Задачи и вопросы по обфускации

Юрий Лифшиц

31 октября 2004 г.

1 Спрятанный if - демо-код

Неформальная постановка. Предположим у вас есть программа P которая реализует вычисление некоторой функции $f : \{0,1\}^* \to \{0,1\}^*$. Будем считать, что эта функция не восстанавливается по нескольким наугат взятым парам $\{x, f(x)\}$. Мы хотим сделать из P демо-версиюn (h-версию) D обладающую следующими свойствами: 1) если $h(x) = 0$ то $D(x) = P(x)$, если $h(x) \neq 0$, то $D(x)$ - произвольно; 2) Анализируя и исполняя код D невозможно за полиномиальное время написать программу эквивалентную P. Уточним, что функция h противнику неизвестна.

Пример 1. Можно взять $f(x) = x^a(modp)$, где p - всем известное простое число, а показатель a является секретом. Интересно построить нетривиальный демо-код хотя-бы для какого-нибудь h.

Пример 2. Взять в качестве f какой-нибудб псевдослучайный генератор. В интернете можно найти хорошие lecture notes по этой теме (pseudorandom generators).

2 Неаппроксимируемый выход

Неформальная постановка. Задача заключается в написании программы, множество результатов которой невозможно было бы аппроксимировать (накрыть большим множеством) даже с полиномиальным (относительно размера n входа и выхода программы) приближением. Выяснить какие выходные множества обладают таким свойством. Иначе говоря нет алгоритма угадывающего "является ли строка возможным выходом" с вероятностью хотя бы $1/p(n)$ для какого-нибудь полинома p.

Пример. Придумать семейства функций $F \subseteq \{f : B^n \to B^{n+m}\}$ и $H \subseteq \{h : B^{n+m} \to \{0,1\}\}$ такие, что для каждой $f \in F$ найдется $h \in H$ такая что $h(f(x)) \equiv 0$, но подобрать эту h анализируя код f очень сложно.

3 Расширение операции

Неформальная постановка. Требуется придумать операцию на подмножестве строк

определенной длины, так чтобы ее можно было расширить на все строки так, чтобы исходную операцию нельзя было бы восстановить.

Пример. Пусть $f \subset \{f : B^n \to \{0,1\}\}$ – семейство таких функций, что имея несколько пар $(x_1, f(x_1)), \ldots, (x_{p(n)}, f(x_{p(n)}))$ нельзя эффективно выяснить о каком представителе F идет речь (можно сказать по-другому - владея такой последовательностью трудно самостоятельно угадать значение $f(x_{p+1})$). Требуется придумать семейства функций $P \subset \{p : B^n \to B^{n+m}\}$ и $Q \subset \{q : B^{n+m} \to \{0,1\}\}$, что для $\forall f \in F \exists p \in P, q \in Q : f(x) \equiv q(p(x))$ и обладая описанием p и q, нельзя получить описание f.

4 Операции над запутанным кодом

Неформальная постановка. Требуется придумать общий формат запутывания программ, который бы поддавался следующим операциям: 1) композиция - способ из обфускаций $O(f)$ и $O(g)$ построить еще более запутанную обфускацию $O(f \circ g)$; 2) слияние - построить из обфускаций $O(f_1)$ и $O(f_2)$ обфускацию $O(F)$, где $F(b, x) = f_b(x)$.