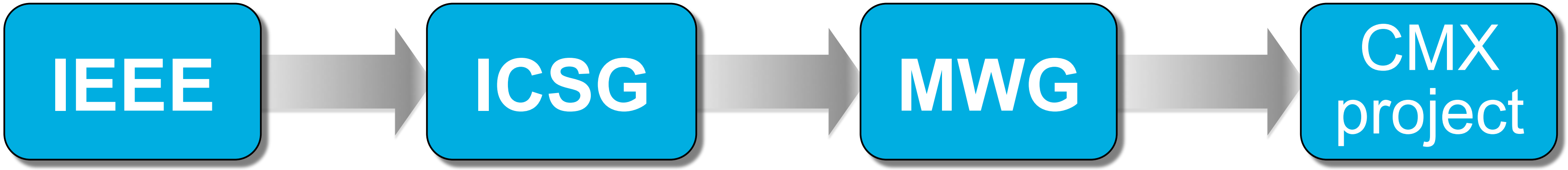
# CMX: IEEE Clean File Metadata Exchange

Dr. Igor Muttik, McAfee

Mark Kennedy, Symantec

# Who We Represent

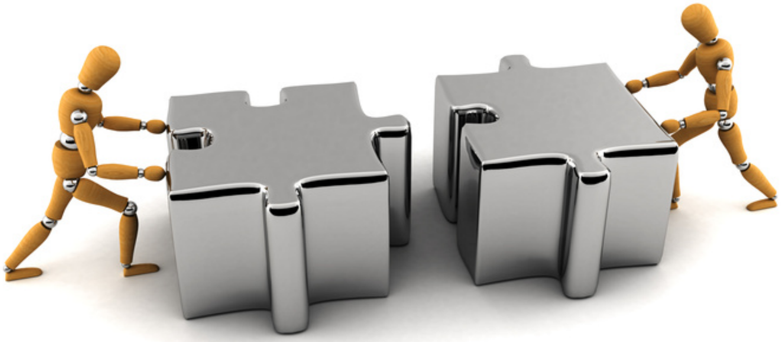


- IEEE Industry Connections Security Group (ICSG)
  - Many security companies take part
- ICSG has multiple Working Groups
  - Malware Working Group (MWG) is one of them
- Clean Metadata eXchange (CMX) system is the child of ICSG MWG

# Background

- ❑ Malware problem is constantly growing
  - Quantity and complexity
  - Evasion Techniques
  - Size and High-level language use
- ❑ Better heuristics are needed
  - To detect 0-day threats
- ❑ False Positives
  - Heuristics can lead to more false positives (FPs)
  - If there are too many FPs the solution will be turned off

# Issues with Whitelists



- Difficult to collect
  - Trusted sources can be compromised
  - Some sources may be operated by malware authors
- Delay between discovery and whitelist updating
- Certain programs are intrinsically variable (.NET with JIT)
- Whitelists are black or white classification
  - There are shades of grey
  - Some legit software can be misused (e.g. remote access tools)
  - Trusted software might contain hidden functionality (“Easter Egg”)

# Current Approaches

- ❑ On-machine whitelists (existed for years)
- ❑ Cloud whitelists (relatively new)
- ❑ CMX helps this tremendously
  - ■ Currently, vendors must each seek out clean files
  - ■ Some 3rd parties work with multiple vendors – leading to extra work
- ❑ CMX provides a single point of contact
  - ■ Simplifies exchange for both vendors and 3rd parties

# The CMX System

- Provide timely information about clean files
  - Currently Windows files: PE/DLL executables (e.g. inside CAB, MSI)
  - Only files for public release
- What metadata is gathered?
  - Hashes (MD5, SHA-1, SHA-256)
    - SHA-512 and SHA-3 can also be considered
  - Filename: the name as it will appear once installed
  - Path: the path where the file will appear once installed, using CSIDL normalized paths
  - Signature information: if the file is digitally signed, information about the signing certificate
  - File version information: the various fields from the file version record

