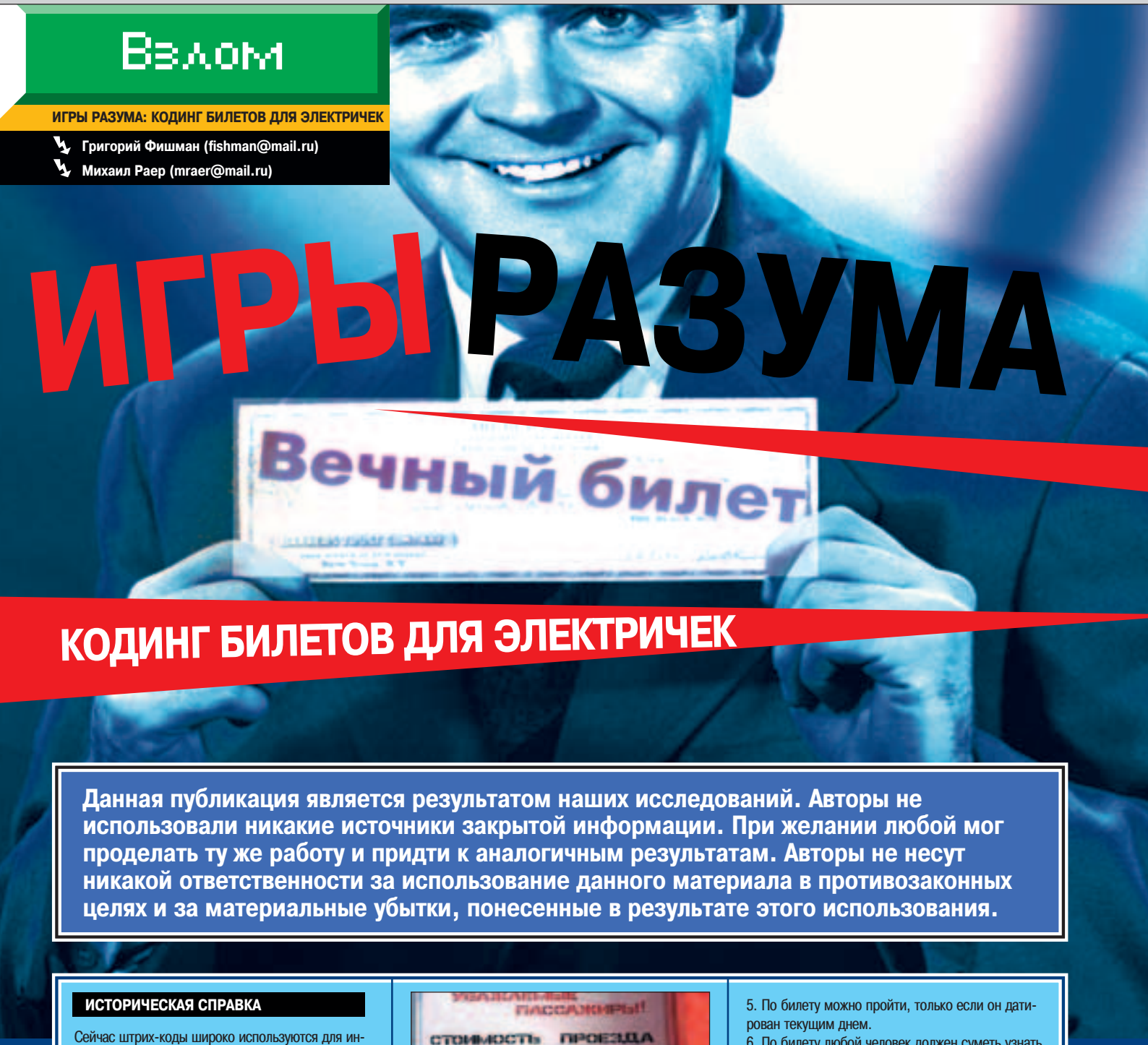
ИГРЫ РАЗУМА: КОДИНГ БИЛЕТОВ ДЛЯ ЭЛЕКТРИЧЕК

