(/CATEGORY/BUSINESS/)
OFFICE HARDWARE (/CATEGORY/OFFICE-HARDWARE) printing (/tag/printing/)

Xerox scanners found to sometimes alter numbers

Martyn Williams, IDG News Service
@martyn_williams   Aug 6, 2013 10:25 AM

Xerox scanners have been found to randomly alter numbers on documents when reproducing them if a certain combination of image quality and compression setting is used.

The problem first came to light last week when David Kriesel, a computer scientist pursuing a PhD at the University of Bonn, posted results of several scans (http://www.dkriesel.com/en/blog/2013/0802_xerox-workcentres_are_switching_written_numbers_when_scanning) on his website.

On Tuesday, Xerox acknowledged the problem and advised customers to use a higher quality scanner setting if it wanted to avoid the problem.

Kriesel said he first noticed the problem when he used a Xerox WorkCentre to scan to PDF some building construction documents. The documents were of a building floorplan and each room was marked with a small box that contained a room name and the area in

square meters: 14.13m2, 21.11m2 and 17.42m2.

On first glance, the PDF reproductions of the plans appeared to be identical to the originals—as anyone would expect them to be—but closer examination revealed that wasn't actually the case.

The areas of the three rooms in the reproduced version were incorrect.

Investigation

Kriesel set out to investigate the problem. When scanned in TIFF mode, a pixel-for-pixel reproduction, the copy was identical to the original. But when image compression was used, things started getting weird.

A Xerox WorkCentre 7535 (shown above) reproduced an image where every room was labeled 14.13m2 in area. The same thing happened on a Xerox WorkCentre 7556 on one scan. A second scan on the same machine had two rooms at 17.42m2 and one at 21.11m2 and a third scan produced two rooms labeled 14.13m2 and one at 17.42m2.

Kriesel had switched off optical character recognition so it wasn't related to that, he wrote.

"There seems to be a correlation between font size, scan dpi used. I was able to reliably reproduce the error for 200 DPI PDF scans without OCR, of sheets with Arial 7pt and 8pt numbers," he said on his blog.

Once publicized, he says he began receiving emails from other Xerox users who were able to replicate the problem and also offer a few clues. He narrowed the problem down to the way the scanner's JBIG2 image compression works—subsequently confirmed by Xerox as being at the root of the problem.

In order to reduce file space, the compression system looks for areas of an image that are similar and, when it finds them, makes one compressed version and reuses it across all the similar areas. Because the numbers in the document were printed in a small, fine font, the scanner apparently mistook them for identical and reused data resulting in figures for room area getting reproduced.

Explanation

"The problem stems from a combination of compression level and resolution setting," Xerox said in a statement on Tuesday. "The devices mentioned are shipped from the factory with a compression level and resolution that produces scanned files which are optimized for viewing or printing while maintaining a reasonable file size. We do not normally see a character substitution issue with the factory default settings however, the defect may be seen at lower quality and resolution settings."

"For data integrity purposes, we recommend the use of the factory defaults with a quality level set to 'higher,'" the company advised.

Xerox said the machines had warned users for years that character substitution could occur at lower quality and higher compression settings.

A message on a web interface to the copier warns: "The normal quality option produces small file sizes by using advanced compression techniques. Image quality is generally acceptable, however, text quality degradation and character substitution errors may occur with some originals."

On Tuesday, Kriesel said he spoke to two Xerox executives about the problem who explained it only exists at the lowest quality setting—called "normal" on the scanner —because it's the only one that uses JBIG2 compression. The other two settings —"high" and "higher"—use a different compression system, he wrote (http://www.dkriesel.com/en/blog/2013/0806_conference_call_with_xerox).

Updated at 11:54 a.m. PT with additional information.

WE RECOMMEND

(http://www.pcworld.com/article/2024879/displair-gives-you-a-touchscreen-out-of-thin-air.html) (http://www.pcworld.com/article/209079/11_uses_for_an_old_PC.html)

(http://www.pcworld.com/article/2045142/asus-chairman-jonney-shih-nexus-7-almost-broke-us.html)

Displair gives you a touchscreen out of thin air (video)

11 Uses for an Old PC

Asus Chairman

what's this?