Completely Regular semigroups and cryptography and Stratified semigroups

Jim Renshaw

October 30, 2025

University of Southampton
School of Mathematics

Outline

Discrete Log

- Choose a large prime p and a residue e coprime to $p - 1$.
- Encode data using integers in $\mathbb{Z}_p$.
- Encrypt data using the function $x \mapsto x^e \mod p$.
- Decrypt using the function $x \mapsto x^f \mod p$ where $ef \equiv 1 \mod (p-1)$.

Discrete Log

- More generally let S be a group (or a ring) and let e be a unit modulo $|S|$.
- Encode data using elements from S.
- Encrypt data using the function $x \mapsto x^e$
- The value of x is the *plaintext*, e is the *(encryption) key* and x^e is the *ciphertext*.
- Decrypt using the function $x \mapsto x^{e^{-1}}$ where e^{-1} is the inverse of the unit e.
- The *discrete log problem* is the problem of determining e given both x and x^e.
- The usefulness of this system lies in the fact that we know of no efficient, non-quantum algorithms, to solve this particular *discrete log problem* - given x, x^e and S, calculate e.

UNIVERSITY OF Southampton
School of Mathematics

Discrete Log

Given *only* x^e, there is no point in trying to determine x or e.

There are $|S|$ candidates for x and $\phi(|S|)$ candidates for e. Hence we have to check $|S|\phi(|S|) \cong |S|^2$ values.

However, for any potential unit $f, x^e = \left((x^e)^{f^{-1}}\right)^f$.

If, on the other hand, we are given both x and x^e then there is a unique solution to the equation

$$x^e = x^f.$$

Discrete Log

- Two classic examples that are frequently used, particularly with encryption of websites:
- S is the group associated with an Elliptic Curve C;
- S is the ring $\mathbb{Z}_n$ where $n = pq$ and p and q are distinct primes - RSA cryptosystem.

UNIVERSITY OF Southampton
School of Mathematics

Discrete Log

- Although $\mathbb{Z}_n$ is a ring, we are only using one arithmetic operation and so we are effectively working with a semigroup.
- Can we use a more general type of semigroup?
- If S is a semigroup and we use exponentiation as above then we would expect

$$\left(x^e\right)^f = x.$$

- This means that the monogenic semigroup $\langle x \rangle$ has index 1 and so is a cyclic group.
- Hence S is a *completely regular* semigroup.

Discrete Log

For the ring $\mathbb{Z}_{pq}$ used with RSA, we have a completely regular semigroup structure

$$
\begin{tikzcd}
 & U_{pq} \arrow[dl, dash] \arrow[dr, dash] & \\
U_q \arrow[dr, dash] & & U_p \arrow[dl, dash] \\
 & 0 &
\end{tikzcd}
$$

Completely Regular semigroups

So we can't just use our favourite semigroup (unless these happen to be completely regular).

Given plaintext x, then all elements of S that we subsequently work with belong to $\langle x \rangle$, then our arithmetic is essentially restricted to a maximal subgroup of S.

Let us consider the structure of S in more detail.

Completely Regular semigroups

A completely regular semigroup S is a semilattice

$$S = \mathcal{S}[Y; S_e]$$

of completely simple semigroups S_e.

So x belongs to a completely simple semigroup, which we can view as a Rees Matrix semigroup

$$\mathcal{M}[G; I, \Lambda; P].$$

We can then equate x with an element of the form

$$x = (i, g, \lambda)$$

and so

$$x^e = \left(i, (gp_{\lambda i})^{e-1}\, g, \lambda\right).$$

Variants of Semigroups

- Variants were originally introduced by authors such as Lyapin, Magill, Chase in the 60s and 70s.
- John Hickey then published more general papers on variants of semigroups in the 1983 onwards and a number of more recent papers have subsequently appeared.
- If $(S, \cdot)$ is a semigroup and $s \in S$ then we can define a new multiplication, $*_s$, on S by

$$x *_s y = x \cdot s \cdot y.$$

  It is easy to check that this gives an associative operation, and so the system $(S, *_s)$ is a semigroup, referred to as a *variant of S* and often denoted by S^s.

