WMM – Uma ferramenta de extração de vestígios deixados pelo *Windows Live Messenger*

Galileu Batista de Sousa
Departamento de Polícia Federal
galileu at dpf.gov.br

***Abstract*— Personal communications based on Internet are growing continuously, specially using instant messaging tools, such as Windows Live Messenger - *WLM*. So, forensics midia content analysis needs to consider this tendency. This paper offers advances in three fronts in this scenario: add knowledgment about where artefacts are left by *WLM;* describes how to recover these artefacts - typically ciphered and dispersed over many points in the file system; and presents *WMM*, a public-domain Java tool, developed by the author and complaint to forensic methods, to automatically collect, process and show artefacts left by WLM.**

***Index Terms*— Instant Messaging, WLM, MSN Messenger, Conversation content, contact list. FBox. WMM.**

I. Introdução

Com o crescimento das comunicações pessoais usando a Internet, a análise forense de mídias computacionais ganhou mais uma atividade: identificar usuários, listas de contatos e conversações entre usuários dos programas de comunicação instantânea, tais como *MSN Messenger*, *Yahoo! Messenger* e *Skype*. Nesse contexto, identificar os contatos pode ser importante para avanços na investigação de delitos; as conversações, por sua vez, podem ser fundamentais inferir a ocorrência de ilícitos. Dependendo da natureza do crime, a lista de contatos combinada com as conversações formam a prova do crime.

O programa mais usado para comunicação instantânea é o *Windows Live Messenger* (*WLM)* [1,2], nome que designa o Microsoft *MSN Messenger* a partir da sua versão 8.0, com centenas de milhões de usuários. Apesar disso, as principais ferramentas de análise forense no mercado, *Forensic Toolkit* (*FTK)* [3] e *Encase* [4] não oferecem os métodos apropriados de extração de listas de contatos, conversações ou dados trocados entre participantes de conversações. O Encase faz *data carving* das conversações, mas não as vincula aos contatos ou usuários do *WLM*. FTK não considera o tema. O *Paraben Chat Examiner* [5] só trata versões anteriores do *MSN*.

Este artigo apresenta o *WMM*, uma ferramenta de extração de vestígios deixados pelo *WLM*. Os objetivos do autor, ao implementar o *WMM* foram:

- Dar um passo adiante na identificação da localização dos vestígios deixados pelo *WLM* no sistema de arquivos;
- Suportar investigações sobre o significado associado aos vestígios encontrados;
- Evidenciar as estruturas de dados, algoritmos e técnicas de criptografia usados para armazenar alguns dos vestígios.
- Oferecer uma ferramenta, de domínio público e código fonte aberto, para identificar listas e dados de contatos (incluindo imagens e conversações) deixados em um sistema pelo *WLM*.

O texto aqui apresentado pode ser visto como um complemento à referência clássica sobre artefatos forenses associados ao *WLM* [6]. Já o *WMM* pode ser visto como uma versão melhorada (ainda que em estágio de protótipo e sem interface visual) do *Forensic Box* (*FBox)* [7], a ferramenta gratuita, mas de domínio não público, tipicamente usada para extração de lista de contatos do *WLM*.

Muitas das implementações contidas no *WMM* e aqui discutidas resultaram da análise dos códigos binários do *WLM* e do *FBox*. A maior parte das investigações foi conduzida usando a versão 8.1 do *WLM* e a versão 1.43 do *FBox*.

A organização do texto é tal que explica um conceito e descreve como o *WMM* o implementa. A seção 2 discute o quê e onde é possível encontrar vestígios relativos a contatos do *WLM*. O formato dos arquivos contendo listas e dados de contatos, bem como os meios de decifrá-los e interpretá-los são apresentados nas seções III e IV. Nestas seções também estão formalizadas estratégias de recuperação (*carving) de* arquivos de contatos, mesmo que cifrados. A seção V discute como recuperar conversações e vinculá-las aos usuários do *WLM,* A seção VI descreve especificamente os detalhes de uso e as funcionalidades do *WMM*. Os trabalhos futuros estão na seção VII e a conclusão na seção VIII.

II. Usuários *WLM* e suas Listas de Contatos

Toda a base de informações deixada pelo *WLM* gira em torno de usuários e seus contatos. Portanto, descobrir onde e como esses dados estão armazenados é fundamental para atribuir um valor semântico relevante ao conjunto. Inicialmente, é importante mencionar que os dados de contatos podem, de fato, não estar armazenados. Este aspecto é controlado pela chave do "*Windows Registry*" "***HKEY_CURRENT_USER\Software\Microsoft\MSNMessenger\PerPassportSettings\MSN_PassportID\DisableCache***". Igualmente revelante é citar que, pelo

Galileu Batista de Sousa é Perito Criminal Federal, lotado no Setor Técnico-Científico da Superintendência Regional do Departamento de Polícia Federal no Estado do Rio Grande do Norte.

menos durante o primeiro *login* em um sistema, a lista de contatos é armazenada, sendo apagada na conveniência do usuário, *a posteriori*.

