

The mystery of the missing email that wasn't

Tj <hacker@iam.tj>

What happened when thousands of hosted email customers reported their emails going missing!

In late 1999 and early 2000 I was the consultant technical director of a rapidly growing internet service provider in the United Kingdom that was struggling to cope with its rate of growth. The ISP had some unique selling points at the time which were driving customer numbers up rapidly.

It went from 0 to around 200,000 dial-in customers in 18 months. This was before xDSL when customers still had analogue voice-circuit modems and dialled into a local-call-rate point of presence (PoP). A big selling point was “free” email addresses and web space!

The hosted email service was the second-most important service after dial-in. Storage requirements grew exponentially.

Almost the entire operation, but particularly the email service, operated on a cluster of Dell PowerEdge servers with fibre-channel PowerVault storage attached. The operating system was Microsoft Windows NT 4.0. The servers were housed in a small rural data centre about 60 miles away.

The email server application was Infinite InterChangeⁱ from Infinite Technologies, of Owings Mills, Maryland, USA. It was an ISP-grade email server solution that could handle millions of accounts, included a web-mail interface and - very popular at the time - a Wireless Application Protocolⁱⁱ (WAP) interface for access from cellular telephones.

Sometime in early 2000 the 24 hour support began getting reports of email missing from customer's mailboxes. However, when the technicians investigated they couldn't find any evidence of any emails having been deleted. Initially therefore, customers got the blame and were told to try deleting and re-adding the email server settings to their email clients or WAP 'phones, and to be more careful.

Monday

On Monday morning when I arrived at the office there was some light-hearted banter about some customers losing or deleting their emails. During the day more reports came in and it began to be taken seriously. The senior engineers got involved but despite exhaustive examination of the server logs and trying to reproduce the issue they drew a blank.

Tuesday

By Tuesday there were hundreds of complaints and it was clear there was a major incident developing. At this point I joined the incident team.

We began having conference calls with the programmers at Infinite Technologies about possible bugs in InterChange – developing hypotheses we could test against the very little evidence we had. We continued to fail to reproduce the issue ourselves.

I asked all the obvious questions: Any recent software upgrades? No. Recent reboots? No. Service packs up to date? Yes. Event logs clean? Yes. Any patterns in the customer accounts being affected? None we could identify (it affected old and new accounts alike, didn't seem dependent on the numbers of emails in the account, or size, or age).

Some customers volunteered access to their affected accounts in the hopes we might spot something but we were still drawing a blank. Meantime thousands of emails were disappearing.

Spooky Tuesday

Then, things got spooky! Customers started reporting that their missing emails were back! At this point we were doing double-takes, wandering around with raised eyebrows and muttering "wtf" to ourselves.

We were working late into the night trying to spot clues (daytime and evenings were the busy email times) and beginning to suffer sleep deprivation so it began to feel like one massive hallucination.

Wednesday

Then, on Wednesday afternoon I was stood behind JB watching over his shoulder whilst he had an RDP session open to the email server. He was using Windows Explorer to view the directories where emails were stored by InterChange (Interchange stored each email in a separate file rather like Maildir format).

As we watched the one hundred or so files in the Explorer window started to disappear, one by one, until the directory was empty! We must have been a sight - jaws hanging open in bemused bewilderment, looking at each other and exclaiming "WTF!?" repeatedly.

Other technicians were gathering around wanting to know what was up and JB had turned away to explain when I pointed at his monitor and cried "WTF?" once again – the files were re-appearing in front of our eyes!

I think at this point some very strong coffee was called for whilst we tried to rationalise what we'd just seen. Jokes about the "ghost in the machine" abounded.

Our first worry was that the server had been compromised and some attacker was currently logged in and moving files around. We were able to discount that possibility quickly but were at a loss as to what was happening.

We got into another conference call with Infinite Technologies and they suggested the Dell PowerVault storage array might be at fault. We immediately got Dell support in on the call. The issue was rapidly escalated up to their senior engineers. Many ideas were thrown into the ring but each one was dismissed as we tested them and drew a blank.

Our favourite hypothesis was that somehow the storage controller is flipping bits on reads but not on writes – the only way we can conceive of files reappearing apparently none the worse for going AWOL.

