1225 Northmeadow Pkwy, Suite 114,
Roswell, GA 30076
770-449-7704
www.commetrex.com

A Commetrex White Paper

How to Make FoIP Work

By Michael B. Coffee, CEO
Commetrex, Inc.

How to Make FoIP Work
(In Carrier Networks)

So, just what does it take to make FoIP (fax over IP) work (in carrier networks)? The short answer is experience, effective technology, patience, and smart FoIP routing. From that experience must come a practical understanding of how FoIP should work so the way it does work can be brought into line. Effective technology is required to overcome the inherent problems with FoIP. And finally, patience is required to qualify fax-friendly routes and carrier partners, and then include and use those qualified FoIP routes in your call routing.

And what makes FoIP special? Why does it require more attention than VoIP, anyway? Well, voice is judged by the human ear; fax is judged by synchronous modems and a time-critical protocol: T.30. So, for a particular call, FoIP either works or it doesn't. Yes, you can get a partial page, but if the two endpoint fax terminals can't execute the T.30 fax protocol and transfer the image data, you have a fax-transaction failure. There is no Mean Opinion Score (MOS) for a fax. T.30 is a computer-to-computer protocol (yes) that was designed for TDM transport with rather precise timing requirements…not the case with two humans. So, when we put a fax session over an IP network, it better meet those end-to-end protocol timing and image-data-integrity requirements. There's another way to say this: voice is easy; fax is hard.

Not only must the T.30-protocol's timing requirements be met, the SIP call-set-up timing can cause problems. The carrier networks assume each call is a voice call, or at least they usually require that the session begin as such, and then re-Invite the session over to T.38 if a fax terminal answers. (See sidebar for a T.38 backgrounder.) It turns out that the timing of that transition is critical, and maintaining it over multiple tandem connections is, to say the least, problematic.

But if the call-set-up problems are overcome, there's still the problem of image-data-transport integrity. If the transition from G.711 pass-through mode to T.38 is successful, the chances of successful image-data transfer are good due to the data-correction features of T.38, which, by the way, are not necessarily enabled by your carrier (check it out). But if the switch to T.38 does not occur--for whatever reason--the image transfer must be successful in G.711 mode, which can also be problematic.

Today, missing or late packets in the IP networks of advanced-economy countries are becoming increasingly rare, yet G.711 pass-through faxes still fail at a rate that causes customer-satisfaction problems. The problem, it turns out, is something that plagues virtually every carrier since it's inherent in all modem calls. We call it the PCM clock-sync problem, which is caused by small differences between the PCM timing clock in the endpoint fax terminal and the carrier's gateway. If the fax is long enough, this timing difference causes the gateway's jitter buffer—which is fixed for a fax call--to either underflow or overflow, depending on the direction of the difference. Either way, the fax will fail when the jitter buffer either runs dry or overflows.

Then, there are the problems introduced by carriers using least-cost-routing-as-usual, rather than FoIP-qualified fax-aware routes. This problem occurs anywhere, but most often in international calls. In four years of testing with several international carriers we found that FoIP success rates for international calls to be just 50%. And the problem was primarily inter-carrier peering, including multiple TDM-IP transitions.

So, here are the problems:

- SIP negotiations
- T.38 transitions
- G.711 clock-sync
- FoIP-unaware routing

Here are the solutions:

SIP Negotiations and T.38 reINVITEs

SIP is the call-control protocol of the global IP-telephony network. When a call is initiated, the caller uses SIP to indicate media preferences for this call. For a fax call, G.711 should be listed since that is what the carriers want to be listed first in the initial call invitation (SIP Invite) and fax modems are designed to work with the quantization errors introduced by G.711 (as opposed to G.729, for example). If a fax terminal answers, the gateway near the called terminal (except for V.34 calls) sends a SIP message (a reINVITE) to the caller's gateway to start over using T.38, a protocol designed to be used by gateways to support FoIP. Basically, T.38's job is to make the IP network and its timing uncertainties transparent to the two endpoint terminals (see sidebar). But problems can occur along the journey from G.711 to T.38.

