

#OSDFCon, oct 26 2016

# Clearing the fog on cloud forensics

**VASSIL ROUSSEV**

vassil@roussev.net

**SHANE McCULLEY**

smcculle@uno.edu

# whoami

≡ Professor @UNO

≡ DFRWS.org co-founder & organizer

≡ sometimes I even write code

> sdhash

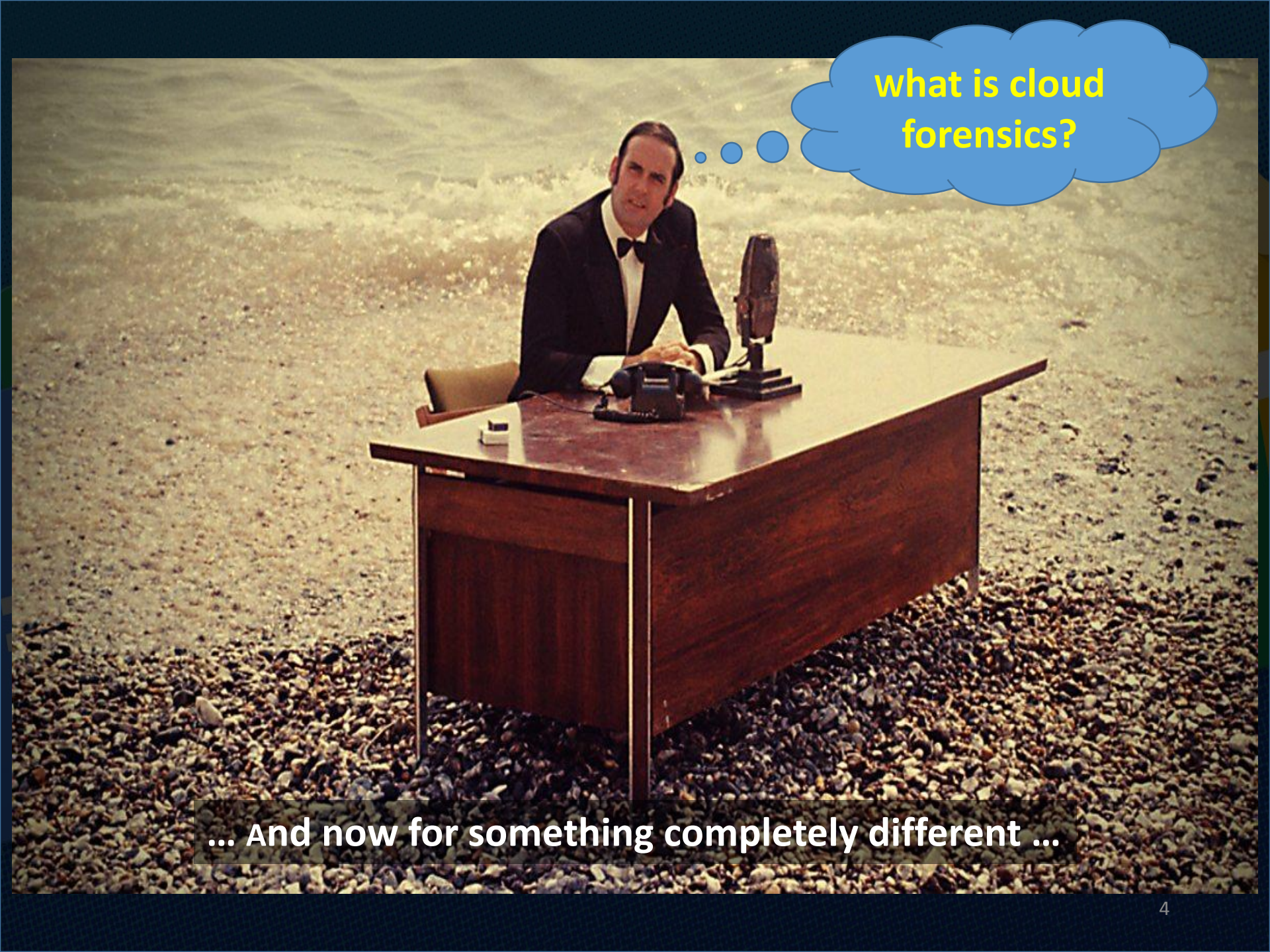


# *Digital forensics* since the 1980s

- ≡ All computations local
- ≡ All storage local; lots of leftovers
  - ➔ All evidence local
  - ➔ All tools designed for HDD/RAM analysis
- ≡ By now, we are pretty good at this!





