Quantificando Vazamento de Informação sobre Estratégias

Mário S. Alvim[1], **Piotr Mardziel**[2], **Michael Hicks**[2]

[1] Universidade Federal de Minas Gerais
msalvim@dcc.ufmg.br

[2]University of Maryland, College Park
{piotrm,mwh}@cs.umd.edu

Abstract. *The field of quantitative information flow concerns the rigorous mathematical assessment of the amount of secret information leaked by computational systems. We report first steps towards a formal model for* strategic leakage. *We generalize the representation of prior adversarial knowledge from a distribution on secrets to a distribution on* strategies *for generating secrets, which we call an* environment. *Applying information-theoretic techniques to environments allows us to disentangle information leakage about a secret from leakage about how users generate secrets, i.e., their* strategy.

Resumo. *Este resumo reporta primeiros passos na formalização do conceito de* vazamento de estratégias. *Nós generalizamos a representação do conhecimento prévio de um adversário de uma distribuição de probabilidade sobre os segredos para uma distribuição sobre* estratégias *que geram segredos, chamando esta nova distribuição de* ambiente. *Aplicando técnicas de teoria da inforamção a ambientes, conseguimos distinguir o vazamento de informação sobre segredos do vazamento de informação sobre a maneira como segredos são gerados, i.e., sobre a* estratégia.

1. Introdução

O campo de *fluxo de informação quantitativo* (QIF, do inglês *quantitative information flow*) diz respeito à mensuração da informação sigilosa vazada por sistemas computacionais. Baseando-se no rigor matemático da teoria da informação, dois princípios fundamentais de QIF são: (i) um segredo é considerado "seguro" na medida em a distribuição de probabilidade sobre os valores sigilosos tem alta entropia; e (ii) o vazamento de informação de um sistema é uma medida de quanto o comportamento observável do sistema, durante o processamento de um valor sigiloso, degrada a entropia desse segredo. Estes princípios têm sido usados para criar modelos de QIF cada vez mais sofisticados para representar sistemas e medir vazamentos. No entanto, pouca atenção tem sido dedicada a entender as fontes que injetam a entropia inicial na distribuição sobre os segredos.

Muitas vezes, o processo de geração de segredos não é completamente aleatório. Por exemplo, usuários podem ter uma "estratégia" para escolher senhas que seja parcialmente previsível (por exemplo, usando sua data de nascimento), e um adversário pode usar isso para prever senhas futuras. Estudos têm mostrado que forçar usuários a alterar frequentemente suas senhas pode não ser tão benéfico quanto se acreditava [Zhang et al. 2010, Chiasson and van Oorschot 2015]. Além disso, um segredo que varia no tempo e sofre repetidas observações, em alguns casos, pode ser aprendido *mais rapidamente* se for alterado (e observado) mais frequentemente [Mardziel et al. 2014].

Neste resumo investigamos em que medida o conhecimento sobre *estratégias* ajuda o adversário a inferir o segredo. Desenvolvemos um modelo que não considera apenas o espaço de segredos possíveis, mas também o espaço de possíveis estratégias que os geram. Nosso modelo nos permite considerar adversários capazes de aprender a estratégia e quantificar a vantagem que eles obtêm, desacoplando a medida do valor de se aprender a estratégia da medida do valor se aprender o segredo. No resto deste resumo, apresentamos o nosso modelo inicial e caracterizações de vazamento de estratégias.

