Boxed Out: Blocking Cellular Interconnect Bypass Fraud at the Network Edge

Bradley Reaves, *University of Florida;* **Ethan Shernan,** *Georgia Institute of Technology;*
Adam Bates, *University of Florida;* **Henry Carter,** *Georgia Institute of Technology;*
Patrick Traynor, *University of Florida*

https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/reaves-boxed

This paper is included in the Proceedings of the 24th USENIX Security Symposium

August 12–14, 2015 • Washington, D.C.

ISBN 978-1-931971-232

Open access to the Proceedings of the 24th USENIX Security Symposium is sponsored by USENIX

Boxed Out: Blocking Cellular Interconnect Bypass Fraud at the Network Edge

Bradley Reaves
University of Florida
reaves@ufl.edu

Ethan Shernan
Georgia Institute of Technology
eshernan3@mail.gatech.edu

Adam Bates
University of Florida
adammbates@ufl.edu

Henry Carter
Georgia Institute of Technology
carterh@gatech.edu

Patrick Traynor
University of Florida
traynor@cise.ufl.edu

Abstract

The high price of incoming international calls is a common method of subsidizing telephony infrastructure in the developing world. Accordingly, international telephone system interconnects are regulated to ensure call quality and accurate billing. High call tariffs create a strong incentive to evade such interconnects and deliver costly international calls illicitly. Specifically, adversaries use VoIP-GSM gateways informally known as "simboxes" to receive incoming calls over wired data connections and deliver them into a cellular voice network through a local call that appears to originate from a customer's phone. This practice is not only extremely profitable for simboxers, but also dramatically degrades network experience for legitimate customers, violates telecommunications laws in many countries, and results in significant revenue loss. In this paper, we present a passive detection technique for combating simboxes at a cellular base station. Our system relies on the raw voice data received by the tower during a call to distinguish errors in GSM transmission from the distinct audio artifacts caused by delivering the call over a VoIP link. Our experiments demonstrate that this approach is highly effective, and can detect 87% of real simbox calls in only 30 seconds of audio with no false positives. Moreover, we demonstrate that evading our detection across multiple calls is only possible with a small probability. In so doing, we demonstrate that fraud that degrades network quality and costs telecommunications billions of dollars annually can easily be detected and counteracted in real time.

1 Introduction

Cellular networks provide digital communications for more than five billion people around the globe. As such, they represent one of the largest, most integral pieces of critical infrastructure in the modern world. Deploying these networks requires billions of dollars in capital by providers, and often necessitates government subsidies in poorer nations where such investments may not produce returns for many decades. As a means of maintaining these systems, international calls destined for such networks are often charged a significant tariff, which distributes the costs of critical but expensive cellular infrastructure to callers from around the world.

Many individuals seek to avoid such tariffs by any means necessary through a class of attacks known as *interconnect bypass fraud*. Specifically, by avoiding the regulated network interconnects and instead finding unintended entrances to the provider network, a caller can be connected while dramatically lowering his or her costs. Such fraud constitutes a "free rider" problem, a term from economics in which some participants enjoy the benefits of expensive infrastructure without paying to support it. The most common implementation of interconnect bypass fraud is known as simboxing. Enabled by VoIP GSM gateways (i.e., "simboxes"), simboxing connects incoming VoIP calls to local cellular voice network via a collection of SIM cards and cellular radios. Such calls appear to originate from a customer phone to the network provider and are delivered at the subsidized domestic rate, free of international call tariffs. Interconnection bypass fraud negatively impacts availability, reliability and quality for legitimate consumers by creating network hotspots through the injection of huge volumes of tunneled calls into underprovisioned cells, and costs operators over $2 Billion annually [28].

In this paper, we present Ammit[1], a system for detecting simboxing designed to be deployed in a cellular network. Our solution relies on the fact that audio transmitted over the Internet before being delivered to the GSM network will be degraded in measurable, distinctive ways. We develop novel techniques and build

[1] Ammit was an Egyptian funerary deity who was believed to separate pure and impure souls, preventing the latter from achieving immortality in the afterlife.

on mechanisms from the Pindr0p call fingerprinting system [25] to measure these degradations by applying a number of light-weight signal processing methods to the received call audio and examining the results for distinguishing characteristics. These techniques rapidly and automatically identify simboxed calls and the SIMs used to make such connections, thereby allowing us to quickly shut down these rogue accounts. In so doing, our approach makes these attacks far less likely to be successful and stable, thereby largely closing these illegal entrances to provider networks.

We make the following contributions:

- **Identify audio characteristics useful for detecting simboxes:** We identify features in simboxed call audio that make it easily differentiable from traditional GSM cellular calls and argue why such features are difficult for adversaries to avoid.
- **Develop rapid detection architecture for the network edge:** We design and implement Ammit, a detection tool that uses signal processing techniques for identifying illicitly tunneled VoIP audio in a GSM network, and demonstrate that our techniques can easily execute in real time. Such performance means that our solution can be practically deployed at the cellular network edge.
- **Demonstrate high detection rate for SIM cards used in simboxes:** Through experimental analysis on a real simbox, we show that Ammit can quickly profile and terminate *87% of simboxed calls with no false positives*. Such a high detection rate arguably makes interconnect bypass fraud uneconomical.

We note that our techniques differ significantly from related work, which requires either large-scale post hoc analysis [42] or serendipitous test calls to network probes [10, 13, 15, 16, 18]. Our approach is intended to be used in real time, allowing for rapid detection and elimination of simboxes. We specifically characterize these techniques in Section 8.

It should be noted that the authors are not attempting to combat the spread of inexpensive VoIP calls in this paper. Traditional VoIP calls, which connect users through IP or a licensed VoIP-PSTN (Public Switched Telephone Network) gateway, are not considered a problem in countries that combat simboxes. Instead, we seek to prevent the creation of unauthorized entry points into private cellular networks that degrade performance for legitimate users and cost providers and governments two billion dollars annually. This is analogous to the problem of rogue Wi-Fi access points; simboxing prevents network administrators from controlling access to the network and can degrade service for other users. Moreover, similar to other economic free rider problems, failure to combat such behavior can lead to both underprovisioning and the overuse of such networks, making quality and stability difficult to achieve [49]. *Failure to combat simbox fraud may ultimately lead to raising prices and lower reliability for subsidized domestic calls in developing nations, where the majority of citizens can rarely afford such cost increases.*

The remainder of this paper is organized as follows: Section 2 provides background information on cellular networks; Section 3 describes simbox operation and their consequences; Section 4 presents our detection methodology; Section 6 describes our experimental methodology; Section 7 discusses our results; Section 8 offers an overview of important related work; and Section 9 presents our final remarks.

2 Background

2.1 Cellular Networks

The Global System for Mobile Communications (GSM) is a suite of standards used to implement cellular communications. It is used by the majority of carriers in the US and throughout Europe, Africa, and Asia. GSM is a "second generation" (2G) cellular network and has evolved into UMTS (3G) and LTE (4G) standards. We focus on GSM because it is the most available for direct experimentation. Note that the methods we present in the paper can easily be ported to other cellular standards.

GSM manages user access to the network by issuing users a small smartcard called a Subscriber Identity Module (SIM card) that contains identity and cryptographic materials. A carrier SIM card can be placed in any device authorized to operate on a carrier's network. Because GSM networks cryptographically authenticate almost every network transaction, cellular network activity can always be attributed to a specific SIM card. In the past, the ability to clone a SIM card negated this guarantee; however, modern SIM cards now have hardware protections that prevent practical key recovery and card cloning.

In addition to describing network functionality, the GSM standards also specify a method for encoding audio known as the GSM Full Rate (GSM-FR) codec [23]. Although designed for mobile networks, it is also used as a general purpose audio codec and is frequently implemented in VoIP software. To avoid ambiguity, we use "GSM" or "air transmission" to mean GSM cellular networks and "GSM-FR" to indicate the audio codec.

2.2 VoIP

Voice over Internet Protocol (VoIP) is a technology that implements telephony over IP networks such as the Internet. Two clients can complete a VoIP call using exclusively the Internet, or calls may also be routed from a VoIP client to a PSTN line (or vice-versa) through a VoIP Gateway. Providers including Vonage, Skype, and Google Voice provide both IP-only and IP-PSTN calls. The majority of VoIP calls are set up using a text-based protocol called the Session Initiation Protocol (SIP). One of the jobs of SIP is to establish which audio codec will be used for the call. Once a call has been established, audio flows between callers using the Realtime Transport Protocol (RTP), which is typically carried over UDP.