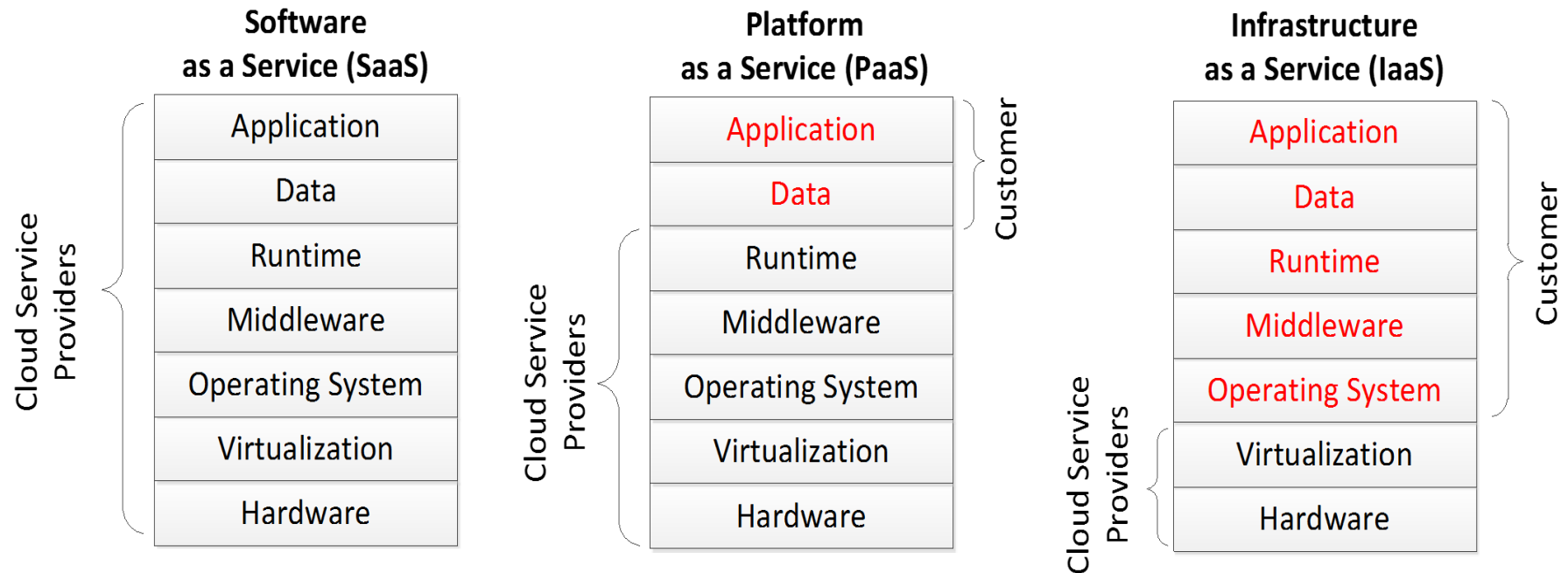


what is cloud forensics?

... And now for something completely different ...



# The cloud troika



Let's take a  
look here

???

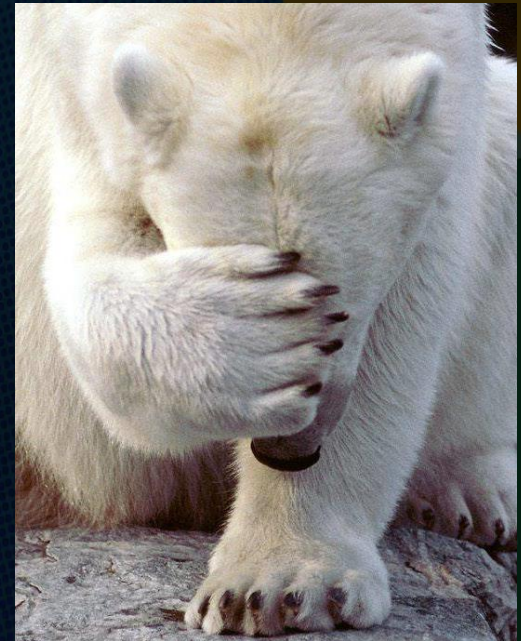
We can (mostly)  
handle this

# Example: cloud drive acquisition

≡ "Traditional" solution

> go to the client, look for the leftovers:

- » [Dropbox] Quick & Choo, 2013
- » [GoogleDrive] Quick & Choo, 2014
- » [ownCloud] Do et al., 2014
- » [SugarSync] Shariati et al., 2016
- » [Mega] Daryabar et al., 2016
- » ...



≡ "small" detail → it is kinda **WRONG**:

- > partial replication (data may not be on device)
- > versions (only one on the client)
- > cloud-native artifacts (e.g. Google Docs)





# Fixing it: kumodd

≡ Approach

> use the API

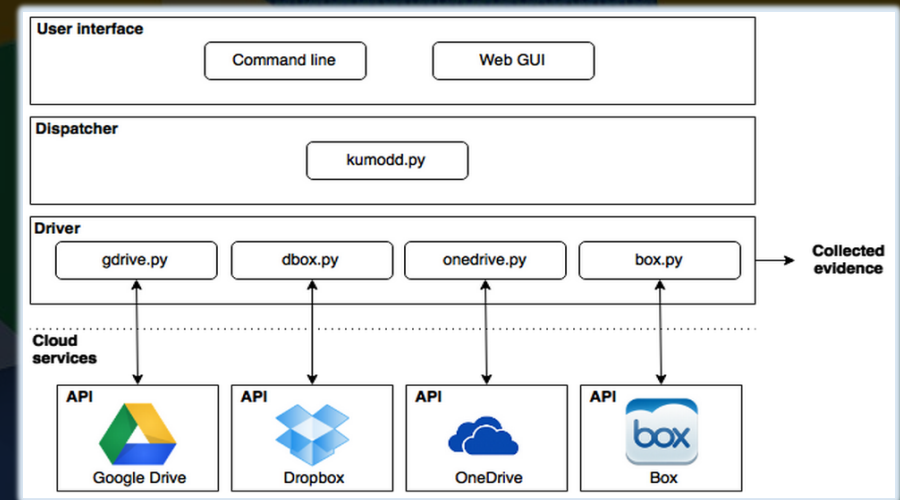
≡ Problems solved

> partial replication ✓

> revision acquisition ✓

> cloud natives? ✓ ✗

≡ cloud natives → kumodocs



# cloud-native artifacts

≡ Definition

data objects which maintain the state of web/SaaS applications, and have no external representation on the client.