Uma questão recorrente na análise de registros do W*LM* é o mapeamento da identificação de um usuário *(*um nome de correio eletrônico, conhecido por *MSN Passport*) em um número de identificação – o *PassportID* - usado como parte do nome do caminho de localização de vários artefatos. A Figura 1 apresenta o pseudocódigo para este mapeamento, onde os caracteres do passaporte são *ASCII,* não *Unicode.*

```
uint  passportToPassportID(String eMail)
  uint passportID = 0;
  for (i = 0; i < strlen(passport); i++)
    passportID = passportID*101 + eMail[i];
  return passportID;
}
```

Figura 1 – Mapeamento de Passaporte para PassaporteID

A.Listas de Contatos

Quando a lista de contatos é armazenada localmente, formando um *cache* e viabilizando o *login* em redes lentas, há pelo menos dois diretórios onde os dados são armazenados: "***Documents and Settings\<windows user>\Contacts\<Passport>***" e '***Documents and Settings\<windows user>\Local Settings\Application Data\Microsoft\Windows Live Contacts\<Passport>\shadow***". Antes de descrever precisamente o conteúdo dos diretórios, deve-se notar que todos os arquivos neles contidos são cifrados. O formato dos arquivos e as cifras usadas serão descritas mais adiante. De antemão é suficiente saber que, depois de decifrados, os arquivos podem ser abstraídos por pares da forma: (ATRIBUTO, VALOR).

A Tabela 1 apresenta um quadro resumo dos principais arquivos (expressos em termos das suas extensões) encontrados nos diretórios acima, bem como descreve as informações de contatos neles armazenados.

Tabela 1 – Arquivos de contatos deixados pelo *WLM*

Arquivo	Descrição
.WindowsLiveContact	Dados de identificação de um contato em particular (cifra dupla).
.CONTACT	Dados de identificação de um contato (cifrado uma vez).
.MeContact	Dados de identificação de um usuário WLM.
.AddressBook	Metadados sobre a lista de contatos de um usuário WLM.
.WindowsLiveGroup	Dados de identificação de um grupo de contatos .
members.stg	Dados sobre todos os contatos de um usuários WLM.

No primeiro diretório há um arquivo para cada um dos contatos de um usuário *WLM. O* "*Passporte"* no nome do caminho indica a que o usuário o diretório se refere. Os arquivos de contatos são nomeados pelo *Global Unique ID* (GUID) do contato, seguido por "***.WindowsLiveContact***" e contêm as informações de caracterização de cada contato, entre elas: o endereço eletrônico, o caminho onde está armazenada a sua imagem, as mensagens de apresentação, além de informações que permitem vincular um contato a um grupo de contatos. O mapeamento entre os dados do contato e o *GUID* não foi determinado pelo autor até o momento da redação deste texto. A Tabela 2 apresenta os principais atributos contidos nos arquivos "***.WindowsLiveContact***" e as descrições dos seus conteúdos associados.

Tabela 2 – Atributos em *.WindowsLiveContact*

ATRIBUTO	Descrição do VALOR
CID	Número de Identificação do Contato.
Groups	Identificação do grupo do contato
MSNAddress	Endereço de correio eletrônico
FriendlyName	Nome de apresentação do contato durante as conversações.
StatusMessage	Nome de apresentação do contato na tela inicial do WLM.
IsMobileEnabled	Comunicações via telefone celular são possíveis para o contato.

Por vezes, dependendo da configuração do *WLM,* os arquivos com extensão "***.WindowsLiveContact***" são substituídos por arquivos com extensão "***.CONTACT***" . As informações contidas em ambos, para um mesmo contato, são exatamente as mesmas, a diferença entre eles reside na cifragem aplicada em cada caso (vide seção III).

No segundo diretório ('***Documents and Settings\<windows user>\Local Settings\Application Data\Microsoft\Windows Live Contacts\<Passport>\shadow***") há dois arquivos, cujos nomes terminam com "***.MeContact***" e "***.AddressBook***".

O arquivo "***.AddressBook***" contém informações gerais sobre a lista de contatos do usuário, incluindo número de contatos e número de grupos. Há também atributos com informações sobre datas de atualização dos demais arquivos, versão do *WLM* que criou a lista e outros atributos cuja semântica necessita ser melhor investigada.

O arquivo "***.MeContact***" armazena, para um usuário, informações similares àquelas contidas em um arquivo "***.WindowsLiveContact***" para um contato, exceto pela adição de informações sobre a imagem do usuário

armazenada no servidor, conforme descrito adiante.

Ainda no segundo diretório, um único arquivo "***members.stg***" guarda mais informações sobre cada um dos contatos do usuário *WLM* e, também, sobre contatos que não são de *WLM*, por exemplo contatos do serviço de correio eletrônico *Live* da *Microsoft.* Nestes casos, dados associados ao contato permitem classificar a sua natureza.

Muitas informações contidas em "***members.stg***" são repetições daquelas contidas nos arquivos "***.WindowsLiveContact***". Entre as informações adicionais especificamente armazenadas em "***members.stg***" estão: o *status* da relação entre o contato e usuário *WLM* (*Pending, Accepted*), quem fez o convite para a relação entre eles, e, quando se deu a última mudança na relação entre eles (aceitar o convite, por exemplo).

