鉴定实践
Forensic Practice

# 一种针对 MongoDB 数据库的证据获取方法

吴 森，倪力舜
（上海辰星电子数据司法鉴定中心 公安部第三研究所信息网络安全公安部重点实验室，上海 201204）

**摘 要**：根据 MongoDB 文件型数据库的特点，提出了一种针对 MongoDB 文件型数据库的证据获取方法。在研究 GridFS 文件系统原理的基础上，详细阐述了 MongoDB 和 GirdFS 文件系统结构，对 MongoDB 存放数据的文件进行了分析，进而获取被删除的文件数据。实验结果表明该方法能有效地从 MongoDB 数据库中获取电子证据。

**关键词**：计算机取证；电子证据；MongoDB
**中图分类号**：TP33；DF794　　**文献标志码**：A　　doi：10.3969/j.issn.1671-2072.2011.03.012
**文章编号**：1671-2072-(2011)03-0054-02

## Electronic Evidence Acquisition from MongoDB

*WU Miao, NI Li-Shun*
*(Shanghai Stars Digital Forensic Center, Information and Network Security Lab,*
*The Third Research Institute of Ministry of Public Security, Shanghai 201204, China)*

**Abstract**: According to the characteristic of MongoDB, an approach of electronic evidence acquisition is proposed. The structure of MongoDB and GirdFS is introduced; the file which MongoDB stores data in is analyzed; and the procedure to recover deleted data in MongoDB is put forward. The experiment shows that the approach can acquire electronic evidence from MongoDB effectively.

**Key words**: forensic computer technology; electronic evidence; MongoDB

## 1 概 述

2009 年计算机数据存储领域提出了一种 NoSQL 类型的文件型数据库，其代表产品 MongoDB 已在国内外被广泛采用，大有取代传统二维 SQL 关系型数据库的趋势。黑客极有可能将可疑数据文件隐藏在 MongoDB 数据库中，从而躲避现有的计算机取证技术的取证。本文针对上述情况，对 MongoDB 文件系统进行分析，找出藏在其中的电子数据证据，提出一种针对 MongoDB 文件型数据库的取证方法。

MongoDB 是一种 NoSQL 文件型分布式数据库，自诞生以来，在性能和稳定性上逐渐显露出优势。相对传统的 SQL 类型的数据库，其在存储大数据和文件方面又有着明显的优势。MongoDB 使用 BSON 格式存储小于 4M 的数据，使用 GridFS 文件系统存储文件。

对目标主机进行证据获取，第一步使用硬盘复制机复制整个目标硬盘，防止在后续的证据获取操作中破坏原始证据。第二步使用文件恢复软件对硬盘中被删除文件进行恢复，尽可能地恢复文件，其中可能也会包括 MongoDB 数据文件。第三步分析 MongoDB 数据文件，找出藏在其中的可疑数据和文件。

收稿日期：2010-09-20
基金项目：国家 863 计划项目(2008AA01Z412)
作者简介：吴森(1982—)，男，研究实习员，硕士，主要从事网络信息安全研究。E-mail：wum383@gmail.com。

## 2 方 法

### 2.1 GridFS 文件系统基本结构

MongoDB 使用 GridFS 文件系统来存储文件，GridFS 文件系统是一种把文件存储在 MongoDB 数据库中的规范。GridFS 文件系统工作原理就是把文件分成几个小的块(chunk)，每个 chunk 大小通常为 256k，每个 chunk 作为独立的记录存放在 chunk 集合中，而所有的关于文件信息的元数据则存放在 file 集合中。对于一个文件，会有一个 file 块和多个 chunk 块。在删除 MongoDB 数据库中的文件数据时，数据并不会立刻被物理删除，只是被打上删除的标记位。GridFS 文件系统结构如图 1 所示。