→ I.e., these are *internal objects* for the app.



# why?

## How I REVERSE ENGINEERED GOOGLE DOCS

To Play Back Any Document

JAMES SOMER

NOVEMBER 5TH, 2014



Draftback

offered by jsomers.net

★★★★★ (33)

[Productivity](#)

20,368 users

ADDED TO CHROME

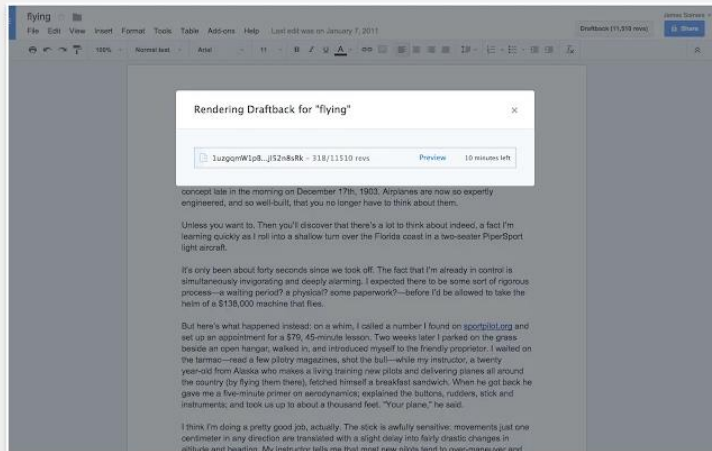
OVERVIEW

REVIEWS

SUPPORT

RELATED

G+ 380



USERS OF THIS EXTENSION HAVE ALSO USED

Compatible with your device

### The archaeology of great writing

Draftback lets you play back the revision history of any Google Doc you can edit. It's like going back in time to look over your own shoulder as you write. Notes:

- With Draftback, your data is kept entirely private. Draftback was purposely designed so that you could play back your own docs without having to share them with a third party. This is -your- data; Draftback just lets you see it in a new way.

- Draftback only needs access to docs.google.com to get the revision data for

[Website](#)

[Report Abuse](#)

Version: 0.0.8

Updated: March 2, 2015

Size: 471KB

Language: English (United States)



# Revisions

## Revision history

Today, 6:43 PM



100% ▾

Total: 1 edit



## Revision history

**Today, 6:43 PM**

■ Vassil Roussev

September 27, 8:32 PM

■ Vassil Roussev

March 18, 2013, 9:12 PM

■ Tom Sires

March 12, 2013, 1:10 AM

■ Tom Sires

March 12, 2013, 12:10 AM

■ Tom Sires

March 11, 2013, 10:59 PM

■ Tom Sires

March 11, 2013, 8:06 PM

■ Tom Sires

March 11, 2013, 7:24 PM

■ Tom Sires

March 11, 2013, 6:38 PM

■ Tom Sires

☒ Show changes

Show more detailed revisions

### Common mistakes from CSCI 4311 PA2:

~~Some test text~~

#### 1a) Using available() to check for the end of a stream

Classes implementing available() return an “estimate” of the number of bytes that can be read without blocking. If this returns zero, it only means there is no data to read **now**, but there may be data to read in the future. Streams backed by sockets or some other source that buffers data often need more time for data to become available.

A more reliable indicator of the end of a stream is a return value of -1 from a read() method, or null from readLine(). This is the only guarantee that no more data will be available, aside from an IOException for some other reason.

#### 1b) Using available() to check whether an input stream has been redirected

Unfortunately, there does not seem to be a portable way to check whether stdin has been redirected in Java. In C, the “isatty” function will let you check whether stdin or stdout is





# More(!) revisions

The screenshot shows a web editor interface for a document titled "CSCI-4311-PA2 Notes". The interface includes a menu bar (File, Edit, View, Insert, Format, Tools, Table, Add-ons, Help), a status bar (Last edit was 24 min...), and a toolbar with various editing tools. A yellow highlight is placed over the "Draftback (2,971 revs)" button. A modal window titled "Rendering Draftback for 'CSCI-4311-PA1 Notes'" is open, displaying a list of revisions. The first revision is highlighted in green and shows the file "1EhfdnBKHw...QssfyN568" with 1717/1717 revisions, a "View" link, and a green checkmark icon.