VoIP call quality is affected by packet loss and jitter. Absent packets, whether they are the result of actual loss or jitter, cause gaps in audio. Such gaps are filled in with silence by default. Some VoIP clients attempt to improve over this standard behavior and implement Packet Loss Concealment (PLC) algorithms to fill in missing packets with repeated or generated audio. Specifically, PLC algorithms take advantage of the fact that speech waveforms are more or less stationary for short time periods, so clients can generate a plausible section of audio from previous packets. Many codecs have mandatory PLCs, although some are optional (as in the case of the G.711 audio codec) or are not implemented (as is frequently the case when GSM-FR is used outside of cellular networks). Some VoIP software (including Asterisk) implements their own PLC algorithms, but do not activate them unless configured by an administrator.

3 What is a Simbox?

A simbox is a device that connects VoIP calls to a GSM voice (*not data*) network. A simple mental model for a simbox is a VoIP client whose audio inputs and outputs are connected to a mobile phone. The term "simbox" derives from the fact that the device requires one or more SIM cards to wirelessly connect to a GSM network.

There is a strong legitimate market for these devices in private enterprise telephone networks. GSM-VoIP gateways are sold to enterprises to allow them to use a cellular calling plan to terminate[2] calls originating in an office VoIP network to mobile devices. This is typically a cost saving measure because the cost of maintaining a mobile calling plan is often lower that the cost of paying termination fees to deliver the VoIP call through a VoIP PSTN provider (as well as the cost to the receiving party). *Such a setup is done with the permission of a licensed telecommunications provider and is only done for domestic calls.* This is in direct opposition to simboxers, who purchase subsidized SIM cards to deliver traffic onto a local network without paying the legally mandated tariffs.

Because there is a high demand for GSM-VoIP gateways, they span a wide range of features and number of concurrent calls supported. Some gateways support limited functionality and only a single SIM card, while others hold hundreds of cards and support many audio codecs. Some simboxes used in simbox fraud rings are actually distributed, with one device holding hundreds of cards in a "SIM server" while one or more radio interfaces connect calls using the "virtual SIM cards" from the server. This allows for simple provisioning of SIM cards, as well as the ability to rotate the cards to prevent high-use or location-based fraud detection.

3.1 How Simbox Fraud Works

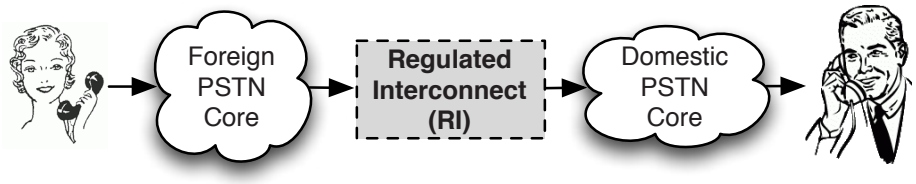
Simboxing is a lucrative attack. Because simboxers can terminate calls at local calling rates, they can significantly undercut the official rate for international calls while still making a handsome profit. In doing so, simboxers are effectively acting as an unlicensed and unregulated telecommunications carrier. Simboxers' principal costs include simbox equipment (which can represent an investment up to $200,000 US in some cases), SIM cards for local cellular networks, airtime, and an Internet connection. Successfully combating this type of fraud can be accomplished by making any of these costs prohibitively high.

Figure 1 demonstrates in greater detail how simboxing compares to typical legitimate international call termination. Figure 1 shows two international call paths: a typical path (Figure 1a) and one simbox path (Figure 1b).

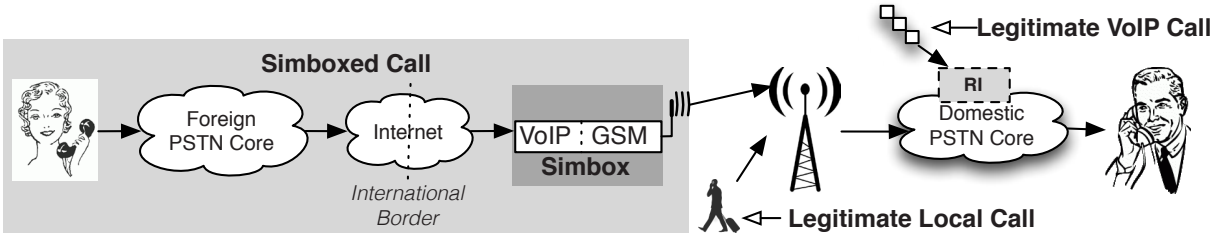
In the typical case, when Alice calls Bob, her call is routed through the telephone network in her country (labeled "Foreign PSTN Core") to an interconnect between her network and Bob's network. The call is passed through the interconnect, routed through Bob's domestic telephone network ("Domestic PSTN Core") to Bob's phone. If Alice and Bob are not in neighboring countries, there may be several interconnects and intermediate networks between Alice and Bob. The process essentially remains the same if Alice or Bob are using mobile phones. The interconnect in this scenario is crucial — interconnects are heavily regulated and monitored to ensure both call quality and billing accuracy (especially for tariffs).

In the simbox case, Alice's call is routed through her domestic telephone network, but rather than passing through a regulated interconnect, her call is routed over IP to a simbox in the destination country. The simbox then places a separate call on the cellular network in the

[2] In cellular and telephone networks, "terminating a call" has the counterintuitive meaning of "establishing a complete circuit from the caller to the callee."



(a) A typical international call is routed through a regulated interconnect. Note that VoIP calls from services such as Skype that terminate on a mobile phone also pass through this regulated interconnect and are not the target of this research.



(b) A simboxed international call (gray box) avoids the regulated interconnect by routing the call to a simbox that completes the call using the local cellular network.

Figure 1: Typical and Simboxed Calls

destination country, then routes the audio from the IP call into the cellular call, which is routed to Bob through the domestic telephone network.

In practice, simboxers execute this attack and profit in one of two ways. The most common method is for the simboxer to present themselves as a legitimate telecommunications company that offers call termination as a service to other telecom companies. As a call is routed through these intermediate networks, neither of the end users is aware that the call is being routed through a simbox. This agreement is analogous to a contract between two ISPs who have agreed to route traffic between their networks. While the end user has no knowledge of how his traffic is routed, the intermediate network owners profit from reduced prices for routed traffic.

The second method simboxers use to profit is to offer discounted call rates directly to end consumers, primarily through the sale of international calling cards. Such cards have a number that the user must dial before she can dial the recipient's number; this number will route to a number provided by a VoIP provider that points to the simbox in the recipient's country. When the user calls the number on her calling card, the simbox will answer, prompt her to dial the recipient's number, then the simbox will connect the call.

3.2 Consequences of Simbox Operation

The consequences of simboxing are significant to users who place simbox calls, users who share the cellular network with simboxers, and to cellular carriers and national governments.

As for the effects on users, Alice is likely unaware of the details of her call routing. However, Alice and Bob may both notice a degradation in quality, and Bob may notice that the Caller ID for Alice does not show her correct number. Bob may blame his local carrier for poor call quality, and so the carrier unfairly suffers in reputation.

Other users in the same cell as the simbox also suffer negative consequences. Cellular networks are provisioned to meet an expected demand of *mobile* users who only use the network a fraction of the time, and accordingly may only be able to support a few dozen simultaneous calls. When a simboxer sets up an unauthorized carrier and routes dozens of calls through a cell provisioned to support only a handful of simultaneous calls, the availability of that cell to service legitimate calls is significantly impaired. Connectivity within the cell may be further impaired by the dramatic increase in control traffic [50].

4 Methodology

Legitimate VoIP calls and other international calls enter a cellular network through a regulated interconnect or network border gateway. To halt simboxed calls, we only need to monitor incoming calls from devices containing a SIM card. Figure 1b shows the path of legitimate and simboxed audio, respectively, from the calling source to the final destination. In both cases, the tower believes it is servicing a voice call from a mobile phone. However, the audio received by the tower from a simboxed call will contain losses, indicating that the audio signal has traveled over an Internet connection, while the audio from a legitimate call will not contain these losses, having been recorded directly on the transmitting mobile phone. As discussed in Section 2, jitter and loss in Internet telephony manifest as unconcealed and concealed gaps of audio to the receiving client (the simbox, in this case). These features are *inherent* to VoIP transmission, and the only variant is the frequency of these events. All simboxed calls will have some amount of packet loss and jitter, so we design Ammit to detect these audio degradations. Because the audio transmitted *to* the mobile device could have originated from a variety of connection types, Ammit only analyzes audio received *from* mobile devices. If the mobile device is a simbox, the characteristics of this audio will exhibit the loss patters consistent with a VoIP connection, making the call distinguishable from audio recorded and sent by a mobile phone.

4.1 Inputs to Ammit

The most common codec supported by simboxes is G.711 [3] (see the Appendix for details). The G.711 codec is computationally simple, royalty-free, and serves as a least common denominator in VoIP systems. It was originally developed in 1972 for digital trunking of audio in the PSTN, and it is still the digital encoding used in PSTN core networks. The original standard indicated that G.711 should insert silence when packets are delayed or lost, so we examine G.711 using this setting.