# Example XML

```
<cleanMetaData xmlns="http://xml/metadataSharing.xsd" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance" xsi:schemaLocation="http://xml/metadataSharing.xsd file:metadataSharing.xsd"
version="1.2">
  <company>TrustedSource</company>
  <author>ZIP 1</author>
  <comment>Test MMDEF v1.2 file generated using genMMDEF</comment>
  <timestamp>2011-08-19T13:50:21.721000</timestamp>

<objects>
<file id="4edc50d3a427566d6390ca76f389be80">
 <md5>4edc50d3a427566d6390ca76f389be80</md5>
 <sha1>9cb1bd5dc93124f526a1033b1b3f37cc0224a77e</sha1>
 <sha256>e942d28c0e835b8384752731f1b430cb3fbd571381666ded7637a2db47fafcc0</sha256>
<sha512>3ceb1bd07af9e470ff453ef3dd4b97f9228856cb78eb5cddb7b81796b4b830368e3ed2f0c6a9ce930
09397e8158c68dba67e398f58df87137d8872cb0bb3b53b</sha512>
 <size>3412856</size>
 <crc32>1119775926</crc32>
 <filename>procexp.exe</filename>
 <filenameWithinInstaller>procexp.exe</filenameWithinInstaller>
 <MIMEType>application/octet-stream</MIMEType>
 </file>

<softwarePackage id="procexp">
 <vendor>Sysinternals</vendor>
 <product>Process Explorer</product>
 <version>14.11</version>
 <language>English</language>
 </softwarePackage>
 ...
```

# Why We Don’t Share Files

- Key difference between clean files and malicious files: **Copyright**
  - It is illegal to share many clean files
  - Sharing metadata solves this problem

- Privacy
  - Large companies like to keep their internal apps internal

- Space and Bandwidth
  - Most cloud systems do not require the file
  - Hashes are sufficient

# How the System Works

- ❑ Two types of users:  Providers and Consumers

- ❑ Providers create the metadata and submit it to CMX
  - ■ Use existing IEEE metadata XML format
  - ■ Python scripts assist in the extraction and formatting of the metadata

- ❑ Consumers pull the data and use them in their ecosystem
  - ■ Trust level can be assigned to each data provider

- ❑ Interfaces to pull the most recent data (UI and command-line tols)
  - ■ Keeps track of data downloaded, can give latest data
  - ■ Offline archive for older data

# Access to the System

- Requires a login be created
- Requires one or more public certs be registered to that user
  - Private certs are used to sign the content as it is created
  - Public cert is used to authenticate the data on the CMX backend.
- Public cert is provided along with the content for Consumer validation.

# Types of Providers

- Direct content creators
  - Two types:  Invited and Self-registered
    - Invited is for large companies
    - Self-registered are for companies with a Class 3 code signing certificate
  - Submit data for the files they create

- 3rd Part Provider
  - Must be approved
  - Provide metadata for others’ files

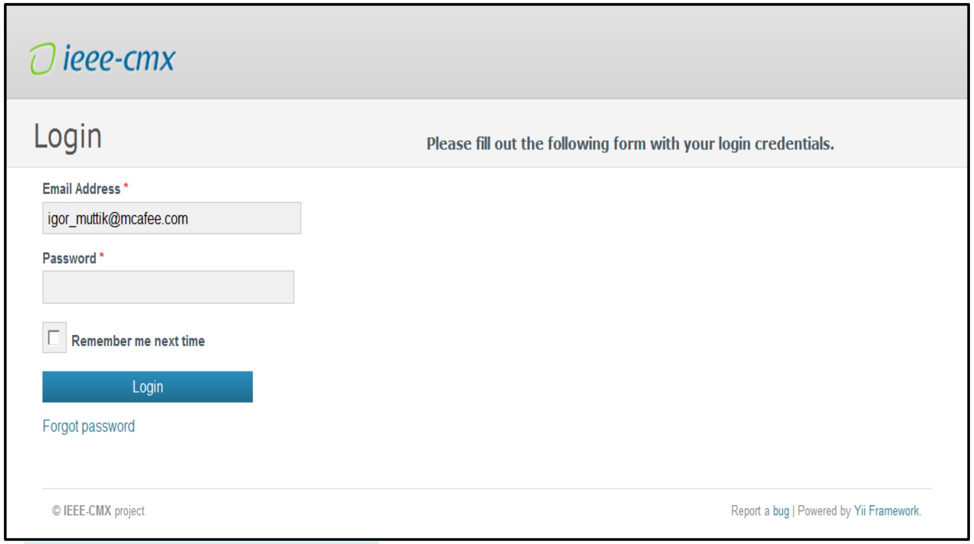
# Current Status

The system is now fully operational and hosted on servers owned by Avira in Germany (https://ieee-cmx.avira.com)

CMX is somewhat similar to the MUTE system, which was implemented by Avira to share malicious URLs

CMX required several modifications (including specific metadata extractors implemented currently in Python), but it is largely based on MUTE

# Screenshots (1)



Login

Please fill out the following form with your login credentials.