It depends on the networks involved, but nearly 5% of the time the T.38 reINVITE either never arrives or it arrives so late that that two fax terminals are far enough into the T.30 protocol so that if they are interrupted that fax fails. And guess what? Nearly every gateway and ATA on the market will do just that. If you want FoIP completion rates that equal the PSTN, you'll need technology in those gateways and ATAs that solves this so-call "late T.38 reINVITE" problem. (Surprise! NetGen's patent-pending Smart FoIP® does just that.)

PCM Clock Sync

Smart FoIP is available to solve the late T.38 reINVITE problem by declining the invitation to make the switch if it arrives too late. This causes the call to remain in G.711 pass-through mode, which is better than an outright failure, but also has an inherent problem. As mentioned, clock differences between the endpoint fax terminal and the carrier's gateways are always different. This means that if the fax is long enough, a gateway's jitter buffer will eventually over-run or under-run, no ifs, ands, or buts. But since it's rare for the clocks to be so far off that the fax fails on the first page, there is technology available that turns all n-page faxes into n one-page faxes, and all transparent to other network elements.

Intelligent FoIP Routing

But even if we eliminate reINVITE and clock-sync problems, there's more work to be done since not all call routes are created equal as far as FoIP is concerned. If you make sure your gateway has “effective technology,” it won’t cause a failure due to its jitter buffer causing a problem. But the more gateways there are in the call path, the better the chance that one of them will cause a G.711 pass-through failure. And a gateway or transcoding SBC usually means that the call is being passed between two carriers, and the more carriers that are involved the bigger the chance that there will be a failure of some kind, such as:

- Plain-old Interop problems,
- Multiple TDM call segments,
- Low-bit-rate coders used in a TDM segment, and
- FoIP-hostile network configurations.

These are just some of the problems. One approach is to take each failure, attempt to determine the cause of the failure, and work with the responsible carrier to resolve the problem. But this is often not cost-effective or even possible. A more effective approach is to take control of the routing, determine which calls are likely to be a fax call, and route the call over an IP-only route, at least until it arrives in the called terminal’s local area. This means working with carriers with proven experience in handling fax calls with T.38 support. And if you want to be successful with international calls, that IP carrier will need a large international footprint with over-the-top routing or has peering arrangements that have the same effect.

But if you know something about FoIP and have been following this discussion, you may be wondering how all the routing magic is performed; after all, an ingress IP provider’s routing entity usually assumes that a session is a voice call, not a fax call. And it’s well known that for T.38 version 0 the re-Invite to T.38 is the responsibility of the off-ramp or called gateway, long after the call was routed. The answer is that the carriers must deploy a way for the on-ramp/calling/emitting gateway or server to signal that the call should be routed over an FoIP-qualified route.

One way to do that is for the service provider to provision the subscriber’s fax number for that special routing. Often, a subscriber is more than happy to pay a little more for an FoIP trunk that performs since his alternative is an even-higher-priced POTS line. Another approach is for the service provider to apply a heuristic algorithm to the originating point of inbound calls. (If it looks like a duck and walks like a duck, it’s a fax call.)

Yet another way is to support RFC 6913. It turns out that there are IETF RFCs, such as RFC 3840 and 3841, that allow a caller to express routing preferences to the routing entity by using IANA-registered media-feature tags in the SIP Invite. But there was a problem: there was no media-feature tag for “fax” that could be used in the SIP-INVITE header that was registered with the IANA. The solution is the IETF’s RFC 6913, which defines the “SIP.fax” media-feature tag and registers it with the IANA, allowing the originating SIP entity to declare that it prefers a “fax-friendly” route for the call.

The service provider should use a combination of these techniques since RFC6913 is not widely deployed. But the ITSP can take matters into her own hands and only deploy ATAs that support RFC6913, then route those FoIP calls over proven routes.

Avoiding FoIP Altogether

This paper is about how a service provider or a channel partner can stop saying, "Keep your POTS for fax," which is a way of avoiding FoIP altogether. Yet another way to avoid FoIP is to terminate the fax on the subscriber's premises with a special ATA and send the resulting file to a cloud-based server that will then take the image file and fax it to the destination fax terminal on behalf of the originating terminal. Although this can be an effective way to avoid FoIP problems by avoiding FoIP, it's a proprietary approach that can result in significant additional costs.

