Optimization of Access Control Policies for Cross-Domain Marine Data Sharing from the Perspective of Dynamic Data Tracing

Zhoujin Deng[1] Shunmiao Zhang[2]

1. Fujian ZhongXin Network Security Information Technology Co, Ltd, Fuzhou, Fujian, 350001, China
2. Fujian University of Technology, Fuzhou, Fujian, 350000, China

Abstract

During cross-domain transmission, marine observation data are continuously derived, aggregated, and redistributed. However, current access control primarily relies on dataset identifiers and pre-defined sensitivity levels for judgment; once authorization is issued, it becomes impossible to restrict the evolution of the derivation chain. From the perspective of dynamic data tracing, this paper constructs a dynamic tracing graph by treating derivation paths as real-time inputs at the decision moment, and proposes a sensitivity calculation method for data flow that incorporates cross-domain amplification, desensitization attenuation, and multi-source fusion emergence increments. Furthermore, it integrates data flow sensitivity, downstream diffusion surface, entity historical compliance, and situational risk into a risk function to form a risk-adaptive access control model driven by data tracing. It also designs an incremental update algorithm that only recomputes the "influence domain," along with conservative rollback rules.

Keywords

marine data; cross-domain sharing; access control; data provenance; risk-adaptive; dynamic policy adjustment

数据动态溯流视角下海洋数据跨域共享的访问控制策略优化

邓宙锦[1] 张顺淼[2]

1.福建中信网安信息科技有限公司，中国·福建 福州 350001
2.福建理工大学，中国·福建 福州 350000

摘要

海洋观测数据在跨域传输过程中会被不断地派生、聚合以及再分发，但是目前的访问控制主要依靠对数据集标识以及预先标注的敏感级别进行判断，并且授权发出之后就无法限制派生链的发展。本文以数据动态溯流的角度出发，把派生路径看作判决时刻的实时输入来构建动态溯流图，然后提出包含跨域放大、脱敏衰减和多源融合涌现增量的流转敏感度计算方法；进而将流转敏感度、下游扩散面、主体历史合规度以及情境风险都纳入风险函数之中，形成溯流驱动的风险自适应访问控制模型，并且设计出只重算“影响域”的增量更新算法和保守回退规则。

关键词

海洋数据；跨域共享；访问控制；数据溯源；风险自适应；策略动态调整

1 引言

潮位资料跨部门调用一般不是一次性完成的，浮标温盐剖面先交由高校进行同化处理后作为产物应用于区域环境评估当中，再将得到的栅格图层传给海事平台，在此基础上与航道、水深数据叠加在一起，虽然各个环节都具有合法授权，但是从整体上看，敏感信息可能会通过派生链向外传播到申请用途之外的系统中，授权的时候数据还没有生成，交付之后派生的产品也无法再受到原授权的限制，这是本文所要讨论的问题。

从制度上来说还是有约束的。《海洋观测资料管理办法》第十四条用共享协议来确定共享的内容和方法，并且禁止超出国别用途或范围使用、禁止擅自扩散；两条规则把“用途”写进授权要件里，但是没有规定用途怎样在派生链上被持续绑定的技术路线。GB/T 43697—2024给出了分类分级的原则、框架以及重要的数据识别指南[4]，但不适用于涉及国家秘密的数据和军事数据，部分涉海资料可能会属于该例外

【基金项目】榕科（2023）118号，科技重大项目（项目编号：2022-ZD-022）。

【作者简介】邓宙锦（1988-），男，中国福建福州人，本科，工程师，从事网络与数据安全研究。

范围。

访问控制的研究脉络比较清楚：李凤华等人对自主访问控制、强制访问控制、基于角色的访问控制以及跨域访问控制和基于安全属性的访问控制等进行了梳理，并指出细粒度多级安全以及策略可伸缩调整是未来的发展方向[1]；周秀等人提出D-RBAC，用角色映射来解决海洋云平台多级多域权限转换的问题[2]；Kandala等人将业务需要、风险和情境因素引入到属性基决策中，并提出了一个框架[6]；Simmhan等把数据溯源定义为从原始来源开始的数据产品自生成后的派生历史[5]。