Simboxers will have a clear incentive to configure their simboxes to evade detection, and an obvious evasion strategy is to ensure that audio is as close as possible to legitimate audio by using the GSM-FR codec for the VoIP link. Therefore, we show how Ammit accounts for this difficult case where GSM-FR is used with and without PLC. We discuss how Ammit addresses other evasion techniques in Section 5.

In summary, Ammit must detect the two audio phenomena characteristic of VoIP transmission: concealed and unconcealed packet losses. The following subsections detail how Ammit detects these phenomena, but first we briefly describe the data that Ammit receives from the tower before detecting audio features.

In GSM, audio encoded with the GSM-FR codec is transmitted between a mobile station (MS, i.e., a phone) and a base transceiver station (BTS, i.e., a cell tower) using a dedicated traffic channel. The encoding used by GSM-FR causes certain bits in a frame to be of greater importance than others. When an audio frame is transmitted, frame bits are separated by their importance. "Class 1" bits containing the most important parameters are protected by a parity check and error correcting codes, while "Class 2" bits are transmitted with no protections because bit error in these bits has only a small effect on the quality of the audio. The approach of only protecting some bits is a compromise between audio quality and the cost of the error correcting code. When Class 1 bit errors cannot be corrected, the receiver erases (i.e., drops) the entire frame. When Class 2 bits are modified, the audio is modified, but the receiver has no mechanism to detect or correct these modifications. This is termed "bit error." It should be noted that bit error and frame erasure are distinct concerns in GSM.

The receiving device (MS or BTS) may use PLC to conceal this frame erasure. When a BTS erases a frame, it conceals the loss before forwarding the audio into the core network. Visibility into frame erasures motivates our choice to place Ammit at the tower. However, there are additional benefits to locating Ammit at a tower. Specifically, this allows for scalable detection of simboxes because a single Ammit instance is only responsible for the dozens of calls that pass through the tower instead of the thousands of concurrent calls in a region or nation. Finally, if Ammit has a high confidence that a call is simboxed (as defined by a network administrator policy), ending a call at the tower is simpler than in other parts of the network. This policy would further frustrate the efforts of simboxers. It is also possible to deploy Ammit closer to the network core, perhaps at BSC or MSC nodes, but GSM loss information would need to be forwarded.

Ammit takes two inputs: a stream of GSM-encoded audio frames and a vector indicating which audio frames were erased (both of which can be collected by the BTS connecting the call). Ammit uses the frame erasure vector to ignore the effects of the air interface on the call audio. Ignoring erased frames ensures that losses on the air interfaces are not misinterpreted as losses caused by VoIP.

4.2 Detecting Unconcealed Losses

Ammit must detect two degradation types: unconcealed packet loss and concealed packet loss. To detect unconcealed loss, Ammit looks for portions of audio where the energy of the audio drops to a minimum value then

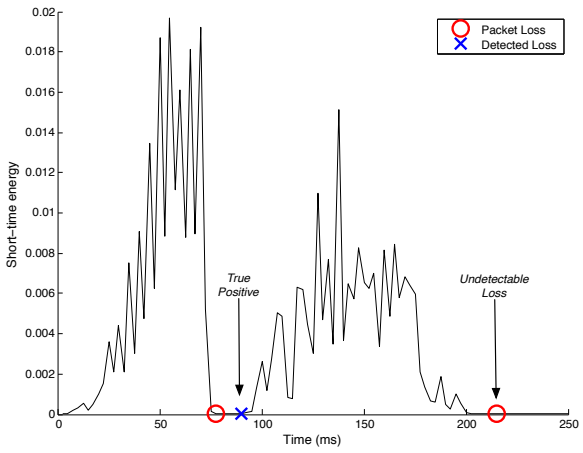


Figure 2: The short-term energy of speech during audio can reveal silence insertion. Packet loss that falls in naturally silent sections of audio is undetectable.

quickly rises again. This technique is also used in the Pindr0p system. The following discussion describes the Pindr0p approach to detecting unconcealed losses, with additional implementation insight and details.

Figure 2 demonstrates unconcealed packet loss in a clip of audio at 78 ms and 215 ms. At 78 ms, a packet is lost and silence begins. A short time later, at 90 ms, the energy rises again, indicating that a new packet has arrived containing speech. Because the time between the energy fall and rise is less than typical in speech, Ammit marks that section of audio as containing a lost packet.

While the intuition is simple, there are several challenges to using this technique to detect losses from simboxed audio. The first challenge is that many packet losses will occur during naturally silent audio — meaning that there will be no significant change in energy. This fact merely limits the amount of detectable loss events. The second challenge is that speech regularly has short pauses (causing false positives). A third challenge is that because there is no guarantee that VoIP frames are fully contained within a single GSM frame, a VoIP loss could begin in the middle of a GSM packet. Finally, uncorrected packet losses will have very low but non-zero energy because the pure silence is altered by bit errors in air transmission or by degradations within the simbox.

The first step of detecting unconcealed packet loss is to compute the energy of the audio signal. Ammit uses Short Time Energy (STE) as its measure of signal energy. Short time energy is a frequently used metric in speech analysis [38]. STE is computed by taking small windows of data and summing the squared values of the signal in the window. More formally, STE can be written as

$$E_n = \sum_{i=n-N+1}^{n} ((x(i))w(n-i))^2$$

where x is the audio signal, w is the window function, n is the frame number and N is the frame size.

Ammit computes STE using a 10ms audio frame, not the 20ms frames used in GSM-FR and many other codecs, because 10ms is the minimum frame size used by a VoIP codec, as standardized in RFC 3551 [47]. We use the standard practice of using a Hamming window half the length of the frame with a 50% overlap. Therefore, each STE measurement covers 5ms of audio and overlaps with 2.5ms of audio with the last window. This fine-grained measurement of energy ensures that Ammit can detect packet loss that begins in the middle of a GSM air frame.

With STE computed, Ammit then computes the lower envelope of the energy. In the presence of noise, the "silence" inserted in the VoIP audio will have non-zero energy. We define the lower envelope as the mean of the minimum energy found in the 10 ms frames. We also determine a tolerance around the minimum energy consisting of 50% of the lower envelope mean (this was determined experimentally).

Once Ammit has determined the lower envelope, it looks for energies that fall within the minimum envelope tolerance but then rise after a short number of energy samples. We experimentally chose 40ms as the maximum value for a sudden drop in packet energy, and our experimental results reflect the fact that this period is lower than the minimum for pauses in standard speech (which is around 50–60ms).

Because this method simply looks for silence, it is effective for both codecs we study, and it is fundamentally suited for all codecs that insert silence in the place of lost packets.

4.3 Detecting Concealed Losses in GSM-FR

Before we describe how Ammit detects GSM-FR packet loss concealment, we first describe GSM-FR PLC [24] at a high level. On the first frame erasure, the erased frame is replaced entirely by the last good frame. On each consecutive frame erasure, the previous frame is attenuated before replacing the erased frame. After 320ms (16 frames) of consecutive frame erasures, silence is inserted. Attenuation of repeated frames is motivated by the fact that while speech is stationary in the short term, longer-term prediction of audio has a high error that users perceive as unnatural.

Repeating frames wholesale has the frequency domain effect of introducing harmonics every $\frac{1}{20ms} = 50Hz$ [43]. Thus, there will be a spike in the cepstrum[3] at the 20ms

[3] A "Cepstrum" is a signal representation defined as the inverse Fourier transform of the logarithm of the Fourier transform. A rough mental model is to think of the "cepstrum" as the "Fourier transform

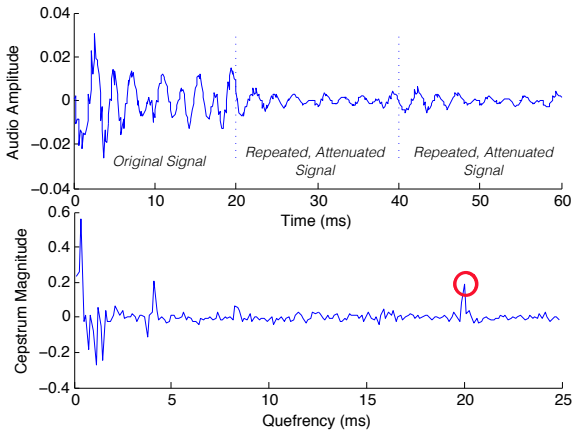


Figure 3: GSM-FR repeats and attenuates the last good frame to conceal packet loss. This results in a clear peak at 20ms in the cepstrum of the audio that can be used to detect a simboxed call.

quefrency. Because 50Hz is well below human pitch, this is a distinctive indicator of GSM packet loss. Figure 3 shows a clip of audio that has had GSM-FR packet loss concealment applied and the corresponding cepstrum. Note that the audio repeats (but is attenuated) every 20ms resulting in a peak at the cepstrum at 20ms. To detect GSM-FR PLC, Ammit computes the cepstrum of a window of three frames of audio and looks for a coefficient amplitude in the 20ms quefrency bin that is double the standard deviation of amplitudes of the other cepstral coefficients and not located in a silent frame.