2. Preliminares

Faremos uma breve revisão de conceitos fundamentais de *fluxo de informação quantitativo* (QIF). Um *adversário* tem apenas informação parcial sobre o valor de um *segredo*, modelada por uma distribuição de probabilidade chamada de *prior*. Denotamos por $\mathcal{X}$ o conjunto de segredos possíveis e por $\mathbb{D}\mathcal{X}$ o conjunto de distribuições de probabilidade sobre $\mathcal{X}$. Normalmente usamos p_X para denotar uma *prior*. Uma medida de informação é uma função $\mathbb{V}{:}\mathbb{D}\mathcal{X}{\to}\mathbb{R}$ com imagem real. Uma medida de informação pode medir *vulnerabilidade* - quanto maior o valor, menos seguro é o segredo - ou *incerteza/entropia* - quanto maior o valor, mais seguro é o segredo. Definições de medidas de informação na literatura variam de acordo com a interpretação operacional da medida. Recentemente, o arcabouço de *g-vulnerability* [Alvim et al. 2012] foi proposto para capturar vários modelos de adversário. Neste artigo vamos utilizar o termo "vulnerabilidade" para medidas em geral, embora as nossas reivindicações também se aplicam a incerteza medidas. Uma *hiper-distribuição* [McIver et al. 2014] (ou *hiper*) é uma distribuição de probabilidade sobre distribuições. Uma hiper sobre o conjunto $\mathcal{X}$ é do tipo $\mathbb{D}^2\mathcal{X}$, significando $\mathbb{D}(\mathbb{D}\mathcal{X})$. Usamos Δ para denotar uma hiper, $\lceil\Delta\rceil$ para seu *suporte* (o conjunto de valores com probabilidade não nula), e $[p_X]$ para hiper puntual associando probabilidade 1 para p_X.

3. Estratégias e sua vulnerabilidade

Esta seção apresenta um modelo que explicita que segredos são gerados pela aplicação de uma estratégia particular, e apresenta generalizações de vulnerabilidade correspondentes.

3.1. Vulnerabilidade contextual do segredo

Uma *estratégia* é uma regra probabilística para a geração de segredos, representada por uma distribuição de probabilidade sobre o conjunto $\mathcal{X}$ de segredos. O conjunto $\mathcal{S}$ de todas as estratégias é, portanto, $\mathbb{D}\mathcal{X}$. Modelamos o *ambiente* em que o segredo é gerado como uma distribuição p_S sobre o conjunto $\mathcal{S}$ de estratégias. O conjunto $\mathbb{D}\mathcal{S}$ de todos os ambientes é o conjunto das distribuições de probabilidade sobre estratégias, ou seja, o conjunto $\mathbb{D}^2\mathcal{X}$ de hipers sobre $\mathcal{X}$. O ambiente representa o espaço de onde estratégias são amostrados. Assumimos que a *prior* p_X sobre segredos é *consistente* com o ambiente p_S, isto é, que a *prior* é obtida como o comportamento esperado do ambiente na falta qualquer conhecimento mais refinado sobre estratégia utilizada: $p_X{=}\,\mathrm{E}\,p_S$. Caso haja uma única estratégia p_X para gerar segredos, o ambiente p_S é a hiper puntual $[p_X]$.

Enquanto em modelos tradicionais de QIF o adversário só sabe a *prior* p_X, consideramos aqui também adversários que conhecem o ambiente p_S. A vulnerabilidade do segredo para este tipo adversário mais poderoso é definida a seguir.

Definition 1. Dada uma medida de informação $\mathbb{V}$, a *vulnerabilidade contextual (do segredo)* é uma função $\mathbb{V}_C{:}\mathbb{D}^2\mathcal{X}{\to}\mathbb{R}$ definida como $\mathbb{V}_C(p_S)\stackrel{\text{def}}{=}\mathrm{E}_{p_S}\,\mathbb{V}$ representando a expectativa da vulnerabilidade do segredo quando o ambiente é conhecido pelo adversário.

Nosso primeiro resultado é que um adversário que conhece o ambiente nunca está em desvantagem em relação a um adversário que só conhece a *prior* sobre os segredos.
Proposition 1. *Se* $\mathbb{V}$ *é uma g-vulnerability, então* $\mathbb{V}_C(p_S) \geq \mathbb{V}(p_X)$.

Além disso, no caso de o ambiente p_S ser uma hiper puntual $[p_X]$, a vulnerabilidade contextual $\mathbb{V}_C(p_S)$ se reduz a $\mathbb{V}(p_X)$.
Proposition 2. *Se* $p_S = [p_X]$, *então* $\mathbb{V}_C(p_S) = \mathbb{V}(p_X)$.

3.2. Segurança real e segurança por agregação

A modelagem do conhecimento do adversário como apenas uma *prior* p_X ignora como o adversário pode explorar o conhecimento do ambiente p_S para inferir segredos. Demonstramos este fato com um exemplo em que segredos distribuídos de acordo com uma mesma *prior* apresentam vulnerabilidades contextuais drasticamente diferentes.