A Tabela 3 apresenta os principais atributos presentes no arquivo "***members.stg***" e as descrições dos seus conteúdos associados.

Tabela 3 – Principais dados em *members.stg*

ATRIBUTO	Descrição do VALOR
CID	Número de Identificação do Contato.
Type	EMAIL ou PASSPORT, caracterizando se o contato é do WLM ou não.
Passport	*Email* do contato, se TYPE=PASSPORT.
Email	*Email* do contato, se TYPE=EMAIL.
Role	(*Allow, Reverse*, ..) – solicitação de relacionamento iniciada pelo usuário ou contato.
Status	(*Accepted*, *Pending*, ...) Situação do relacionamento entre usuário e contato.

WMM combina as informações dos vários arquivos para criar a lista de contatos e o perfil de cada contato. O atributo de junção das várias fontes é o "***CID***". Quando contatos de diferentes fontes (arquivos "***.CONTACT***", "***.WindowsLiveContact***" *e* "***members.stg***") têm o mesmo "***CID***", a união dos seus conjuntos de atributos é realizada. Caso contrário, o atributo alternativo "***Unique_Id***" é usado para garantir a unicidade de um contato em uma lista.

No tocante aos usuários do *WLM*, *WMM* combina os atributos encontrados nos arquivos "***.MeContact***" e "***.AddressBook***", permitindo a visualização de dados inerentes aos contatos, bem como aqueles de características operacionais, mas importantes do ponto de vista forense, como a data em que a lista de contatos foi recuperada do servidor.

B.Grupos de Contatos

A organização dos contatos em grupos é uma funcionalidade bastante utilizada por usuários do *WLM.* Os arquivos que contêm as características dos grupos estão gravados em "***Documents and Settings\<windows user>\Contacts\<Passport>***" e têm nomes formados pelo "***GUID"*** do grupo seguido por "***.WindowsLiveGroup***". A informação preponderante associada a cada grupo é o seu nome (valor associado ao atributo "***Name"***) e um número de identificação (atributo "***SERVER_ID"*).**

A vinculação de um contato a um grupo se dá usando o atributo "***SERVER_ID***" do grupo e o atributo "***GROUPS"*** do contato. Este atributo do contato armazena dois valores separados por "|", onde o segundo correspondente exatamente ao "***SERVER_ID"*** do grupo, viabilizando a vinculação.

WMM faz a vinculação entre contatos e grupos. Um grupo artificial, cujo nome é "Não identificado" é criado para associar os contatos sem grupo definido, mantendo a semântica do *WLM* (de que cada contato pertence a um grupo) e melhorando a apresentação dos resultados.

III. Os Arquivos .*WindowsLiveContact* e *.MeContact*

Os arquivos "***.WindowsLiveContact***" (e "***.MeContact***") são cifrados usando *AES* [8] com chave de 128 *bits,* em modo *CBC* (*Cipher Block Chaining*), com vetor de inicialização formado por *128* zeros.

Uma vez obtida a chave e decifrado o conteúdo dos arquivos "***.WindowsLiveContact***", observa-se tratar de um conteúdo em *XML*. Nesse ponto, é importante caracterizar a diferença entre os arquivos "***.WindowsLiveContact***" e "***.CONTACT***". Naqueles todo o conteúdo é cifrado, tratando-se de um arquivo binário puro. Após uma primeira decifragem, um arquivo "***.WindowsLiveContact***" apresenta o mesmo conteúdo que o seu correspondente "***.CONTACT***". Porém, um arquivo "***.CONTACT***" não é completamente acessível, pois a parte mais significativa dos dados requer uma nova rodada de decifragem.

A.Obtenção da chave de decifragem

A chave de decifragem é derivada a partir do passaporte do usuário *WLM.* Inicialmente os *bytes* formadores do passaporte (em *Unicode low-endian* e incluindo o caractere final ***\u0000***) são submetidos a uma função de *hash SHA-1* [9], gerando uma base (B) para obtenção da chave.

A base obtida acima, de 20 *bytes (*160 *bits),* é usada para obter a chave de decifragem K da seguinte forma.: um *array* Z de 64 *bytes* é assim definido:

$$Z(i) = \begin{cases} 0x36 \ xor \ B(i) & \text{para } 0 \le i < 20 \\ 0x36 & \text{para } 20 \le i < 64 \end{cases}$$

Ao *array* Z é aplicada a função de *hash SHA-1*. Os primeiros 16 *bytes* (128 *bits*) resultantes formam a chave, K, de *decifragem* do arquivo.

A Figura 2, apresenta o pseudocódigo de obtenção da chave de decifragem a partir do passaporte.