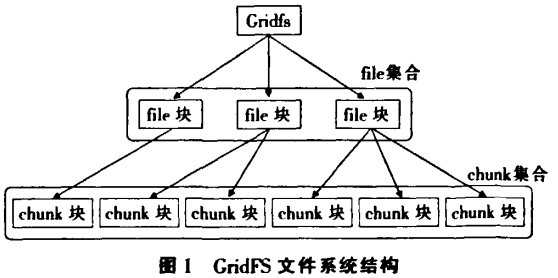


图 1 GridFS 文件系统结构

## 2.2 MongoDB证据获取方法

文中通过一个完整的文件恢复过程来阐述针对MongoDB数据库的证据获取方法。在描述方法之前，使用MongoDB创建一个testdb数据库，并且存入2个文件python.msi和private1.pem。

### 2.2.1 对MongoDB数据文件进行分析

(1)使用十六进制编辑器打开testdb.0文件(可能还会有testdb.1,testdb.2…，视数据量而定)，根据GridFS文件系统结构首先找到file快集合,再从集合中找到文件的file块,一个文件对应一个file块,如图2所示。

```
00003b00h: 44 43 42 41 00 00 00 00 00 00 3B 00 00 00 FF FF FF FF ; DCBA.....;...
00003b10h: 00 00 00 00 00 FF FF FF FF 00 00 00 00 00 74 65 73 74 ; .....          ....test
00003b20h: 64 62 2E 66 73 2E 66 69 6C 65 73 00 00 00 00 00 00 00 ; db.fs.files.....
```

图2 file集合位置

从图2中可以看出testdb数据库的file集合从00003b00h开始。在没有删除MongoDB数据库中文件数据的情况下,2个文件的file块,如图3所示。

```
00003ba0h: 00 00 00 00 B0 3B 00 00 00 00 00 00 4C 3C 00 00 ; ....?......L<..
00003bb0h: 9C 00 00 00 00 00 3B 00 00 4C 3C 00 00 FF FF FF FF ; ?....;..L<..
00003bc0h: 8C 00 00 00 07 5F 69 64 00 4C 91 9B 69 0B 34 00 ; ?..._id.L慱i.4.
00003bd0h: 00 00 00 32 AE 02 66 69 6C 65 6E 61 6D 65 00 0F ; ...2?filename..
00003be0h: 00 00 00 63 3A 5C 5C 70 79 74 68 6F 6E 2E 6D 73 ; ...c:\\python.ms
00003bf0h: 69 00 10 6C 65 6E 67 74 68 00 00 CE DD 00 10 63 ; i..length..暘..c
00003c00h: 68 75 6E 6B 53 69 7A 65 00 00 00 04 00 09 75 70 ; hunkSize......up
00003c10h: 6C 6F 61 64 44 61 74 65 00 64 14 C7 18 2B 01 00 ; loadDate.d.?+..
00003c20h: 00 02 6D 64 35 00 21 00 00 00 34 62 63 61 30 30 ; ..md5.!...4bca00
00003c30h: 31 37 31 61 61 36 34 62 34 38 38 36 62 38 38 ; 171aa614b4886b88
00003c40h: 39 32 39 30 63 34 66 65 64 39 00 00 A0 00 00 00 ; 9290c4fed9..?..
00003c50h: 00 3B 00 00 FF FF FF FF B0 3B 00 00 8E 00 00 00 ; .;...      ?..?..
00003c60h: 07 5F 69 64 00 4C 91 9B 7D 24 5B 00 00 00 00 6F ; ._id.L慱}$[....o
00003c70h: 70 02 66 69 6C 65 6E 61 6D 65 00 11 00 00 00 63 ; p.filename.....c
00003c80h: 3A 5C 5C 70 72 69 76 61 74 65 31 2E 70 65 6D 00 ; :\\private1.pem.
00003c90h: 10 6C 65 6E 67 74 68 00 6B 03 00 00 10 63 68 75 ; .length.k....chu
00003ca0h: 6E 6B 53 69 7A 65 00 00 00 04 00 09 75 70 6C 6F ; nkSize......uplo
00003cb0h: 61 64 44 61 74 65 00 A6 62 C7 18 2B 01 00 00 02 ; adDate.  ?+....
00003cc0h: 6D 64 35 00 21 00 00 00 31 37 38 63 30 30 32 36 ; md5.!...178c0026
00003cd0h: 32 30 65 39 64 38 33 34 33 62 38 66 61 61 34 32 ; 20e9d8343b8faa42
00003ce0h: 35 33 30 33 37 30 33 66 00 00 00 00 14 1E 00 00 ; 5303703f........
00003cf0h: 00 3B 00 00 FF FF FF FF 00 00 00 00 EE EE EE EE ; .;..          ....
```