Sejam $\mathcal{X} = \{x_1, x_2\}$ um conjunto de valores secretos binários, e $\mathcal{S} = \{s_1, s_2, s_3\}$, um conjunto de possíveis estratégias, onde $s_1 = [1, 0]$ sempre gera o valor x_1, $s_2 = [0, 1]$ sempre gera o valor x_2 e $s_3 = [1/2, 1/2]$ gera cada valor com igual probabilidade. Considere dois ambientes $p_S^1 = [1/2, 1/2, 0]$, em que as estratégias s_1 e s_2 pode ser adotadas com igual probabilidade, e $p_S^2 = [0, 0, 1]$, em que a estratégia s_3 é sempre adotada. Retratamos tais estratégias e ambientes na Tabela 1. Colunas listam estratégias; o primeiro agrupamento de linhas contém a definição da estratégia (isto é, a probabilidade de cada segredo ser escolhido), o agrupamento seguinte de linhas contém a definição de cada ambiente, um por linha, fornecendo a probabilidade de cada estratégia. Ambos os ambientes produzem a mesma *prior* $p_X = \mathrm{E}\, p_S^1 = \mathrm{E}\, p_S^2 = [1/2,\ 1/2]$, assim um adversário que não pode observar a estratégia utilizada obteria a mesma vulnerabilidade prévia em ambos os ambientes. Por exemplo, para a vulnerabilidade de Bayes $\mathbb{V}^{\text{Bayes}}(p_X) \stackrel{\text{def}}{=} \max_{x \in \mathcal{X}} p_X(x)$, o adversário obteria uma vulnerabilidade de $\mathbb{V}^{\text{Bayes}}(p_X) = 1/2$. No entanto, um adversário que possa aprender a estratégia utilizada obterá diferentes vulnerabilidades do segredo em cada ambiente. No ambiente p_S^1 a vulnerabilidade contextual é $\mathbb{V}_C^{\text{Bayes}}(p_S^1) = 1/2 \cdot p(s_1) + 1/2 \cdot p(s_2) = 1/2 \cdot 1 + 1/2 \cdot 1 = 1$, enquanto no ambiente p_S^2 a vulnerabilidade contextual é $\mathbb{V}_C^{\text{Bayes}}(p_S^2) = 1 \cdot p(s_3) = 1 \cdot 1/2 = 1/2$. Note que no ambiente p_S^2, o valor para a vulnerabilidade contextual e vulnerabilidade prévia é o mesmo ($\mathbb{V}_C^{\text{Bayes}}(p_S^2) = \mathbb{V}^{\text{Bayes}}(p_X) = 1/2$), assim um adversário que aprende a estratégia utilizada não terá vantagem sobre um adversário que só conhece a *prior*. Por outro lado, no ambiente p_S^1, a vulnerabilidade contextual excede a vulnerabilidade prévia ($\mathbb{V}_C^{\text{Bayes}}(p_S^1) = 1 > 1/2 = \mathbb{V}^{\text{Bayes}}(p_X)$), e um adversário que aprende a estratégia utilizada obterá sucesso duas vezes mais frequentemente que um adversário que só conhece a *prior*.

	S_1	s_2	s_3
x_1	1	0	$1/2$
x_2	0	1	$1/2$
p_S^1	$1/2 e 1/2 e 0$		
p_S^2	0	0	1

Tabela 1.

Podemos obter dois *insights* a partir deste exemplo. Em primeiro lugar, *segurança por agregação* ocorre quando a vulnerabilidade contextual excede a vulnerabilidade prévia: $\mathbb{V}_C(p_S) \gg \mathbb{V}(p_X)$. Neste caso, o segredo está protegido pela falta de conhecimento por parte do adversário sobre a estratégia utilizada, e se o adversário aprender a estratégia a vulnerabilidade do segredo pode aumentar significativamente. Um exemplo de segurança por agregação é um cenário em que todos usuários escolhem senhas com estratégias determinísticas, mas o adversário não sabe qual usuário está gerando a a senha. Se houver um grande número de usuários, e se são as suas estratégias são variada o suficiente, as senhas podem ser considerados "seguras" apenas no sentido de que o adversário não pode usar o conhecimento sobre o ambiente para identificar a estratégia utilizada.