Este comportamento foi determinado a partir de análise do código do *WLM* e de manuais da função "***CryptDeriveKey"*** pertencente à API de criptografia do *Microsoft Windows*. Especificamente, os passos para obtenção da chave, a partir da base são uma simplificação desta função.

```
keyFromPassport(String passport) {
   byte  K[] = new byte[16];
   byte  *B, *Z, *hash2;

   // Unicode-16LE + \0
   B = SHA1(passport);

   memset (Z, (byte) 0x36, 64);
   for (int i=0; i < 20; i++)
     Z[i] = Z[i] ^ B[i];

   hash2 = SHA1(Z, 64);
   for (int i=0; i < 16; i++)
     K[i] = hash2[i];
   return K;
}
```

Figura 2 – Obtenção da chave de decifragem dos contatos.

B. Conteúdo dos arquivos

Após uma primeira decifragem dos arquivos "***.WindowsLiveContact***" usando *AES-128-CBC* com a chave obtida conforme algoritmo precedente, o conteúdo se apresenta em *XML,* na forma umas poucas *tags* de identificação, duas delas com valores codificados em Base64 - ***<WL:WLPropBlob>*** e ***<WL:WLPropBlob2> -*** e, em alguns casos, a *tag* ***<c:Url>*** que armazena o caminho completo da imagem do contato.

Após decodificar o conteúdos de cada um dos *blobs,* observa-se que os dados são binários e estão cifrados. Uma segunda rodada de decifragem é então requerida, para finalmente obter os dados do contato de forma legível. Esta decifragem usa o mesmo algoritmo e a mesma chave que a primeira.

```
...
<prop>
<name>FriendlyName</name>
<value>XXX: "Cansado!!!"</value>
</prop>
...
```

Figura 3 – Fragmento de arquivos de contatos decifrado.

Os dados resultantes da decifragem são novamente conteúdos em XML formados de várias *tags* que seguem o padrão apresentado na Figura 3. A associação com os pares (ATRIBUTO, VALOR) definidos anteriormente é imediata. A maior parte dos atributos significativos aparece vinculado ao primeiro *blob. N*os casos analisados pelo autor, apenas um atributo, sem uso aparente, aparece no segundo *blob*.

WMM implementa todas as funcionalidades descritas acima, decifrando os arquivos, e gerando as propriedades associadas aos contatos, na forma de (ATRIBUTO, VALOR). Além das rotinas de criptografia, foi necessário usar *parsers* de *XML* para converter os arquivos em formatos internos viáveis às demais operações do programa.

Toda a discussão para os arquivos "***.WindowsLiveContact***" também se aplica aos arquivos "***.MeContact***", "***.AddressBook***" e "***.WindowsLiveGroup***". Em verdade, na implementação do *WMM,* um usuário de *WLM* também é um contato, pois aquele herda da classe que implementa este, adicionando-lhe propriedades e (muitos) outros métodos.

C.Imagens dos Contatos

O *WLM* armazena a imagem associada a um contato sempre que uma conversação com este é estabelecida, ou ainda quando há uma notificação de disponibilidade do contato para conversações. Como mencionado na seção precedente e diferentemente do argumentado em [6], a imagem dos contatos não foi encontrada vinculada ao atributo "***UserTitleUrl***", mas ao atributo "***Url***" nos arquivos "***.WindowsLiveContact***". De fato, nos arquivos "***.MeContact***" há uma atributo de nome similar ("***StaticUserTileUrl***"), mas este armazena a URL da foto do usuário no servidor *WLM*, não no cliente. O atributo ***UserTileFriendlyName*** faz referência ao nome original da imagem do usuário, antes de ser armazenada no servidor.

Localmente, as imagens de contatos são armazenadas em "***Documents and Settings\<windowsuser>\Local Settings\Temp\MessengerCache***" sem extensão e com o nome sendo o valor (em Base64) resultante da aplicação de um *hash SHA-1* sobre o conteúdo do arquivo.

WLM também mantém um histórico das imagens utilizadas pelos usuários. Estas imagens são mantidas no diretório "***Documents and Settings\<windows user>\Local Settings\Application Data\Microsoft\Messenger\<Passport>\ObjectStore\UserTile***", juntamente com imagens padrões, disponibilizadas quando o software é instalado.

WMM processa e vincula imagens aos contatos e também considera que, quando executado sobre um sistema cujo mapeamento de discos rígidos é diferente do originalmente presente no sistema alvo (o caso típico durante uma análise forense), o caminho associado a imagens deve modificado apropriadamente na apresentação dos resultados.

Com respeito às imagens dos usuários, *WMM* varre o diretório onde elas são armazenadas e vincula todas as imagens já usadas ao respectivo usuário. Neste processo, as imagens padrões da instalação do *WLM* são ignoradas, tendo por base os nomes das imagens.

D.Carving de ".WindowsLiveContact" e ".CONTACT"

Considerando que, mesmo com o *cache* de contatos desabilitado, pelo menos em um momento os arquivos de contatos *"****.WindowsLiveContact****" ou "****.CONTACT****"* são armazenados em um sistema onde um usuário se conectou ao *WLM, o carving* deles pode ser frutífero, quando não forem encontrados diretamente.

Uma propriedade interessante destes arquivos é o seu tamanho, normalmente inferior a 4KB, o que impõe um controle imediato ao tamanho e significa que se o *header* do arquivo for encontrado, é muito provável que o conteúdo

estará completo.