4.4 Simbox Decision and SIM Detection

While concealed and unconcealed packet loss are measurable indicators of simboxing, there is a small false positive rate caused by the imperfection of our signal processing techniques. Accordingly, a single instance of a detected loss or concealed loss is not sufficient to consider a call to be originated from a simbox. Instead, we normalize the counts of loss events by the number of total frames in a call and consider a call as simboxed if the loss event percentage is much higher than the average loss event percentage for legitimate audio. We show in the following section that this approach is effective for all but the highest quality VoIP links, which provide few loss events to detect.

Even with this thresholding, some legitimate calls will occasionally be marked as a simbox. To ensure detection of simboxes with even improbably low loss rates, and to reduce the impact of false positives, we propose that the network should keep track of the number of times a call placed from a SIM is marked as a simboxed call. We term this technique "SIM detection" and show in the following sections that by using this technique we can further discriminate the legitimate subscribers from simboxers.

of the Fourier Transform of a signal. "The domain of the function is termed "quefrency" and has the units of seconds

4.5 Efficiency of Ammit

Ammit is designed to analyze call audio in real time as it is received by the cellular tower. So, the system must be designed to function efficiently using minimal computation and network resources. To accomplish this, we avoid using costly analysis associated with machine learning or complex signal feature analysis, and instead apply simple threshold checks to processed audio signals. For each time window collected by Ammit, we apply two iterations of the Fast Fourier Transform (FFT) and a comparison operation to the distinguishing criteria noted above. The FFT is a well-known algorithm that can be run with $O(n \log n)$ complexity, and is used to analyze audio in real time for applications such as audio visualizers. We further verify empirically that these operations can be executed in real time in Section 7.

In addition, any added load on the network will cause a minimal impact on the overall throughput. While Traynor et al. [50] demonstrated that added signaling within the cellular network can cause a DDoS effect, Ammit sends only a single message to the HLR for any call flagged as a simboxed call. For this added messaging to cause an effect on the internal cellular network, a cell containing a simbox would have to simultaneously send significantly more messages than there are channels to handle cellular calls, which is not possible.

5 Threat Model and Evasion

To evade Ammit, simboxers must either compromise Ammit's measurement abilities or successfully prevent or hide VoIP losses. While simboxers will take every economically rational action to preserve their profitability, attempting to evade Ammit will be difficult and likely expensive. This will hold true even if simboxers are aware of Ammit's existence and detection techniques, and even if simboxers are able to place arbitrary numbers of calls to test evasion techniques. In this section, we outline basic assumptions about our adversary. We then provide details about how Ammit can be expanded to address stronger adversaries that could defeat the prototype described in this paper.

5.1 Security Assumptions

The effectiveness of Ammit relies on four reasonable assumptions to ensure that Ammit cannot be trivially evaded by simboxers. First, we assume that the Ammit system (hardware and software) is no more accessible to the attacker than any other core network system (including routing and billing mechanisms). Second, we assume that Ammit will be used to analyze all call audio so that simboxers cannot evade a known evaluation period. We show in Section 7.4 that Ammit can efficiently analyze calls. Third, we assume that Ammit will report measurements to a single location (like the HLR) so that simboxers cannot evade Ammit by frequently changing towers. Finally, we recommend that Ammit be widely deployed throughout a carrier's infrastructure because a wider deployment will provide fewer places for simboxes to operate.

5.2 Evasion

If the simboxer cannot avoid Ammit analysis, he must hide or prevent VoIP packet loss and jitter. Hiding packet loss and jitter was the very goal of over two decades of intense academic and industrial research that has so far only provided good *but algorithmically detectable* solutions, including jitter buffering and loss concealment.

Extreme jitter buffering VoIP clients (including simboxes) routinely use short audio buffers to prevent low levels of jitter from causing delays in playback. Simboxers could set the jitter buffer to a large value (say, several seconds of audio) to prevent jitter from causing noticeable audio artifacts. However, this would be intrusive to users, and Ammit could still detect true losses as well as the added false starts and double talk. While we leave the testing of this approach to future work, we briefly describe how high jitter buffers could be detected by measuring the incidence of double talk. Double talk is the phenomenon where, after a lull in conversation, two users begin to talk (apparently) simultaneously. Because double talk increases with audio latency, increased double talk will be indicative of increased latency. Because an increased jitter buffer (combined with the already high call latency from an international call) will lead to higher than "normal" latency, detecting anomalous double talk will help in detecting simboxing. Detecting double talk is an important task in equipment quality testing, and ITU-T standard P.502 provides an off-the-shelf method for measuring it. Feasibility and appropriate thresholds can be determined using call data through simboxes and from legitimate subscribers. While such data is unavailable to outside researchers (including the authors of this paper), it is available to the carriers who would be fielding such a system.

Alternative PLC approaches Ammit looks for brief silences as one signal of VoIP loss, so simboxers could replace silence with noise or other audio. This is a well known form of Packet Loss Concealment. In general, PLC algorithms (like the GSM-FR PLC) fall into three categories: insertion, interpolation, and regeneration [44]. Although there are a number of algorithms in each category, the majority are published (and those that are not are often similar to those that are). All will have some artifacts that can lead to detection, and because the Pindr0p project has developed techniques to identify other codecs[25], we leave detecting other PLCs as future engineering work not essential to confirming our hypothesis that audio features can identify simboxes.

Improved link quality In addition to jitter and loss concealment, simboxers could reduce losses and jitter with high-quality network links or a redundant transmission scheme, but there are several barriers to this. First, finding a reliable provider may not be possible given the low connectivity conditions in simboxing nations. If a provider is available, the costs will likely be prohibitive. For example, in Kenya one can expect to pay $200,000 US *per month* for a high-quality 1 Gbps link[40]. This connection also guarantees little beyond the first routing hop. Beyond the costs, having a better quality connection than many universities and businesses may raise undesirable scrutiny and attention to the simboxers. Even if a high-quality link is available, it would not remove degradations from the call that occur before the call arrives at the entry point to the simbox.

Garbled frame transmission Finally, Simboxers could evade Ammit detection by failing to transmit valid GSM air frames when an IP frame is lost. In effect, Ammit would believe that all VoIP losses were air losses and would not detect VoIP losses. Ammit could detect this evasion by noting anomalous air loss patterns.

Currently, conducting a simboxing operation requires the technical sophistication of systems administrator. This evasion technique will require significant engineering resources (with expertise in embedded system design, implementation, and production) because GSM modems are typically sold as packages that accept an audio stream and high-level control commands (e.g. "place a call" or "send an SMS"). These tightly integrated chips are not capable of sending damaged packets on command. While the Osmocom baseband project [20] could provide a start for a custom radio, Osmocom targets inexpensive (though relatively rare) feature phone variants

and would not be a turnkey GSM baseband for a custom simbox[4]. Finally, even if the simboxers develop such a modem, they would have to conceal all detectable artifacts from both the final VoIP step as well as any intermediate networks (like a caller's mobile network). For these reasons, this strategy would only be effective for the most motivated and very well-funded simboxers.

However, in the event that simboxers do pursue this strategy, we propose the following methodology to detect such an attack. Given the considerable difficulty in developing the attack as well as constructing a suitable test environment, we leave testing this detection methodology to future work. We hypothesize that this a garbled packet evasion strategy can be detected from anomalous air interface loss patterns because simboxed calls will see the "typical" amount of loss *plus* the loss created by the simboxer. Loss patterns may be anomalous for improbable amounts of loss, or for improbably bursty sequences of lost frames. These anomalies could be determined on a tower-by-tower basis to take into account local transmission conditions (like a tunnel affecting signal quality). Because mobile stations (i.e. phones) do not know which frames are erased when they arrive at the tower, simboxers will not be able to tune their loss rate to be within the bounds used by this strategy.

6 Experimental Setup

In this section, we describe how we characterize Ammit through the use of simulation and test its effectiveness against a real simbox.

We simulate simboxed calls by taking a corpus of recorded audio and passing them first through a VoIP simulator then through a GSM air simulator (again, we use the term "air" to distinguish GSM cellular transmission). The GSM air simulator provides Ammit with both audio and a vector of GSM frame errors. To simulate legitimate calls, we pass the audio corpus through the air simulator only. We motivate the use of simulation in Section 6.6.