CSCI-4311-PA2 Notes ☆

File Edit View Insert Format Tools Table Add-ons Help Last edit was 24 min... Draftback (2,971 revs) Comments Share

100% Normal text Arial 11 B I U A More

Rendering Draftback for "CSCI-4311-PA1 Notes"

1EhfdnBKHw...QssfyN568 - 1717/1717 revs View

**Common mistakes from CSCI 4311 PA2:**

**1a) Using available() to check for the end of a stream**

Classes implementing available() return an "estimate" of the number of bytes that can be read without blocking. If this returns zero, it only means there is no data to read now, but there may



# p(l)ayback

Finish & Publish 387 revisions

Common issues from CSCI 4311 PA1:

Using line-buffered I/O.  
BufferedReader.readLine() "A line is considered to be terminated by any one of a line feed ('\n'), a carriage return ('\r'), or a carriage return followed immediately by a linefeed."  
For text data, this is generally not a problem, as long as platform-independent linebreaks are added to output, e.g. with BufferedWriter.newLine().

Relevant docs:  
[http://docs.oracle.com/javase/7/docs/api/java/io/BufferedReader.html#readLine\(\)](http://docs.oracle.com/javase/7/docs/api/java/io/BufferedReader.html#readLine())  
[http://docs.oracle.com/javase/7/docs/api/java/io/BufferedWriter.html#newLine\(\)](http://docs.oracle.com/javase/7/docs/api/java/io/BufferedWriter.html#newLine())

Sun, 3/3/2013, 3:51:48 PM  
[document graphs and statistics](#)

☐ play at actual speed?

Common issues from CSCI 4311 PA1:

Using line-buffered I/O.  
BufferedReader.readLine() "A line is considered to be terminated by any one of a line feed ('\n'), a carriage return ('\r'), or a carriage return followed immediately by a linefeed."  
For text data, this is generally not a problem, as long as platform-independent linebreaks are added to output, e.g. with BufferedWriter.newLine().

# Behind the scenes

≡ clearly, gDocs stores everything you did!

≡ why?

- > why not?

- » bandwidth & storage on the house!

- > user analytics

- » "if you are not paying ... you *are* the product"

- > user convenience?

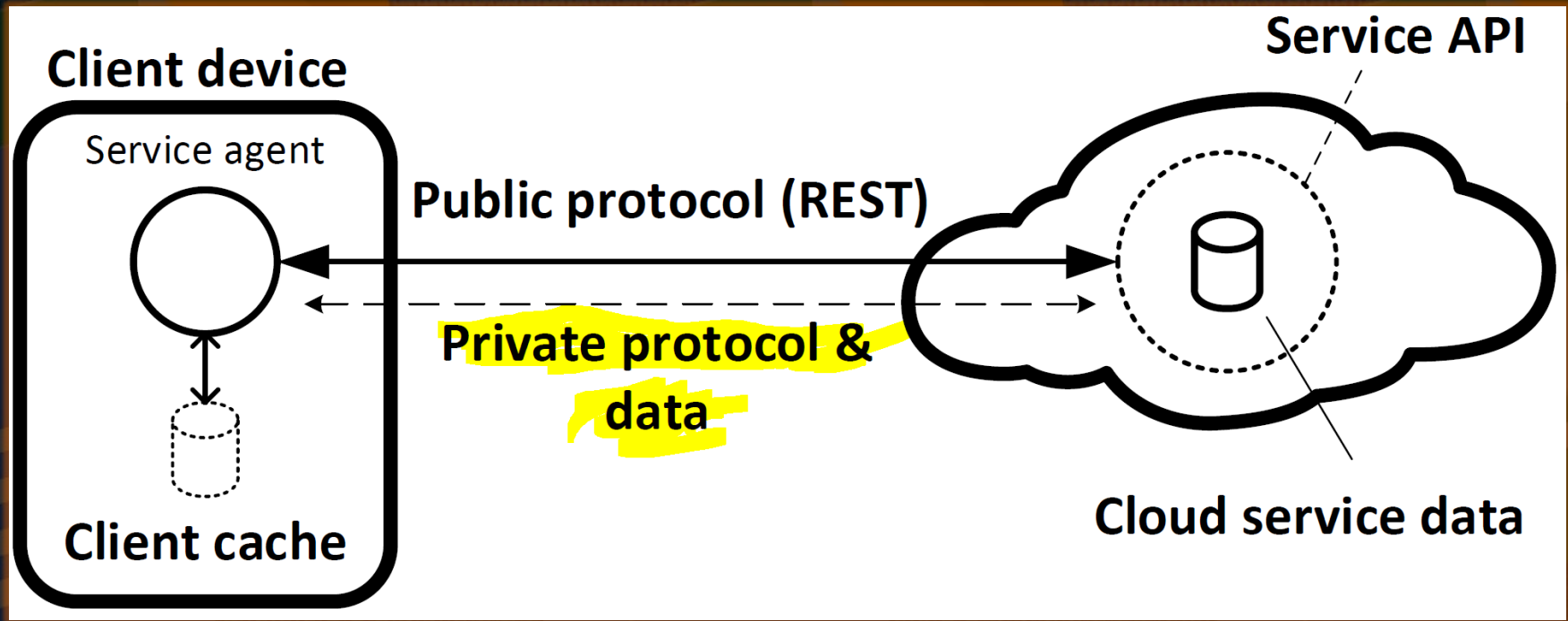
- » can you handle 10k revisions?

- > programmer convenience?