Todos os arquivos "***.CONTACT***" iniciam pelo padrão apresentado na Figura 4. Assim, a busca é um procedimento simples. Uma vez selecionado um conjunto de arquivos candidatos, tenta-se fazer as *decifragens* dos *blobs* nele presentes usando as chaves derivadas dos vários passaportes encontrados no sistema (seja no *Windows Registry* ou nos diretórios padrões que permanecem no sistema de arquivos mesmo quando as listas de contatos são apagadas [6}).

O *carving* arquivos "***.WindowsLiveContact***", por sua vez, requer uma inversão no processo. Inicialmente, o texto da Figura 4 deve ser cifrado, usando *AES-128-CBC* com cada uma das chaves referentes a passaportes encontrados. São estes padrões que devem ser procurados durante o *carving*. Ao encontrar um padrão, automaticamente é possível decifrar os dados do contato e vinculá-lo ao respectivo usuário *WLM.*

```
<?xml version="1.0" encoding="UTF-8" ?>
<c:contact c:Version="1" xmlns:
c="http://schemas.microsoft.com/Contact"
xmlns:xsi="http://www.w3.org/2001/XMLSche
ma-instance"
xmlns:WL="http://schemas.microsoft.com/Co
ntact/Extended/WL">
```

Figura 4 – Padrão para *carving* de arquivos "*.Contact*"

As mesmas justificativas e estratégias acima aplicadas podem ser usadas para buscar arquivos "***.MeContact***", "***.AddressBook***" e "***.WindowsLiveGroup***".

WMM ainda não implementa o *carving* destes arquivos.

IV. O Arquivo *members.stg*

O arquivo "***members.stg***" é organizado na forma de um "*Microsoft Structured Storage"*, também conhecido por *"OLE 2 Compound Document*". Trata-se de um arquivo que contém vários registros, cada um representando um arquivo, como em um arquivo *ZIP*. Tecnicamente, este formato representa um sistema de arquivos, onde o volume que o armazena é um arquivo de outro sistema de arquivos. No caso de "***members.stg***", cada registro/arquivo no volume contém informações sobre um contato.

Embora o arquivo "***members.stg***" não esteja cifrado, o conteúdo associado a cada contato (um *stream*, no jargão "*Structured Storage*") o é. Felizmente, a forma de cifragem é a mesma descrita anteriormente. O conteúdo original é cifrado uma única vez, usando *AES-128-CBC* com a chave de cifragem derivada do passaporte, seguindo o algoritmo da Figura 2. A decifragem é, portanto, direta.

O formato final resultante da decifragem é basicamente o mesmo já descrito na seção III.B, um conteúdo *XML* com pares na forma da Figura 3.

WMM executa o processamento do arquivo "***members.stg***", empregando uma biblioteca externa para fazer a leitura de cada um dos registros de contatos. Como esperado, após a decifragem, um *parser* de *XML* é usado para converter as chaves e seus valores no formato interno do *WMM.*

A. Carving de "members.stg"

O *carving* do arquivo "***members.stg***" também pode ser uma forma de encontrá-lo, quando não acessível diretamente.

Várias aspectos confluem para implementação de um processo de *carving* neste arquivo:

- O *carving* de um só arquivo permite recuperar informações de vários contatos de uma só vez;
- O arquivo tem uma assinatura (número mágico) conhecida (***0xE11AB1A1E011CFD0L***). No seu *header* de 512 *bytes* há informações para calcular o tamanho total [10];
- A busca pode ser executada eficientemente, limitando-se ao início das unidades de alocação, buscando um padrão pequeno e bem definido;
- O tamanho dos *streams* (dados de um contato) são normalmente menores que 4096 *bytes* e o tamanho total do arquivo menor que 6.8 MB, o que permite verificações e validações adicionais na estrutura do arquivo e simplifica o seu processamento;
- A decifragem correta por uma das várias chaves associadas a usuários *WLM* permite a vinculação dos contatos aos usuários.

Note-se também que diferentemente das estratégias sugeridas na referência [6], as aqui propostas permitem uma implementação automatizada do processo de recuperação de contatos e vinculação a usuários. Isso é possível porque as propriedades da cifragem e da organização do arquivo devem ser necessariamente considerados.

Em suma, mesmo sendo a abordagem proposta plenamente viável, *WMM* ainda não implementa o *carving* deste arquivo..

V. Rastreamento de Conversações

O registro de conversações pelo *WLM* se dá em um formato simples, embora tenha que ser configurado explicitamente pelo usuário. Se os registros serão ou não gravados é uma configuração constante na entrada do R*egistry* "***HKEY_CURRENT_USER\Software\Microsoft\MSNMessenger\PerPassportSettings\<MSN_PassportID>\MessageLoggingEnabled***". O diretório de gravação dos registros é configurado na entrada "***HKEY_CURRENT_USER\Software\Microsoft\MSNMessenger\PerPassportSettings\MSN_PassportID\MessageLogPath***".

Quando um usuário configura o diretório de gravação de conversações, o *WLM* adiciona automaticamente ao "***MessageLogPath***" (*default*[1] "***Documents and Settings\<windows user>\My Documents\My Received Files\<QuickName + Passport_ID>\History***") duas propriedades identificadoras do usuário: o seu *PassportID*, obtido de acordo com a Figura 1, e o atributo "***QuickName***" - basicamente o seu *email* sem o domínio.