Email Address *

igor_muttik@mcafee.com

Password *

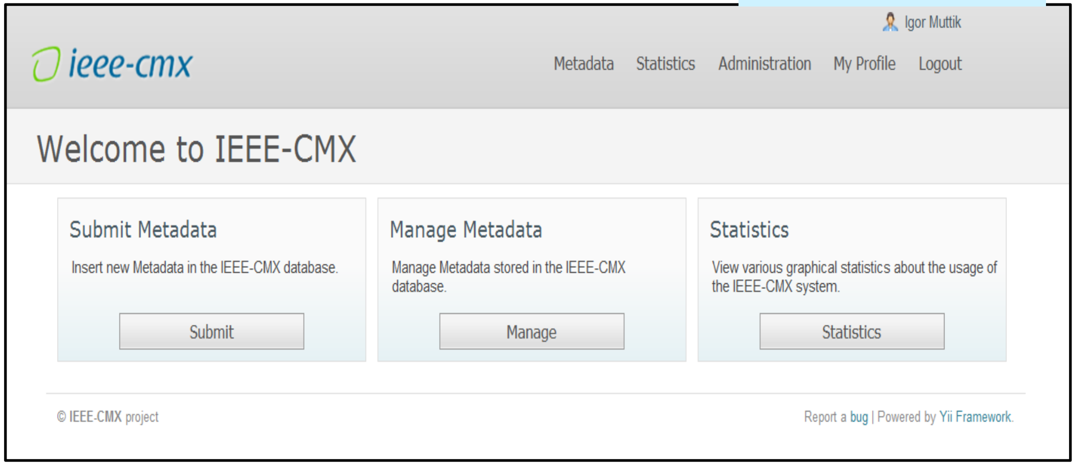
☐ Remember me next time

© IEEE-CMX project

Report a bug | Powered by Yii Framework.

Login Screen

Main menu



Welcome to IEEE-CMX

Submit Metadata

Insert new Metadata in the IEEE-CMX database.

Manage Metadata

Manage Metadata stored in the IEEE-CMX database.