By the end of the day we estimated almost 10% of 120,000 customers had reported the problem.

Thursday

Overnight I made the decision to replace the Dell PowerVault storage array since we had no clear idea of the cause and we couldn't afford to let the incident continue through the weekend and into the next week. At the office early on Thursday I had discussions with the CEO and we decided to buy another brand of storage array in case this was a Dell-specific issue.

I then began putting together a plan of attack for how we were going to handle a seamless migration of the email from one storage array to the other with minimal customer down-time.

The first major challenge was how to tell InterChange the email had moved. It wasn't as simple as swapping out the arrays since the new array had to be attached at the same time as the existing so we would have to do something like "copy D:\InterChange*E:\InterChange\" but then InterChange would have to rebuild it's indexes by deleting the existing index and scanning every email in the new location.

Apparently the absolute file path to each email was hashed in the indexes. The indexes contained email header fields.

Infinite Technologies estimated InterChange would need around 4 hours to re-index.

We knew we had to minimise the time the email service would be closed to customers and to warn them about the impending disruption as soon as possible. We decided to do the migration Friday night into Saturday morning when the mail system was at its quietest.

Infinite Technologies programmers agreed to be on hand to advise us whilst the migration was performed.

A banner was added to the webmail log-in page, the front-page of the ISP web-site, and (ironically) emails were sent to every customer warning them of the "planned maintenance" window from 22:00 Friday to 10:00 Saturday.

Our next problem turned out to be securing a replacement storage array. The person tasked with that was failing to find any supplier that could ship it to us next day. Most had one week lead times. Our base was in the north and most suppliers and distributors were based in the south of England.

By the close of business on Thursday we still had not secured a replacement storage array.

Friday 04:00

Early start to continue planning. At 08:30 we began telephoning every quasi-local hardware distributor and supplier in search of the elusive storage array. Eventually, mid-morning, we found a local distributor only five miles away that had a single IBM storage array sitting on a pallet in their warehouse!

At that point the pressure was transferred to the finance director who had to arrange an immediate £10,000 bank transfer before we could take possession.

Whilst the FD was sorting the payment out we finalised our migration plan. I arranged overnight access to the (usually unmanned) data centre. I would collect the new storage array late afternoon and then drive myself and AK, the senior sysadmin (a very talented Indian ex-pat), to the data center. The technical manager RS, and JB, the other senior engineer, would join us separately later. Because they would all be working over the weekend I arranged that they would take Monday and Tuesday off to recover.

Infinite Technologies would be on the other end of a telephone line if we needed them.

Friday 17:00

We were in high spirits during the drive to the data centre; looking forward to a smooth migration and getting some well-earned rest. We'd bought lots of food and drink to keep us provisioned and awake since the data-centre was in a very rural village with few amenities.

After arriving we moved the storage array into the DC and commandeered the attached office/canteen as our operations centre and began our preparations.

Friday 19:00

The first task was to shutdown the servers and attach the new fibre channel storage array, restart, and ensure it was working correctly. We needed to keep this outage as short as possible because the "maintenance window" wasn't due to start until 22:00.

This being the first time I'd visited the DC my understanding of the cluster configuration was based on what I'd been told by the engineers that had originally installed it.

This was when I got my first surprise. I was on the KVM console preparing to shutdown each server in the cluster but the KVM wouldn't connect to any but 'number 1'. Myself and AK spent a fruitless 15 minutes in the hot isle tracing KVM cables to the servers expecting them to be disconnected – but no, they were all correctly connected, and to the correct ports!

Eventually in the cold isle after opening the cabinet door one of us spotted that the activity LEDs on the other servers in the cluster weren't lit!

There were some worried and knowing glances exchanged at that point as we realised how fragile the configuration was. We quickly consulted with the technical manager and were shocked to discover that the cluster had never actually been configured – it had fallen off the todo list and been forgotten!

Friday 20:00

AK begins the task of starting the other servers, applying service packs, and correctly configuring them for clustering. The rest of us sit back and let him get on with it.