图3 文件file块

从图3的文件file块中可以读出文件的元数据，例如灰底的部分即为private1.pem文件的长度，为0x036B(875字节)，还可以读出该文件的上传日期、md5值等等。元数据中较为重要的是文件在MongoDB中的_id值,该值唯一标识一个文件,该值如图4灰底部分所示,private1.pem文件的_id值为4C919B7D245B000000006F70。

```
00003c60h: 07 5F 69 64 00 4C 91 9B 7D 24 5B 00 00 00 00 6F ; ._id.L慱}$[....o
00003c70h: 70 02 66 69 6C 65 6E 61 6D 65 00 11 00 00 00 63 ; p.filename.....c
00003c80h: 3A 5C 5C 70 72 69 76 61 74 65 31 2E 70 65 6D 00 ; :\\private1.pem.
```

图4 文件_id值

(2)根据_id值找出private1.pem文件file块对应的chunk块,chunk块中存放的是该文件的内容，如图5所示。

图5中的灰底部分即是文件内容的开头,可以根据file块中找到的文件长度读出该文件。

```
00045ba0h: 00 00 00 00 B0 5B 04 00 00 00 00 00 B0 5B 04 00 ; ....?......?..
00045bb0h: C0 03 00 00 00 00 5B 04 00 FF FF FF FF FF FF FF ; ?...[..
00045bc0h: AD 03 00 00 07 5F 69 64 00 4C 91 9B 7D D3 09 00 ; ?..._id.L慱}?.
00045bd0h: 00 00 00 59 51 07 66 69 6C 65 73 5F 69 64 00 4C ; ...YQ.files_id.L
00045be0h: 91 9B 7D 24 5B 00 00 00 00 6F 70 10 6E 00 00 00 ; 慱}$[....op.n...
00045bf0h: 00 00 05 64 61 74 61 00 6F 03 00 00 02 6B 03 00 ; ...data.o....k..
00045c00h: 00 2D 2D 2D 2D 42 45 47 49 4E 20 52 53 41 20 ; .----BEGIN RSA
00045c10h: 50 52 49 56 41 54 45 20 4B 45 59 2D 2D 2D 2D 2D ; PRIVATE KEY-----
00045c20h: 0A 4D 49 49 43 58 51 49 42 41 41 4B 42 67 51 44 ; .MIICXQIBAAKBgQD
00045c30h: 44 4A 79 4F 5A 2B 57 67 41 32 77 6C 68 39 6A 5A ; DJyOZ+WgA2wlh9jZ
00045c40h: 4B 71 64 70 6B 56 61 62 5A 47 56 71 6F 43 36 61 ; KqdpkVabZGVqoC6a
00045c50h: 56 47 58 70 37 53 42 2B 53 47 4F 52 43 79 6B 51 ; VGXp7SB+SGORCykQ
00045c60h: 7A 53 32 47 72 34 55 4E 59 56 71 62 71 4C 69 4C ; zS2Gr4UNYVqbqLiL
```

图5 文件内容

(3)通过使用mongofiles -d testdb delete命令删除testdb数据库中的private1.pem和python.msi文件,删除之后,通过正常的访问渠道是无法找回文件的,可以通过mongofiles -d testdb list命令尝试查看数据库中是否还有文件,结果如图6所示。

```
E:\db\mongodb\bin>mongofiles -d testdb list
connected to: 127.0.0.1

E:\db\mongodb\bin>_
```

图6 正常渠道读取MongoDB数据库中文件

从图6中可以看出数据库中所有的文件都被删除了。事实上,python.msi和private1.pem这2个文件并没有立刻被物理删除。通过分析MongoDB数据文件仍可以找回。