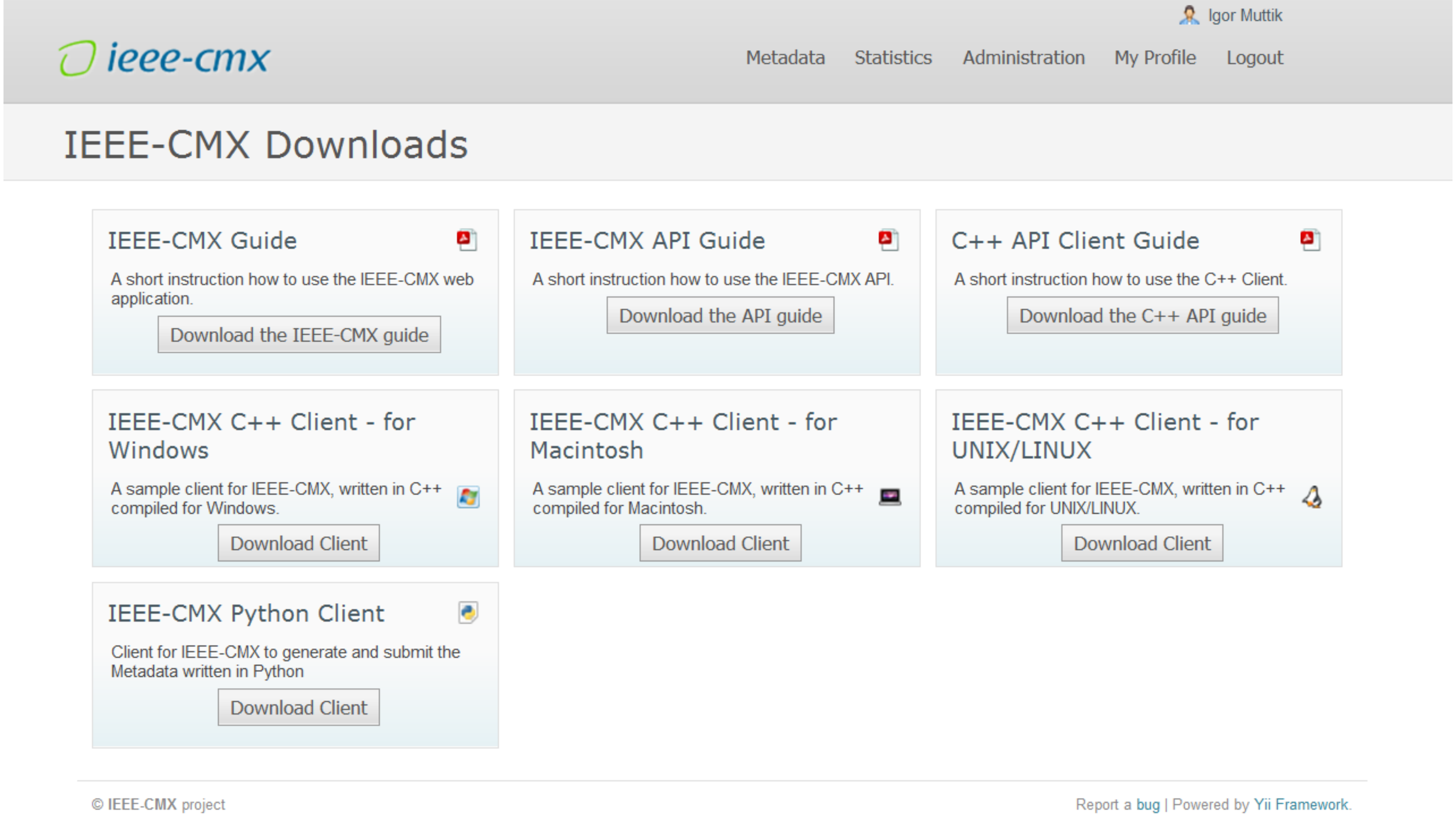
Statistics

View various graphical statistics about the usage of the IEEE-CMX system.

© IEEE-CMX project

Report a bug | Powered by Yii Framework.

# Screenshot (2)



Igor Muttik

ieee-cmx

Metadata | Statistics | Administration | My Profile | Logout

## IEEE-CMX Downloads

### IEEE-CMX Guide

A short instruction how to use the IEEE-CMX web application.

Download the IEEE-CMX guide

### IEEE-CMX API Guide

A short instruction how to use the IEEE-CMX API.

Download the API guide

### C++ API Client Guide

A short instruction how to use the C++ Client.

Download the C++ API guide

### IEEE-CMX C++ Client - for Windows

A sample client for IEEE-CMX, written in C++ compiled for Windows.

Download Client

### IEEE-CMX C++ Client - for Macintosh

A sample client for IEEE-CMX, written in C++ compiled for Macintosh.

Download Client

### IEEE-CMX C++ Client - for UNIX/LINUX

A sample client for IEEE-CMX, written in C++ compiled for UNIX/LINUX.