Friday 23:00

Already one hour into the “maintenance window” we’re finally ready to shutdown ‘number 1’ server. At this point only so AK can ensure the cluster comes up correctly. Each start takes about five minutes until the OS is ready, and AK has to do it a few times before he is happy. AK triggers a “chkdsk /f D:” on the PowerVault file-system and then sits down for ½ hour whilst it chugs through the data. No errors.

We decide it is too risky to bring up the cluster at the same time as migrating storage so we stick with only having “number 1” operating. Clustering is scheduled in for a couple of weeks time once we’ve recovered from this incident.

Saturday 00:00

The new storage array is connected and being formatted (NTFS). We’ve been practising the InterChange move on an empty spare system and are confident of the steps. AK gets the right to start the file copy operation from D: to E:

Saturday 01:00

File transfer is much slower than we expected. It turns out that due to the millions of small email files NTFS is heavily fragmented and the array cannot get near to saturating the fibre channel link because the disk heads are spending most of their time seeking rather than reading. At least the write should be sequential!

Saturday 05:00

Everyone is starting to sweat at how far behind schedule the operation is. File copy finally finishes and InterChange is told to start re-indexing on the new array. Fifteen minutes in it becomes clear it isn’t going nearly as fast as Infinite Technologies predicted. Fortunately they’ve been with us overnight and begin to investigate. They discover they had overlooked something: InterChange’s internal indexing is done on a low-priority thread to avoid TCP connection failures. As a result instead of indexing thousands of emails each minute it is not managing much more than 100.

Infinite suggest some configuration tweaks. For each we have to delete the existing indexes and start again.

Saturday 09:00

Indexing still going very slowly; Infinite Technologies programmers are working on a custom indexing tool for us. Meanwhile we realise the “maintenance window” is going to have to be extended, and update the log-in banner warnings and status pages with a new estimated end time of 18:00.

Anticipating elevated levels of calls to the free support line we contact several support call handlers at home and ask them to go to the office to cope with the call volume.

Saturday 12:00

The custom indexing tool is ready. We're told it can index at least ten thousand emails a minute and should take around 6 ½ hours. The downside is the currently running InterChange indexer has reached 15% and we have to lose that partial index and start afresh.

Saturday 17:00

Indexing not progressing as fast as expected. The custom indexer is slowing down as the index gets larger. We move the estimated end time again to 22:00.

Saturday 19:00

Order in pizzas for everyone!

Saturday 20:00

Custom indexing finished! AK gets the honour of restarting InterChange.

Crash! Crash! Crash! Panic!

Event logs reveal the process is failing to start. Infinite Technologies advise how to enable some developer debug options. These reveal the new index format is incorrect! More panic! Infinite eventually discover there is a bug in the custom indexer and set about re-writing it and running some bigger stress tests their side!

Support phone lines are jammed, more call handlers called in to handle night shift. Estimated end time moved again to Sunday 10:00.

Cold pizza rarely tasted this good.

Saturday 22:00

Custom indexer version 2 arrives. Set it running with a feeling of grim foreboding. Decide to interrupt it after 30 minutes and check if InterChange can read the indexes rather than letting it run to completion again and risk it failing. Version 2 cannot be interrupted. Cancel the run and wait for version 3 to arrive.

Custom indexer version 3 started.

Saturday 23:00

Stop the custom indexer and test InterChange. Starts but then hangs, won't talk to us, isn't listening on any sockets. Have to kill it. Infinite programmers pour over the debug logs and discover some records have missing fields.

Back to the drawing board.

Sunday 00:30

Custom indexer version 4 arrives and is started running.

Sunday 01:00

Whilst the others find somewhere to nap I decide to start hacking the index format and writing my own C-language tool to do the indexing in case Infinite Technologies fails us.

Stop the custom indexer and check InterChange. InterChange starts! Log-in to the first email account (we knew which email accounts the indexer had processed) but there's no email! Check the logs: no clues. Teleconference with the now very-tired Infinite Technologies programmers (not the only ones!). Send them a large sample of emails from our system to test the tool against.

Continue hacking.

Sunday 02:00

Custom indexer version 5 arrives, several more bugs fixed. Set it running.

Sunday 03:00

Stop the indexer and check InterChange – again. InterChange starts! Log-in to email and, glory be, there are all the expected emails! Do a happy dance in the cold isle and wake up the others.