No diretórios de gravação de conversações de um usuário

[1] Valor *default* para *Microsoft Windows* e *WLM* em inglês. O sinal de "+" representa a concatenação das cadeias de caracteres.

WLM há um arquivo *XML* que armazena todas as conversações de um usuário com um contato. O nome do arquivo para um contato segue o padrão: "***<QuickName + Passport_ID>.xml***". Neste caso ambos os atributos referem-se ao contato.

A. Formato dos arquivos de conversações

O arquivo *XML* que armazena as conversações entre contatos contém, principalmente, um conjunto de *tags* com o formato apresentado na Figura 6.

```
<Message Date="..." Time="..."
         DateTime="..." SessionID="...">
<From><User FriendlyName="..." /></From>
<To><User FriendlyName="..." /></To>
<Text Style="...">A mensagem </Text>
</Message>
```

Figura 5 - Registro de conversações entre usuários *WLM*.

A vinculação das conversações aos usuários e seus contatos é imediata em função dos dados de identificação presentes nos nomes de diretórios e arquivos. É importante notar, contudo, que os valores nos atributos *XML* "***FriendlyName***" em "***From***"/"***To***" presente nos arquivos de conversações são os mesmos que nos atributos "***FriendlyName***" encontrados nos arquivos de caracterização de usuários e contatos.

O "***SessionID***" é um número crescente que identifica conversações ocorridas durantes *logins* distintos no *WLM*. Todas as mensagens e eventos são englobados pela *tag* "***Log***", cujo conteúdo sumariza as sessões presentes no arquivo: "***<Log FirstSessionID="..." LastSessionID = "...">***".

Além de mensagens, é possível observar registro de outros eventos ocorridos durante conversações, tais como transferências de arquivos, ativação de câmeras e adição de terceiros a conversações. Os registros seguem o formato da Figura 6, exceto que a *tag* "***Invitation***" substitui "***Message***" para caracterizar o início do evento e *tag* "***InvitationResponse***" é usada para final do mesmo.

WMM identifica os arquivos de conversações e os vincula a usuários *WLM* e seus contatos . Na versão corrente, sessões não são consideradas.

B.Carving de conversações

O armazenamento de conversações não é ativo por padrão no *WLM* e dificilmente é configurado pelo usuário. Portanto, *carving* pode ser a única forma de acesso a conversações. Implementar *carving* se resume a fazer uma busca por expressão regular em um volume (ou mídia física completa), preferencialmente na região não alocada. Uma expressão regular simplificada para busca de mensagens é (quebra de linhas usada apenas para apresentação):

```
<Message
Date=\"[0-9/]+\"
Time=\"[0-9:]+\"
DateTime=\"[0-9TZ:\\.\\-]+\"
.+
</Message>
```

O processamento da expressão deve ser tal que a cadeia mínima que atenda ao padrão seja retornado e o processo repetido.

WMM implementa *carving* de conversações, recuperando mensagens e eventos ocorridos durante a conversação. Após recuperar um registro de conversação é importante vinculá-lo ao usuário *WLM* e seu contato. A única forma de fazer a vinculação é através do atributo "***FriendlyName".*** *WMM* tenta fazer a vinculação em vários níveis. Primeiro, determina se o "***FriendlyName***" da fonte e do destino correspondem a usuários *WLM* já reconhecidos.

Quando um registro de conversação é vinculado a um usuário, parte-se para identificar o seu interlocutor (um contato), buscando o "***FriendlyName***" na sua lista de contatos. Pode não ser possível, contudo, identificar o contato, visto que "***FriendlyName***" é um atributo mutável. Para estes casos *WMM* cria um "contato falso" e a ele vincula estas conversações.

Quando nem a fonte, nem o destino de registro de conversação encontrado pode ser estabelecido, *WMM* os coloca no lista de conversações órfãs, mostrando isso no seu relatório final.

Igualmente importante no tratamento de conversações recuperadas é a sua organização cronológica e o tratamento de repetições. *WMM* usa o valor no atributo "***DateTime***" (Figura 5) como chave de classificação e o par de atributos ("***DateTime***", "***Text***") para tratamento de repetições de mensagens relativas a um mesmo usuário.

Dada a importância de conversações, pode-se tentar recuperá-las através de *traces* do protocolo deixados no sistema de arquivos, ou ainda de arquivos no *cache* de navegadores, conforme descrito em [6]. *WMM* ainda não implementa estas estratégias.

VI.O *WMM*

Ainda que as motivações para escrever *WMM* tenham sido estabelecidas na seção I, dois requisitos de operação foram adicionados: preservação do sistema alvo, através de acessos somente de leitura, e, processamento automático de todas as evidências deixadas pelo *WLM* em um sistema.