Григорий Фишман (fishman@mail.ru)
Михаил Раер (mraer@mail.ru)

ИГРЫ РАЗУМА



КОДИНГ БИЛЕТОВ ДЛЯ ЭЛЕКТРИЧЕК

Данная публикация является результатом наших исследований. Авторы не использовали никакие источники закрытой информации. При желании любой мог проделать ту же работу и придти к аналогичным результатам. Авторы не несут никакой ответственности за использование данного материала в противозаконных целях и за материальные убытки, понесенные в результате этого использования.

ИСТОРИЧЕСКАЯ СПРАВКА

Сейчас штрих-коды широко используются для индексации различных товаров. Их можно встретить на обертках продуктов питания, на книгах, тетрадях, библиотечных читательских билетах и еще много где. Мы же расскажем об их использовании на железнодорожных билетах.

Ирония судьбы состоит в том, что штриховое кодирование изобрел железнодорожник, инженер Давид Коллинз. В 50-х годах он окончил Массачусетский технологический институт и пошел работать на Пенсильванскую железную дорогу. Там он занимался сортировкой вагонов. Это был тяжкий труд. Вагоны нужно было пересчитать, проконтролировать по документам каждый вагон, определить путь следования. Тогда он придумал записывать номера вагонов кодом, состоящим из красных и синих полос. Длина такого кода достигала полуметра. Шло время. Цветные полосы превратились в черно-белые, штрих-код умещается на билете шириной 4,5 см, да и цели перестали быть такими бескорыстными и благородными...

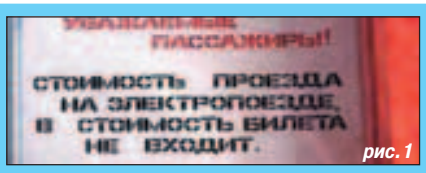


рис. 1

АВСТРАЛИЙСКОЕ МПС ПРОТИВ КРОЛИ

Совсем недавно МПС жилось плохо. Ему мешали собирать капусту многочисленные зайчики и кролики. Поэтому, для сохранения бизнеса, МПС поставило перед собой задачу: оградить свое поле высокой стеной и ввести систему платных пропусков для обитателей леса. Эта система должна была удовлетворять следующим требованиям:

1. Позволять проходить как по абонементам, так и по одноразовым билетам.
2. Автоматически контролировать доступ пассажиров на платформы.
3. Низкая себестоимость билета (стоимость клочка бумаги).
4. Билет должен быть защищен от подделок.
5. По билету можно пройти, только если он датирован текущим днем.
6. По билету любой человек должен суметь узнать дату билета, исходную и конечную зоны маршрута, направление, цену и тип билета. Эта информация должна быть доступна контролерам и пассажирам без использования технических средств.
7. На билете должны содержаться номер кассы и номер билета для проверки фальсификаций, предупреждения повторного использования и контроля денежных средств.

Учитывая всеобщую любовь к овощам, на возможностях этой системы было решено сэкономить. В результате была сформулирована концепция автономности станций, по которой, например, по билету, купленному и использованному на одном вокзале, можно пройти на другом вокзале той же зоны.

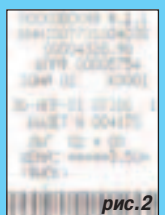


рис.2

Теперь рассмотрим разные варианты реализации одноразовых билетов:

1. Магнитная или смарт-карта отпадают по п.3 и п.6.
2. Жетоны отпадают по п.5 и п.6.
3. Бумажка с водяными знаками и напечатанными данными отпадает по п.2.

Единственное, что осталось - это клочок бумаги с напечатанными на нем данными и секретным штрих-кодом, по которому турникет сможет определять параметры билета. Именно такой вариант и был выбран (см. рис. 2). Далее расскажем о методе кодирования используемого штрих-кода.