Download Client

### IEEE-CMX Python Client

Client for IEEE-CMX to generate and submit the Metadata written in Python

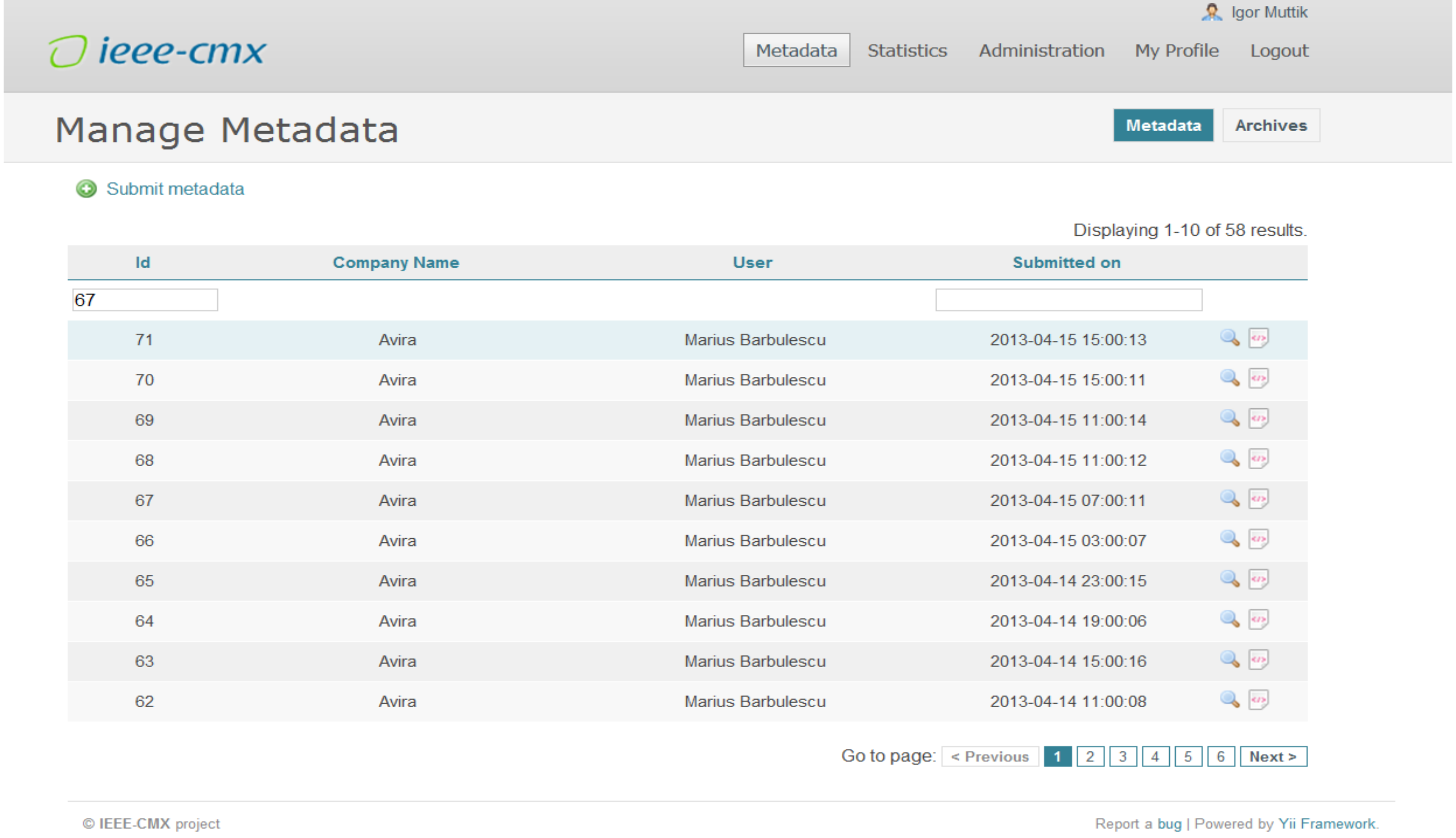
Download Client

© IEEE-CMX project

Report a bug | Powered by Yii Framework.

Downloads (documentation, tools and examples)

# Screenshot (3)



Manage Metadata

Displaying 1-10 of 58 results.

| Id | Company Name | User | Submitted on |
|---|---|---|---|
| 67 | | | |
| 71 | Avira | Marius Barbulescu | 2013-04-15 15:00:13 |
| 70 | Avira | Marius Barbulescu | 2013-04-15 15:00:11 |
| 69 | Avira | Marius Barbulescu | 2013-04-15 11:00:14 |
| 68 | Avira | Marius Barbulescu | 2013-04-15 11:00:12 |
| 67 | Avira | Marius Barbulescu | 2013-04-15 07:00:11 |
| 66 | Avira | Marius Barbulescu | 2013-04-15 03:00:07 |
| 65 | Avira | Marius Barbulescu | 2013-04-14 23:00:15 |
| 64 | Avira | Marius Barbulescu | 2013-04-14 19:00:06 |
| 63 | Avira | Marius Barbulescu | 2013-04-14 15:00:16 |
| 62 | Avira | Marius Barbulescu | 2013-04-14 11:00:08 |

© IEEE-CMX project

Report a bug | Powered by Yii Framework.

Metadata Web page (also accessible from cmd-line and/or APIs)

# Revocation

- Sometimes providers make a mistake
- More common with 3rd party providers
- Will go out at regular CMX content
    - Special tag will flag this as revocation
- As with all CMX data, the consumers decide what to do with the data

# Takeaways

- If you are a software producer
  - You will benefit from being a provider
  - Benefit: reduced FP rates from AV products
- If you are an enterprise administrator
  - You will benefit from being a provider
  - Benefit: you do not have to send actual software
- If you are a security/AV company
  - You may become a consumer
  - Benefit: reduces support costs due to lower FP rate

# Acknowledgements

We are extremely grateful to the Avira team for their efforts in implementing the CMX system and hosting it on their servers.

Special thanks go to Philipp Wolf and Thomas Wegele,
who organized the development, coded the system and provided documentation.

Danke