WMM materializou-se como uma ferramenta escrita em Java e que implementa os aspectos discutidos nas seções precedentes. A opção por Java trouxe as vantagens associadas à execução em múltiplas plataformas, além de minimizar o uso de bibliotecas externas, dada a disponibilidade de *software* de suporte a criptografia e *XML* no *JRE* (*Java Runtime Environment*). O único pacote externo usado foi o "Apache POI - *Java API To Access Microsoft Format Files*", disponível em http://www.apache.org/dyn/closer.cgi/poi/

WMM é uma ferramenta de linha de comando que faz o procedimento de rastreamento sobre todo o sistema alvo, sem necessidade de repetição para cada usuário *WLM*. O resultado da execução é um conjunto de arquivos *HTML* contendo uma árvore de usuários, seus grupos de contatos e demais propriedades. A ferramenta tenta descobrir configurações e idiomas do sistema original, sem necessidade de configurações complexas. A Figura 6 mostra as opções de processamento do

WMM, supostamente auto-explicativas. Estão omitidos as configurações de pacotes e diretórios necessários

```
# java WMM -h
  -dN Debug level
     default 0 (todas as mensagens)
  -sX: Disco para busca de dados de
       usuários e contatos WLM
     default c:
  -cD Disco (Lógico ou físico) para
     carving de conversações
     0  - disco físico 0
     N  - disco físico N
     X: - disco lógico X:
  -CN Máximo de conversações a buscar
      0  - Ilimitado - busca todas.
     -1  - (default) - não busca.
      N  - Busca as N primeiras
  -rdir Diretório para gerar relatório
     default .
  -lN Detalhes no relatório
     0  - normal (default)
     1  - estendido
  -h Mostra esta mensagem
```

Figura 6 – Opções de processamento do WMM.

Observe-se que para aderir ao princípio forense de preservação das evidências, *WMM* funciona sem que seja necessária a execução do sistema operacional onde o *WLM* estava originalmente instalado. Contudo, a mídia deve ser mapeada como uma unidade de disco do sistema hospedeiro.

O resultado da execução do *WMM* pode ser observado na Figura 7, onde nomes de usuários foram modificados e imagens de fundo removidos.

USUÁRIOS DE MSN

Usuário Windows	Usuários de MSN
joao.teste	c2k@hotmail.com - Contatos Claudiaista14@hotmail.com - Contatos avogo_kert@hotmail.com - Contatos avoamagio@hotmail.com - Contatos abcga7@hotmail.com - Contatos AAAA2k@hotmail.com - Contatos gabrielabela1@hotmail.com - Contatos

CONVERSAÇÕES ENTRE USUÁRIOS IDENTIFICADOS

Usuário MSN	Participantes	Ação
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> Paulo de Tarso	Visualizar
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> .. Toni .:.	Visualizar
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> A @ Manaus	Visualizar
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> Armando	Visualizar
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> Claudia ista	Visualizar
AAAA2k@hotmail.com	AAAA: "Cansado!!!" <-> Carla	Visualizar

CONVERSAÇÕES - USUÁRIOS NÃO IDENTIFICADOS

Participantes	Ação
BBB: "Cansado!!!" <-> João Carlos	Visualizar

Figura 7 – Resultado de um processamento do WMM.

Ao clicar em um dos *links* relativos a contatos, visualiza-se algo como o apresentado na Figura 8, com os contatos apresentados em grupos e com seus *emails* e mensagens padrões. Particularmente, o usuário apresentado não possui foto associada, de outro modo ela seria exibida automaticamente, logo abaixo do cabeçalho. De toda sorte, pode-se notar que alguns dos contatos têm *links* para as respectivas fotos e/ou conversações.

Lista de contatos para abcga7@hotmail.com (Giselda)

Grupo	Contatos
Amigos (3)	franc_monteiro@hotmail.com - Narinha fksb@hotmail.com - fabio rdiia@hotmail.com - Ruth
Colegas de trabalho (6)	fat@hotmail.com - Txia Fatima (Foto) (Conversações) welima_39@hotmail.com - Wellington raedo@hotmail.com - RAIRA LEITE canuo@hotmail.com - Canuto (Foto) (Conversações) humber42@hotmail.com - ta onde??? gugd@hotmail.com - GUSTAVO
Família (4)	gabriea1@hotmail.com - GABRIELA (Foto) asta14@hotmail.com - - Amanda wg7@hotmail.com - Wagner André tiagio@hotmail.com - janilda (Foto)
	lis32@hotmail.com - (F)Lis(F) mromao@hotmail.com - Marli mapioli@hotmail.com - map@hotmail.com lfxju@hotmail.com - França Junior mapi@hotmail.com - marilene

Figura 8 – Lista de contatos de um usuárioWLM.

Um fragmento de conversação está apresentado na Figura 9. Neste caso com fotos de ambos os partícipes.


[redacted] @ Rio Branco/AC [redacted]: "Cansado!!!"