С ЧЕМ ЕДЯТ ШТРИХ-КОД

Штрих-код - это не магнитная или смарт-карта, это всего лишь последовательность вертикальных полосок, которую может напечатать каждый. Поэтому непонятно, почему билет так слабо защищен! Существует несколько разных форматов штрих-кодов. Они бывают линейные и двухмерные. Нас будет интересовать линейный код ITF (Interleaved 2 из 5). Код состоит из последовательности чередующихся черных и белых вертикальных полосок, начинающейся и заканчивающейся черными. Полоски бывают широкие (будем обозначать "1") и узкие (будем обозначать "0"). ITF предполагает наличие стартового и стопового символов, которые кодируются "0000" и "100" соответственно. Между этими символами и расположена полезная информация. Свое название код получил, исходя из того, что информация кодируется расположением двух широких полосок среди пяти. Применяя элементарную комбинаторику, получаем, что таким блоком можно закодировать C<!!!! upperindex:2 lowerindex:5 !!!!!> = 10 различных значений, т.е. все цифры от 0 до 9. Interleaved (перемежающийся - англ.) же он потому, что рассматриваются отдельно черные и отдельно белые полоски. На четных позициях записывается цифра, образованная черными полосками (см. рис. 3). Алфавит кода можно найти на http://www.sbarcode.com/encoding/int25.shtml или на нашем сайте http://tickers.da.ru.

рис.3

КРОЛИКИ РЕШИЛИ ПОДУМАТЬ

Штрих-код железнодорожного билета состоит из 87 полосок. Семь из них - это стартовый и стоповый символы. Остается 80, которые несут информацию. Делим на пять и получаем, что там записано 16 десятичных цифр. Что же это за 16 цифр? Что там записано? Может быть, это номер записи в базе данных проданных билетов? Нет, это невозможно из-за автономности касс. Там закодирована вся информация о билете! (см. таблицу1)

Такая вот странная нумерация с нуля сложилась из-за наличия Си'шников в нашей команде :).
Сразу возникают следующие вопросы:
Q: Почему не кодируются все цифры номеров касс и билета?
A: Ответ прост - а зачем? Этого и так вполне достаточно.
Q: Какая же позиция в штрих-коде у каждой аббревиатуры (у каждого поля)?
A: Все не так просто :). Во-первых, недостаточно знать только это, по той причине, что турникет пропускает только билеты сегодняшней даты, поля которой у нас в таблице не отражены. Во-вторых, эти позиции переставляются каждый день. И, в-третьих, к каждому значению поля каждый день прибавляется по модулю 10, определенное для данного дня и данного поля число (которое мы называем смещением).
Рассмотрим перестановку и смещение определенной даты. Сделаем это, например, для 17 марта 2001г. (см. таблицу2)

Всю таблицу мы называем маской (mask).
Верхнюю строчку - перестановкой (transposition).
Нижнюю - вектором смещения (offset vector) или просто вектором.
Давай подведем первый итог:
Чтобы узнать шифр, т.е. научиться по штрих-коду узнавать все данные о билете и наоборот, генерировать штрих-код по любым желаемым данным, достаточно знать соответствующие дате маску и значения констант. Т.е. по приведенной маске можно создавать и распознавать любой билет, датированный 17-м марта 2001г...