Variants of Semigroups

If $(G, \cdot)$ is a group and $p \in G$ then $G^p = (G, *_p)$ is also a group, with identity p^{-1} and where the inverse of x, in G^p, is given by the element $p^{-1} \cdot x^{-1} \cdot p^{-1}$ in $(G, \cdot)$.

[p^{-1} and x^{-1} are the inverses of p and x in $(G, \cdot)$]

The map $(G, \cdot) \to (G, *_p)$ given by

$$x \mapsto x \cdot p^{-1}$$

is a group isomorphism.

Hence if $e \in \mathbb{N}$ then within G^p, the element g^e, the e^{th} power of the element g, is represented by the element

$$(gp)^{e-1}g$$

in $(G, \cdot)$.

Variants of Semigroups

The discrete log problem in this case would involve finding e given both g and $(gp)^{e-1}g$.

If we knew p we could of course compute $(gp)^e$ and we have the classic discrete problem over the original group.

Otherwise, it is a different story.

It is perfectly possible to solve the equation

$$(gp)^{e-1}\,g = (gq)^{f-1}\,g$$

in a non-trivial way.

If there are large numbers of such solutions then solving the discrete log problem is much harder/infeasible.

UNIVERSITY OF Southampton
School of Mathematics

Euler's totient function

The number of units modulo n is $\phi(n)$ where ϕ is Euler's totient function. It is well known that

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

where p runs through the prime divisors of n.

So for the classic case, if G is a group and $x \in G$ and you know that $x = g^e$ for some $g \in G$, and some unit e, then

$$x = \left(x^{f^{-1}}\right)^f$$

and there are $\phi(|G|)$ possible 'solutions'.

UNIVERSITY OF Southampton
School of Mathematics

Schemmel's totient

For our group variant problem, if we are given $x = (gp)^{e-1}g$ but we don't know p or e, how many solutions are there to the equation

$$(gp)^{e-1}\, g = (gq)^{f-1}\, g?$$

Schemmel's totient

Schemmel's totient number, $S_r(n)$ counts the number of consecutive terms $1 \leq m, m+1, \ldots, m+(r-1) \leq n$ which are all coprime to n, or in other words, the number of r consecutive units in $\mathbb{Z}_n$. It is easily shown that this function is also multiplicative and that

$$S_r(n) = n \prod_{p|n} \left(1 - \frac{r}{p}\right).$$

Schemmel's totient

In particular

$$S(n) = S_2(n) = n \prod_{p|n} \left(1 - \frac{2}{p}\right).$$

counts the number of units m such that $m - 1$ is also a unit.

Schemmel's totient

Suppose g, p and e are fixed and we wish to find solutions (q, f) to the equation

$$(gp)^{e-1} = (gq)^{f-1}\,.$$

Suppose first that $|G|$ is odd. We know that f has to be a unit, but if $f-1$ is also a unit and k is the inverse of $f-1$ then

$$\left((gp)^{e-1}\right)^{k}$$

provides us with a solution. Here $q = g^{-1}\left((gp)^{e-1}\right)^{k}$.

There are then at least $S(|G|)$ such solutions.

A new totient

Suppose now that $|G|$ is even. Then f must be odd and so $f - 1$ can't be a unit.

Notice that e is also odd and so

$$(gp)^{e-1} = h^2$$

for some $h \in G$.

Suppose that f is a unit such that $(f - 1)/2$ is also a unit with inverse k. Then

$$(h^k)^{f-1} = (h^k)^{2(f-1)/2} = (h^2)^{k(f-1)/2} = h^2 = (gp)^{e-1}$$

and so $q = g^{-1}h^k$ provides us with a solution.

So how many such solutions are there?

A new totient

Let

$$T(n) = |\{f | f \text{ and } (f-1)/2 \text{ are both units mod } n\}|$$

As an example, let $p = 2q + 1$ be a *safe* prime, where q is also prime. The prime q is often referred to as a *Sophie Germaine prime*.
If we encode our data from the group of units of $\mathbb{Z}_p$, then the number of solutions to our previous problem is $T(p-1)$.