上述的研究多以被保护的客体在做出决策的时候已经给定、敏感级别预先标注并且比较稳定为前提，而派生链场景却使该前提无法成立，目前的研究存在三个方面的问题：一是派生品不能继承母本所受的用途限制；二是跨域跳数、下游扩散以及异质融合没有被纳入到判定中来；三是策略调整需要人工重新配置或者全部重算，本文提出动态溯流图和敏感度传导规则，构建包含流转状态判决的风险自适应模型及增量更新算法，并用合成算例来验证内部一致性以及相对性能。本文的工作内容主要有：①动态溯流图的定义以及跨域放大、脱敏衰减和多源融合涌现增量流转敏感度计算；②构建风险函数包括流转敏感度、下游扩散面、主体历史合规度、情境风险以及三级决策策略；③设计出只更新影响域的增量算法以及保守回退规则，并用合成算例来对比两者相对性能。

2 动态溯流图与敏感度传导

2.1 从溯源到溯流

溯源回答“数据来自何处”，一般用于事后的审计；溯流则把派生链作为判决时刻的输入来询问，“现在这串信息流向了什么地方以及将来还能流往什么地方”。两者可以共享元数据，但是消费的时间点不一样：前者是在争议之后进行检索的时候用到的，后者则是每次请求到达之前计算，所以流转记录不光是日志，还具有增量更新的能力，并且是以图的方式存在。

定义1（动态溯流图）时刻t时的溯流图记作$G(t) = (V(t),E(t),\Lambda)$，其中V为数据对象节点，包含原始观测、中间派生集以及信息产品；$E \subseteq V\times V\times A\times T\times\{0,1\}$表示流转事件；$e=(u,v,a,t,d)$，t时刻由u通过操作a生成或交付给v，d=1表示跨域；A={原样转发、加密导出、聚合、脱敏、融合}。Λ 提供固有敏感等级 $L(v)\in\{1,2,3,4\}$、要素门类、空间范围、采样海域和用途标签，四级刻度为算例所设定的，不是国家标准级别的映射。

2.2 流转敏感度

设 $P(v_k)$ 为以 v_k 为终点的溯流路径集，$P=(v_0\rightarrow\cdots\rightarrow v_k)$。流转敏感度定义为：

$$S(vk, t) = \min\{ Lmax, \max P \in (v_k) [L(v0P) \cdot \prod i=1k \gamma(di) \cdot \varphi(ai)] + \Delta(vk, t) \}$$

其中$\gamma(0)=1$、$\gamma(1)=\gamma>1$ 跨域放大因子；$\phi(a)\in(0,1]$ 是操作衰减系数，算例取原样转发1.00、加密导出0.90、聚合0.75、脱敏0.60；Lmax=4 用于截断，取所有的入路径的最大值的原因是，如果存在一个高敏感的来源链，那么就不用考虑低敏感的来源对它造成的影响了，归一化记为s=S/4。

Δ 用于处理路径累乘难以解释的情况：温盐、声速剖面单独看大多是普通数据，但是同海域高精度海底地形、水下声场和长时间的水文要素结合之后，可以反演出来单份资料无法得出的信息，本文把它叫做涌现敏感增量取。

$$\Delta(v, t) = \lambda \log2(m + 1) \kappa(P)$$

其中，m表示异质门类的数量，κ代表元素的互补程度，κ与所选领域的判断有关，并且是该模型里具有很强主观性的变量，在第四章也作了一定的论述。

3 溯流驱动的策略优化模型

3.1 判决属性与风险函数

把请求记作q=(u,o,a,c,t)，依次代表主体、客体、操作、情境和时间，除了主体域、角色、客体门类及其用途标签之外，模型还保存了流转敏感度s(o,t)、下游可达规模B(o,t)=|Reach(o,t)|/|V|以及主体历史合规度Hu∈[0,1]这三个溯流状态量，风险函数是：

$$R(q, t) = w1\ s(o, t) + w2\ (1 - Hu) + w3\ E(c, t) + w4\ B(o, t) - w5\ N(c, t)$$