### 2.2.2 文件恢复

(1)虽然通过mongofiles命令删除了数据,但是文件的file块还在,这就说明文件的元数据完好,并没有被后续的数据覆盖。这是一种比较理想的情况,删除文件之后的file块如图7所示。

```
00003bc0h: EE EE EE EE 07 5F 69 64 00 4C 91 9B 69 0B 34 00 ; ????._id.L慱i.4.
00003bd0h: 00 00 00 32 AE 02 66 69 6C 65 6E 61 6D 65 00 0F ; ...2?filename..
00003be0h: 00 00 00 63 3A 5C 5C 70 79 74 68 6F 6E 2E 6D 73 ; ...c:\\python.ms
00003bf0h: 69 00 10 6C 65 6E 67 74 68 00 00 CE DD 00 10 63 ; i..length..暘..c
00003c00h: 68 75 6E 6B 53 69 7A 65 00 00 00 04 00 09 75 70 ; hunkSize......up
00003c10h: 6C 6F 61 64 44 61 74 65 00 64 14 C7 18 2B 01 00 ; loadDate.d.?+..
00003c20h: 00 02 6D 64 35 00 21 00 00 00 34 62 63 61 30 30 ; ..md5.!...4bca00
00003c30h: 31 37 31 61 61 36 34 62 34 38 38 36 62 38 38 ; 171aa614b4886b88
00003c40h: 39 32 39 30 63 34 66 65 64 39 00 00 A0 00 00 00 ; 9290c4fed9..?..
00003c50h: 00 3B 00 00 00 00 00 00 B0 3B 00 00 EE EE EE EE ; .;.......?..锘锘
00003c60h: 07 5F 69 64 00 4C 91 9B 7D 24 5B 00 00 00 00 6F ; ._id.L慱}$[....o
00003c70h: 70 02 66 69 6C 65 6E 61 6D 65 00 11 00 00 00 63 ; p.filename.....c
00003c80h: 3A 5C 5C 70 72 69 76 61 74 65 31 2E 70 65 6D 00 ; :\\private1.pem.
00003c90h: 10 6C 65 6E 67 74 68 00 6B 03 00 00 10 63 68 75 ; .length.k....chu
00003ca0h: 6E 6B 53 69 7A 65 00 00 00 04 00 09 75 70 6C 6F ; nkSize......uplo
00003cb0h: 61 64 44 61 74 65 00 A6 62 C7 18 2B 01 00 00 02 ; adDate.  ?+....
00003cc0h: 6D 64 35 00 21 00 00 00 31 37 38 63 30 30 32 36 ; md5.!...178c0026
00003cd0h: 32 30 65 39 64 38 33 34 33 62 38 66 61 61 34 32 ; 20e9d8343b8faa42
00003ce0h: 35 33 30 33 37 30 33 66 00 00 00 00 14 1E 00 00 ; 5303703f........
00003cf0h: 00 3B 00 00 FF FF FF FF 00 00 00 00 EE EE EE EE ; .;..          ....锘锘
```

图7 删除文件之后的file块

从图7中可以看出虽然用户删除了MongoDB数据库中的文件,但是文件的file块还在,只是前面4个字节被标记成EE EE EE EE。接下来读出文件的_id和length值，然后通过_id值和chunk集合中每个chunk块的files_id进行匹配,从而找到该文件的chunk块集合，最后读出chunk块中的data属性中的数据。本例中，提取出private1.pem文件file块中的_id值和length,找到chunk块,结果如图8所示。(下转第62页)

行为时，应予惩戒和处罚，涉及触犯刑律的，应报公安机关处理。其次，在行业自律方面，司法鉴定(人)协会等行业组织应承担更多的自律职能，搞好业务培训和本地区业务规范指导工作，强化法医临床学鉴定业务培训，提高鉴定人的专业水平。同时对标准中存在的过于原则的规定加以细化，形成地区性乃至全国性共识，杜绝少数鉴定人钻标准条款的漏洞。

针对上述问题，还应当强化重新鉴定职能及其相关配套制度如严格执业纪律、建立错鉴责任追究等制度的建设。通过重新鉴定，可以对鉴定机构和鉴定人的业务水平和职业道德进行确实评价，聚焦某些业务水平低下、职业道德败坏的鉴定机构和鉴定人。通过对其进行严肃处理，可以净化司法鉴定行业，逐步真正实现司法鉴定科学、客观、公正的目标，为社会主义法治社会、和谐社会建设做出应有的贡献。

**参考文献：**

[1] 霍先丹. 规范重新鉴定条件保障司法鉴定质量——对《司法鉴定程序通则》关于重新鉴定规定的解读[J]. 中国司法，2007，(12)：82-86.