NEXT

K1 8	Z2 1	N2 5	D 4	CZ0 1	Z0 2	CN1 0	N1 5	K2 0	N3 6	Z1 3	Z3 2	CN0 0	N0 2	CK0 4	K0 0

таб. 2

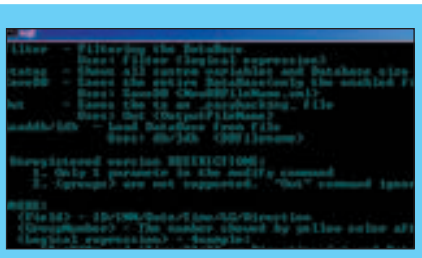
АББРЕВИАТУРА ДЕСЯТЕРИЧНОЙ ЦИФРЫ Нумерация производится справа налево	ЗНАЧЕНИЕ	ПРИМЕР С РИС.1
K0, K1, K2	Последние три цифры номера кассы (строка "БПМФ..." в билете).	K0 = 4; K1 = 5; K2 = 7
N0, N1, N2, N3	Последние четыре цифры номера билета (строка "БИЛЕТ N..." в билете).	N0 = 0; N1 = 7; N2 = 1; N3 = 4
Z0, Z1	Зона прибытия	Z0 = 0; Z1 = 0
Z2, Z3	Зона отправления	Z2 = 2; Z3 = 0
D	Тип билета и направление (расшифровка далее)	
CZ0, CN0, CK0, CN1	Мы их называем "четыре константы" (скоро будет пояснено)	

таб. 1

Взлом

ИГРЫ РАЗУМА: КОДИНГ БИЛЕТОВ ДЛЯ ЭЛЕКТРИЧЕК

Григорий Фишман (fishman@mail.ru)
Михаил Раер (mraer@mail.ru)



ОНО БЫЛО ИХ СЕКРЕТОМ. ПРЕЛЮДИЯ

Все началось в марте 2001-го... На большинстве лекций было скучно, дома печально, а в аквариуме (наш компьютерный класс) грустно. И вдруг, как гром среди ясного неба, Миша предложил новый увлекательный проект... И мы, почти всей группой, стали ходить на вокзал, собирать билетики и анализировать их. Наш аквариум был сразу же перепрофилирован из места, в котором происходят великие битвы в Starcraft и поля для сражения в шахматы по интернету с сокурсниками из соседней части аквы, в центр обработки билетов. Треть людей вносила их в базу данных, треть рядом писала различные программки для их анализа, а другая треть сидела на лекциях, вглядываясь в бесконечные потоки цифр и аббревиатур. Постепенно начала выявляться некая логическая структура. И вдруг, на одной из лекций, наш чемпион Александр Штучкин уловил определенную закономерность (вспомните фильм "Игры разума" со сценой в Пентагоне :)). Это был поворотный момент, после которого наши исследования перешли на новый качественный уровень.
Однако вскоре наступила сессия, и приоритеты сдвинулись. Летом '01 была написана программа, которая по известным нескольким билетам одной даты выдавала все варианты масок, отбрасывая те, которые не подчиняются парному закону. О них будет сказано чуть позже. Вскоре был сделан новый (и на данный момент последний) качественный скачек, и в программу была добавлена функция, по которой стало достаточно лишь одного билета для каждой даты.

ЧЕТЫРЕ КОНСТАНТЫ

Константы так названы потому, что их значение не изменяется в течение дня. Возникает даже здоровое подозрение, что это дата... Однако самое смешное, что одна из констант все-таки меняется. Но мы заметили это слишком поздно, и потому поле сохранило свое название. В науке, а особенно в физике, существует масса примеров, когда первоначальное неправильное представление о чем-либо приводило к неправильной терминологии, которая продолжает использоваться в силу привычки. Мы же, подобно физикам-экспериментаторам, следуя их дурному примеру (а он, как известно, заразителен), сохранили за этим полем название константа. Повторим константы: CZ0, CN0, CK0 и та самая заблудившаяся потеряшка, псевдоконстанта CN1. Как корабль назовешь, так он и поплывет. Внимательный читатель обратит внимание на то, что в названия констант входят имена других полей. Почему так, будет объяснено дальше, в следующем абзаце...