Carved	Data	Hora	De	Para	Mensagem
Não	2/5/2008	13:52:52	[redacted] @ Rio Branco/AC	[redacted]: "Cansado!!!"	
Não	2/5/2008	13:53:01	[redacted]: "Cansado!!!"	[redacted] @ Rio Branco/AC	diga [redacted]
Não	2/5/2008	13:53:53	[redacted] @ Rio Branco/AC	[redacted]: "Cansado!!!"	Falei com [redacted] ... pra uma possivel misso em [redacted] ... Ele chegou a falar com vc ?
Não	2/5/2008	13:53:59	[redacted]: "Cansado!!!"	[redacted] @ Rio Branco/AC	sim...
					Devo passar 1 ms em

Figura 9 – Uma conversação entre usuários WLM.

Não está mostrado, mas em qualquer ponto de referência a contatos há um *link* para acesso aos dados a ele vinculados.

WMM já foi usado com sucesso para obter vestígios de várias mídias. Em todos os casos, seus resultados foram confrontados com aqueles gerados pelo *FBox*, tendo sido os resultados iguais, nas partes comuns aos programas. Para os dados não tratados pelo *FBox*, o próprio *WLM* como padrão de comparação onde possível. Os resultados foram como esperado, porém testes mais exaustivos são requeridos.

VII. Trabalhos Futuros

O estágio do *WMM* permite a identificação de vários vestígios, porém não estão implementados o *carving* de listas de contatos, fundamentais para vinculação dos demais artefatos. Além disso, o *carving* de conversações pode ser melhorado, seguindo técnicas descritas ao longo do texto e rastreando vestígios do próprio protocolo do *MSN* [11].

É natural pensar-se em estender *WMM* para tratar vestígios deixados por outros programas de comunicação instantânea. De imediato, os vestígios deixados pelo *Skype* poderiam ser tratados, dada a disponibilidade de informação sobre o formato de armazenamento dos mesmos. Isto envolve um redesenho da arquitetura do *WMM* para manter a qualidade do código fonte, isolando o tratamento de cada uma das ferramentas um conjunto independente de classes.

A reescrita do *WMM* como parte de uma ferramenta genérica de análise forense é um trabalho motivador. O Encase, por suas características de extensibilidade através de uma linguagem de programação similar a Java é a alternativa a ser analisada.

Já em fase de conclusão está uma interface visual para o WMM.

Um outro horizonte de trabalho que se poderia vislumbrar seriam as versões anteriores do *WLM*, conhecidas coletivamente por *Microsoft Messenger* [12]. Estas versões oferecem um desafio a parte, pois a criptografia das listas de contatos baseia-se em uma função nativa do Windows ("***CryptProtectData***") implementada no nível do Sistema Operacional e dependente da senha do usuário *Windows* onde acessos ao *Messenger* foi executado. Estes aspectos tornam uma implementação fora do ambiente de execução original muito difícil e a operação complicada.

VIII. Conclusões

Este artigo apresenta uma ferramenta para extração de vestígios deixados do *Windows Live Messenger* em um sistema. Porém seu escopo vai além, ao mencionar onde encontrar e como ter acesso a tais vestígios. Embora a ferramenta careça de vários melhoramentos, já tem sido usada como forma de obtenção de dados, que de outra forma seriam operacionalmente difíceis de obter. Nos casos em que foi usada, os resultados do *WMM* foram confrontados contra aqueles obtidos por ferramentas similares, oferecendo resultados esperados. Porém, uma bateria maior de testes necessita ser realizada.

WMM se apresenta, portanto, como uma ferramenta que, com pouca configuração e de forma forense, extrai automaticamente um número significativo de artefatos deixados *WLM* em um sistema, o que pode ser útil tanto em investigação, quanto em materialização de delitos.

Referências

[1] Microsoft,. *Windows Live Messenger*, Disponível online em messenger.live.com.

[2] Nate Mook. MSN Messenger Most Used IM Client. Disponível online em betanews.com/article/MSN_messenger_Most_Used_IM_Client/1144778820

[3] Access Data. The Forensic Toolkit. Disponível online www.accessdata.com/media/en_US/print/papers/FTK2.0_cutsheet_Core_Print.pdf

[4] Guidance, Encase Forensic, disponível online em www.encase.com/downloads/EnCase_Forensic.pdf

[5] Paraben Corporation, Chat Examiner 1.0.2. Disponível online www.paraben-forensics.com/catalog/

[6] Wouter S. Van Dongen, Forensic artefacts left by Windows Live Messenger 8.0. Digital Investigation (4). 2007.

[7] _, Forensic Box, disponível por requisição a forensicbox@gmail.com.

[8] NIST. Announcing the ADVANCED ENCRYPTION STANDARD (AES). Disponível online em: csrc.nist.gov/publications/fips/fips197/fips-197.pdf

[9] NIST, NIST SECURE HASH STANDARD. Disponível online em: csrc.nist.gov/publications/ fips/fips180-2/fips-180-2.pdf

[10] Apache POI, POIFS File System Internals. Disponível online em: poi.apache.org/poifs/fileformat.html .

[11] Apache POI, POIFS File System Internals. Disponível online em: poi.apache.org/poifs/fileformat.html .

[12] Mike Mintz. MSN Messenger Protocol, general – HTTP connections. Disponível online em www.hypothetic.org/docs/msn.

[13] Mike Dickson. An examination into MSN Messenger 7.5 contact identification. Digital Investigation (3). 2006.