We test Ammit against three simbox codec choices: G.711 with no packet loss concealment and GSM-FR with and without packet loss concealment (we discussed this choice in Section 4. We evaluate single simbox call detection and SIM detection at 1%, 2%, and 5% loss rates (we justify this choice later in this section).

6.1 Speech corpus

The source of voice data for our experiments was the TIMIT Acoustic-Phonetic Continuous Speech corpus [33]. This is a *de facto* standard dataset for call audio testing. The TIMIT corpus consists of recordings of audio of 630 English speakers from 8 distinct regions each reading 10 "phonetically rich" standard sentences [5]. The recordings are 16kHz 16-bit signed Pulse Code Modulation (PCM), which are downsampled to 8kHz to conform to telephone quality. For the single call detection tests, we concatenate the 10 sentences for each of the 462 speakers into 1 call per speaker, creating a dataset of 462 calls [6]. Each call is approximately 30 seconds in length. The SIM detection test requires a larger call corpus, so for 98 randomly selected speakers we generate 20 calls for each speaker using permutations of the 10 sentences for each speaker (for a total of 1960 calls). Calls consist of only one speaker because Ammit analyzes each direction of the call separately.

6.2 VoIP Degradation and Loss

VoIP simulation takes TIMIT call audio as input and outputs audio that has been degraded by VoIP transmission. The simulator must convert the input audio from its original format (PCM) to the VoIP codec simulated (GSM or G.711), simulate loss, implement packet loss concealment in the case of GSM-FR, and output the final degraded audio. We examine these steps in greater detail in this subsection.

Audio conversions: The input audio files, encoded using PCM, must either be converted to G.711 or GSM-FR. We use the widely-used open source utility `sox` [8] for all codec transitions throughout the Ammit testing infrastructure. Note that these codec transitions are standard practice throughout PSTN and VoIP networks.

Packet Loss Modeling: We model Internet losses with the widely-used [39] Gilbert-Elliot packet loss model [34]. The Gilbert-Elliot model is a 2-state Markov model that models packet losses with bursty tendencies. A given channel can be in either a "good" state or a "bad" state. If the channel is in the "bad" state, packets are dropped. The Gilbert-Elliot model can be described with two parameters: p, the likelihood that the channel enters the "bad" state, and r, the likelihood that the channel leaves the bad state. p controls the frequency of loss events while r controls how long bursts last. We parameterize the model such that p is the target loss rate (for these experiments, 1%,2%, and 5%) and $r = 1 - p$. This means that the higher the loss rate, the greater the tendency of losses to be bursty.

Although jitter is a source of audio artifacts, we do not model jitter explicitly. Instead, because the audio symp-

[4] We pursued this line of research ourselves before finally purchasing a commercial simbox

[5] N.B. We use a subset of 462 male and female speakers from all 8 regions

[6] We set aside 12 of these calls as a training set to develop and verify our algorithms and set detection thresholds. These calls were not used for testing.

toms of jitter and packet loss are the same (i.e., audio is not present when needed), we simply consider jitter as a special case of packet loss, as is done by Jiang and Schulzrinne [39].

Loss Rate Justification: The reader may note that we are modeling loss rates that are considered high for Internet loss. Our model is justified for several reasons. The first consideration is that the typical Internet connection conditions in simboxing countries are of much lower quality than what most of Europe, East Asia, or even North America experiences [40, 51], with loss rates *often exceeding 10%*. Second, because conditions can vary from hour to hour or even moment to moment, examining performance at higher loss rates than typical is justified [39].

G.711 processing: To implement VoIP loss in G.711 audio, we use a packet loss simulation tool from the G.711 reference implementation available in the ITU Software Tools Library [7]. This tool implements concealed and unconcealed loss on 16-bit 8kHz PCM audio. We use `sox` to encode our input files to G.711 and back to 16-bit PCM before processing by the tool. This step is required because G.711 is a lossy codec, and the act of encoding and decoding irreversibly changes the audio. The tool takes a frame error vector as input, allowing us to use the Gilbert-Elliot Model described above.

GSM-FR processing: We developed our own GSM-FR VoIP loss simulator in Matlab. All audio processing in this tool is done on GSM-FR encoded audio. The tool implements the previously discussed packet loss model, the GSM-FR PLC as defined in 3GPP Standard 46.011 [24], and unconcealed packet loss by inserting GSM-FR silent frames.

6.3 GSM Air Loss

As we discussed in Section 6.6, we simulate simbox calls out of necessity. To simulate GSM cellular transmission (i.e., "air loss") we modify a GSM Traffic Channel simulation model for Simulink [41]. This model takes frames of GSM-encoded audio and encodes them as transmission frames for transmission over a GSM traffic channel as specified in 3GPP Standard 45.003 [21]. The transmission encoding includes interleaving as well as the error correcting codes and parity checks applied to Class 1 bits (as discussed in Section 4).

The model then simulates the modulation and transmission of the encoded frame using GMSK (Gaussian Minimum Shift Keying) in the presence of Gaussian white noise in the RF channel. This white noise is the source of random transmission errors in the model.

The model then demodulates the transmitted channel frame, evaluates the error correcting codes, and computes the parity check to determine if the frame is erased or not. Finally, the model outputs the received audio and a vector indicating which frames were erased.

The channel model signal-to-noise ratio is tuned to produce a frame erasure rate (FER) of 3% at the receiver, which is considered nominal according to 3GPP Standard 45.005 [22].

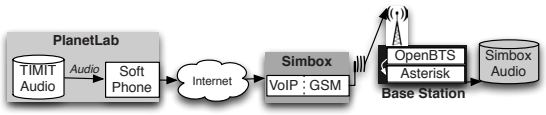


Figure 4: Our detection mechanisms are run against a real simbox deployment (Hybertone GoIP-1) communicating to a modified Range Networks OpenBTS base station.

6.4 Simboxing SIM Detection Test

Our SIM detection mechanism is tailored to reduce the effect of a single false positive or false negative call judgment by examining multiple calls.

To measure the effectiveness of this mechanism, we use 20 audio files from 98 unique speakers (for a total of 1960 calls) to simulate legitimate and simboxed calls using our GSM and VoIP simulators. We examine legitimate calls as well as simboxes covering all three codecs (GSM-FR, GSM-FR with PLC, and G.711) at 1%, 2%, and 5% loss rates. We model individual SIM cards as groups of 20 calls. For legitimate SIM cards, all calls from a particular speaker are assigned to a single SIM card, while simbox SIM cards consist of groups of randomly selected calls. This models the fact that simbox SIMs will rarely be used to provide service for the same user twice.

We analyzed all legitimate and simboxed calls with Ammit, then computed the percentage of calls in each SIM card group that were marked as simboxed. We consider a SIM to be used in a simbox if at least 25% of the calls it makes are marked as simboxed by Ammit call analysis.

6.5 Real Simbox Tests

We collect audio traces from calls made through a real simbox to validate our simulation experiments.

Figure 4 shows a schematic diagram of our experimental setup. We use 100 randomly selected audio files from the single call detection corpus (discussed in Section 6.1) to model the original call source. The call path begins at a PJSIP soft phone at a PlanetLab node located in Thai-

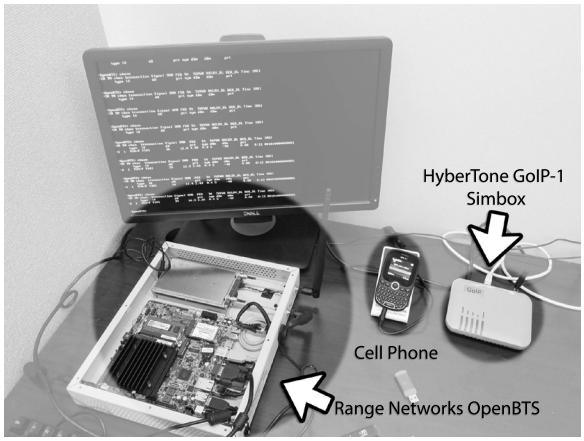


Figure 5: Our simbox experimental apparatus, including our OpenBTS GSM base station, mobile phone to model legitimate calls, and our GoIP-1 simbox.

land, a country with major simboxing problems [46][7]. This step emulates the arrival of a call to a simboxer.

The call originates from a soft phone and is routed through an Asterisk PBX[8] (not shown in the figure) to our Hybertone GoIP-1 simbox in the United States. Hybertone simboxes offer useful features to simboxing, including the ability to automatically change the IMEI number broadcast to evade filtering and detection systems like those presented in prior work [42]. Hybertone products have been advertised for sale specifically for simboxing [2], and entrepreneurs even sell value-added management consoles specifically for simboxers [1]. While the GoIP-1 supports several incoming codecs, it does not disclose which PLC algorithm it uses. We have determined experimentally that it is using a variant of the GSM-FR PLC.