Use the recommendations in this paper and work with a vendor that has the experience, knowledge, technology, patience to work with you to optimize your FoIP routing, and you'll be able to achieve fax-transaction success rates on a par with the PSTN.

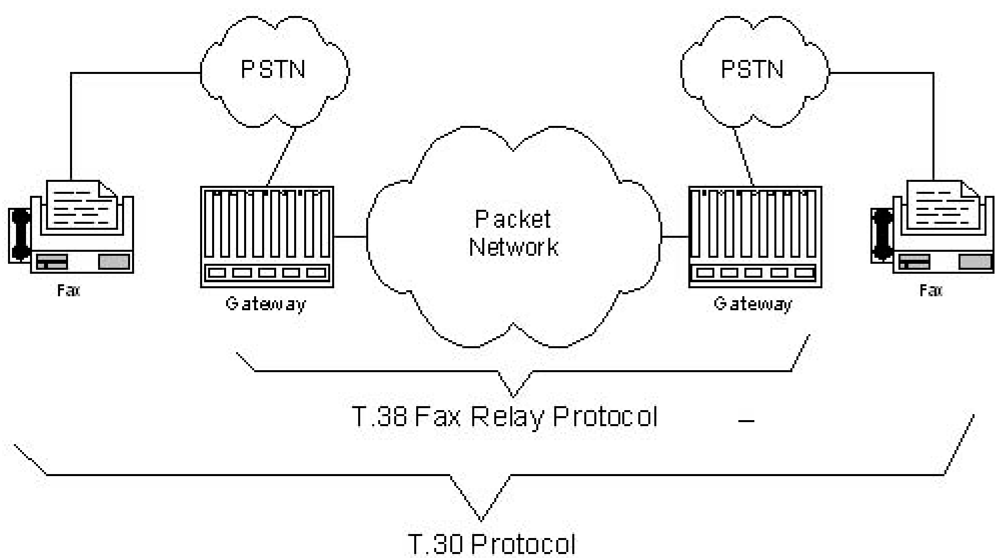
Sidebar

Just What is T.38?

The ITU published (or "determined," to use ITU speak) *T.38, Procedures for Real-Time Group 3 Facsimile Communication over IP networks*, in October 1998, and it's taken 16 years for deployments to begin to reach the international carrier industry, which is now taking up the special set of challenges of making FoIP work in international calls.

The purpose of T.38 is to make the packet network and its timing and data-transport uncertainties, transparent to the endpoint terminals executing the T.30 fax protocol. (See the diagram, below.)

Since the amount of infrastructure involved was relatively small and the networks self contained, initial deployments of FolP were in enterprise IP networks, typically as an outgrowth of VoIP deployments. Since there were no FoIP-capable carriers until the mid 2000s, enterprise FoIP stopped at the enterprise-network edge where IP-PSTN gateways converted the FoIP call to TDM for the PSTN. Interop problems were easy to solve due to the few vendors involved.



At the same time (the first half of the 2000 decade) the IP carriers that came out of the late-‘90s technology expansion were enlarging their footprint for Internet and VoIP services. SIP began to take over IP-telephony call control, setting the stage for SIP trunking to be offered by Internet-telephony service providers (ITSPs). But for FoIP to be a part of their networks, the major IP carriers, such as XO Communications and Global Crossing, had to deploy T.38-capable infrastructure and offer T.38 service agreements, which they began to do just after mid-decade. Then, in 2008, the SIP Forum came out with its SIPconnect specification, spurring ITSPs to offer SIP-trunking services, handing calls off to those major IP carriers. After much trial and error, this began to work out for some service providers and carriers as long as the FoIP calls were within national borders. To be more specific, these FoIP calls worked when the SIP-trunking provider handed off to the IP carrier, which then peered with the incumbent TDM network hosting the destination fax terminal. If it got more complicated than that, additional problems arose. This paper addresses those problems.