- » works with the real-time collaboration concept



# web app architecture



# Meet the changelog

☐ load?id=1IdObEjEPRwAfmYoaSc6vVZVYgrM3oag\_mU-Y-3Mj4FQ&start=4943&end=4960&token=AC4w5...

```
▼ changelog: [[{ty: "mlti", mts: [{ty: "ds", si: 14776, ei: 14776}, {ty: "ds", si: 14775, ei: 14775}]],...],...]  
  ▶ 0: [{ty: "mlti", mts: [{ty: "ds", si: 14776, ei: 14776}, {ty: "ds", si: 14775, ei: 14775}]],...]  
  ▶ 1: [{ty: "mlti", mts: [{ty: "ds", si: 14774, ei: 14774}, {ty: "is", s: ".", ibi: 14774}]], 1443757842902,...]  
  ▶ 2: [{ty: "is", s: " ", ibi: 14775}, 1443757843127, "18178839968700900856", 4945, "df36183a2f26250", 660,...]  
  ▶ 3: [{ty: "is", s: "Afte", ibi: 14777}, 1443757843660, "18178839968700900856", 4946, "df36183a2f26250",...]  
  ▶ 4: [{ty: "is", s: "r a", ibi: 14781}, 1443757843854, "18178839968700900856", 4947, "df36183a2f26250", 662,...]  
  ▶ 5: [{ty: "is", s: "n h", ibi: 14784}, 1443757844440, "18178839968700900856", 4948, "df36183a2f26250", 663,...]  
  ▶ 6: [{ty: "is", s: "our ", ibi: 14787}, 1443757844751, "18178839968700900856", 4949, "df36183a2f26250",...]  
  ▶ 7: [{ty: "is", s: "and ", ibi: 14791}, 1443757845081, "18178839968700900856", 4950, "df36183a2f26250", 1
```



# The chunkedSnapshot

```
"chunkedSnapshot":[
  [{"ty":"is","s":"Test document","ibi":1},
  {"ty":"as","sm":{"hs_h1":{"sdef_ts":{"ts_fs":18.0,"ts_fs_i":false}},
    ...
    "hs_h6":{"sdef_ts":{"ts_bd":false,"ts_bd_i":true,"ts_fg": "#000000"},
  {"ty":"as","sm":{"lgs_l":"en"},"ei":0,"st":"language","fm":false,"si":0},
  {"ty":"as","sm":{"ts_bd":false,"ts_bd_i":true,"ts_bgc":null,"ts_bgc_i":true,"ts_fg": "#000000",
    "ts_fg_i":true,"ts_fs":11.0,"ts_fs_i":true,"ts_sc":false,"ts_sc_i":true,"ts_st":false,"ts_st_i":true,"ts_va": "nor",
    "ts_va_i":true},"ei":12,"st":"text","fm":false,
```

# Drawing/slides

```
{ "changelog": [
  [[3,"g27de7cf84_0_0",108,[2.292,0.0,0.0,0.2674,63984.0,37722.0],[44,0,45,1],"p"],
    1444063509783,"08413168629437028300",2,"f13f7456cc71754",0,n
  [[15,"g27de7cf84_0_0",null,0,"T"],1444063511799,"08413168629437028300",3,"f13f7456cc71754",1,n
  [[15,"g27de7cf84_0_0",null,1,"e"],1444063512119,"08413168629437028300",4,"f13f7456cc71754",2,n
  [[15,"g27de7cf84_0_0",null,2,"s"],1444063512448,"08413168629437028300",5,"f13f7456cc71754",3,n
  [[15,"g27de7cf84_0_0",null,3,"t"],1444063512448,"08413168629437028300",6,"f13f7456cc71754",3,n
  [[3,"g27de7cf84_0_1",99,[0.1432,0.0,0.0,0.3263,174285.0,78309.0],[22,381,15,"#CFE2F3",19,"#0000
    1444063520352,"08413168629437028300",12,"f13f7456cc71754",7,n
  ,"chunkedSnapshot": [
    [[1,[365760,274320],[302400,427680]], [45,[],[0,"en"]], [13,0,0,null,"p"], [13,0,1,"m","l"], [13,0,
    [12,"m",0,2,[],[12,"l",0,1,[],[12,"p",0,0,[]]]
  ]}]
```



# Embedded images

≡ Upon upload

- > a temporary Google CDN link is provided ([googleusercontent.com](https://googleusercontent.com))

- » lasts ~1 hour

- > a permanent CDN link is also provided

≡ CDN link is like a dead drop

- > if you know the address, you can access it

- » no authentication, no "nonsense"

# CDN meets changelog

≡ what should happen if we delete an embedded image?

- a) CDN image is deleted and becomes unrecoverable, or
- b) just keep it around, in case the change is rolled back and need to reinsert it

≡ well, b) of course!

≡ That is, as long as **any** document revision references the image, the public CDN link will remain live!

≡ if the whole document is deleted, embeds are garbage collected

> ... after ~1 hour ...



# Reversions

Q: what happens when we REVERT to a prior version?

A:

- > a snapshot of the desired revision is created,
- > a REVERT entry w/ the snapshot is *appended* to the log.

# Privacy

≡ How to edit-share a doc without sharing the history?

> you cannot → the history is the document

≡ Workaround

> create a copy (which zaps the history) and share that

≡ Privacy audit

> extract all "deleted" embeds → make sure you still need them

≡ CDN link as a dead man's switch

> remember the librarian ...



# Forensics

≡ changelogs cannot be spoiled

- > Google will only **add** things to the log; no way to permanently modify prior state

≡ The golden CDN hour

- > could potentially recover "SWAT-triggered" deletions

≡ collecting changelogs is easy ...