The simbox delivers the call to a cellular base station under our control. Our base station is a Range Networks Professional Development Kit running the OpenBTS 5.0 open-source base station software and Asterisk 11.7. This base station is a low-power research femtocell and allows us to record call audio digitally as the base station receives it – including frame erasure information. To determine false positives, we create control calls by playing the same 100 randomly-selected audio files into a BLU Samba Jr. Plus feature phone and capturing the call audio at the base station. Figure 4 shows our base station and simbox experimental apparatus.

[7]Note that Thailand is the only major simboxing country with functional PlanetLab node at the time of writing

[8]A Private Branch Exchange (PBX) is a telephony switch analogous to an intelligent router in the Internet

6.6 Technical Considerations

Our experimental setup uses both simulation and real simbox data we collect ourselves for several reasons. First, simulations provide the best way to examine the effects of codec choice, packet loss concealment, and loss rates reproducibly and accurately. Second, they allow us to build generic models of simboxes so that our detection mechanism is not tied to any particular simbox model. Third, because we use tools and models that are extensively studied, verified, and frequently used throughout the literature [25, 52, 34, 39, 7, 41], we can have confidence that our results are correct. We supplement our simulations with data collected through a commonly used simbox to support and confirm our simulation results.

The reader will note that our real simbox calls were originated in a simboxing country, not terminated there. While simboxing is a global problem [42], we wanted to focus on areas where the problem is endemic and has a substantial impact. However, logistical, economic, and legal considerations prevented us from placing our simbox and research base station abroad. Instead, we capture the exact loss and jitter characteristics of the Internet connections in a simboxing country by originating the call there while terminating the call in our lab.

Legal and privacy concerns prevent us from receiving simbox audio from mobile operators (since the audio would be from callers who could not give their consent for such use). However, we note that there are no additional privacy concerns created when an operator deploys Ammit in a real network. Operators regularly use automated techniques to monitor call quality of ongoing conversations, and Ammit does no analysis that could be used to identify either the speakers or the semantic content of the call.

Finally, we note that the use of TIMIT audio is extremely conservative; it presumes pristine audio quality before the call transits an IP link. In fact, there will be detectable degradations from the PSTN even before the VoIP transmission. Chief among these will be GSM-FR PLC applied if Alice calls from a mobile phone. Because mobile phones regularly see high loss rates[9], simboxers carrying mobile-originated traffic will be even more vulnerable to detection by Ammit than this methodology reflects.

7 Detection Results

This section demonstrates how Ammit detects simbox fraud. We first discussed Ammit's effectiveness at identifying a real simbox, followed by a discussion of the

[9]Recall from 3GPP standard 45.005 [22] that 3% loss is considered nominal

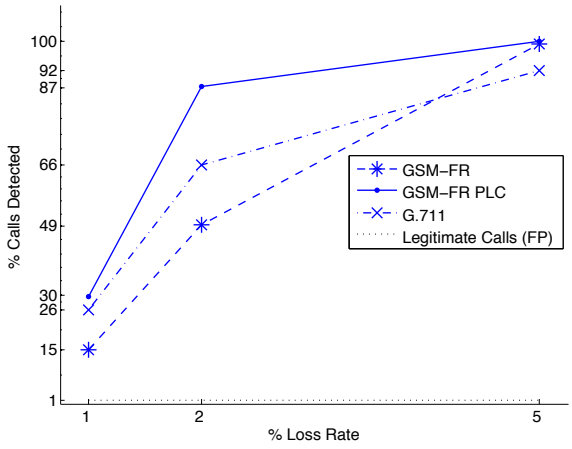


Figure 6: Ammit detection depends on the loss rate and Simbox codec used. For a 2% loss rate, Ammit detects over 55% of simboxed calls with less than a 1% false positive rate. This performance makes SIM detection (shown in Fig. 7) very reliable.

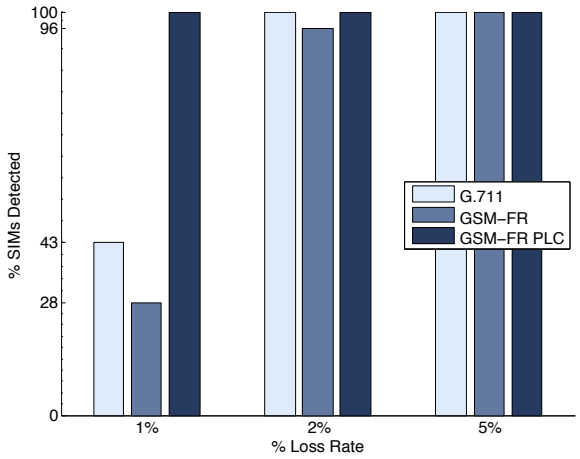


Figure 7: Even with unusually high-quality network connections, Ammit can be used to identify SIM cards used for simboxing.

results of detecting simulated simboxed calls. We then examine how Ammit can be used to identify SIM cards used in simboxing fraud. Finally, we show that Ammit is fast enough to be effective in real networks.

7.1 Simulated Call Analysis

In this subsection, we evaluate Ammit's ability to detect individual simboxed calls and SIM cards used in simboxing.

Figure 6 presents the percentage of simboxed calls detected for three simbox types at three different loss rates. At the still plausible 5% loss rate, Ammit detects from 87% to 100% of simboxed calls. Lower detection rates for low loss rates are simply a result of fewer loss events for Ammit to detect. However, in the case of no packet loss concealment, Ammit still detects from 15–66% of the simboxed calls for 1 and 2% loss. As discussed in the previous section, these loss rates include the effect of jitter, so loss rates as low as 1% and 2% are unlikely to be encountered often in practice [40, 51].

Third, the lowest dotted line in Figure 6 shows the low (but non-zero) detection rate for the control group of simulated legitimate calls — less than 1% (0.87% to be exact).

Figure 7 shows the percentage of simbox SIM cards that can be automatically disabled at the threshold of 25% of calls. For a 5% loss rate, our policy can identify *100%* of SIM cards used in simboxes. For calls using GSM-FR with packet loss concealment our policy can also detect 100% of SIM cards. As the loss rates decrease, we identify fewer SIM cards for codecs without packet concealment. In the case of 2% loss, we identify 96% and 100% of SIMs used in GSM-FR and G.711 simboxes, respectively. In the case of 1% loss, we still identify 43% of G.711 SIMs and 28% of GSM-FR SIMs. Our threshold results in a false positive rate of 1% and was determined experimentally from a ROC curve (omitted for space reasons). To counter the effects of false positives, the operator could implement a simple policy step allowing users to reactivate canceled SIM after some verification. One possibility is requiring flagged users to verify the National ID numbers used to register the SIM card over the phone or in person at a sales agent.

7.2 Detection of Real Simboxes

We begin with the most important result that Ammit is effective at detecting real simboxes. We find Ammit can detect 87% of real simboxed calls with *zero* false positives on the call set. These figures are the result of running our GSM-FR packet loss concealment after tuning on simulated individual call data; improved detection may be possible at a cost of a low false positive rate. While simulations produce useful insights about Ammit's performance in a wide range of conditions, these results confirm our hypothesis that call audio can be used to effectively combat simbox fraud.

7.3 Discussion

We make three observations from the individual call simulations. First, the results show a clear relationship between the loss rate of a call and Ammit's ability to detect a call. Second, Figure 6 shows the counterintuitive result that using GSM-FR packet loss concealment makes calls

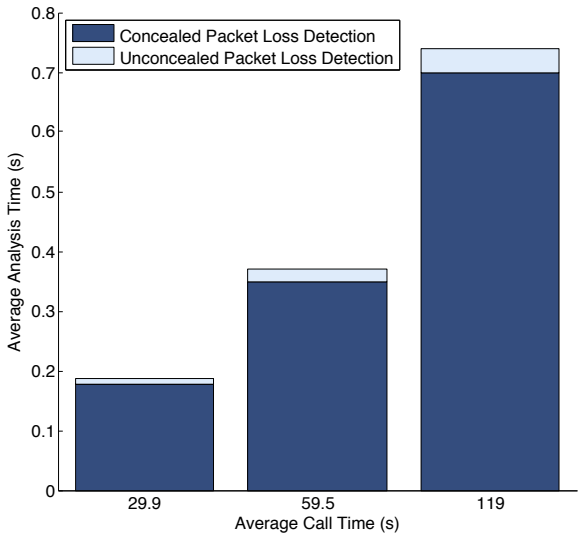


Figure 8: Ammit analyzes audio much faster than real time and is efficient enough to deploy in cell towers.

easier to detect. Even at a 1% loss rate, Ammit detects 30% of simboxed calls using GSM-FR PLC. Ammit is so effective at detecting concealed packet loss events because the GSM-FR PLC cepstral peak is distinctive and rare in speech. The corollary to this finding is that simboxers will have an incentive to disable packet loss concealment. This will noticeably impair call quality and user acceptability. Third, the non-zero false positive rate means that discretion will be required when Ammit indicates a positive simbox call.