Sad moment as I delete the partial index and restart the custom indexer version 5 for a complete run.

Time to take a nap. Everyone crashes.

Sunday 08:00

Wake up bleary-eyed, sweaty and hungry. Wash in the kitchen sink. JB drives out to the local garage shop and brings back piles of pre-packed sandwiches, Mars bars and cans of drink.

Indexer still going strong, reports 60% done.

Sunday 10:00

Indexing complete! InterChange starts and emails are visible. Start to allow incoming TCP connections and monitor performance.

Sunday 12:00

Declare incident over. Pack up our mess, make sure everything is secure, and start on the drive back to HQ.

Sunday 16:00

Arrive home and go straight to bed!

Monday

On Monday morning I arrived at the office mid-morning. There was a lot of interest in the tale of our horrendous weekend. The three senior engineers have today and Tuesday off – lucky bees! Early afternoon some of the call handlers start some light-hearted banter about customers losing or deleting their emails. I'm not in the mood for joking.

Then more reports come in and it quickly becomes clear the issue is still happening!

I place a call at home to AK, sysadmin, and we consult. We're at a loss. He offers to come in but I tell him to continue resting until there's something concrete he can do.

Get on the telephone to Infinite Technologies. Much dismay and scratching of heads. This proves it wasn't a storage array fault so we're back to pointing the finger at InterChange. CEO calls their CEO and verbally lambasts him.

Get Dell senior support and Infinite Technologies on a conference call where we try to figure out just what is happening.

AK, RS and JB turn up despite them supposed to be having time off to recover.

Lacking anything better I start searching the Microsoft Knowledge Base for any articles that may give a clue. At one point I find "Q229607: File Corruption on an NTFS Volume with More Than 4 Million Files" but that was rolled up in service pack 6a.

Decide to go back to basics and check everything myself, with Infinite Technologies engineers following along and double-checking every step.

I open an RDP session to the email server and go through the basics. Anything unusual in the event logs? No. Any processes consuming too much CPU? No. Is it the most recent version of InterChange? Yes. Service Pack? 6

... 6 ? not 6a ? You have got to be kidding me!? I asked this question a week ago and was told the server is on the latest service pack!

Lots of uncomfortable shuffling and clearing of throats behind me.

Monday 15:30

Quickly move SP6a.exe onto the server and prepare to apply it. Better put up an emergency maintenance warning on the web-mail log-in page since this is going to require a reboot.

Start installing service pack 6a. Takes a while. Whilst it is running read the KB article:

Subject: File Corruption on an NTFS Volume with More Than 4 Million Files
Article: Q229607

Product(s): Microsoft Windows NT
Version(s): winnt:4.0
Operating System(s):
Keyword(s): kbWinNT4sp6fix

The information in this article applies to:

- Microsoft Windows NT Server version 4.0, Terminal Server Edition
 - Microsoft Windows NT Server version 4.0
 - Microsoft Windows NT Workstation version 4.0
 - Microsoft Windows NT Server, Enterprise Edition version 4.0
-

SYMPTOMS

=====

When you create and delete files on an NTFS volume that holds more than 4 million files, you see file corruption that may show up in one of the following ways:

- Deleted files continue to be displayed on the drive.
- Files that you have not deleted are no longer accessible.
- A pop-up message is displayed reporting that corruption has been detected on the drive and requesting that you run CHKDSK.

CAUSE

=====

This problem occurs when the Master File Table (MFT) has grown larger than 4 GB, which may happen when you have more than 4 million files on your computer. When you delete a file whose MFT entry is beyond the 4 GB point under these conditions, an error in calculations causes the wrong entry to be marked as available. If this entry contains information for another file, and new files are added to the volume shortly after the deletion occurs, the entry could be re-used causing the file it actually referenced to be lost.

RESOLUTION

=====

Windows NT Server or Workstation 4.0

To resolve this problem, obtain the latest service pack for Windows NT 4.0 or the individual software update.

- i <https://web.archive.org/web/19990429084259/http://www.ihub.com/>
- ii https://en.wikipedia.org/wiki/Wireless_Application_Protocol