≡ storing & replaying them, much less so

- > what format should they be in?
- > how do you render them (years from now)?
  - » changelog is an internal data structure, it can change at any time

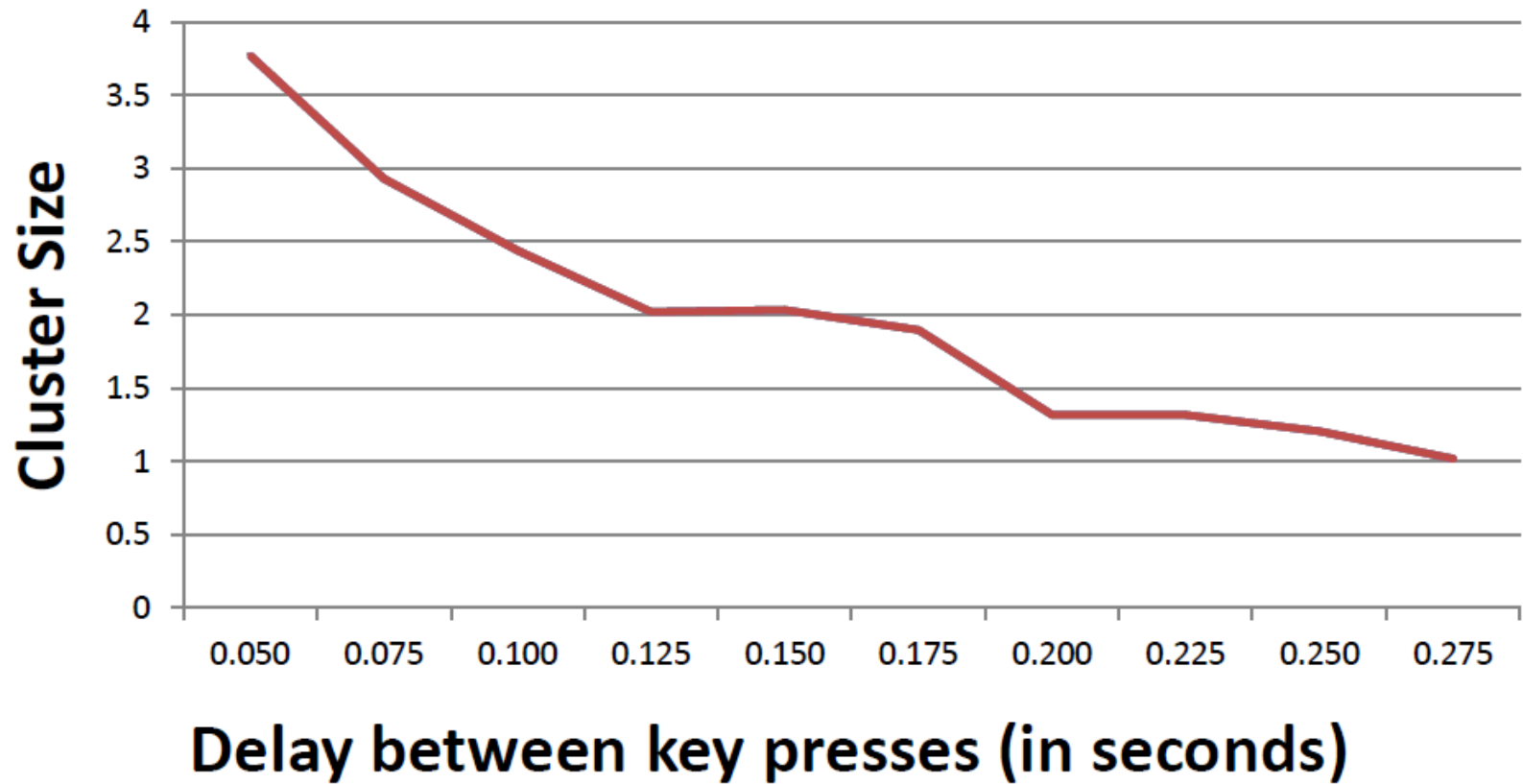
# keystroke biometrics

```
▼ changelog: [[{ty: "mlti", mts: [{ty: "ds", si: 14776, ei: 14776}],
  ►0: [{ty: "mlti", mts: [{ty: "ds", si: 14776, ei: 14776}, {ty: "ds", si: 14774, ei: 14774}],
  ►1: [{ty: "mlti", mts: [{ty: "ds", si: 14774, ei: 14774}, {ty: "ds", si: 14775, ei: 14775}],
  ►2: [{ty: "is", s: " ", ibi: 14775}, 1443757843127, "18178839968",
  ►3: [{ty: "is", s: "Afte", ibi: 14777}, 1443757843660, "18178839968",
  ►4: [{ty: "is", s: "r a", ibi: 14781}, 1443757843854, "18178839968",
  ►5: [{ty: "is", s: "n h", ibi: 14784}, 1443757844440, "18178839968",
  ►6: [{ty: "is", s: "our ", ibi: 14787}, 1443757844751, "18178839968",
  ►7: [{ty: "is", s: "and ", ibi: 14791}, 1443757845081, "18178839968"]]]
```