Por outro lado, *segurança real* ocorre quando vulnerabilidades contextual e prévia têm valores semelhantes: $\mathbb{V}_C(p_S) \approx \mathbb{V}(p_X)$. Neste caso, o segredo está protegido pela imprevisibilidade (ou incerteza) no âmbito das estratégias que geram o segredo, e mesmo que a estratégia se torne conhecida pelo adversário, a vulnerabilidade do segredo não irá aumentar significativamente. Um exemplo de segurança real é um sistema de banco em que senhas são escolhidas automaticamente para cada usuário de modo uniforme. Mesmo que o algoritmo seja conhecido do adversário, a vulnerabilidade do segredo não aumenta.

3.3. Vulnerabilidade da estratégia

Consideramos agora como o conhecimento de um ambiente afeta o conhecimento do adversário sobre a estratégia a ser utilizada, definindo uma medida $\mathbb{V}_S(p_S) : \mathbb{D}\mathcal{S} \to \mathbb{R}$ da *vulnerabilidade da estratégia*. A ideia central é que $\mathbb{V}_S(p_S)$ deve refletir a "semelhança" entre as estratégias no suporte de p_S. Do ponto de vista do adversário, duas estratégias devem ser consideradas "semelhantes" se levarem a "semelhantes" vulnerabilidades do segredo, conforme medido de acordo com uma $\mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}$ de interesse. Assim, a vulnerabilidade de um ambiente deve ser mais elevada quando o seu suporte consistir em estratégias semelhantes, e mais baixa quando seu suporto consistir em estratégias muito diferentes. Uma estratégia é, *para fins práticos*, conhecida pelo adversário quando $\mathbb{V}(p_X) \approx \mathbb{V}_C(p_S)$, ou, equivalente, quando $\mathbb{V}(\mathrm{E}\, p_S) \approx \mathrm{E}_{p_S}\, \mathbb{V}$.

Definition 2. Se $\mathbb{V}$ é uma medida de vulnerabilidade, então a *vulnerabilidade da estratégia* é a relação $\mathbb{V}_S(p_S) \stackrel{\text{def}}{=} {}^{\mathbb{V}(\mathrm{E}\, p_S)}/_{\mathbb{V}_C(p_S)}$. A razão é invertida para medidas de incerteza.

Note que temos sempre $0 \leq \mathbb{V}_S(p_S) \leq 1$ (pela Proposição 1), e $\mathbb{V}_S(p_S)$ é máxima quando $\mathbb{V}(\mathrm{E}\, p_S) = \mathbb{V}_C(p_S)$. Além disso, note que a definição é consistente com a de decomposição vulnerabilidade prévia no produto de vulnerabilidade estratégica e vulnerabilidade contextual: $\mathbb{V}(p_X) = \mathbb{V}_S(p_S) \cdot \mathbb{V}_C(p_S)$. Este é o desacoplamento que buscávamos.

4. Conclusão

Atualmente estamos trabalhando na caracterização de níveis intermediários de conhecimento do adversário, em que seu conhecimento é mais refinado do que uma *prior*, mas não tão refinado quanto o conhecimento completo do ambiente.

Referências

[Alvim et al. 2012] Alvim, M. S., Chatzikokolakis, K., Palamidessi, C., and Smith, G. (2012). Measuring information leakage using generalized gain functions. In *Proceedings of the IEEE Computer Security Foundations Symposium (CSF)*.

[Chiasson and van Oorschot 2015] Chiasson, S. and van Oorschot, P. C. (2015). Quantifying the security advantage of password expiration policies. *Journal of Designs, Codes, and Cryptography*, 77(2-3):401–408.

[Mardziel et al. 2014] Mardziel, P., Alvim, M. S., Hicks, M., and Clarkson, M. (2014). Quantifying information flow for dynamic secrets. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*.

[McIver et al. 2014] McIver, A., Meinicke, L., and Morgan, C. (2014). Compositional closure for Bayes risk in probabilistic noninterference. In *Proc. ICALP'10*.

[Zhang et al. 2010] Zhang, Y., Monrose, F., and Reiter, M. K. (2010). The security of modern password expiration: an algorithmic framework and empirical analysis. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*.