Our SIM detection results show that Ammit can be used not only to detect single calls but as a larger initiative against simboxing. At 2% and 5% loss, we can detect and disable a single SIM card after at most 20 calls. Even at 1% loss, we can still detect and disable many SIM cards. Given that SIM cards come at a non-trivial cost (either at a legitimate point of sale or on a black market), by reducing the lifetime of a SIM card we make simboxers unable to operate.

Finally, we make two observations from the real simbox results. First, we note that our simulations were effective for tuning Ammit before applying real data. This validates our methodological strategy. Second, our simulation false positive rates were conservatively high; while we saw 1% false positives on our simulated data, we saw no false positives on our actual data.

7.4 Ammit Performance

To show that Ammit is scalable and performant, we examine the amount of time Ammit requires to analyze a call for concealed and unconcealed packet losses. Although in the previous subsections we analyzed Ammit's performance for 30 second calls, we hypothesize that longer analyses would lead to even better results, especially for lower loss rate calls. We tested Ammit's performance on a set of 10 calls of approximately 30s, 60s, and 120s; we present the averages of 10 analyses of each call in Figure 8.

We test Ammit on a late 2011 iMac with a quadcore 3.4 GHz Intel Core i7, 16GB RAM, and a 1TB solid state disk running OS X 10.9. Although this is capable hardware, the detection is done entirely with Matlab in a single thread, and the detection code is correct but far from optimal. Optimizing the Matlab code for efficiency would likely reduce analysis time. Beyond that, implementing Ammit in a more performant language like C could reduce analysis time further. For a commercial implementation, code customized for a digital signal processor could further improve performance. Ammit may be deployed directly as a BTS or BSC software update or as inexpensive standalone hardware.

As Figure 8 shows, the majority of analysis time is spent detecting concealed packet loss. Nevertheless, calls can be analyzed 150 times faster than real time, indicating that a single thread of execution could analyze approximately 150 calls per unit time. Even our unoptimized code would be able to analyze all traffic at a tower in real time.

8 Related work

Although this work is concerned with detecting simbox fraud, the techniques used belong to the long tradition of non-intrusive call quality measurement. Non-intrusive measurements are taken passively and without a reference audio; this is in opposition to intrusive measurements [4, 6] which measure the degradation of a known reference signal. Traditional call quality metrics measure listener experience, and imperceptible degradations do not significantly affect these scores. These scores have been shown to vary widely based on random conditions, language choice [48] or VoIP client [26]. The most widely used non-intrusive measurement standard is ITU specification P.563 [5], but other metrics have been developed for holistic quality measurements [32, 37, 30] and for individual artifacts like robotization [43] and temporal clipping [36]. Because call quality metrics like P.563 are only concerned with perceptible degradation and vary widely in results, they are unsuitable for detection of simbox fraud.

Telephony fraud detection is a well-studied problem, and efforts to fight telecommunications fraud have primarily depended on call records. Machine learning and data mining have been used extensively to detect fraudulent activity using call records [27, 29, 35, 45].

Given the importance of the simboxing problem in affected countries, there are a number of commercial simbox detection products, as well as two published research papers [31, 42]. Most simbox detection systems use one of two techniques: test call generation and call record analysis. A few products use hybrid techniques [14, 17]. Test call generation approaches [10, 13, 15, 16, 18] use probes widely deployed in many networks to verify that the CLI (i.e. Caller-ID) records on calls are correct — if a simbox is used, the CLI record would indicate the MSISDN (i.e. the phone number) of the SIM card routing the call and not the originating probe. Test call methods only work for certain kinds of simboxing (when a simboxer sells services to another telecom, not through the common case of selling calling cards to consumers). By contrast, call record analysis detect all types of simboxing. Those approaches rely on the fact that SIMs used in simboxes have usage patterns distinct from legitimate customers [31, 9, 11, 12, 19]. These techniques are prone to false positives and active evasion by simboxers. In recent work, Murynets et al. published a call record analysis approach that used machine learning to identify IMEIs (device identifiers) used by simboxes [42]. The authors' published accuracy rates measure identifying individual calls (not simbox devices) only after simboxes are identified, and thus are not directly comparable to the accuracy figures for Ammit. Additionally, that work identifies IMEIs (which are an asserted — and thus spoofable — identifier) of devices only after a simbox makes dozens or hundreds of calls with a single SIM card; even if the work described in that paper is deployed, simboxing will continue to be profitable. Our work is an improvement over the state of the art because we can reliably detect simboxed calls using features inherent to simboxing *at the time of the call*, thus making simboxing unprofitable.

While Ammit is the first system to combat simboxing using call audio, our system is a refinement of the ideas used in the Pindr0p system developed by Balasubramanian et al˙ [25].

The Pindr0p system combats telephony fraud by identifying callers using audio "fingerprints." These fingerprints consist of noise characteristics and indicators of different codecs used by the different PSTN and VoIP networks that route a call. For Pindr0p, capturing characteristics of end-to-end call path is essential to identify repeat callers. For Ammit, it is sufficient to hear audio that has been degraded by any prior network.

Ammit's techniques are tailored to better combat simboxing in several important ways.

First, for simbox detection, Pindr0p's greatest weakness is that it is designed to identify codec transitions. Accordingly, Pindr0p would fail to detect a simbox call that uses a single codec (i.e. the GSM codec) end to end. If Pindr0p were employed to combat simboxing, simbox operators would simply migrate to single-codec solutions. Accordingly, Ammit will detect simboxed calls that Pindr0p would fail to detect. In fact, by using loss information from the tower (and not endpoint audio) Ammit excels at this case better than any other tested.

Second, we showed that Pindr0p techniques used on individual calls may result in unacceptable false positives. This was especially true for the silence insertion detection methods proposed in that work. For Pindr0p, a classification error may prompt a call center worker to request additional authentication. For Ammit, a classification error will prompt a dropped call or suspended account — a much higher burden to the user. Ammit enhances Pindr0p by developing and verifying a SIM detection method that reduces false positives and increases confidence in classification.

Third, Pindr0p was designed to quickly fingerprint audio based on a short segment using a large number of features fed to a machine learning classifier. While we were unable to obtain access to Pindr0p for a direct comparison, we *do* show in this work that Ammit's techniques will enable real-time processing of all call audio at mobile network base stations.

9 Conclusions

Cellular networks in developing nations rely on tariffs collected at regulated interconnects in order to subsidize the cost of their deployment and operation. These charges can result in significant expense to foreign callers and create incentive for such callers to find less expensive, albeit unlawful, means of terminating their calls. Simboxes enable such interconnect bypass fraud by tunneling traffic from a VoIP connection into a provider network without proper authorization. In this paper, we develop the Ammit tool, which allows us to detect simboxes based on measurable differences between true GSM and tunneled VoIP audio. Ammit uses fast signal processing techniques to identify whether individual calls are likely made by a simbox and then to develop profiles of SIM cards. This approach allows a provider to deactivate the associated SIMs rapidly, and virtually eliminates the economic incentive to conduct such fraud. In so doing, we demonstrate that the subsidized rates that allow much of the developing world to be connected can be protected against the impact of this fraud.

10 Acknowledgments

The authors are grateful for the help of our shepherd, Srdjan Capkun, and for comments from our anonymous reviewers. Michael Good was instrumental in early work

on this project that was not published in this paper, and conversations with Charles Lever were especially helpful to work out the kinks.

This work was supported in part by the US National Science Foundation under grant numbers CNS-1464088 and CNS-1318167. This work was also supported in part by the Harris Corporation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Harris Corporation.

References

[1] GoAntifraud.com.

[2] Goip For Grey Route SIM Box. http://www.alibaba.com/product-detail/16-ports-gsm-gateway-goip-for_862885942.html.

[3] ITU-T recommendation G.711.

[4] ITU standard P.800 methods for subjective determination of transmission quality.

[5] ITU standard P.563:single-ended method for objective speech quality assessment.

[6] ITU standard P.862:perceptual evaluation of speech quality (PESQ).

[7] *ITU Software Tool Library Manual*. ITU, Geneva, 2009.

[8] sox. http://sox.sourceforge.net/Main/HomePage, 2013.

[9] Agilis international simbox detection. http://www.agilisinternational.com/solutions/customer-analytics/risk-and-fraud-management/, 2014.

[10] Araxxe SIM box detection. http://www.araxxe.com/SIM-box-detection.html, 2014.

[11] CxB solutions SIM box detection. http://www.cxbsolutions.com/html/sim_box_detection.html, 2014.

[12] FraudBuster SIMBuster. http://www.fraudbuster.mobi/new-simbuster-and-trafficchecker-deployment-in-africa/, 2014.