E 表示情境风险，N 代表业务必要性，应急场景选为最高值，负号项表示 RAdAC 中运营上可以抵消一部分风险的观点[6]，但是抵消有上限：N 不得使 R 下降到 $w_1 s$ 的一半以下，以免应急标记被当作通用绕过渠道使用，算例权重取 $(w_1,\cdots,w_5)=(0.38,0.14,0.12,0.16,0.20)$。

3.2 三级判决与策略表达

$\pi(q)\in\{allow,deny\}$，其中$\theta_1=0.42$，$\theta_2=0.68$，当$\pi=deny$或$R>\theta_2$时，判定为Deny；当$\pi=allow$且$R\leqslant\theta_1$时，判定为Permit；当$\pi=allow$且$\theta_1<R\leqslant\theta_2$时，判定为Step-up，Step-up能够组合以下操作：降低精度、嵌水印、减小令牌TTL、二次确认，并且声明用途写进派生节点标签中。策略以属性表达式而不是“主体×客体”矩阵的形式存储，例如：

IF o.门类∈{海底地形地貌，海洋声学} AND s(o)≧0.60 AND u.域≠o.属主域 AND c.应急≡false THEN Step-up（降精度，TTL=15 min，水印标识=0）WITH 用途=声明用途 ∧ 禁止再分发。

S(o) 的计算会在运行过程中进行，并且根据流的流转情况而变化，因此无需对每一个子集都编写对应的规则来处理。

3.3 增量更新算法

全量重算整个地图的敏感度不能接受，一个事件只会对下游可到达的子图产生影响，并且这种影响会沿着路径衰减，因此定义影响域 I(e)：以 v 为根、深度不超过 h、沿途 $|\Delta S| \geq \varepsilon$ 的可达节点集，小于 ε 即剪枝，具体做法是怎样的。

（1）追加写入事件 e 与节点标注，形成可审计的追加式流水。

（2）自 v 起进行受限广度优先遍历，按 h 与 ε 确定 I(e)。

（3）重算 I(e) 内节点的 S 与 B，其余节点保持缓存值。

（4）仅对状态量变化超过 ε 的节点重绑策略谓词中的溯流参数。

（5）对仍有效且落在 I(e) 内的令牌按新 R 复判：$R>\theta_2$ 立即撤销；$\theta_1<R\leqq\theta_2$ 降级为Step-up。

（6）输出判决差异与审计记录。

复杂度由 $O(|V|+|E|)$ 降为 $O(|I(e)|\cdot\bar{h})$。3.3节显示，h 取3后改善趋缓。

3.4 冲突消解与保守回退

多域策略按照三条优先级消解：法定禁止及属主显式deny优先；用途绑定优先于权限放宽；同层冲突依照风险从严的原则处理，Step-up不能被宽松规则覆盖，更主要的是图不完整时：如果某个区域没有上报流转事件的话，溯流链就会断裂，在已经观测到的部分算出的 S 就会被系统性地低估，此时对缺失段采用上界 $\gamma^{h}L_{max}$ 进行估计，并且做出保守判决即为fail-closed，代价就是误拒率增加。

4 仿真算例与结果

4.1 数据与参数说明

仿真包含4个逻辑域（观测、海区管理、高校科研、海事业务），320个主体，1200个原始对象，12000条流转事件和2600个派生节点；注入了240次异常请求，其中92次是合法身份超范围二次转发，68次是跨域串联拉取，80次是情境异常。见表1，静态角色映射、静态属性策略和本文的SFD-RAdAC三个方法同时对一个合成事件序列进行操作。