A new totient

Theorem

If $p = 2q + 1$ is a safe prime and $G = U_p$ then

$$T(p-1) = \begin{cases} (p-3)/4 & q \equiv 1 \mod (4) \\ (p-7)/4 & q \equiv 3 \mod (4) \end{cases}.$$

UNIVERSITY OF Southampton
School of Mathematics

A new totient

To compute $T(2n)$ we must remove from the set of residues $R = \{1, \ldots, 2n\}$ numbers f of the form

1. $f = 2x$;
2. $f = xp$ where $p|n$ and x is odd;
3. $f = 1$ or $f = 1 + 2xp$ where $p|n$ and $\gcd(f, n) = 1$;
4. $f = 1 + 4x$ where $\gcd(x, n) = \gcd(f, n) = 1$.

UNIVERSITY OF Southampton
School of Mathematics

A new totient

In counting the values in (4), we eventually arrive at a simple diophantine equation of the form

$$1 + 4x = ry \leq 2n$$

where r is an odd divisor of n.

Counting the solutions for this equation is relatively easy except that there are 2 possible cases depending on whether r is congruent to 1 or 3 mod 4.

A new totient

Theorem

Let $n > 1$ be an odd integer with ω distinct prime divisors. Then

$$\left| T(2n) - \frac{S(n) - 1}{2} \right| \leq \frac{3^\omega - 2^\omega + 1}{2}.$$

If $n = q^m$ where q is prime and $m > 0$, then
$T(2n) = (S(n) - 1)/2$ when $q \equiv 3 \mod (4)$ and
$T(2n) = (S(n) + 1)/2$ when $q \equiv 1 \mod (4)$.

A new totient

For the specific example that we gave where $2n = p - 1 = 2q$, the situation is even more interesting.

Proposition

If $G = U_p$ where $p = 2q + 1$ is a safe prime, then there are $p - 5 = 2(q - 2)$ solutions.

UNIVERSITY OF Southampton
School of Mathematics

Proof

- Each unit $f > 1$ with the property that $(f-1)/2$ is also a unit, provides a solution, w say.
- Notice that in this case, $-w$ is also a solution.
- If $(f-1)/2$ is not a unit it is because it is even, in which case $(f-1)/4$ may be a unit.
- If not then $(f-1)/4$ is even and so $(f-1)/8$ may be a unit.
- Continuing in this fashion we see that there is a positive integer m such that $(f-1)/2^m$ is a unit.

Proof

- Now it is well known that if $p \equiv 3 \mod (4)$ is a prime and if $y \in \mathbb{Z}_p$ then either y or $-y$, but not both, has a square root modulo p (the square root is in fact $y^{(p+1)/4}$).
- If $c = h^2$ then either h or $-h$ will have a square root, h_1 say, and so $c = h_1^4$.
- But then either h_1 or $-h_1$ will have a square root, h_2 say, and so $c = h_2^8$.
- Continuing in this fashion we see that $c = h_{m-1}^{2^m}$.
- So if k is the multiplicative inverse of $(f-1)/2^m$ then $w = \pm h_{m-1}^k$ will be solutions
- Hence every unit $f > 1$ in $\mathbb{Z}_p$ provides two solutions and since $|U_{p-1}| = q - 1$ the result follows.

A new totient

So use of a completely simple semigroup rather than a group would appear to give more protection from ‘brute force’ attacks.

A new totient

If $n = 2^e m$ for $e \geq 0$ and $m > 1$ is odd and if ω is the number of distinct prime factors of m, then

- $T(1) = T(2) = 0$ and $T(2^e) = 2^{e-2}$ for $e \geq 2$;
- $T(n) = 2^{e-2} S(m)$ for $e \geq 2$;
- $\left| T(n) - \frac{S(m)-1}{2} \right| \leq (3^\omega - 2^\omega + 1)/2$ for $e = 1$;
- $\left| T(n) - \frac{S(m)-1}{2} \right| \leq (3^\omega - 2^{\omega+1} + 1)/2$ when $e = 0$;
- $T(n) = (S(m) - 1)/2$ when $e = 0$ and $\omega = 1$;
- $T(n) = (S(m) - 1)/2 \pm 1$ when $e = 0$ and $\omega = 2$.

