoke method of the IDispatch object for DIID\_HTMLDocumentEvents2 and DIID\_HTMLTextContainerEvents2”

IDispatch is not a C++ object, it is COM object that is declared as a C-style struct. Getting to this point requires calling multiple COM functions to get an COM object instance of the desired application – in this case Internet Explorer. Additionally, the IDispatch struct/object/class is more accurately defined in Microsoft documentation as an Interface. Establishing that this is an interface is significant because the Invoke method is declared, but not defined. In other words, we must define the functionality of the Invoke method.

Assuming that the technique was not dated, the documentation provided by Microsoft is sparse at best and completely absent at worst. Much of the documentation referred us to a Microsoft article from 2005 that documents a small portion of COM functionality.

Implementation in the future would not be possible since the methods are already deprecated. Despite this, the correct implementation would be to enumerate all Internet Explorer Browser Objects and to attach an IDispatch object to each of them.

In summary, due to the outdated technique and esoteric implementation discovered during research, Blackbird recommends moving on from this PoC.