4.2 结果

拦截率提高主要是因为两类静态方案无法覆盖的请求：在合法身份下进行的二次转发以及多源融合使得敏感度提高的派生产品，策略条目相对属性策略减少了33.3%，而相对于角色映射则减少了69.9%，这是由于动态状态量变为运行时求值所致；更新耗时从之前的312毫秒缩短到现在的27毫秒，主要得益于影响域剪枝带来的节省。代价方面：相比于静态属性策略，决策时间增加了1.7ms；合法请求直接通过率比角色映射低了1.6个百分点，原因是Step-up将一些请求加入到了二次确认队列中。如果加上了二次确认之后放行的情况，最后通过率就是98.1%，减少的只是“无摩擦放行”的占比而不是可用性。毫秒级开销并不影响秒级交付，高频API直调场景需要单独考虑。

表1 仿真主要参数

参数	取值	参数	取值
逻辑域数/主体数	4/320	跨域放大因子 γ	1.25
原始对象/派生节点	1200/2600	涌现系数 λ	0.35
流转事件/异常请求	12000/240	影响域深度 h	3
敏感级 L	1—4	剪枝阈值 ε	0.02
判决阈值 (θ_1, θ_2)	(0.42, 0.68)	权重 ($w_1,\cdots,w_5$)	(0.38, 0.14, 0.12, 0.16, 0.20)

4.3 敏感性

从原来的3增加到现在的5之后，漏拦截率只提升了0.4个百分点，而所花费的时间也增加了近1.3倍，所以取值为h=3，当γ由原来的1.10提高到1.40时，跨域串联类拦截率由82.7%上升至96.4%，合法直接通过率则从原来的96.0%下降到了92.3%，因此，γ 应该根据域间信任协定来分档而不是在全域内统一规定。

5 结论

把派生链带入判决时刻就可以弥补静态授权在海洋数据跨域共享中所存在的结构缺陷，本文提出了一种动态溯流图、包含跨域放大和融合涌现的敏感度计

表2 三种方案的对比结果（合成数据）

方案	越权拦截率/%	合法直接通过率/%	平均决策时延/ms	有效策略条目/条	流转后策略更新耗时/ms
静态角色映射	61.3	96.8	1.8	4120	人工重配
静态属性策略	78.8	93.5	4.6	1860	312（全量）
SFD-RAdAC	94.6	95.2	6.3	1240	27（增量）

表3 影响域深度 h 的敏感性（合成数据）

h	增量更新耗时/ms	漏拦截率/%
1	11	12.4
2	19	6.1
3	27	3.5
4	41	3.2
5	63	3.1

算以及三级判决与影响域增量更新，从而形成了从状态刻画到策略执行之间的联系；合成算例表明了拦截能力和策略规模方面的好处，并真实地体现了时延以及无摩擦放行比例的代价，模型不适用于离线交付及事件缺报的情况，κ的标定还需要用到真实的语料。下一步按照域间信任协定对γ和κ进行分级标识，在具备事件记录功能的平台上线后进行测试，用作代替模拟实验的结果。

参考文献

[1] 李凤华, 苏铠, 史国振, 等. 访问控制模型研究进展及发展趋势[J]. 电子学报, 2012, 40(4): 805-813.

[2] 周秀, 刘培顺, 刘加标, 等. 海洋环境云平台访问控制系统研究[J]. 微型机与应用, 2015, 34(7): 9-12, 23.

[3] 国土资源部. 海洋观测资料管理办法: 国土资源部令第74号[Z]. 2017-06-07.

[4] 国家市场监督管理总局, 国家标准化管理委员会. 数据安全技术 数据分类分级规则: GB/T 43697—2024[S]. 北京: 中国标准出版社, 2024.

[5] SIMMHAN Y L, PLALE B, GANNON D. A survey of data provenance in e-science[J]. ACM SIGMOD Record, 2005, 34(3): 31-36.

[6] KANDALA S, SANDHU R, BHAMIDIPATI V. An attribute based framework for risk-adaptive access control models[C]//2011 Sixth International Conference on Availability, Reliability and Security. IEEE, 2011: 236-241.