A new totient

In practice, there are certain *known plaintext attacks* that can sometimes reduce this discrete log problem to the more classic case and choosing a random value for p for each encryption will help mitigate this.

A new totient

For example, choose $e \in U_n$ and $s \in \mathbb{Z}_m$ for some $m \geq n$ and let the pair (s, e) be the secret key.

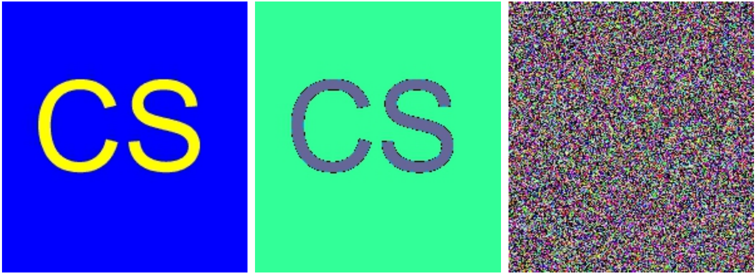
Then given plaintext $g \in G$, let i be a random value in $\mathbb{Z}_m$ and define $p_i = H(i \oplus s)$ where $\oplus$ is the bitwise XOR operator and H is a suitably chosen cryptographic hash function whose image coincides with G.

The ciphertext is then the pair $(i, (gp_i)^{e-1}g)$, and anyone with access to the key, can replicate p_i and decrypt.

However, even if an attacker can identify the correct value of p_i amongst all the $T(n)$ or more solutions , it will be relatively ineffective as we shall use a different value of p_i for each encryption.

UNIVERSITY OF Southampton
School of Mathematics

A new totient



(a) no encryption (b) group (c) semigroup

Figure: discrete log encryption on similar blocks

$\mathbb{Z}_{pq}$

For the ring $\mathbb{Z}_{pq}$ used with RSA, we have a completely regular semigroup structure

$$\begin{tikzcd}[column sep=small] & U_{pq} \arrow[dl, dash] \arrow[dr, dash] & \\ U_q \arrow[dr, dash] & & U_p \arrow[dl, dash] \\ & 0 & \end{tikzcd}$$

$\mathbb{Z}_n$

Let m be a positive integer and let $p_1, \ldots, p_m$ be distinct primes and let $I = \{1, \ldots, m\}$. Let $n = \prod_{i \in I} p_i$ and for any non-empty subset $S \subseteq I$, let $n_S = \prod_{i \in S} p_i$ and denote by $\overline{S} = I \setminus S$, so that $n = n_{\overline{S}} n_S$. Define

$$U_S = \{n_{\overline{S}} x | x \in U_{n_S}\} = n_{\overline{S}}(U_{n_S}, *, n_{\overline{S}}) \cong U_{n_S},$$

where $U_{n_S} = \{1, \ldots, n_S - 1\}$ is the group of units modulo n_S and let $U_{\emptyset} = \{0\}$.

$\mathbb{Z}_n$

Theorem

For any non-empty subset $S \subseteq I$, U_S is a subgroup of the multiplicative semigroup $\mathbb{Z}_n$, and is isomorphic to U_{n_S}. Moreover

$$\mathbb{Z}_n = \dot{\bigcup}_{S \subseteq I} U_S$$

is a strong semilattice of groups, $\mathcal{S}[Y; U_S]$ in which Y is the boolean algebra $\mathcal{P}(I)$.
The structure maps are given by $\phi_T^S : U_S \to U_T$ for $T \subseteq S \subseteq I$

$$\phi_T^S(x) = \left(n_{\overline{T}}\right)^{-1} x$$

where $\left(n_{\overline{T}}\right)^{-1}$ is the inverse of $n_{\overline{T}}$ in U_{n_T}.

Non square-free case

Joint work with Will Warhurst.