[2] 胡锡庆. 略论重新鉴定[J]. 中国司法鉴定，2010，(2)：11-15.

(本文编辑：夏文涛)

(上接第 55 页)

```
00045bc0h: ■■EE EE EE 07 5F 69 64 00 4C 91 ■■ 7D D3 09 00 ; 锘锘._id.L锟)?.
00045bd0h: 00 00 00 59 51 07 66 69 6C 65 73 5F 69 64 00 4C ; ...YQ.files_id.L
00045be0h: 91 9B 7D 24 5B 00 00 00 00 6F 70 10 6E 00 00 00 ; 锟)$[....op.n...
00045bf0h: 00 00 05 64 61 74 61 00 6F 03 00 00 02 6B 03 00 ; ...data.o....k..
00045c00h: 00 2D 2D 2D 2D 2D 42 45 47 49 4E 20 52 53 41 20 ; .-----BEGIN RSA
00045c10h: 50 52 49 56 41 54 45 20 4B 45 59 2D 2D 2D 2D 2D ; PRIVATE KEY-----
00045c20h: 0A 4D 49 49 43 58 51 49 42 41 41 4B 42 67 51 44 ; .MIICXQIBAAKBgQD
00045c30h: 44 4A 79 4F 5A 2B 57 67 41 32 77 6C 68 39 6A 5A ; DJyOZ+WgA2wlh9jZ
00045c40h: 4B 71 64 70 6B 56 61 62 5A 47 56 71 6F 43 36 61 ; KqdpkVabZGVqoC6a
00045c50h: 56 47 58 70 37 53 42 2B 53 47 4F 52 43 79 6B 51 ; VGXp7SB+SGORCykQ
00045c60h: 7A 53 32 47 72 34 55 4E 59 56 71 62 71 4C 69 4C ; zS2Gr4UNYVqbqLiL
```

图 8 被删除文件的 chunk 块

根据之前从 file 块中提取的文件长度值，读出从 data 关键字后面 10 个字节(不同版本的 MongoDB 可能有所不同)开始的 length 字节的数据，保存到一个文件，完成文件恢复。

(2) 通过 mongofiles 命令删除了数据，后继的数据也覆盖了文件的 file 块，但 chunk 块仍完好。

在这种情况下就不能从文件的 file 块着手恢复文件，因为文件的 file 块已经被其它文件的 file 块覆盖，只能从 chunk 块着手。任何被用户删除的文件的 chunk 块前 4 个字节都被标记为 EE EE EE EE。查找连续字节为 EE EE EE EE 07 5F 69 64 的 chunk 块，把 files_id 相同的归在一组中，表示它们属于同一个文件。接下来，拷贝出每组中所有 chunk 块的 data 属性部分，每组 chunk 块的 data 内容是从 data 关键字后 10 个字节(视不同版本的 MongoDB 而定)一直读到接近下一个 chunk 块的开始部分，并且是 00 00 字节前面的一个字节，最后合并在一起就完成了文件数据恢复。

## 3 结 论

本文分析了一种针对 MongoDB 文件型数据库的证据获取方法。通过分析 MongoDB 数据库数据文件结构获得已删除的文件数据。实验表明该方法能有效地恢复 MongoDB 数据库中被删除的文件数据，为计算机取证提供有力的支持。

(本文编辑：施少培)

# 一种针对MongoDB数据库的证据获取方法



作者： 吴淼， 倪力舜， WU Miao， NI Li-Shun
作者单位： 上海展星电子数据司法鉴定中心,公安部第三研究所信息网络安全公安部重点实验室,上海,201204

刊名： 中国司法鉴定 PKU
英文刊名： CHINESE JOURNAL OF FORENSIC SCIENCES
年，卷(期)： 2011(3)