[13] Meucci solutions SIM box detection. http://www.meucci-solutions.com/solutions/fraud-and-revenue/sim-box-detection/, 2014.

[14] Mobius fraud mangement. http://www.mobiusws.com/solutions/fraud-management/, 2014.

[15] Mocean SIM box detector. http://www.mocean.com.my/SIM_box_detector_solution.php, 2014.

[16] Roamware SIM box detector. http://www.roamware.com/predictive_intelligence_sim_box_detector.php, 2014.

[17] ROC fraud management. http://www.subex.com/pdf/bypass-fraud.pdf, 2014.

[18] Telenor simbox detection. http://www.telenorglobal.com/wp-content/uploads/sites/4/2013/09/Global-SIM-Box-Detection1.pdf, 2014.

[19] XINTEC SIM box detector. http://www.xintec.com/fraud-management/sim-box-detector/, 2014.

[20] OsmocomBB GSM baseband. http://bb.osmocom.org/trac/, 2015.

[21] 3RD GENERATION PARTNERSHIP PROJECT. 3GPP TS 45.003 v12.0.0. Tech. Rep. Channel coding.

[22] 3RD GENERATION PARTNERSHIP PROJECT. 3GPP TS 45.005 v12.1.0. Tech. Rep. Radio transmission and reception.

[23] 3RD GENERATION PARTNERSHIP PROJECT. 3GPP TS 46.010 v11.1.0. Tech. Rep. Full rate speech; Transcoding.

[24] 3RD GENERATION PARTNERSHIP PROJECT. 3GPP TS 46.011 v11.1.0. Tech. Rep. Full rate speech; Transcoding.

[25] BALASUBRAMANIYAN, V. A., POONAWALLA, A., AHAMAD, M., HUNTER, M. T., AND TRAYNOR, P. PinDr0p: using single-ended audio features to determine call provenance. In *Proceedings of the 17th ACM conference on Computer and communications security* (New York, NY, USA, 2010), CCS 10, ACM, p. 109120.

[26] BROOM, S. VoIP quality assessment: Taking account of the edge-device. *IEEE Transactions on Audio, Speech, and Language Processing 14*, 6 (2006).

[27] BURGE, P., AND SHAWE-TAYLOR, J. An unsupervised neural network approach to profiling the behavior of mobile phone users for use in fraud detection. *Journal of Parallel and Distributed Computing 61*, 7 (July 2001), 915–925.

[28] COMMUNICATIONS FRAUD CONTROL ASSOCIATION (CFCA). 2013 Global Fraud Loss Survey. http://www.cvidya.com/media/62059/global-fraud_loss_survey2013.pdf, 2013.

[29] COX, K. C., EICK, S. G., WILLS, G. J., AND BRACHMAN, R. J. Visual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery 1*, 2 (June 1997), 225–231.

[30] DING, L., LIN, Z., RADWAN, A., EL-HENNAWEY, M. S., AND GOUBRAN, R. A. Non-intrusive single-ended speech quality assessment in VoIP. *Speech Communication 49*, 6 (June 2007), 477–489.

[31] ELMI, A. H., IBRAHIM, S., AND SALLEHUDDIN, R. Detecting SIM box fraud using neural network. In *IT Convergence and Security 2012*, K. J. Kim and K.-Y. Chung, Eds., no. 215 in Lecture Notes in Electrical Engineering. Springer Netherlands, Jan. 2013, pp. 575–582.

[32] FALK, T., AND CHAN, W.-Y. Single-ended speech quality measurement using machine learning methods. *IEEE Transactions on Audio, Speech, and Language Processing 14*, 6 (2006), 1935–1947.

[33] GAROFOLO, J. S., LAMEL, L. F., FISHER, W. M., FISCUS, J. G., PALLETT, D. S., DAHLGREN, N. L., AND ZUE, V. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Linguistic Data Consortium, Philadelphia, 1993.

[34] HASSLINGER, G., AND HOHLFELD, O. The Gilbert-Elliott model for packet loss in real time services on the Internet. In *Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB), 2008 14th GI/ITG Conference* (March 2008), pp. 1–15.

[35] HILAS, C. S., AND MASTOROCOSTAS, P. A. An application of supervised and unsupervised learning approaches to telecommunications fraud detection. *Knowledge-Based Systems 21*, 7 (Oct. 2008), 721–726.

[36] HINES, A., SKOGLUND, J., KOKARAM, A., AND HARTE, N. Monitoring the effects of temporal clipping on VoIP speech quality. In *14th Annual Conference of the International Speech Communication Association* (2013), ISCA.

[37] HOENE, C., KARL, H., AND WOLISZ, A. A perceptual quality model intended for adaptive VoIP applications: Research articles. *Int. J. Commun. Syst. 19*, 3 (Apr. 2006), 299316.

[38] JALIL, M., BUTT, F., AND MALIK, A. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In *Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2013 International Conference on* (May 2013), pp. 208–212.

[39] JIANG, W., AND SCHULZRINNE, H. Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss. In *Proceedings of the 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video* (New York, NY, USA, 2002), NOSSDAV '02, ACM, pp. 73–81.

[40] LES COTTRELL, R. Pinging Africa - a decade long quest aims to pinpoint the Internet bottlenecks holding Africa back. *Spectrum, IEEE 50*, 2 (Feb 2013), 54–59.

[41] MOLINA, J. A. S. GSM traffic channel simulator. `http://www.mathworks.com/matlabcentral/fileexchange/11078-gsm-traffic-channel-simulator`, 2006.

[42] MURYNETS, I., ZABARANKIN, M., JOVER, R., AND PANAGIA, A. Analysis and detection of SIMbox fraud in mobility networks. In *2014 Proceedings IEEE INFOCOM* (Apr. 2014), pp. 1519–1526.

[43] PAPING, M., AND FAHNLE, T. Automatic detection of disturbing robot voice and ping-pong effects in GSM transmitted speech. In *EUROSPEECH* (1997).

[44] PERKINS, C., HODSON, O., AND HARDMAN, V. A survey of packet loss recovery techniques for streaming audio. *IEEE Network 12*, 5 (1998), 40–48.

[45] QAYYUM, S., MANSOOR, S., KHALID, A., KHUSHBAKHT, K., HALIM, Z., AND BAIG, A. Fraudulent call detection for mobile networks. In *2010 International Conference on Information and Emerging Technologies (ICIET)* (2010), pp. 1–5.

[46] RATANA, U. Telcos lose money to SIM fraud. *Phnom Penh Post* (Feb. 2014).

[47] SCHULZRINNE AND CASNER. RFC 3551. Tech. Rep. RTP Profile for Audio and Video Conferences with Minimal Control.

[48] TAKAHASHI, A., KURASHIMA, A., AND YOSHINO, H. Objective assessment methodology for estimating conversational quality in VoIP. *IEEE Transactions on Audio, Speech, and Language Processing 14*, 6 (2006), 1984–1993.

[49] TRAYNOR, P. Characterizing the Security Implications of Third-Party EAS Over Cellular Text Messaging Services. *IEEE Transactions on Mobile Computing (TMC) 11*, 6 (2012), 983–994.

[50] TRAYNOR, P., LIN, M., ONGOING, M., RAO, V., JAEGER, T., MCDANIEL, P., AND LA PORTA, T. On cellular botnets: Measuring the impact of malicious devices on a cellular network core. In *Proceedings of the 16th ACM conference on Computer and communications security* (2009).

[51] WANG, Y., HUANG, C., LI, J., AND ROSS, K. Queen: Estimating packet loss rate between arbitrary internet hosts. In *Passive and Active Network Measurement*, S. Moon, R. Teixeira, and S. Uhlig, Eds., vol. 5448 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2009, pp. 57–66.

[52] WHITE, A. M., MATTHEWS, A. R., SNOW, K. Z., AND MONROSE, F. Phonotactic reconstruction of encrypted voip conversations: Hookt on fon-iks. In *Proceedings of the 2011 IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2011), SP '11, IEEE Computer Society, pp. 3–18.

A Appendix

Table 1: Commercial VoIP GSM Gateway Survey

Brand	Selected Codecs
2N VoiceBlue Next	G.711, G.729ab
2N StarGate	G.711, G.729a
OpenVox VoxStack	G.711, G.729, GSM
Dinstar DWG2000B	G.711, G.729ab
ElGato K32	G.711, G.729
Gempro GP-708	G.711, GSM
iQsim M400	G.711, G.729(a&ab)
Nicherons SpoGSM-G4	G.711, GSM, G729
PORTech MV-378	G.711, G.729(a&ab)
SunComm SC-024-S	G.711, G.729ab
Hybertone GoIP-1	G.711, G.729ab, GSM
Yeastar NeoGate TG800	G.711, G.729a