ТАНЕЦ ДАТСКОГО КОРОЛЯ ИЛИ "ЗАКОН ПАР"

Изучая первые полученные маски, мы обнаружили, что расположение полей день изо дня "скачет". И вскоре, изучая эти "скачки", заметили, что все перестановки осуществляются только парами. То есть существуют 8 неразлучных пар, которые и переставляются. Чем руководствовались разработчики, нам непонятно. Вряд ли у них была цель нас запутать, потому что так нам стало гораздо легче - мы смогли различать константы. Просто назвали константу именем поля, которое находится с ней в одной связке. Далее мы призадумались над тем, как осуществляются перестановки пар. Искали период, но безуспешно. Казалось бы, вот расположение пар повторяется на девятый день. Ждем еще "9 дней, 9 ночей", ожидая ту же маску... как бы не так - другая. И только много месяцев спустя мы нашли период. Он оказался равным 16 дням. Притом, что масок всего 10. То есть внутри периода тоже бывают совпадения масок. Это и вводило нас в заблуждение. Разработчики же добились своего - выиграли некоторое время.

"О ПОЛЕ D ЗАМОЛВИТЕ СЛОВО..."

В этом абзаце мы расскажем о типах билетов и направлениях. Четыре типа билетов имеют два направления, и еще два типа не имеют направлений. Первые четыре - это "ПОЛНЫЙ", "ДЕТСКИЙ", "ЛЬГОТНЫЙ" и "БЕСПЛАТНЫЙ". Каждый из них может быть как "ТУДА" (->), так и "ТУДА И ОБРАТНО" (<->). Остались типы, которые идентифицируют себя просто цифрой "9" и словом "абонемент" :).

КОЧЕРЫЖКА ЦВЕТНОЙ КАПУСТЫ

Что же такое константы? Так как 3 константы в течение дня не изменяются, мы предположили, что это дата или функция от даты. Логично? Семантика четвертой константы остается вообще загадкой. На билетах она принимает за день всего два значения - "i" и "i + 1". Поэтому-то мы и зачислили по ошибке ее в ряды констант. Но на самом деле она может быть равна чему угодно, и валидность билета от этого не теряется.
А теперь поговорим о том, что нам еще предстоит открыть. Поля не хранятся в штрих-коде в явном виде. К каждому полю в маске прибавляется по модулю 10 определенное число, которое мы назвали смещением. Именно сумма реального значения поля и смещения записывается в штрих-код. Слава Богу, что смещения не меняются в течение дня. Таким образом, 16 смещений в маске можно представить как 16 функций от даты. Осталось их найти. Может, они периодичны? А, может, просто удастся их экстраполировать? В любом случае, имея билет на какое-то число, зная его реальные характеристики, маску и значения в штрих-коде, можно без труда, с помощью операции вычитания узнать все 16 смещений на сегодня. А потом, зная эти 16 чисел, изготовить билет с любыми характеристиками, но опять же только на сегодня.

АВТОНОМНЫЙ ТУРНИКЕТ

Как при этом турникет узнает, билет какого числа ему подсунули? Нам кажется, все происходит так. Билет попадает в турникет. Турникет предполагает, что билет сегодняшний. Смотрит в то место, где сегодня должны лежать 3 константы (дата), вычитает смещение. И если полученное значение не совпало с сегодняшней датой, значит его предположение неверно, и билет не сегодняшний. Далее он предполагает, что он вчерашний и т.д., пока не наткнется на совпадение.
Загадкой всей системы осталась функция смещения от даты. Но мы думаем, что это всего лишь вопрос времени. Так что дерзайте, "лед тронулся, господа присяжные заседатели..."
Особую благодарность выражаем Анатолию Абрамовичу Шалыто, который вдохновил нас на написание статей своими зажигательными лекциями о "Switch-технологии".

Ссылки:

1. http://tickers.da.ru (сайт, посвященный данной тематике)
2. tickers@yandex.ru (авторы)
3. http://www.sbarcode.com (сайт, посвященный штриховому кодированию). На этом же сайте можно найти демо-версию программы LabelBar для печати штрих-кодов
4. http://www.barcode.kiev.ua/history.html (история штрихового кодирования)