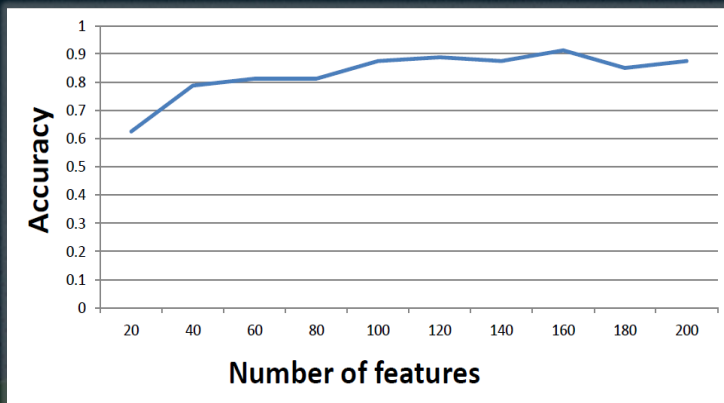
≡ Q: Do online collaboration tools collect (implicitly) a keystroke biometric signature?



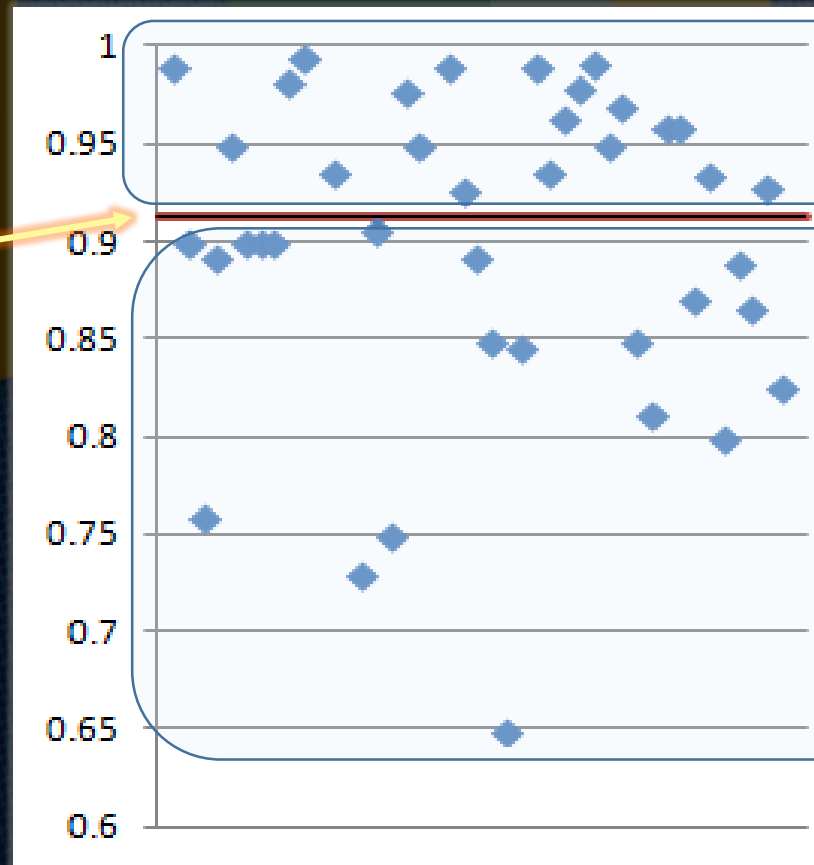
# GDocs keystroke clusters



A: GDocs  $\rightarrow$  Probably



Identification  
rate (us)



20 classifiers (them)

20 classifiers (them)



# How similar are other services?

≡ Min update interval per service

- > Dropbox Paper → 100ms
- > Zoho → 100ms
- > Word Online → 50ms (!)
- > Zoho sends a client-side timestamp

➔ Results for gDocs should hold

# ex: zoho writer

```
"ver":3,
"ops":{
  ...
  "17":{ "sid":"1e13c77a-fd82-4ae9-869d-9235bab4b1b8",
    "zuid":"5267691",
    "time":"1468965917546",
    "op":["{"r":84}, {"is":"t"}]",
    "mc":["{"ei":85,"si":85}]",
    "18":{"sid":"..","zuid":"..","time":"..",
      "op":["{"r":85}, {"is":"e"}]",
      "mc":["{"ei":86,"si":86}]",
      "19":{"sid":"..","zuid":"..","time":"..",
        "op":["{"r":86}, {"is":"s"}]",
        "mc":["{"ei":87,"si":87}]",
        "20":{"sid":"..","zuid":"..","time":"..",
          "op":["{"r":87}, {"is":"t"}]",
          "mc":["{"ei":88,"si":88}]",
          ... }
  }
```



# ex: Dropbox Paper

```
{"text":{"0":"%0A*ab%0A*cd%0A*ef%0A*gi%0A*jk%0A*lk%0A%0A"},
"attrs":{"0":"*0
|1+1*0*1*2*3*4*5+1*0
|1+3*0*1*2*6*4*7+1*0
|1+3*0*1*2*8*4*9+1*0
|1+3*0*1*2*a*4*b+1*0
|1+3*0*1*2*c*4*d+1*0
|1+3*0*1*2*e*4*f+1*0
|1+4
|1+1"}}}
```