Suppose now that

$$n = p_1^{e_1} \dots p_m^{e_m}$$

with each $p_i > 0$.

What does $\mathbb{Z}_n$ look like in this case?

Non square-free case

Joint work with Will Warhurst.

Suppose now that

$$n = p_1^{e_1} \dots p_m^{e_m}$$

with each $p_i > 0$.
What does $\mathbb{Z}_n$ look like in this case?

Theorem

If $n = p_1^{e_1} \dots p_m^{e_m}$ where each p_i is prime and each $m_i > 0$, then $\mathbb{Z}_n$ is a semilattice of stratified extensions of groups

$$\mathbb{Z}_n = \mathcal{S}[\mathcal{P}(I); R_e]$$

where $I = \{1, \dots, m\}$ and each R_e is a stratified extension of a group.

Non square-free case

Define the *base* of a semigroup S to be the subset

$$\mathrm{Base}(S) = \bigcap_{m>0} S^m.$$

If $\mathrm{Base}(S) = \{0\}$ or $\mathrm{Base}(S) = \emptyset$ then Grillet called this a

stratified semigroup.

A semigroup S is then called a *stratified extension* of $\mathrm{Base}(S)$ if $\mathrm{Base}(S) \neq \emptyset$.
The name signifying the fact that in this case S is an ideal extension of $\mathrm{Base}(S)$ by a stratified semigroup with zero.

Non square-free case

Let S be a stratified extension of $\mathsf{Base}(S)$.

- The *layers* of S are defined to be the sets $S_m = S^m \setminus S^{m+1}$, $m \geq 1$;
- Every element of S lies either in the base of S or in exactly one layer of S, and if $s \in S_m$ then m is the *depth* of s.
- If S has finitely many layers then the numbers of layers is called the *height* of S.
- The layer S_1 generates every element of $S \setminus \mathsf{Base}(S)$ and is contained in any generating set of S.

University of Southampton School of Mathematics

Non square-free case

Proposition

Suppose that S is a semigroup.

1. $\mathrm{Reg}(S) \subseteq \mathrm{Base}(S)$. *Hence if S is regular,* $\mathrm{Base}(S) = S$.
2. $E(S) = E(\mathrm{Base}(S))$.
3. *If $s \in S \setminus \mathrm{Base}(S)$ then $|J_s| = 1$, where J_s is the $\mathcal{J}-$class of s.*

Non square-free case

Proposition

Let T and R be any semigroups.Then there exists a stratified extension S such that $T \subseteq \mathsf{Base}(S)$ and $S/T \cong R$. Moreover, if R is stratified without a zero then $T = \mathsf{Base}(S)$.

Non square-free case

Let $n = p_1^{e_1} \dots p_m^{e_m}$ where each p_i is prime and each $m_i > 0$.

$$\mathbb{Z}_n = \mathcal{S}[\mathcal{P}(I); R_e]$$

Let $e = \prod_{i \in K} p_i^{e_i}$ and let $R_e = \{x \in \mathbb{Z}_n | x^m = e \text{ for some } m\}$.

This is a stratified extension of $U_{n/e}$ where if $x \in R_e$ is in the i^{th} layer then

$$x = \prod_{j \in K} p_j^{g_j} u$$

where $u \in U_n$ and $0 < g_j \leq e_j$ and $\min\{g_j | g_j \neq e_j\} = i$.

Non square-free case

As an example, if $n = 12 = 2^2 \times 3$, then

$$\mathcal{P}(I) = \{\{2,3\},\{2\},\{3\},\emptyset\}$$

and we have four subsemigroups
$R_{\{2,3\}} = \{6,12\}$ where $\mathsf{Base}(R_{\{2,3\}}) = \{12\}$.
$R_{\{2\}} = \{2,4,8,10\}$ where $\mathsf{Base}(R_{\{2\}}) = \{4,8\}$ and $\{2,10\}$ forms layer 1.
$R_{\{3\}} = \{3,9\}$ which is a group.
$R_{\emptyset} = \{1,5,7,11\}$ which is the group of units mod 12.

Non square-free case

The semilattice structure can be pictured as