Give me a name

1. ab
2. cd
3. ef
4. gi
5. jk
6. lkm



```
{"numToAttrib":{"0":["author","d.16Vtz98QRU..40M"],
"1":["insertorder","first"],
"2":["list","number1"],
"3":["start","1"],
"4":["taskcreated","Fri Jul 08 .. "],
"5":["usuallyUniqueId","698635090"],
"6":["start","2"],
"7":["usuallyUniqueId","82066825"],
"8":["start","3"],
"9":["usuallyUniqueId","326744561"],
"10":["start","4"],
"11":["usuallyUniqueId","1018504667"],
"12":["start","5"],
"13":["usuallyUniqueId","369272341"],
"14":["start","6"],
"15":["usuallyUniqueId","494339320"]}}
```

# kumofs

≡ The problem:

POSIX vs. cloud drive API

≡ Similar, yet different

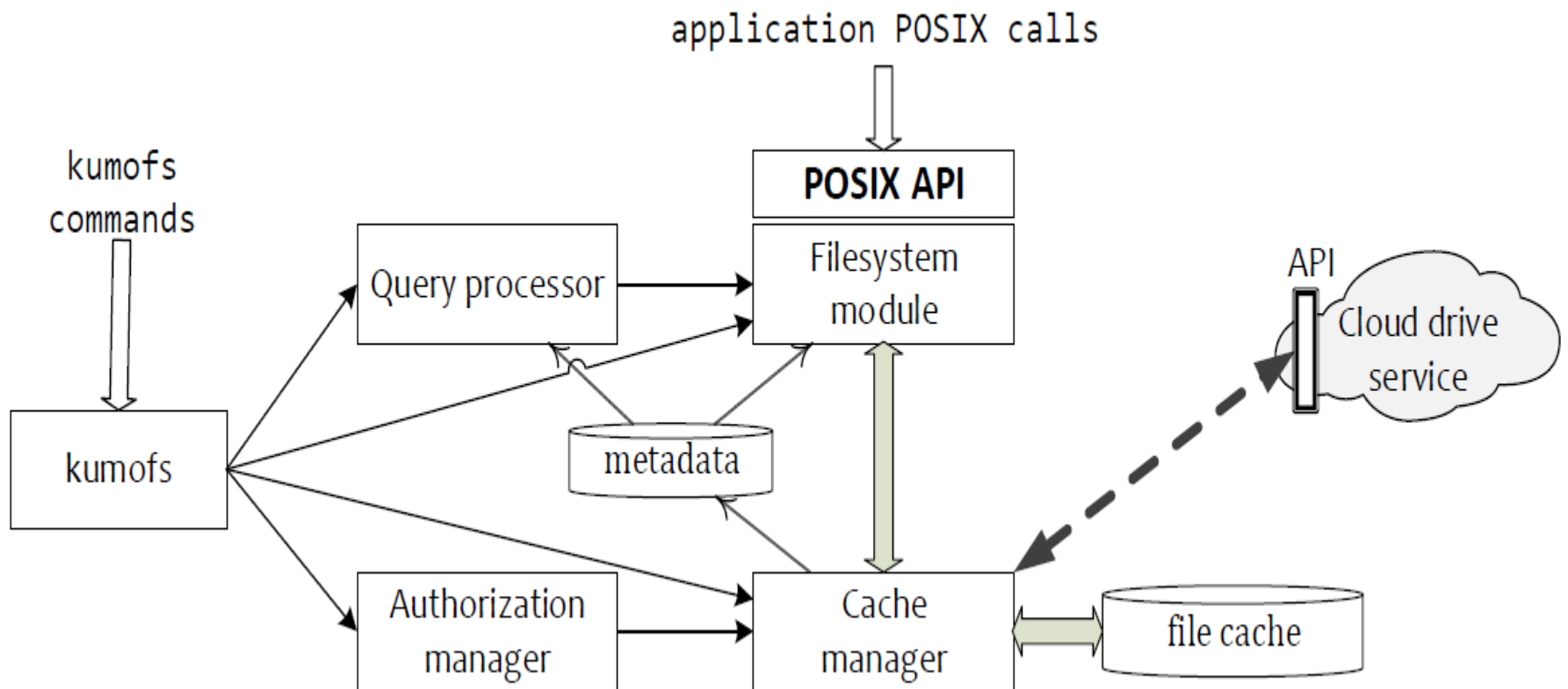
≡ We can build API-based tools

> ... but we want to use old ones too





# kumofs



# Features

≡ kumofs mount|umount

≡ .DELETED/.VERS

≡ kumofs get | dl

```
kumofs tt "Aug -31 -2011 5:00 p" ./state/Aug -31
```

```
kumofs diff "Aug -31 -2011 5:00 p"
```

```
"Sep -30 -2011 5:00p" ./diff/Sep
```

```
kumofs mq 'labels .starred == "True"' show '
```

```
id ,title ,labels ' ./starred
```

≡ gDocs

```
summary .gdoc .docx
```

```
summary .gdoc .odt
```

```
summary .gdoc .txt
```

```
summary .gdoc .pdf
```



# Thank you!

≡ [vassil@roussev.net](mailto:vassil@roussev.net)

≡ github:

[kumofx/kumodd](#)

[kumofx/kumodocs](#)

[kumofx/kumofs](#)

≡ Questions?